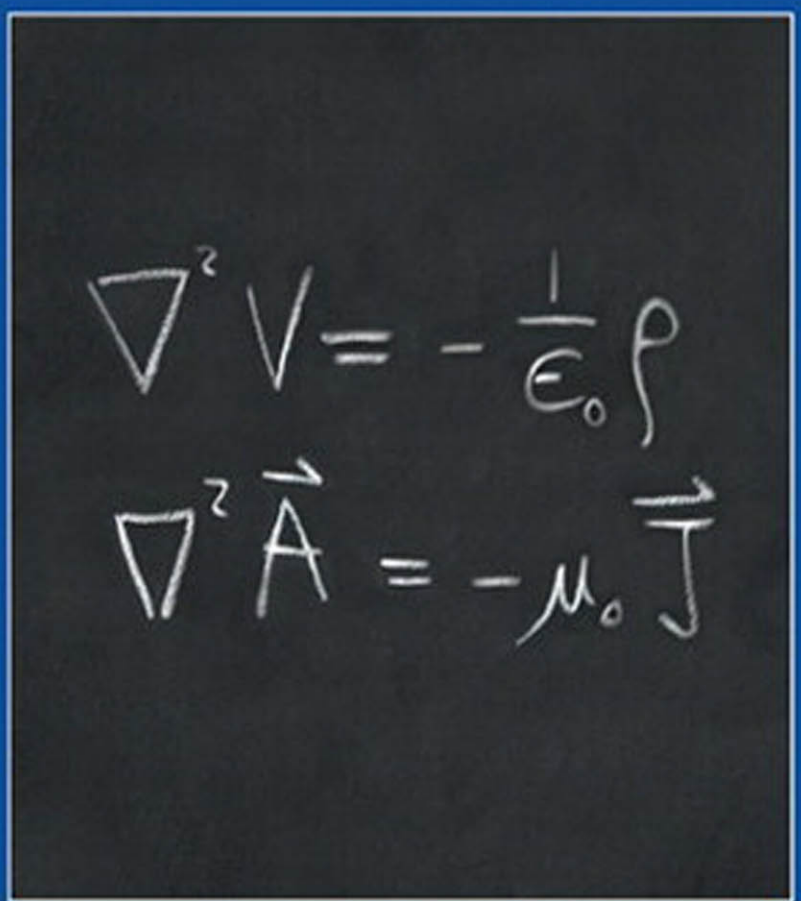
INTRODUCTION TO ELECTRODYNAMICS

Fourth Edition



DAVID J. GRIFFITHS

INTRODUCTION TO ELECTRODYNAMICS

This page intentionally left blank

INTRODUCTION TO ELECTRODYNAMICS

Fourth Edition

David J. Griffiths
Reed College

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Executive Editor: Jim Smith
Senior Project Editor: Martha Steele
Development Manager: Laura Kenney
Managing Editor: Corinne Benson
Production Project Manager: Dorothy Cox
Production Management and Composition: Integra
Cover Designer: Derek Bacchus
Manufacturing Buyer: Dorothy Cox
Marketing Manager: Will Moore

Credits and acknowledgments for materials borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within the text.

Library of Congress Cataloging-in-Publication Data
Griffiths, David J. (David Jeffery), 1942-
Introduction to electrodynamics/
David J. Griffiths, Reed College. – Fourth edition.
pages cm
Includes index.
ISBN-13: 978-0-321-85656-2 (alk. paper)
ISBN-10: 0-321-85656-2 (alk. paper)
1. Electrodynamics–Textbooks. I. Title.
QC680.G74 2013
537.6–dc23
2012029768

ISBN 10: 0-321-85656-2
ISBN 13: 978-0-321-85656-2

PEARSON

www.pearsonhighered.com 1 2 3 4 5 6 7 8 9 10—CRW—16 15 14 13 12

Contents

Preface

This is a textbook on electricity and magnetism, designed for an undergraduate course at the junior or senior level. It can be covered comfortably in two semesters, maybe even with room to spare for special topics (AC circuits, numerical methods, plasma physics, transmission lines, antenna theory, etc.) A one-semester course could reasonably stop after Chapter 7. Unlike quantum mechanics or thermal physics (for example), there is a fairly general consensus with respect to the teaching of electrodynamics; the subjects to be included, and even their order of presentation, are not particularly controversial, and textbooks differ mainly in style and tone. My approach is perhaps less formal than most; I think this makes difficult ideas more interesting and accessible.

For this new edition I have made a large number of small changes, in the interests of clarity and grace. In a few places I have corrected serious errors. I have added some problems and examples (and removed a few that were not effective). And I have included more references to the accessible literature (particularly the *American Journal of Physics*). I realize, of course, that most readers will not have the time or inclination to consult these resources, but I think it is worthwhile anyway, if only to emphasize that electrodynamics, notwithstanding its venerable age, is very much alive, and intriguing new discoveries are being made all the time. I hope that occasionally a problem will pique your curiosity, and you will be inspired to look up the reference—some of them are real gems.

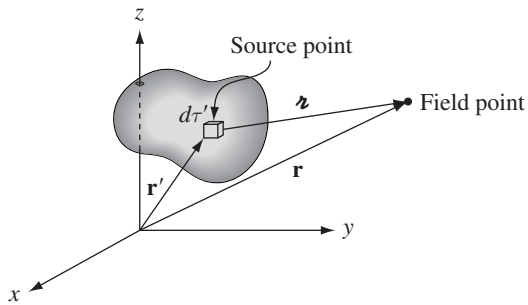
I have maintained three items of unorthodox notation:

- The Cartesian unit vectors are written $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ (and, in general, all unit vectors inherit the letter of the corresponding coordinate).

- The distance from the z axis in cylindrical coordinates is designated by s, to avoid confusion with r (the distance from the *origin*, and the radial coordinate in spherical coordinates).

- The script letter $\boldsymbol{\mathcal{r}}$ denotes the vector from a source point $\mathbf{r}'$ to the field point $\mathbf{r}$ (see Figure). Some authors prefer the more explicit $(\mathbf{r} - \mathbf{r}')$. But this makes many equations distractingly cumbersome, especially when the unit vector $\hat{\boldsymbol{\mathcal{r}}}$ is involved. I realize that unwary readers are tempted to interpret $\boldsymbol{\mathcal{r}}$ as $\mathbf{r}$—it certainly makes the integrals easier! *Please take note:* $\boldsymbol{\mathcal{r}} \equiv (\mathbf{r} - \mathbf{r}')$, which is *not* the same as $\mathbf{r}$. I think it's good notation, but it does have to be handled with care.[1]

[1] In MS Word, $\boldsymbol{\mathcal{r}}$ is "Kaufmann font," but this is very difficult to install in TeX. TeX users can download a pretty good facsimile from my web site.



As in previous editions, I distinguish two kinds of problems. Some have a specific pedagogical purpose, and should be worked immediately after reading the section to which they pertain; these I have placed at the pertinent point within the chapter. (In a few cases the solution to a problem is used later in the text; these are indicated by a bullet (•) in the left margin.) Longer problems, or those of a more general nature, will be found at the end of each chapter. When I teach the subject, I assign some of these, and work a few of them in class. Unusually challenging problems are flagged by an exclamation point (**!**) in the margin. Many readers have asked that the answers to problems be provided at the back of the book; unfortunately, just as many are strenuously opposed. I have compromised, supplying answers when this seems particularly appropriate. A complete solution manual is available (to instructors) from the publisher; go to the Pearson web site to order a copy.

I have benefitted from the comments of many colleagues. I cannot list them all here, but I would like to thank the following people for especially useful contributions to this edition: Burton Brody (Bard), Catherine Crouch (Swarthmore), Joel Franklin (Reed), Ted Jacobson (Maryland), Don Koks (Adelaide), Charles Lane (Berry), Kirk McDonald[2] (Princeton), Jim McTavish (Liverpool), Rich Saenz (Cal Poly), Darrel Schroeter (Reed), Herschel Snodgrass (Lewis and Clark), and Larry Tankersley (Naval Academy). Practically everything I know about electrodynamics—certainly about teaching electrodynamics—I owe to Edward Purcell.

David J. Griffiths

[2]Kirk's web site, http://www.hep.princeton.edu/~mcdonald/examples/, is a fantastic resource, with clever explanations, nifty problems, and useful references.

Advertisement

WHAT IS ELECTRODYNAMICS, AND HOW DOES IT FIT INTO THE GENERAL SCHEME OF PHYSICS?

Four Realms of Mechanics

In the diagram below, I have sketched out the four great realms of mechanics:

Classical Mechanics (Newton)	**Quantum Mechanics** (Bohr, Heisenberg, Schrödinger, et al.)
Special Relativity (Einstein)	**Quantum Field Theory** (Dirac, Pauli, Feynman, Schwinger, et al.)

Newtonian mechanics is adequate for most purposes in "everyday life," but for objects moving at high speeds (near the speed of light) it is incorrect, and must be replaced by special relativity (introduced by Einstein in 1905); for objects that are extremely small (near the size of atoms) it fails for different reasons, and is superseded by quantum mechanics (developed by Bohr, Schrödinger, Heisenberg, and many others, in the 1920's, mostly). For objects that are both very fast *and* very small (as is common in modern particle physics), a mechanics that combines relativity and quantum principles is in order; this relativistic quantum mechanics is known as quantum field theory—it was worked out in the thirties and forties, but even today it cannot claim to be a completely satisfactory system. In this book, save for the last chapter, we shall work exclusively in the domain of classical mechanics, although electrodynamics extends with unique simplicity to the other three realms. (In fact, the theory is in most respects *automatically* consistent with special relativity, for which it was, historically, the main stimulus.)

Four Kinds of Forces

Mechanics tells us how a system will behave when subjected to a given *force.* There are just *four* basic forces known (presently) to physics: I list them in the order of decreasing strength:

1. Strong
2. Electromagnetic
3. Weak
4. Gravitational

The brevity of this list may surprise you. Where is friction? Where is the "normal" force that keeps you from falling through the floor? Where are the chemical forces that bind molecules together? Where is the force of impact between two colliding billiard balls? The answer is that *all* these forces are *electromagnetic.* Indeed, it is scarcely an exaggeration to say that we live in an electromagnetic world—virtually every force we experience in everyday life, with the exception of gravity, is electromagnetic in origin.

The **strong forces**, which hold protons and neutrons together in the atomic nucleus, have extremely short range, so we do not "feel" them, in spite of the fact that they are a hundred times more powerful than electrical forces. The **weak forces**, which account for certain kinds of radioactive decay, are also of short range, and they are far weaker than electromagnetic forces. As for gravity, it is so pitifully feeble (compared to all of the others) that it is only by virtue of huge mass concentrations (like the earth and the sun) that we ever notice it at all. The electrical repulsion between two electrons is 10^{42} times as large as their gravitational attraction, and if atoms were held together by gravitational (instead of electrical) forces, a single hydrogen atom would be much larger than the known universe.

Not only are electromagnetic forces overwhelmingly dominant in everyday life, they are also, at present, the *only* ones that are completely understood. There is, of course, a classical theory of gravity (Newton's law of universal gravitation) and a relativistic one (Einstein's general relativity), but no entirely satisfactory quantum mechanical theory of gravity has been constructed (though many people are working on it). At the present time there is a very successful (if cumbersome) theory for the weak interactions, and a strikingly attractive candidate (called **chromodynamics**) for the strong interactions. All these theories draw their inspiration from electrodynamics; none can claim conclusive experimental verification at this stage. So electrodynamics, a beautifully complete and successful theory, has become a kind of paradigm for physicists: an ideal model that other theories emulate.

The laws of classical electrodynamics were discovered in bits and pieces by Franklin, Coulomb, Ampère, Faraday, and others, but the person who completed the job, and packaged it all in the compact and consistent form it has today, was James Clerk Maxwell. The theory is now about 150 years old.

The Unification of Physical Theories

In the beginning, **electricity** and **magnetism** were entirely separate subjects. The one dealt with glass rods and cat's fur, pith balls, batteries, currents, electrolysis, and lightning; the other with bar magnets, iron filings, compass needles, and the North Pole. But in 1820 Oersted noticed that an *electric* current could deflect

a *magnetic* compass needle. Soon afterward, Ampère correctly postulated that *all* magnetic phenomena are due to electric charges in motion. Then, in 1831, Faraday discovered that a moving *magnet* generates an *electric* current. By the time Maxwell and Lorentz put the finishing touches on the theory, electricity and magnetism were inextricably intertwined. They could no longer be regarded as separate subjects, but rather as two *aspects* of a *single* subject: **electromagnetism**.

Faraday speculated that light, too, is electrical in nature. Maxwell's theory provided spectacular justification for this hypothesis, and soon **optics**—the study of lenses, mirrors, prisms, interference, and diffraction—was incorporated into electromagnetism. Hertz, who presented the decisive experimental confirmation for Maxwell's theory in 1888, put it this way: "The connection between light and electricity is now established . . . In every flame, in every luminous particle, we see an electrical process . . . Thus, the domain of electricity extends over the whole of nature. It even affects ourselves intimately: we perceive that we possess . . . an electrical organ—the eye." By 1900, then, three great branches of physics–electricity, magnetism, and optics–had merged into a single unified theory. (And it was soon apparent that visible light represents only a tiny "window" in the vast spectrum of electromagnetic radiation, from radio through microwaves, infrared and ultraviolet, to x-rays and gamma rays.)

Einstein dreamed of a further unification, which would combine gravity and electrodynamics, in much the same way as electricity and magnetism had been combined a century earlier. His **unified field theory** was not particularly successful, but in recent years the same impulse has spawned a hierarchy of increasingly ambitious (and speculative) unification schemes, beginning in the 1960s with the **electroweak** theory of Glashow, Weinberg, and Salam (which joins the weak and electromagnetic forces), and culminating in the 1980s with the **superstring** theory (which, according to its proponents, incorporates all four forces in a single "theory of everything"). At each step in this hierarchy, the mathematical difficulties mount, and the gap between inspired conjecture and experimental test widens; nevertheless, it is clear that the unification of forces initiated by electrodynamics has become a major theme in the progress of physics.

The Field Formulation of Electrodynamics

The fundamental problem a theory of electromagnetism hopes to solve is this: I hold up a bunch of electric charges *here* (and maybe shake them around); what happens to some *other* charge, over *there?* The classical solution takes the form of a **field theory**: We say that the space around an electric charge is permeated by electric and magnetic **fields** (the electromagnetic "odor," as it were, of the charge). A second charge, in the presence of these fields, experiences a force; the fields, then, transmit the influence from one charge to the other—they "mediate" the interaction.

When a charge undergoes *acceleration,* a portion of the field "detaches" itself, in a sense, and travels off at the speed of light, carrying with it energy, momentum, and angular momentum. We call this **electromagnetic radiation**. Its exis-

tence invites (if not *compels*) us to regard the fields as independent dynamical entities in their own right, every bit as "real" as atoms or baseballs. Our interest accordingly shifts from the study of forces between charges to the theory of the fields themselves. But it takes a charge to *produce* an electromagnetic field, and it takes another charge to *detect* one, so we had best begin by reviewing the essential properties of electric charge.

Electric Charge

1. *Charge comes in two varieties,* which we call "plus" and "minus," because their effects tend to *cancel* (if you have $+q$ and $-q$ at the same point, electrically it is the same as having no charge there at all). This may seem too obvious to warrant comment, but I encourage you to contemplate other possibilities: what if there were 8 or 10 different species of charge? (In chromodynamics there are, in fact, *three* quantities analogous to electric charge, each of which may be positive or negative.) Or what if the two kinds did not tend to cancel? The extraordinary fact is that plus and minus charges occur in *exactly* equal amounts, to fantastic precision, in bulk matter, so that their effects are almost completely neutralized. Were it not for this, we would be subjected to enormous forces: a potato would explode violently if the cancellation were imperfect by as little as one part in 10^{10}.

2. *Charge is conserved:* it cannot be created or destroyed—what there is now has always been. (A plus charge can "annihilate" an equal minus charge, but a plus charge cannot simply disappear by itself—*something* must pick up that electric charge.) So the total charge of the universe is fixed for all time. This is called **global** conservation of charge. Actually, I can say something much stronger: Global conservation would allow for a charge to disappear in New York and instantly reappear in San Francisco (that wouldn't affect the *total*), and yet we know this doesn't happen. If the charge *was* in New York and it *went* to San Francisco, then it must have passed along some continuous path from one to the other. This is called **local** conservation of charge. Later on we'll see how to formulate a precise mathematical law expressing local conservation of charge—it's called the **continuity equation**.

3. *Charge is quantized.* Although nothing in classical electrodynamics requires that it be so, the *fact* is that electric charge comes only in discrete lumps—integer multiples of the basic unit of charge. If we call the charge on the proton $+e$, then the electron carries charge $-e$; the neutron charge zero; the pi mesons $+e$, 0, and $-e$; the carbon nucleus $+6e$; and so on (never $7.392e$, or even $1/2e$).[3] This fundamental unit of charge is extremely small, so for practical purposes it is usually appropriate to ignore quantization altogether. Water, too, "really" consists of discrete lumps (molecules); yet, if we are dealing with reasonably large

[3] Actually, protons and neutrons are composed of three **quarks,** which carry fractional charges ($\pm\frac{2}{3}e$ and $\pm\frac{1}{3}e$). However, *free* quarks do not appear to exist in nature, and in any event, this does not alter the fact that charge is quantized; it merely reduces the size of the basic unit.

quantities of it we can treat it as a continuous fluid. This is in fact much closer to Maxwell's own view; he knew nothing of electrons and protons—he must have pictured charge as a kind of "jelly" that could be divided up into portions of any size and smeared out at will.

Units

The subject of electrodynamics is plagued by competing systems of units, which sometimes render it difficult for physicists to communicate with one another. The problem is far worse than in mechanics, where Neanderthals still speak of pounds and feet; in mechanics, at least all equations *look* the same, regardless of the units used to measure quantities. Newton's second law remains $\mathbf{F} = m\mathbf{a}$, whether it is feet-pounds-seconds, kilograms-meters-seconds, or whatever. But this is not so in electromagnetism, where Coulomb's law may appear variously as

$$\mathbf{F} = \frac{q_1 q_2}{\imath^2}\hat{\boldsymbol{\imath}} \text{ (Gaussian)}, \quad \text{or} \quad \mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{\imath^2}\hat{\boldsymbol{\imath}} \text{ (SI)}, \quad \text{or} \quad \mathbf{F} = \frac{1}{4\pi}\frac{q_1 q_2}{\imath^2}\hat{\boldsymbol{\imath}} \text{ (HL)}.$$

Of the systems in common use, the two most popular are **Gaussian** (cgs) and **SI** (mks). Elementary particle theorists favor yet a third system: **Heaviside-Lorentz**. Although Gaussian units offer distinct theoretical advantages, most undergraduate instructors seem to prefer SI, I suppose because they incorporate the familiar household units (volts, amperes, and watts). In this book, therefore, I have used SI units. Appendix C provides a "dictionary" for converting the main results into Gaussian units.

CHAPTER

1 Vector Analysis

1.1 ■ VECTOR ALGEBRA

1.1.1 ■ Vector Operations

If you walk 4 miles due north and then 3 miles due east (Fig. 1.1), you will have gone a total of 7 miles, but you're *not* 7 miles from where you set out—you're only 5. We need an arithmetic to describe quantities like this, which evidently do not add in the ordinary way. The reason they don't, of course, is that **displacements** (straight line segments going from one point to another) have *direction* as well as *magnitude* (length), and it is essential to take both into account when you combine them. Such objects are called **vectors**: velocity, acceleration, force and momentum are other examples. By contrast, quantities that have magnitude but no direction are called **scalars**: examples include mass, charge, density, and temperature.

I shall use **boldface** (**A**, **B**, and so on) for vectors and ordinary type for scalars. The magnitude of a vector **A** is written $|\mathbf{A}|$ or, more simply, A. In diagrams, vectors are denoted by arrows: the length of the arrow is proportional to the magnitude of the vector, and the arrowhead indicates its direction. *Minus* **A** ($-\mathbf{A}$) is a vector with the same magnitude as **A** but of opposite direction (Fig. 1.2). Note that vectors have magnitude and direction but *not location:* a displacement of 4 miles due north from Washington is represented by the same vector as a displacement 4 miles north from Baltimore (neglecting, of course, the curvature of the earth). On a diagram, therefore, you can slide the arrow around at will, as long as you don't change its length or direction.

We define four vector operations: addition and three kinds of multiplication.

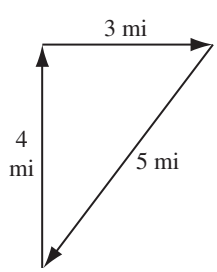


FIGURE 1.1

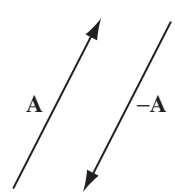


FIGURE 1.2

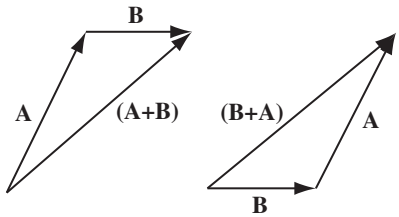


FIGURE 1.3

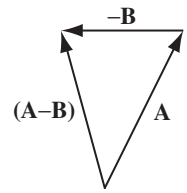


FIGURE 1.4

(i) Addition of two vectors. Place the tail of **B** at the head of **A**; the sum, $\mathbf{A} + \mathbf{B}$, is the vector from the tail of **A** to the head of **B** (Fig. 1.3). (This rule generalizes the obvious procedure for combining two displacements.) Addition is *commutative:*

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A};$$

3 miles east followed by 4 miles north gets you to the same place as 4 miles north followed by 3 miles east. Addition is also *associative:*

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}).$$

To subtract a vector, add its opposite (Fig. 1.4):

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}).$$

(ii) Multiplication by a scalar. Multiplication of a vector by a positive scalar a multiplies the *magnitude* but leaves the direction unchanged (Fig. 1.5). (If a is negative, the direction is reversed.) Scalar multiplication is *distributive:*

$$a(\mathbf{A} + \mathbf{B}) = a\mathbf{A} + a\mathbf{B}.$$

(iii) Dot product of two vectors. The dot product of two vectors is defined by

$$\mathbf{A} \cdot \mathbf{B} \equiv AB\cos\theta, \tag{1.1}$$

where θ is the angle they form when placed tail-to-tail (Fig. 1.6). Note that $\mathbf{A} \cdot \mathbf{B}$ is itself a *scalar* (hence the alternative name **scalar product**). The dot product is *commutative,*

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A},$$

and *distributive,*

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}. \tag{1.2}$$

Geometrically, $\mathbf{A} \cdot \mathbf{B}$ is the product of A times the projection of **B** along **A** (or the product of B times the projection of **A** along **B**). If the two vectors are parallel, then $\mathbf{A} \cdot \mathbf{B} = AB$. In particular, for any vector **A**,

$$\mathbf{A} \cdot \mathbf{A} = A^2. \tag{1.3}$$

If **A** and **B** are perpendicular, then $\mathbf{A} \cdot \mathbf{B} = 0$.

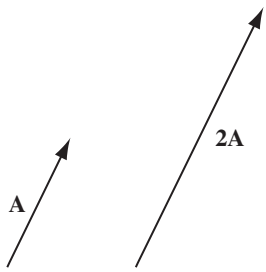


FIGURE 1.5

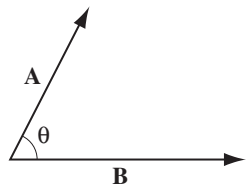


FIGURE 1.6

Example 1.1. Let $\mathbf{C} = \mathbf{A} - \mathbf{B}$ (Fig. 1.7), and calculate the dot product of $\mathbf{C}$ with itself.

Solution

$$\mathbf{C} \cdot \mathbf{C} = (\mathbf{A} - \mathbf{B}) \cdot (\mathbf{A} - \mathbf{B}) = \mathbf{A} \cdot \mathbf{A} - \mathbf{A} \cdot \mathbf{B} - \mathbf{B} \cdot \mathbf{A} + \mathbf{B} \cdot \mathbf{B},$$

or

$$C^2 = A^2 + B^2 - 2AB\cos\theta.$$

This is the **law of cosines**.

(iv) Cross product of two vectors. The cross product of two vectors is defined by

$$\mathbf{A} \times \mathbf{B} \equiv AB\sin\theta\,\hat{\mathbf{n}}, \tag{1.4}$$

where $\hat{\mathbf{n}}$ is a **unit vector** (vector of magnitude 1) pointing perpendicular to the plane of $\mathbf{A}$ and $\mathbf{B}$. (I shall use a hat ($\hat{}$) to denote unit vectors.) Of course, there are *two* directions perpendicular to any plane: "in" and "out." The ambiguity is resolved by the **right-hand rule**: let your fingers point in the direction of the first vector and curl around (via the smaller angle) toward the second; then your thumb indicates the direction of $\hat{\mathbf{n}}$. (In Fig. 1.8, $\mathbf{A} \times \mathbf{B}$ points *into* the page; $\mathbf{B} \times \mathbf{A}$ points *out* of the page.) Note that $\mathbf{A} \times \mathbf{B}$ is itself a *vector* (hence the alternative name **vector product**). The cross product is *distributive,*

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) + (\mathbf{A} \times \mathbf{C}), \tag{1.5}$$

but *not commutative.* In fact,

$$(\mathbf{B} \times \mathbf{A}) = -(\mathbf{A} \times \mathbf{B}). \tag{1.6}$$

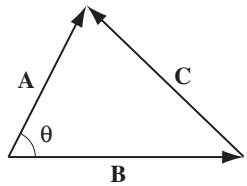


FIGURE 1.7

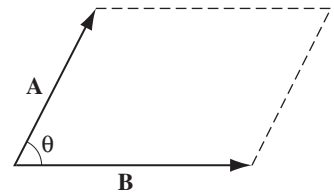


FIGURE 1.8

Geometrically, $|\mathbf{A} \times \mathbf{B}|$ is the area of the parallelogram generated by $\mathbf{A}$ and $\mathbf{B}$ (Fig. 1.8). If two vectors are parallel, their cross product is zero. In particular,

$$\mathbf{A} \times \mathbf{A} = \mathbf{0}$$

for any vector $\mathbf{A}$. (Here $\mathbf{0}$ is the **zero vector**, with magnitude 0.)

Problem 1.1 Using the definitions in Eqs. 1.1 and 1.4, and appropriate diagrams, show that the dot product and cross product are distributive,

a) when the three vectors are coplanar;

! b) in the general case.

Problem 1.2 Is the cross product associative?

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} \stackrel{?}{=} \mathbf{A} \times (\mathbf{B} \times \mathbf{C}).$$

If so, *prove* it; if not, provide a counterexample (the simpler the better).

1.1.2 ■ Vector Algebra: Component Form

In the previous section, I defined the four vector operations (addition, scalar multiplication, dot product, and cross product) in "abstract" form—that is, without reference to any particular coordinate system. In practice, it is often easier to set up Cartesian coordinates x, y, z and work with vector **components**. Let $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ be unit vectors parallel to the x, y, and z axes, respectively (Fig. 1.9(a)). An arbitrary vector $\mathbf{A}$ can be expanded in terms of these **basis vectors** (Fig. 1.9(b)):

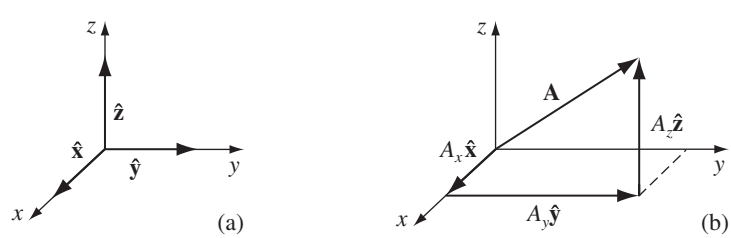


FIGURE 1.9

$$\mathbf{A} = A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}.$$

The numbers A_x, A_y, and A_z, are the "components" of $\mathbf{A}$; geometrically, they are the projections of $\mathbf{A}$ along the three coordinate axes ($A_x = \mathbf{A} \cdot \hat{\mathbf{x}}$, $A_y = \mathbf{A} \cdot \hat{\mathbf{y}}$, $A_z = \mathbf{A} \cdot \hat{\mathbf{z}}$). We can now reformulate each of the four vector operations as a rule for manipulating components:

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= (A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}) + (B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}} + B_z\hat{\mathbf{z}}) \\ &= (A_x + B_x)\hat{\mathbf{x}} + (A_y + B_y)\hat{\mathbf{y}} + (A_z + B_z)\hat{\mathbf{z}}. \end{aligned} \tag{1.7}$$

Rule (i): *To add vectors, add like components.*

$$a\mathbf{A} = (aA_x)\hat{\mathbf{x}} + (aA_y)\hat{\mathbf{y}} + (aA_z)\hat{\mathbf{z}}. \tag{1.8}$$

Rule (ii): *To multiply by a scalar, multiply each component.*

Because $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are mutually perpendicular unit vectors,

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1; \quad \hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{x}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{z}} = 0. \tag{1.9}$$

Accordingly,

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} &= (A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}) \cdot (B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}} + B_z\hat{\mathbf{z}}) \\ &= A_xB_x + A_yB_y + A_zB_z. \end{aligned} \tag{1.10}$$

Rule (iii): *To calculate the dot product, multiply like components, and add.*
In particular,

$$\mathbf{A} \cdot \mathbf{A} = A_x^2 + A_y^2 + A_z^2,$$

so

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}. \tag{1.11}$$

(This is, if you like, the three-dimensional generalization of the Pythagorean theorem.)

Similarly,[1]

$$\begin{aligned}\hat{\mathbf{x}} \times \hat{\mathbf{x}} &= \quad \hat{\mathbf{y}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}} \times \hat{\mathbf{z}} = \mathbf{0}, \\ \hat{\mathbf{x}} \times \hat{\mathbf{y}} &= -\hat{\mathbf{y}} \times \hat{\mathbf{x}} = \hat{\mathbf{z}}, \\ \hat{\mathbf{y}} \times \hat{\mathbf{z}} &= -\hat{\mathbf{z}} \times \hat{\mathbf{y}} = \hat{\mathbf{x}}, \\ \hat{\mathbf{z}} \times \hat{\mathbf{x}} &= -\hat{\mathbf{x}} \times \hat{\mathbf{z}} = \hat{\mathbf{y}}. \end{aligned} \tag{1.12}$$

[1]These signs pertain to a *right-handed* coordinate system (x-axis out of the page, y-axis to the right, z-axis up, or any rotated version thereof). In a *left-handed* system (z-axis down), the signs would be reversed: $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = -\hat{\mathbf{z}}$, and so on. We shall use right-handed systems exclusively.

Therefore,

$$\mathbf{A} \times \mathbf{B} = (A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}) \times (B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}} + B_z\hat{\mathbf{z}}) \tag{1.13}$$
$$= (A_yB_z - A_zB_y)\hat{\mathbf{x}} + (A_zB_x - A_xB_z)\hat{\mathbf{y}} + (A_xB_y - A_yB_x)\hat{\mathbf{z}}.$$

This cumbersome expression can be written more neatly as a determinant:

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}. \tag{1.14}$$

Rule (iv): *To calculate the cross product, form the determinant whose first row is* $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$, *whose second row is* $\mathbf{A}$ *(in component form), and whose third row is* $\mathbf{B}$.

Example 1.2. Find the angle between the face diagonals of a cube.

Solution

We might as well use a cube of side 1, and place it as shown in Fig. 1.10, with one corner at the origin. The face diagonals $\mathbf{A}$ and $\mathbf{B}$ are

$$\mathbf{A} = 1\,\hat{\mathbf{x}} + 0\,\hat{\mathbf{y}} + 1\,\hat{\mathbf{z}}; \qquad \mathbf{B} = 0\,\hat{\mathbf{x}} + 1\,\hat{\mathbf{y}} + 1\,\hat{\mathbf{z}}.$$

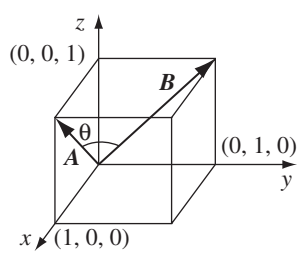


FIGURE 1.10

So, in component form,

$$\mathbf{A} \cdot \mathbf{B} = 1 \cdot 0 + 0 \cdot 1 + 1 \cdot 1 = 1.$$

On the other hand, in "abstract" form,

$$\mathbf{A} \cdot \mathbf{B} = AB\cos\theta = \sqrt{2}\sqrt{2}\cos\theta = 2\cos\theta.$$

Therefore,

$$\cos\theta = 1/2, \quad \text{or} \quad \theta = 60°.$$

Of course, you can get the answer more easily by drawing in a diagonal across the top of the cube, completing the equilateral triangle. But in cases where the geometry is not so simple, this device of comparing the abstract and component forms of the dot product can be a very efficient means of finding angles.

Problem 1.3 Find the angle between the body diagonals of a cube.

Problem 1.4 Use the cross product to find the components of the unit vector $\hat{\mathbf{n}}$ perpendicular to the shaded plane in Fig. 1.11.

1.1.3 ■ Triple Products

Since the cross product of two vectors is itself a vector, it can be dotted or crossed with a third vector to form a *triple* product.

(i) Scalar triple product: $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$. Geometrically, $|\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})|$ is the volume of the parallelepiped generated by $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, since $|\mathbf{B} \times \mathbf{C}|$ is the area of the base, and $|\mathbf{A} \cos\theta|$ is the altitude (Fig. 1.12). Evidently,

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}), \tag{1.15}$$

for they all correspond to the same figure. Note that "alphabetical" order is preserved—in view of Eq. 1.6, the "nonalphabetical" triple products,

$$\mathbf{A} \cdot (\mathbf{C} \times \mathbf{B}) = \mathbf{B} \cdot (\mathbf{A} \times \mathbf{C}) = \mathbf{C} \cdot (\mathbf{B} \times \mathbf{A}),$$

have the opposite sign. In component form,

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}. \tag{1.16}$$

Note that the dot and cross can be interchanged:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}$$

(this follows immediately from Eq. 1.15); however, the placement of the parentheses is critical: $(\mathbf{A} \cdot \mathbf{B}) \times \mathbf{C}$ is a meaningless expression—you can't make a cross product from a *scalar* and a vector.

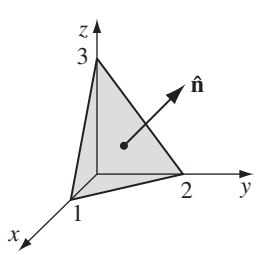


FIGURE 1.11

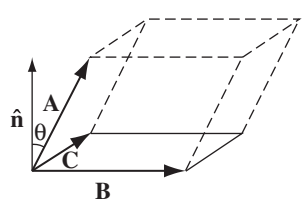


FIGURE 1.12

(ii) Vector triple product: $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$. The vector triple product can be simplified by the so-called **BAC-CAB** rule:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}). \tag{1.17}$$

Notice that

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = -\mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = -\mathbf{A}(\mathbf{B} \cdot \mathbf{C}) + \mathbf{B}(\mathbf{A} \cdot \mathbf{C})$$

is an entirely different vector (cross-products are not associative). All *higher* vector products can be similarly reduced, often by repeated application of Eq. 1.17, so it is never necessary for an expression to contain more than one cross product in any term. For instance,

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C});$$

$$\mathbf{A} \times [\mathbf{B} \times (\mathbf{C} \times \mathbf{D})] = \mathbf{B}[\mathbf{A} \cdot (\mathbf{C} \times \mathbf{D})] - (\mathbf{A} \cdot \mathbf{B})(\mathbf{C} \times \mathbf{D}). \tag{1.18}$$

Problem 1.5 Prove the **BAC-CAB** rule by writing out both sides in component form.

Problem 1.6 Prove that

$$[\mathbf{A} \times (\mathbf{B} \times \mathbf{C})] + [\mathbf{B} \times (\mathbf{C} \times \mathbf{A})] + [\mathbf{C} \times (\mathbf{A} \times \mathbf{B})] = \mathbf{0}.$$

Under what conditions does $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \times \mathbf{C}$?

1.1.4 ■ Position, Displacement, and Separation Vectors

The location of a point in three dimensions can be described by listing its Cartesian coordinates (x, y, z). The vector to that point from the origin ($\mathcal{O}$) is called the **position vector** (Fig. 1.13):

$$\mathbf{r} \equiv x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}. \tag{1.19}$$

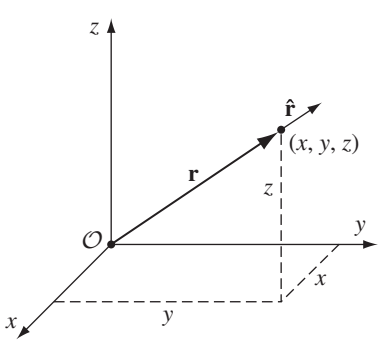


FIGURE 1.13

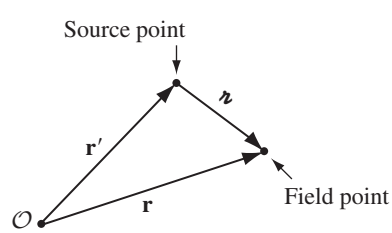


FIGURE 1.14

I will reserve the letter $\mathbf{r}$ for this purpose, throughout the book. Its magnitude,

$$r = \sqrt{x^2 + y^2 + z^2}, \tag{1.20}$$

is the distance from the origin, and

$$\hat{\mathbf{r}} = \frac{\mathbf{r}}{r} = \frac{x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}}{\sqrt{x^2 + y^2 + z^2}} \tag{1.21}$$

is a unit vector pointing radially outward. The **infinitesimal displacement vector**, from (x, y, z) to $(x + dx, y + dy, z + dz)$, is

$$d\mathbf{l} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}. \tag{1.22}$$

(We could call this $d\mathbf{r}$, since that's what it *is*, but it is useful to have a special notation for infinitesimal displacements.)

In electrodynamics, one frequently encounters problems involving *two* points—typically, a **source point**, $\mathbf{r}'$, where an electric charge is located, and a **field point**, $\mathbf{r}$, at which you are calculating the electric or magnetic field (Fig. 1.14). It pays to adopt right from the start some short-hand notation for the **separation vector** from the source point to the field point. I shall use for this purpose the script letter $\boldsymbol{\mathscr{r}}$:

$$\boldsymbol{\mathscr{r}} \equiv \mathbf{r} - \mathbf{r}'. \tag{1.23}$$

Its magnitude is

$$\mathscr{r} = |\mathbf{r} - \mathbf{r}'|, \tag{1.24}$$

and a unit vector in the direction from $\mathbf{r}'$ to $\mathbf{r}$ is

$$\hat{\boldsymbol{\mathscr{r}}} = \frac{\boldsymbol{\mathscr{r}}}{\mathscr{r}} = \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}. \tag{1.25}$$

In Cartesian coordinates,

$$\boldsymbol{\mathscr{r}} = (x - x')\hat{\mathbf{x}} + (y - y')\hat{\mathbf{y}} + (z - z')\hat{\mathbf{z}}, \tag{1.26}$$

$$\mathscr{r} = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}, \tag{1.27}$$

$$\hat{\boldsymbol{\mathscr{r}}} = \frac{(x - x')\hat{\mathbf{x}} + (y - y')\hat{\mathbf{y}} + (z - z')\hat{\mathbf{z}}}{\sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}} \tag{1.28}$$

(from which you can appreciate the economy of the script-$\mathscr{r}$ notation).

Problem 1.7 Find the separation vector $\boldsymbol{\mathscr{r}}$ from the source point (2,8,7) to the field point (4,6,8). Determine its magnitude ($\mathscr{r}$), and construct the unit vector $\hat{\boldsymbol{\mathscr{r}}}$.

1.1.5 ■ How Vectors Transform[2]

The definition of a vector as "a quantity with a magnitude and direction" is not altogether satisfactory: What precisely does "direction" *mean*? This may seem a pedantic question, but we shall soon encounter a species of derivative that *looks* rather like a vector, and we'll want to know for sure whether it *is* one.

You might be inclined to say that a vector is anything that has three components that combine properly under addition. Well, how about this: We have a barrel of fruit that contains N_x pears, N_y apples, and N_z bananas. Is $\mathbf{N} = N_x\hat{\mathbf{x}} + N_y\hat{\mathbf{y}} + N_z\hat{\mathbf{z}}$ a vector? It has three components, and when you add another barrel with M_x pears, M_y apples, and M_z bananas the result is $(N_x + M_x)$ pears, $(N_y + M_y)$ apples, $(N_z + M_z)$ bananas. So it does *add* like a vector. Yet it's obviously *not* a vector, in the physicist's sense of the word, because it doesn't really have a direction. What exactly is wrong with it?

The answer is that $\mathbf{N}$ *does not transform properly when you change coordinates.* The coordinate frame we use to describe positions in space is of course entirely arbitrary, but there is a specific geometrical transformation law for converting vector components from one frame to another. Suppose, for instance, the $\overline{x}, \overline{y}, \overline{z}$ system is rotated by angle ϕ, relative to x, y, z, about the common $x = \overline{x}$ axes. From Fig. 1.15,

$$A_y = A\cos\theta, \qquad A_z = A\sin\theta,$$

while

$$\begin{aligned}\overline{A}_y &= A\cos\overline{\theta} = A\cos(\theta - \phi) = A(\cos\theta\cos\phi + \sin\theta\sin\phi)\\ &= \cos\phi A_y + \sin\phi A_z,\\ \overline{A}_z &= A\sin\overline{\theta} = A\sin(\theta - \phi) = A(\sin\theta\cos\phi - \cos\theta\sin\phi)\\ &= -\sin\phi A_y + \cos\phi A_z.\end{aligned}$$

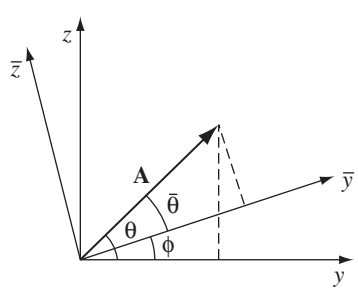


FIGURE 1.15

[2]This section can be skipped without loss of continuity.

We might express this conclusion in matrix notation:

$$\begin{pmatrix} \overline{A}_y \\ \overline{A}_z \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} A_y \\ A_z \end{pmatrix}. \tag{1.29}$$

More generally, for rotation about an *arbitrary* axis in three dimensions, the transformation law takes the form

$$\begin{pmatrix} \overline{A}_x \\ \overline{A}_y \\ \overline{A}_z \end{pmatrix} = \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}, \tag{1.30}$$

or, more compactly,

$$\overline{A}_i = \sum_{j=1}^{3} R_{ij} A_j, \tag{1.31}$$

where the index 1 stands for x, 2 for y, and 3 for z. The elements of the matrix R can be ascertained, for a given rotation, by the same sort of trigonometric arguments as we used for a rotation about the x axis.

Now: *Do* the components of **N** transform in this way? Of *course* not—it doesn't matter what coordinates you use to represent positions in space; there are still just as many apples in the barrel. You can't convert a pear into a banana by choosing a different set of axes, but you *can* turn A_x into $\overline{A}_y$. Formally, then, a *vector is any set of three components that transforms in the same manner as a displacement when you change coordinates.* As always, displacement is the *model* for the behavior of all vectors.[3]

By the way, a (second-rank) **tensor** is a quantity with *nine* components, T_{xx}, T_{xy}, T_{xz}, T_{yx}, …, T_{zz}, which transform with *two* factors of R:

$$\begin{aligned} \overline{T}_{xx} = {} & R_{xx}(R_{xx}T_{xx} + R_{xy}T_{xy} + R_{xz}T_{xz}) \\ & + R_{xy}(R_{xx}T_{yx} + R_{xy}T_{yy} + R_{xz}T_{yz}) \\ & + R_{xz}(R_{xx}T_{zx} + R_{xy}T_{zy} + R_{xz}T_{zz}), \ldots \end{aligned}$$

or, more compactly,

$$\overline{T}_{ij} = \sum_{k=1}^{3}\sum_{l=1}^{3} R_{ik}R_{jl}T_{kl}. \tag{1.32}$$

[3]If you're a mathematician you might want to contemplate generalized vector spaces in which the "axes" have nothing to do with direction and the basis vectors are no longer $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ (indeed, there may be more than three dimensions). This is the subject of **linear algebra**. But for our purposes all vectors live in ordinary 3-space (or, in Chapter 12, in 4-dimensional space-time.)

In general, an nth-rank tensor has n indices and 3^n components, and transforms with n factors of R. In this hierarchy, a vector is a tensor of rank 1, and a scalar is a tensor of rank zero.[4]

Problem 1.8

(a) Prove that the two-dimensional rotation matrix (Eq. 1.29) preserves dot products. (That is, show that $\overline{A}_y\overline{B}_y + \overline{A}_z\overline{B}_z = A_yB_y + A_zB_z$.)

(b) What constraints must the elements (R_{ij}) of the three-dimensional rotation matrix (Eq. 1.30) satisfy, in order to preserve the length of **A** (for all vectors **A**)?

Problem 1.9 Find the transformation matrix R that describes a rotation by 120° about an axis from the origin through the point (1, 1, 1). The rotation is clockwise as you look down the axis toward the origin.

Problem 1.10

(a) How do the components of a vector[5] transform under a **translation** of coordinates ($\overline{x} = x$, $\overline{y} = y - a$, $\overline{z} = z$, Fig. 1.16a)?

(b) How do the components of a vector transform under an **inversion** of coordinates ($\overline{x} = -x$, $\overline{y} = -y$, $\overline{z} = -z$, Fig. 1.16b)?

(c) How do the components of a cross product (Eq. 1.13) transform under inversion? [The cross-product of two vectors is properly called a **pseudovector** because of this "anomalous" behavior.] Is the cross product of two pseudovectors a vector, or a pseudovector? Name two pseudovector quantities in classical mechanics.

(d) How does the scalar triple product of three vectors transform under inversions? (Such an object is called a **pseudoscalar.**)

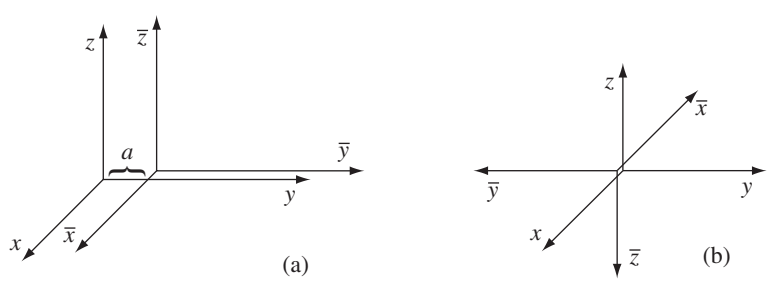


FIGURE 1.16

[4] A scalar does not change when you change coordinates. In particular, the components of a vector are *not* scalars, but the magnitude is.

[5] *Beware:* The vector **r** (Eq. 1.19) goes from a specific point in space (the origin, $\mathcal{O}$) to the point $P = (x, y, z)$. Under translations the *new* origin ($\bar{\mathcal{O}}$) is at a different location, and the arrow from $\bar{\mathcal{O}}$ to P is a completely different vector. The original vector **r** still goes from $\mathcal{O}$ to P, regardless of the coordinates used to label these points.

1.2 ■ DIFFERENTIAL CALCULUS

1.2.1 ■ "Ordinary" Derivatives

Suppose we have a function of one variable: $f(x)$. *Question:* What does the derivative, df/dx, do for us? *Answer:* It tells us how rapidly the function $f(x)$ varies when we change the argument x by a tiny amount, dx:

$$df = \left(\frac{df}{dx}\right) dx. \tag{1.33}$$

In words: If we increment x by an infinitesimal amount dx, then f changes by an amount df; the derivative is the proportionality factor. For example, in Fig. 1.17(a), the function varies slowly with x, and the derivative is correspondingly small. In Fig. 1.17(b), f increases rapidly with x, and the derivative is large, as you move away from $x = 0$.

Geometrical Interpretation: The derivative df/dx is the *slope* of the graph of f versus x.

1.2.2 ■ Gradient

Suppose, now, that we have a function of *three* variables—say, the temperature $T(x, y, z)$ in this room. (Start out in one corner, and set up a system of axes; then for each point (x, y, z) in the room, T gives the temperature at that spot.) We want to generalize the notion of "derivative" to functions like T, which depend not on *one* but on *three* variables.

A derivative is supposed to tell us how fast the function varies, if we move a little distance. But this time the situation is more complicated, because it depends on what *direction* we move: If we go straight up, then the temperature will probably increase fairly rapidly, but if we move horizontally, it may not change much at all. In fact, the question "How fast does T vary?" has an infinite number of answers, one for each direction we might choose to explore.

Fortunately, the problem is not as bad as it looks. A theorem on partial derivatives states that

$$dT = \left(\frac{\partial T}{\partial x}\right) dx + \left(\frac{\partial T}{\partial y}\right) dy + \left(\frac{\partial T}{\partial z}\right) dz. \tag{1.34}$$

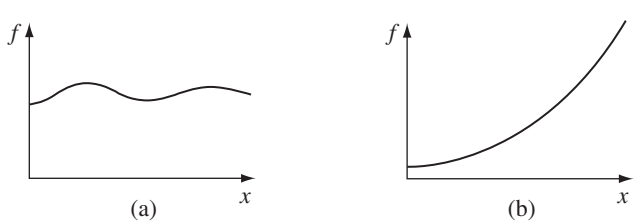


FIGURE 1.17

This tells us how T changes when we alter all three variables by the infinitesimal amounts dx, dy, dz. Notice that we do *not* require an infinite number of derivatives—*three* will suffice: the *partial* derivatives along each of the three coordinate directions.

Equation 1.34 is reminiscent of a dot product:

$$dT = \left(\frac{\partial T}{\partial x}\hat{\mathbf{x}} + \frac{\partial T}{\partial y}\hat{\mathbf{y}} + \frac{\partial T}{\partial z}\hat{\mathbf{z}}\right) \cdot (dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}})$$
$$= (\boldsymbol{\nabla} T) \cdot (d\mathbf{l}), \tag{1.35}$$

where

$$\boldsymbol{\nabla} T \equiv \frac{\partial T}{\partial x}\hat{\mathbf{x}} + \frac{\partial T}{\partial y}\hat{\mathbf{y}} + \frac{\partial T}{\partial z}\hat{\mathbf{z}} \tag{1.36}$$

is the **gradient** of T. Note that $\boldsymbol{\nabla} T$ is a *vector* quantity, with three components; it is the generalized derivative we have been looking for. Equation 1.35 is the three-dimensional version of Eq. 1.33.

Geometrical Interpretation of the Gradient: Like any vector, the gradient has *magnitude* and *direction*. To determine its geometrical meaning, let's rewrite the dot product (Eq. 1.35) using Eq. 1.1:

$$dT = \boldsymbol{\nabla} T \cdot d\mathbf{l} = |\boldsymbol{\nabla} T||d\mathbf{l}| \cos\theta, \tag{1.37}$$

where θ is the angle between $\boldsymbol{\nabla} T$ and $d\mathbf{l}$. Now, if we *fix* the *magnitude* $|d\mathbf{l}|$ and search around in various *directions* (that is, vary θ), the *maximum* change in T evidentally occurs when $\theta = 0$ (for then $\cos\theta = 1$). That is, for a fixed distance $|d\mathbf{l}|$, dT is greatest when I move in the *same direction* as $\boldsymbol{\nabla} T$. Thus:

> *The gradient $\boldsymbol{\nabla} T$ points in the direction of maximum increase of the function T.*

Moreover:

> *The magnitude $|\boldsymbol{\nabla} T|$ gives the slope (rate of increase) along this maximal direction.*

Imagine you are standing on a hillside. Look all around you, and find the direction of steepest ascent. That is the *direction* of the gradient. Now measure the *slope* in that direction (rise over run). That is the *magnitude* of the gradient. (Here the function we're talking about is the height of the hill, and the coordinates it depends on are positions—latitude and longitude, say. This function depends on only *two* variables, not *three*, but the geometrical meaning of the gradient is easier to grasp in two dimensions.) Notice from Eq. 1.37 that the direction of maximum *descent* is opposite to the direction of maximum *ascent*, while at right angles ($\theta = 90°$) the slope is zero (the gradient is perpendicular to the contour lines). You can conceive of surfaces that do not have these properties, but they always have "kinks" in them, and correspond to nondifferentiable functions.

What would it mean for the gradient to vanish? If $\boldsymbol{\nabla} T = \mathbf{0}$ at (x, y, z), then $dT = 0$ for small displacements about the point (x, y, z). This is, then, a **stationary point** of the function $T(x, y, z)$. It could be a maximum (a summit),

a minimum (a valley), a saddle point (a pass), or a "shoulder." This is analogous to the situation for functions of *one* variable, where a vanishing derivative signals a maximum, a minimum, or an inflection. In particular, if you want to locate the extrema of a function of three variables, set its gradient equal to zero.

Example 1.3. Find the gradient of $r = \sqrt{x^2 + y^2 + z^2}$ (the magnitude of the position vector).

Solution

$$\begin{aligned}\nabla r &= \frac{\partial r}{\partial x}\,\hat{\mathbf{x}} + \frac{\partial r}{\partial y}\,\hat{\mathbf{y}} + \frac{\partial r}{\partial z}\,\hat{\mathbf{z}} \\ &= \frac{1}{2}\frac{2x}{\sqrt{x^2+y^2+z^2}}\,\hat{\mathbf{x}} + \frac{1}{2}\frac{2y}{\sqrt{x^2+y^2+z^2}}\,\hat{\mathbf{y}} + \frac{1}{2}\frac{2z}{\sqrt{x^2+y^2+z^2}}\,\hat{\mathbf{z}} \\ &= \frac{x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}}{\sqrt{x^2+y^2+z^2}} = \frac{\mathbf{r}}{r} = \hat{\mathbf{r}}.\end{aligned}$$

Does this make sense? Well, it says that the distance from the origin increases most rapidly in the radial direction, and that its *rate* of increase in that direction is 1…just what you'd expect.

Problem 1.11 Find the gradients of the following functions:

(a) $f(x, y, z) = x^2 + y^3 + z^4$.

(b) $f(x, y, z) = x^2 y^3 z^4$.

(c) $f(x, y, z) = e^x \sin(y) \ln(z)$.

Problem 1.12 The height of a certain hill (in feet) is given by

$$h(x, y) = 10(2xy - 3x^2 - 4y^2 - 18x + 28y + 12),$$

where y is the distance (in miles) north, x the distance east of South Hadley.

(a) Where is the top of the hill located?

(b) How high is the hill?

(c) How steep is the slope (in feet per mile) at a point 1 mile north and one mile east of South Hadley? In what direction is the slope steepest, at that point?

• **Problem 1.13** Let $\boldsymbol{\imath}$ be the separation vector from a fixed point (x', y', z') to the point (x, y, z), and let $\imath$ be its length. Show that

(a) $\nabla(\imath^2) = 2\boldsymbol{\imath}$.

(b) $\nabla(1/\imath) = -\hat{\boldsymbol{\imath}}/\imath^2$.

(c) What is the *general* formula for $\nabla(\imath^n)$?

! **Problem 1.14** Suppose that f is a function of two variables (y and z) only. Show that the gradient $\nabla f = (\partial f/\partial y)\hat{\mathbf{y}} + (\partial f/\partial z)\hat{\mathbf{z}}$ transforms as a vector under rotations, Eq. 1.29. [*Hint:* $(\partial f/\partial \overline{y}) = (\partial f/\partial y)(\partial y/\partial \overline{y}) + (\partial f/\partial z)(\partial z/\partial \overline{y})$, and the analogous formula for $\partial f/\partial \overline{z}$. We know that $\overline{y} = y\cos\phi + z\sin\phi$ and $\overline{z} = -y\sin\phi + z\cos\phi$; "solve" these equations for y and z (as functions of $\overline{y}$ and $\overline{z}$), and compute the needed derivatives $\partial y/\partial \overline{y}$, $\partial z/\partial \overline{y}$, etc.]

1.2.3 ■ The Del Operator

The gradient has the formal appearance of a vector, ∇, "multiplying" a scalar T:

$$\nabla T = \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right) T. \tag{1.38}$$

(For once, I write the unit vectors to the *left,* just so no one will think this means $\partial\hat{\mathbf{x}}/\partial x$, and so on—which would be zero, since $\hat{\mathbf{x}}$ is constant.) The term in parentheses is called **del**:

$$\boxed{\nabla = \hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}.} \tag{1.39}$$

Of course, del is *not* a vector, in the usual sense. Indeed, it doesn't mean much until we provide it with a function to act upon. Furthermore, it does not "multiply" T; rather, it is an instruction to *differentiate* what follows. To be precise, then, we say that ∇ is a **vector operator** that *acts upon* T, not a vector that multiplies T.

With this qualification, though, ∇ mimics the behavior of an ordinary vector in virtually every way; almost anything that can be done with other vectors can also be done with ∇, if we merely translate "multiply" by "act upon." So by all means take the vector appearance of ∇ seriously: it is a marvelous piece of notational simplification, as you will appreciate if you ever consult Maxwell's original work on electromagnetism, written without the benefit of ∇.

Now, an ordinary vector $\mathbf{A}$ can multiply in three ways:

1. By a scalar a : $\mathbf{A}a$;
2. By a vector $\mathbf{B}$, via the dot product: $\mathbf{A}\cdot\mathbf{B}$;
3. By a vector $\mathbf{B}$ via the cross product: $\mathbf{A}\times\mathbf{B}$.

Correspondingly, there are three ways the operator ∇ can act:

1. On a scalar function T : ∇T (the gradient);
2. On a vector function $\mathbf{v}$, via the dot product: $\nabla\cdot\mathbf{v}$ (the **divergence**);
3. On a vector function $\mathbf{v}$, via the cross product: $\nabla\times\mathbf{v}$ (the **curl**).

We have already discussed the gradient. In the following sections we examine the other two vector derivatives: divergence and curl.

1.2.4 ■ The Divergence

From the definition of ∇ we construct the divergence:

$$\nabla \cdot \mathbf{v} = \left(\hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \right) \cdot (v_x \hat{\mathbf{x}} + v_y \hat{\mathbf{y}} + v_z \hat{\mathbf{z}})$$
$$= \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}. \tag{1.40}$$

Observe that the divergence of a vector function[6] $\mathbf{v}$ is itself a *scalar* $\nabla \cdot \mathbf{v}$.

Geometrical Interpretation: The name **divergence** is well chosen, for $\nabla \cdot \mathbf{v}$ is a measure of how much the vector $\mathbf{v}$ spreads out (diverges) from the point in question. For example, the vector function in Fig. 1.18a has a large (positive) divergence (if the arrows pointed *in*, it would be a *negative* divergence), the function in Fig. 1.18b has zero divergence, and the function in Fig. 1.18c again has a positive divergence. (Please understand that $\mathbf{v}$ here is a *function*—there's a different vector associated with every point in space. In the diagrams, of course, I can only draw the arrows at a few representative locations.)

Imagine standing at the edge of a pond. Sprinkle some sawdust or pine needles on the surface. If the material spreads out, then you dropped it at a point of positive divergence; if it collects together, you dropped it at a point of negative divergence. (The vector function $\mathbf{v}$ in this model is the velocity of the water at the surface—this is a *two*-dimensional example, but it helps give one a "feel" for what the divergence means. A point of positive divergence is a source, or "faucet"; a point of negative divergence is a sink, or "drain.")

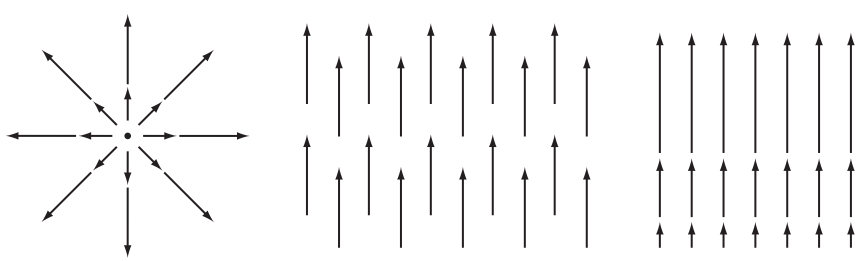


FIGURE 1.18

[6] A vector function $\mathbf{v}(x, y, z) = v_x(x, y, z)\,\hat{\mathbf{x}} + v_y(x, y, z)\,\hat{\mathbf{y}} + v_z(x, y, z)\,\hat{\mathbf{z}}$ is really *three* functions—one for each component. There's no such thing as the divergence of a scalar.

Example 1.4. Suppose the functions in Fig. 1.18 are $\mathbf{v}_a = \mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$, $\mathbf{v}_b = \hat{\mathbf{z}}$, and $\mathbf{v}_c = z\,\hat{\mathbf{z}}$. Calculate their divergences.

Solution

$$\boldsymbol{\nabla} \cdot \mathbf{v}_a = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) + \frac{\partial}{\partial z}(z) = 1 + 1 + 1 = 3.$$

As anticipated, this function has a positive divergence.

$$\boldsymbol{\nabla} \cdot \mathbf{v}_b = \frac{\partial}{\partial x}(0) + \frac{\partial}{\partial y}(0) + \frac{\partial}{\partial z}(1) = 0 + 0 + 0 = 0,$$

as expected.

$$\boldsymbol{\nabla} \cdot \mathbf{v}_c = \frac{\partial}{\partial x}(0) + \frac{\partial}{\partial y}(0) + \frac{\partial}{\partial z}(z) = 0 + 0 + 1 = 1.$$

Problem 1.15 Calculate the divergence of the following vector functions:

(a) $\mathbf{v}_a = x^2\,\hat{\mathbf{x}} + 3xz^2\,\hat{\mathbf{y}} - 2xz\,\hat{\mathbf{z}}$.

(b) $\mathbf{v}_b = xy\,\hat{\mathbf{x}} + 2yz\,\hat{\mathbf{y}} + 3zx\,\hat{\mathbf{z}}$.

(c) $\mathbf{v}_c = y^2\,\hat{\mathbf{x}} + (2xy + z^2)\,\hat{\mathbf{y}} + 2yz\,\hat{\mathbf{z}}$.

• **Problem 1.16** Sketch the vector function

$$\mathbf{v} = \frac{\hat{\mathbf{r}}}{r^2},$$

and compute its divergence. The answer may surprise you... can you explain it?

! **Problem 1.17** In two dimensions, show that the divergence transforms as a scalar under rotations. [*Hint:* Use Eq. 1.29 to determine $\overline{v}_y$ and $\overline{v}_z$, and the method of Prob. 1.14 to calculate the derivatives. Your aim is to show that $\partial\overline{v}_y/\partial\overline{y} + \partial\overline{v}_z/\partial\overline{z} = \partial v_y/\partial y + \partial v_z/\partial z$.]

1.2.5 ■ The Curl

From the definition of $\boldsymbol{\nabla}$ we construct the curl:

$$\boldsymbol{\nabla} \times \mathbf{v} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ v_x & v_y & v_z \end{vmatrix}$$

$$= \hat{\mathbf{x}}\left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) + \hat{\mathbf{y}}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) + \hat{\mathbf{z}}\left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right). \quad (1.41)$$

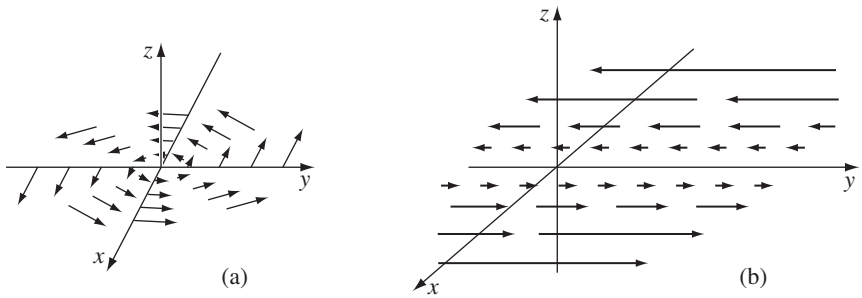


FIGURE 1.19

Notice that the curl of a vector function[7] $\mathbf{v}$ is, like any cross product, a *vector*.

Geometrical Interpretation: The name **curl** is also well chosen, for $\boldsymbol{\nabla} \times \mathbf{v}$ is a measure of how much the vector $\mathbf{v}$ swirls around the point in question. Thus the three functions in Fig. 1.18 all have zero curl (as you can easily check for yourself), whereas the functions in Fig. 1.19 have a substantial curl, pointing in the z direction, as the natural right-hand rule would suggest. Imagine (again) you are standing at the edge of a pond. Float a small paddlewheel (a cork with toothpicks pointing out radially would do); if it starts to rotate, then you placed it at a point of nonzero *curl.* A whirlpool would be a region of large curl.

Example 1.5. Suppose the function sketched in Fig. 1.19a is $\mathbf{v}_a = -y\hat{\mathbf{x}} + x\hat{\mathbf{y}}$, and that in Fig. 1.19b is $\mathbf{v}_b = x\hat{\mathbf{y}}$. Calculate their curls.

Solution

$$\boldsymbol{\nabla} \times \mathbf{v}_a = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ -y & x & 0 \end{vmatrix} = 2\hat{\mathbf{z}},$$

and

$$\boldsymbol{\nabla} \times \mathbf{v}_b = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ 0 & x & 0 \end{vmatrix} = \hat{\mathbf{z}}.$$

As expected, these curls point in the $+z$ direction. (Incidentally, they both have zero divergence, as you might guess from the pictures: nothing is "spreading out"... it just "swirls around.")

---

[7] There's no such thing as the curl of a scalar.

Problem 1.18 Calculate the curls of the vector functions in Prob. 1.15.

Problem 1.19 Draw a circle in the xy plane. At a few representative points draw the vector $\mathbf{v}$ tangent to the circle, pointing in the clockwise direction. By comparing adjacent vectors, determine the *sign* of $\partial v_x/\partial y$ and $\partial v_y/\partial x$. According to Eq. 1.41, then, what is the direction of $\nabla \times \mathbf{v}$? Explain how this example illustrates the geometrical interpretation of the curl.

Problem 1.20 Construct a vector function that has zero divergence and zero curl everywhere. (A *constant* will do the job, of course, but make it something a little more interesting than that!)

1.2.6 ■ Product Rules

The calculation of ordinary derivatives is facilitated by a number of rules, such as the sum rule:

$$\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx},$$

the rule for multiplying by a constant:

$$\frac{d}{dx}(kf) = k\frac{df}{dx},$$

the product rule:

$$\frac{d}{dx}(fg) = f\frac{dg}{dx} + g\frac{df}{dx},$$

and the quotient rule:

$$\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{g\dfrac{df}{dx} - f\dfrac{dg}{dx}}{g^2}.$$

Similar relations hold for the vector derivatives. Thus,

$$\nabla(f+g) = \nabla f + \nabla g, \qquad \nabla \cdot (\mathbf{A}+\mathbf{B}) = (\nabla \cdot \mathbf{A}) + (\nabla \cdot \mathbf{B}),$$

$$\nabla \times (\mathbf{A}+\mathbf{B}) = (\nabla \times \mathbf{A}) + (\nabla \times \mathbf{B}),$$

and

$$\nabla(kf) = k\nabla f, \qquad \nabla \cdot (k\mathbf{A}) = k(\nabla \cdot \mathbf{A}), \qquad \nabla \times (k\mathbf{A}) = k(\nabla \times \mathbf{A}),$$

as you can check for yourself. The product rules are not quite so simple. There are two ways to construct a scalar as the product of two functions:

fg (product of two scalar functions),
$\mathbf{A} \cdot \mathbf{B}$ (dot product of two vector functions),

and two ways to make a vector:

$$f\mathbf{A} \quad \text{(scalar times vector)},$$
$$\mathbf{A} \times \mathbf{B} \quad \text{(cross product of two vectors)}.$$

Accordingly, there are *six* product rules, two for gradients:

(i) $$\boldsymbol{\nabla}(fg) = f\boldsymbol{\nabla}g + g\boldsymbol{\nabla}f,$$

(ii) $$\boldsymbol{\nabla}(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\times(\boldsymbol{\nabla}\times\mathbf{B}) + \mathbf{B}\times(\boldsymbol{\nabla}\times\mathbf{A}) + (\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B} + (\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A},$$

two for divergences:

(iii) $$\boldsymbol{\nabla}\cdot(f\mathbf{A}) = f(\boldsymbol{\nabla}\cdot\mathbf{A}) + \mathbf{A}\cdot(\boldsymbol{\nabla}f),$$

(iv) $$\boldsymbol{\nabla}\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\boldsymbol{\nabla}\times\mathbf{A}) - \mathbf{A}\cdot(\boldsymbol{\nabla}\times\mathbf{B}),$$

and two for curls:

(v) $$\boldsymbol{\nabla}\times(f\mathbf{A}) = f(\boldsymbol{\nabla}\times\mathbf{A}) - \mathbf{A}\times(\boldsymbol{\nabla}f),$$

(vi) $$\boldsymbol{\nabla}\times(\mathbf{A}\times\mathbf{B}) = (\mathbf{B}\cdot\boldsymbol{\nabla})\mathbf{A} - (\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B} + \mathbf{A}(\boldsymbol{\nabla}\cdot\mathbf{B}) - \mathbf{B}(\boldsymbol{\nabla}\cdot\mathbf{A}).$$

You will be using these product rules so frequently that I have put them inside the front cover for easy reference. The proofs come straight from the product rule for ordinary derivatives. For instance,

$$\begin{aligned}\boldsymbol{\nabla}\cdot(f\mathbf{A}) &= \frac{\partial}{\partial x}(fA_x) + \frac{\partial}{\partial y}(fA_y) + \frac{\partial}{\partial z}(fA_z)\\ &= \left(\frac{\partial f}{\partial x}A_x + f\frac{\partial A_x}{\partial x}\right) + \left(\frac{\partial f}{\partial y}A_y + f\frac{\partial A_y}{\partial y}\right) + \left(\frac{\partial f}{\partial z}A_z + f\frac{\partial A_z}{\partial z}\right)\\ &= (\boldsymbol{\nabla}f)\cdot\mathbf{A} + f(\boldsymbol{\nabla}\cdot\mathbf{A}).\end{aligned}$$

It is also possible to formulate three quotient rules:

$$\boldsymbol{\nabla}\left(\frac{f}{g}\right) = \frac{g\boldsymbol{\nabla}f - f\boldsymbol{\nabla}g}{g^2},$$
$$\boldsymbol{\nabla}\cdot\left(\frac{\mathbf{A}}{g}\right) = \frac{g(\boldsymbol{\nabla}\cdot\mathbf{A}) - \mathbf{A}\cdot(\boldsymbol{\nabla}g)}{g^2},$$
$$\boldsymbol{\nabla}\times\left(\frac{\mathbf{A}}{g}\right) = \frac{g(\boldsymbol{\nabla}\times\mathbf{A}) + \mathbf{A}\times(\boldsymbol{\nabla}g)}{g^2}.$$

However, since these can be obtained quickly from the corresponding product rules, there is no point in listing them separately.

Problem 1.21 Prove product rules (i), (iv), and (v).

Problem 1.22

(a) If **A** and **B** are two vector functions, what does the expression $(\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B}$ mean? (That is, what are its x, y, and z components, in terms of the Cartesian components of **A**, **B**, and $\boldsymbol{\nabla}$?)

(b) Compute $(\hat{\mathbf{r}}\cdot\boldsymbol{\nabla})\hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the unit vector defined in Eq. 1.21.

(c) For the functions in Prob. 1.15, evaluate $(\mathbf{v}_a\cdot\boldsymbol{\nabla})\mathbf{v}_b$.

Problem 1.23 (For masochists only.) Prove product rules (ii) and (vi). Refer to Prob. 1.22 for the definition of $(\mathbf{A}\cdot\boldsymbol{\nabla})\mathbf{B}$.

Problem 1.24 Derive the three quotient rules.

Problem 1.25

(a) Check product rule (iv) (by calculating each term separately) for the functions

$$\mathbf{A} = x\,\hat{\mathbf{x}} + 2y\,\hat{\mathbf{y}} + 3z\,\hat{\mathbf{z}}; \qquad \mathbf{B} = 3y\,\hat{\mathbf{x}} - 2x\,\hat{\mathbf{y}}.$$

(b) Do the same for product rule (ii).

(c) Do the same for rule (vi).

1.2.7 ■ Second Derivatives

The gradient, the divergence, and the curl are the only first derivatives we can make with $\boldsymbol{\nabla}$; by applying $\boldsymbol{\nabla}$ *twice*, we can construct five species of *second* derivatives. The gradient $\boldsymbol{\nabla}T$ is a *vector*, so we can take the *divergence* and *curl* of it:

(1) Divergence of gradient: $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}T)$.

(2) Curl of gradient: $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}T)$.

The divergence $\boldsymbol{\nabla}\cdot\mathbf{v}$ is a *scalar*—all we can do is take its *gradient:*

(3) Gradient of divergence: $\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{v})$.

The curl $\boldsymbol{\nabla}\times\mathbf{v}$ is a *vector*, so we can take its *divergence* and *curl:*

(4) Divergence of curl: $\boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}\times\mathbf{v})$.

(5) Curl of curl: $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{v})$.

This exhausts the possibilities, and in fact not all of them give anything new. Let's consider them one at a time:

$$(1)\qquad \boldsymbol{\nabla}\cdot(\boldsymbol{\nabla}T) = \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right)\cdot\left(\frac{\partial T}{\partial x}\hat{\mathbf{x}} + \frac{\partial T}{\partial y}\hat{\mathbf{y}} + \frac{\partial T}{\partial z}\hat{\mathbf{z}}\right)$$
$$= \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}. \tag{1.42}$$

This object, which we write as $\nabla^2 T$ for short, is called the **Laplacian** of T; we shall be studying it in great detail later on. Notice that the Laplacian of a *scalar* T is a *scalar*. Occasionally, we shall speak of the Laplacian of a *vector,* $\nabla^2\mathbf{v}$. By this we mean a *vector* quantity whose x-component is the Laplacian of v_x, and so on:[8]

$$\nabla^2\mathbf{v} \equiv (\nabla^2 v_x)\hat{\mathbf{x}} + (\nabla^2 v_y)\hat{\mathbf{y}} + (\nabla^2 v_z)\hat{\mathbf{z}}. \tag{1.43}$$

This is nothing more than a convenient extension of the meaning of ∇^2.

(2) The curl of a gradient is always zero:

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} T) = \mathbf{0}. \tag{1.44}$$

This is an important fact, which we shall use repeatedly; you can easily prove it from the definition of $\boldsymbol{\nabla}$, Eq. 1.39. *Beware*: You might think Eq. 1.44 is "obviously" true—isn't it just $(\boldsymbol{\nabla} \times \boldsymbol{\nabla})T$, and isn't the cross product of *any* vector (in this case, $\boldsymbol{\nabla}$) with itself always zero? This reasoning is suggestive, but not quite conclusive, since $\boldsymbol{\nabla}$ is an *operator* and does not "multiply" in the usual way. The proof of Eq. 1.44, in fact, hinges on the equality of cross derivatives:

$$\frac{\partial}{\partial x}\left(\frac{\partial T}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial T}{\partial x}\right). \tag{1.45}$$

If you think I'm being fussy, test your intuition on this one:

$$(\boldsymbol{\nabla} T) \times (\boldsymbol{\nabla} S).$$

Is *that* always zero? (It *would* be, of course, if you replaced the $\boldsymbol{\nabla}$'s by an ordinary vector.)

(3) $\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{v})$ seldom occurs in physical applications, and it has not been given any special name of its own—it's just **the gradient of the divergence.** Notice that $\boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{v})$ is *not* the same as the Laplacian of a vector: $\nabla^2\mathbf{v} = (\boldsymbol{\nabla} \cdot \boldsymbol{\nabla})\mathbf{v} \neq \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{v})$.

(4) The divergence of a curl, like the curl of a gradient, is always zero:

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{v}) = 0. \tag{1.46}$$

You can prove this for yourself. (Again, there is a fraudulent short-cut proof, using the vector identity $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}$.)

(5) As you can check from the definition of $\boldsymbol{\nabla}$:

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{v}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{v}) - \nabla^2\mathbf{v}. \tag{1.47}$$

So curl-of-curl gives nothing new; the first term is just number (3), and the second is the Laplacian (of a vector). (In fact, Eq. 1.47 is often used to *define* the

[8] In curvilinear coordinates, where the unit vectors themselves depend on position, they too must be differentiated (see Sect. 1.4.1).

Laplacian of a vector, in preference to Eq. 1.43, which makes explicit reference to Cartesian coordinates.)

Really, then, there are just two kinds of second derivatives: the Laplacian (which is of fundamental importance) and the gradient-of-divergence (which we seldom encounter). We could go through a similar ritual to work out *third* derivatives, but fortunately second derivatives suffice for practically all physical applications.

A final word on vector differential calculus: It *all* flows from the operator ∇, and from taking seriously its vectorial character. Even if you remembered *only* the definition of ∇, you could easily reconstruct all the rest.

Problem 1.26 Calculate the Laplacian of the following functions:

(a) $T_a = x^2 + 2xy + 3z + 4$.

(b) $T_b = \sin x \sin y \sin z$.

(c) $T_c = e^{-5x} \sin 4y \cos 3z$.

(d) $\mathbf{v} = x^2\,\hat{\mathbf{x}} + 3xz^2\,\hat{\mathbf{y}} - 2xz\,\hat{\mathbf{z}}$.

Problem 1.27 Prove that the divergence of a curl is always zero. *Check* it for function $\mathbf{v}_a$ in Prob. 1.15.

Problem 1.28 Prove that the curl of a gradient is always zero. *Check* it for function (b) in Prob. 1.11.

1.3 ■ INTEGRAL CALCULUS

1.3.1 ■ Line, Surface, and Volume Integrals

In electrodynamics, we encounter several different kinds of integrals, among which the most important are **line** (or **path**) **integrals**, **surface integrals** (or **flux**), and **volume integrals**.

(a) **Line Integrals.** A line integral is an expression of the form

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{v} \cdot d\mathbf{l}, \tag{1.48}$$

where $\mathbf{v}$ is a vector function, $d\mathbf{l}$ is the infinitesimal displacement vector (Eq. 1.22), and the integral is to be carried out along a prescribed path $\mathcal{P}$ from point $\mathbf{a}$ to point $\mathbf{b}$ (Fig. 1.20). If the path in question forms a closed loop (that is, if $\mathbf{b} = \mathbf{a}$), I shall put a circle on the integral sign:

$$\oint \mathbf{v} \cdot d\mathbf{l}. \tag{1.49}$$

At each point on the path, we take the dot product of $\mathbf{v}$ (evaluated at that point) with the displacement $d\mathbf{l}$ to the next point on the path. To a physicist, the most familiar example of a line integral is the work done by a force $\mathbf{F}$: $W = \int \mathbf{F} \cdot d\mathbf{l}$.

Ordinarily, the value of a line integral depends critically on the path taken from $\mathbf{a}$ to $\mathbf{b}$, but there is an important special class of vector functions for which the line

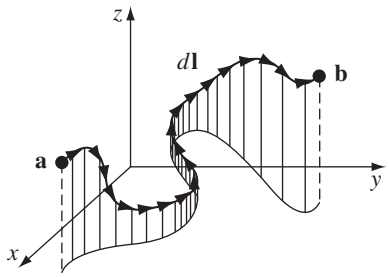


FIGURE 1.20

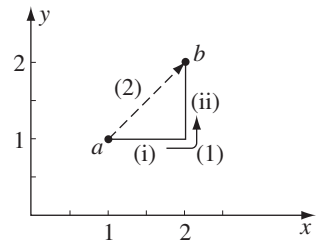


FIGURE 1.21

integral is *independent* of path and is determined entirely by the end points. It will be our business in due course to characterize this special class of vectors. (A *force* that has this property is called **conservative**.)

Example 1.6. Calculate the line integral of the function $\mathbf{v} = y^2\,\hat{\mathbf{x}} + 2x(y+1)\,\hat{\mathbf{y}}$ from the point $\mathbf{a} = (1, 1, 0)$ to the point $\mathbf{b} = (2, 2, 0)$, along the paths (1) and (2) in Fig. 1.21. What is $\oint \mathbf{v} \cdot d\mathbf{l}$ for the loop that goes from **a** to **b** along (1) and returns to **a** along (2)?

Solution

As always, $d\mathbf{l} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}$. Path (1) consists of two parts. Along the "horizontal" segment, $dy = dz = 0$, so

$$\text{(i)}\ \ d\mathbf{l} = dx\,\hat{\mathbf{x}},\ \ y = 1,\ \ \mathbf{v} \cdot d\mathbf{l} = y^2\,dx = dx,\ \text{so}\ \int \mathbf{v} \cdot d\mathbf{l} = \int_1^2 dx = 1.$$

On the "vertical" stretch, $dx = dz = 0$, so

$$\text{(ii)}\ \ d\mathbf{l} = dy\,\hat{\mathbf{y}},\ \ x = 2,\ \ \mathbf{v} \cdot d\mathbf{l} = 2x(y+1)\,dy = 4(y+1)\,dy,\ \text{so}$$

$$\int \mathbf{v} \cdot d\mathbf{l} = 4\int_1^2 (y+1)\,dy = 10.$$

By path (1), then,

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{v} \cdot d\mathbf{l} = 1 + 10 = 11.$$

Meanwhile, on path (2) $x = y$, $dx = dy$, and $dz = 0$, so
$d\mathbf{l} = dx\,\hat{\mathbf{x}} + dx\,\hat{\mathbf{y}}$, $\mathbf{v} \cdot d\mathbf{l} = x^2\,dx + 2x(x+1)\,dx = (3x^2 + 2x)\,dx$,
and

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{v} \cdot d\mathbf{l} = \int_1^2 (3x^2 + 2x)\,dx = (x^3 + x^2)\Big|_1^2 = 10.$$

(The strategy here is to get everything in terms of one variable; I could just as well have eliminated x in favor of y.)

For the loop that goes *out* (1) and *back* (2), then,

$$\oint \mathbf{v} \cdot d\mathbf{l} = 11 - 10 = 1.$$

(b) **Surface Integrals.** A surface integral is an expression of the form

$$\int_{\mathcal{S}} \mathbf{v} \cdot d\mathbf{a}, \tag{1.50}$$

where $\mathbf{v}$ is again some vector function, and the integral is over a specified surface $\mathcal{S}$. Here $d\mathbf{a}$ is an infinitesimal patch of area, with direction perpendicular to the surface (Fig. 1.22). There are, of course, *two* directions perpendicular to any surface, so the *sign* of a surface integral is intrinsically ambiguous. If the surface is *closed* (forming a "balloon"), in which case I shall again put a circle on the integral sign

$$\oint \mathbf{v} \cdot d\mathbf{a},$$

then tradition dictates that "outward" is positive, but for open surfaces it's arbitrary. If $\mathbf{v}$ describes the flow of a fluid (mass per unit area per unit time), then $\int \mathbf{v} \cdot d\mathbf{a}$ represents the total mass per unit time passing through the surface—hence the alternative name, "flux."

Ordinarily, the value of a surface integral depends on the particular surface chosen, but there is a special class of vector functions for which it is *independent* of the surface and is determined entirely by the boundary line. An important task will be to characterize this special class of functions.

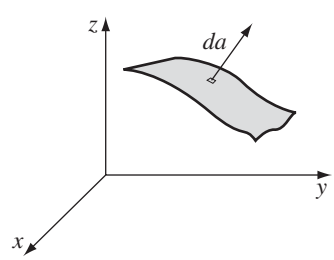


FIGURE 1.22

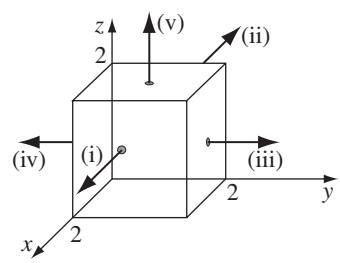


FIGURE 1.23

Example 1.7. Calculate the surface integral of $\mathbf{v} = 2xz\,\hat{\mathbf{x}} + (x+2)\,\hat{\mathbf{y}} + y(z^2-3)\,\hat{\mathbf{z}}$ over five sides (excluding the bottom) of the cubical box (side 2) in Fig. 1.23. Let "upward and outward" be the positive direction, as indicated by the arrows.

Solution

Taking the sides one at a time:

(i) $x = 2,\ d\mathbf{a} = dy\,dz\,\hat{\mathbf{x}},\ \mathbf{v}\cdot d\mathbf{a} = 2xz\,dy\,dz = 4z\,dy\,dz$, so

$$\int \mathbf{v}\cdot d\mathbf{a} = 4\int_0^2 dy\int_0^2 z\,dz = 16.$$

(ii) $x = 0,\ d\mathbf{a} = -dy\,dz\,\hat{\mathbf{x}},\ \mathbf{v}\cdot d\mathbf{a} = -2xz\,dy\,dz = 0$, so

$$\int \mathbf{v}\cdot d\mathbf{a} = 0.$$

(iii) $y = 2,\ d\mathbf{a} = dx\,dz\,\hat{\mathbf{y}},\ \mathbf{v}\cdot d\mathbf{a} = (x+2)\,dx\,dz$, so

$$\int \mathbf{v}\cdot d\mathbf{a} = \int_0^2 (x+2)\,dx\int_0^2 dz = 12.$$

(iv) $y = 0,\ d\mathbf{a} = -dx\,dz\,\hat{\mathbf{y}},\ \mathbf{v}\cdot d\mathbf{a} = -(x+2)\,dx\,dz$, so

$$\int \mathbf{v}\cdot d\mathbf{a} = -\int_0^2 (x+2)\,dx\int_0^2 dz = -12.$$

(v) $z = 2,\ d\mathbf{a} = dx\,dy\,\hat{\mathbf{z}},\ \mathbf{v}\cdot d\mathbf{a} = y(z^2-3)\,dx\,dy = y\,dx\,dy$, so

$$\int \mathbf{v}\cdot d\mathbf{a} = \int_0^2 dx\int_0^2 y\,dy = 4.$$

The *total* flux is

$$\int_{\text{surface}} \mathbf{v}\cdot d\mathbf{a} = 16 + 0 + 12 - 12 + 4 = 20.$$

(c) **Volume Integrals.** A volume integral is an expression of the form

$$\int_{\mathcal{V}} T\,d\tau, \tag{1.51}$$

where T is a scalar function and $d\tau$ is an infinitesimal volume element. In Cartesian coordinates,

$$d\tau = dx\,dy\,dz. \tag{1.52}$$

For example, if T is the density of a substance (which might vary from point to point), then the volume integral would give the total mass. Occasionally we shall encounter volume integrals of *vector* functions:

$$\int \mathbf{v}\,d\tau = \int (v_x\,\hat{\mathbf{x}} + v_y\,\hat{\mathbf{y}} + v_z\,\hat{\mathbf{z}})d\tau = \hat{\mathbf{x}}\int v_x d\tau + \hat{\mathbf{y}}\int v_y d\tau + \hat{\mathbf{z}}\int v_z d\tau; \tag{1.53}$$

because the unit vectors ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$) are constants, they come outside the integral.

Example 1.8. Calculate the volume integral of $T = xyz^2$ over the prism in Fig. 1.24.

Solution

You can do the three integrals in any order. Let's do x first: it runs from 0 to $(1 - y)$, then y (it goes from 0 to 1), and finally z (0 to 3):

$$\int T\, d\tau = \int_0^3 z^2 \left\{ \int_0^1 y \left[\int_0^{1-y} x\, dx \right] dy \right\} dz$$

$$= \frac{1}{2} \int_0^3 z^2\, dz \int_0^1 (1-y)^2 y\, dy = \frac{1}{2}\,(9)\left(\frac{1}{12}\right) = \frac{3}{8}.$$

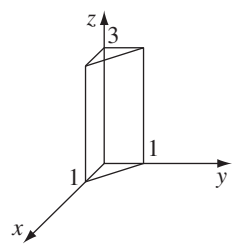


FIGURE 1.24

Problem 1.29 Calculate the line integral of the function $\mathbf{v} = x^2\,\hat{\mathbf{x}} + 2yz\,\hat{\mathbf{y}} + y^2\,\hat{\mathbf{z}}$ from the origin to the point (1,1,1) by three different routes:

(a) $(0, 0, 0) \to (1, 0, 0) \to (1, 1, 0) \to (1, 1, 1)$.

(b) $(0, 0, 0) \to (0, 0, 1) \to (0, 1, 1) \to (1, 1, 1)$.

(c) The direct straight line.

(d) What is the line integral around the closed loop that goes *out* along path (a) and *back* along path (b)?

Problem 1.30 Calculate the surface integral of the function in Ex. 1.7, over the *bottom* of the box. For consistency, let "upward" be the positive direction. Does the surface integral depend only on the boundary line for this function? What is the total flux over the *closed* surface of the box (*including* the bottom)? [*Note:* For the *closed* surface, the positive direction is "outward," and hence "down," for the bottom face.]

Problem 1.31 Calculate the volume integral of the function $T = z^2$ over the tetrahedron with corners at (0,0,0), (1,0,0), (0,1,0), and (0,0,1).

1.3.2 ■ The Fundamental Theorem of Calculus

Suppose $f(x)$ is a function of one variable. The **fundamental theorem of calculus** says:

$$\int_a^b \left(\frac{df}{dx}\right) dx = f(b) - f(a). \tag{1.54}$$

In case this doesn't look familiar, I'll write it another way:

$$\int_a^b F(x)\, dx = f(b) - f(a),$$

where $df/dx = F(x)$. The fundamental theorem tells you how to integrate $F(x)$: you think up a function $f(x)$ whose *derivative* is equal to F.

Geometrical Interpretation: According to Eq. 1.33, $df = (df/dx)dx$ is the infinitesimal change in f when you go from (x) to $(x + dx)$. The fundamental theorem (Eq. 1.54) says that if you chop the interval from a to b (Fig. 1.25) into many tiny pieces, dx, and add up the increments df from each little piece, the result is (not surprisingly) equal to the total change in f: $f(b) - f(a)$. In other words, there are two ways to determine the total change in the function: *either* subtract the values at the ends *or* go step-by-step, adding up all the tiny increments as you go. You'll get the same answer either way.

Notice the basic format of the fundamental theorem: the *integral* of a *derivative* over some *region* is given by the *value of the function* at the end points *(boundaries)*. In vector calculus there are three species of derivative (gradient, divergence, and curl), and each has its own "fundamental theorem," with essentially the same format. I don't plan to prove these theorems here; rather, I will explain what they *mean*, and try to make them *plausible*. Proofs are given in Appendix A.

1.3.3 ■ The Fundamental Theorem for Gradients

Suppose we have a scalar function of three variables $T(x, y, z)$. Starting at point **a**, we move a small distance $d\mathbf{l}_1$ (Fig. 1.26). According to Eq. 1.37, the function T will change by an amount

$$dT = (\nabla T) \cdot d\mathbf{l}_1.$$

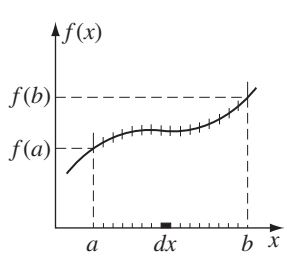


FIGURE 1.25

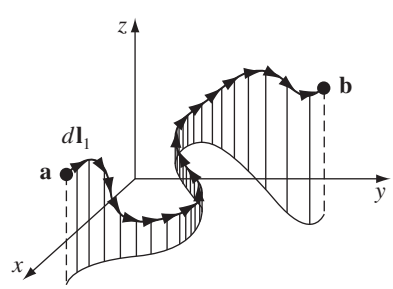


FIGURE 1.26

Now we move a little further, by an additional small displacement $d\mathbf{l}_2$; the incremental change in T will be $(\nabla T) \cdot d\mathbf{l}_2$. In this manner, proceeding by infinitesimal steps, we make the journey to point **b**. At each step we compute the gradient of T (at that point) and dot it into the displacement $d\mathbf{l}$. . . this gives us the change in T. Evidently the *total* change in T in going from **a** to **b** (along the path selected) is

$$\boxed{\int_{\mathbf{a}}^{\mathbf{b}} (\nabla T) \cdot d\mathbf{l} = T(\mathbf{b}) - T(\mathbf{a}).} \tag{1.55}$$

This is the **fundamental theorem for gradients**; like the "ordinary" fundamental theorem, it says that the integral (here a *line* integral) of a derivative (here the *gradient*) is given by the value of the function at the boundaries (**a** and **b**).

Geometrical Interpretation: Suppose you wanted to determine the height of the Eiffel Tower. You could climb the stairs, using a ruler to measure the rise at each step, and adding them all up (that's the left side of Eq. 1.55), or you could place altimeters at the top and the bottom, and subtract the two readings (that's the right side); you should get the same answer either way (that's the fundamental theorem).

Incidentally, as we found in Ex. 1.6, line integrals ordinarily depend on the *path* taken from **a** to **b**. But the *right* side of Eq. 1.55 makes no reference to the path—only to the end points. Evidently, *gradients* have the special property that their line integrals are path independent:

Corollary 1: $\int_{\mathbf{a}}^{\mathbf{b}} (\nabla T) \cdot d\mathbf{l}$ is independent of the path taken from **a** to **b**.

Corollary 2: $\oint (\nabla T) \cdot d\mathbf{l} = 0$, since the beginning and end points are identical, and hence $T(\mathbf{b}) - T(\mathbf{a}) = 0$.

Example 1.9. Let $T = xy^2$, and take point **a** to be the origin (0, 0, 0) and **b** the point (2, 1, 0). Check the fundamental theorem for gradients.

Solution

Although the integral is independent of path, we must *pick* a specific path in order to evaluate it. Let's go out along the x axis (step i) and then up (step ii) (Fig. 1.27). As always, $d\mathbf{l} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}$; $\nabla T = y^2\,\hat{\mathbf{x}} + 2xy\,\hat{\mathbf{y}}$.

(i) $y = 0$; $d\mathbf{l} = dx\,\hat{\mathbf{x}}$, $\nabla T \cdot d\mathbf{l} = y^2\,dx = 0$, so

$$\int_{\mathrm{i}} \nabla T \cdot d\mathbf{l} = 0.$$

(ii) $x = 2$; $d\mathbf{l} = dy\,\hat{\mathbf{y}}$, $\nabla T \cdot d\mathbf{l} = 2xy\,dy = 4y\,dy$, so

$$\int_{\mathrm{ii}} \nabla T \cdot d\mathbf{l} = \int_0^1 4y\,dy = 2y^2\Big|_0^1 = 2.$$

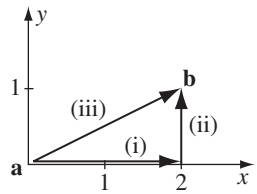


FIGURE 1.27

The total line integral is 2. Is this consistent with the fundamental theorem? Yes: $T(\mathbf{b}) - T(\mathbf{a}) = 2 - 0 = 2$.

Now, just to convince you that the answer is independent of path, let me calculate the same integral along path iii (the straight line from **a** to **b**):

(iii) $y = \frac{1}{2}x,\ dy = \frac{1}{2}\,dx,\ \nabla T \cdot d\mathbf{l} = y^2\,dx + 2xy\,dy = \frac{3}{4}x^2\,dx$, so

$$\int_{\text{iii}} \nabla T \cdot d\mathbf{l} = \int_0^2 \tfrac{3}{4}x^2\,dx = \tfrac{1}{4}x^3\Big|_0^2 = 2.$$

Problem 1.32 Check the fundamental theorem for gradients, using $T = x^2 + 4xy + 2yz^3$, the points $\mathbf{a} = (0, 0, 0)$, $\mathbf{b} = (1, 1, 1)$, and the three paths in Fig. 1.28:

(a) $(0, 0, 0) \to (1, 0, 0) \to (1, 1, 0) \to (1, 1, 1)$;

(b) $(0, 0, 0) \to (0, 0, 1) \to (0, 1, 1) \to (1, 1, 1)$;

(c) the parabolic path $z = x^2$; $y = x$.

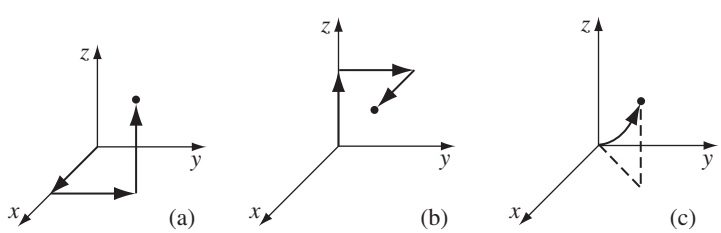


FIGURE 1.28

1.3.4 ■ The Fundamental Theorem for Divergences

The fundamental theorem for divergences states that:

$$\boxed{\int_{\mathcal{V}} (\nabla \cdot \mathbf{v})\,d\tau = \oint_{\mathcal{S}} \mathbf{v} \cdot d\mathbf{a}.} \tag{1.56}$$

In honor, I suppose, of its great importance, this theorem has at least three special names: **Gauss's theorem**, **Green's theorem**, or simply the **divergence theorem**. Like the other "fundamental theorems," it says that the *integral* of a *derivative* (in this case the *divergence*) over a *region* (in this case a *volume*, $\mathcal{V}$) is equal to the value of the function at the *boundary* (in this case the *surface* $\mathcal{S}$ that bounds the volume). Notice that the boundary term is itself an integral (specifically, a surface integral). This is reasonable: the "boundary" of a *line* is just two end points, but the boundary of a *volume* is a (closed) surface.

Geometrical Interpretation: If $\mathbf{v}$ represents the flow of an incompressible fluid, then the *flux* of $\mathbf{v}$ (the right side of Eq. 1.56) is the total amount of fluid passing out through the surface, per unit time. Now, the divergence measures the "spreading out" of the vectors from a point—a place of high divergence is like a "faucet," pouring out liquid. If we have a bunch of faucets in a region filled with incompressible fluid, an equal amount of liquid will be forced out through the boundaries of the region. In fact, there are *two* ways we could determine how much is being produced: (a) we could count up all the faucets, recording how much each puts out, or (b) we could go around the boundary, measuring the flow at each point, and add it all up. You get the same answer either way:

$$\int (\text{faucets within the volume}) = \oint (\text{flow out through the surface}).$$

This, in essence, is what the divergence theorem says.

Example 1.10. Check the divergence theorem using the function

$$\mathbf{v} = y^2\,\hat{\mathbf{x}} + (2xy + z^2)\,\hat{\mathbf{y}} + (2yz)\,\hat{\mathbf{z}}$$

and a unit cube at the origin (Fig. 1.29).

Solution
In this case

$$\boldsymbol{\nabla} \cdot \mathbf{v} = 2(x + y),$$

and

$$\int_{\mathcal{V}} 2(x + y)\,d\tau = 2\int_0^1 \int_0^1 \int_0^1 (x + y)\,dx\,dy\,dz,$$

$$\int_0^1 (x + y)\,dx = \tfrac{1}{2} + y, \quad \int_0^1 (\tfrac{1}{2} + y)\,dy = 1, \quad \int_0^1 1\,dz = 1.$$

Thus,

$$\int_{\mathcal{V}} \boldsymbol{\nabla} \cdot \mathbf{v}\,d\tau = 2.$$

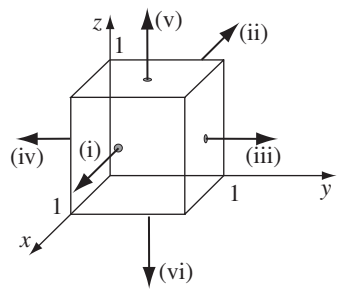


FIGURE 1.29

So much for the left side of the divergence theorem. To evaluate the surface integral we must consider separately the six faces of the cube:

(i) $$\int \mathbf{v} \cdot d\mathbf{a} = \int_0^1 \int_0^1 y^2 dy\, dz = \tfrac{1}{3}.$$

(ii) $$\int \mathbf{v} \cdot d\mathbf{a} = -\int_0^1 \int_0^1 y^2\, dy\, dz = -\tfrac{1}{3}.$$

(iii) $$\int \mathbf{v} \cdot d\mathbf{a} = \int_0^1 \int_0^1 (2x + z^2)\, dx\, dz = \tfrac{4}{3}.$$

(iv) $$\int \mathbf{v} \cdot d\mathbf{a} = -\int_0^1 \int_0^1 z^2\, dx\, dz = -\tfrac{1}{3}.$$

(v) $$\int \mathbf{v} \cdot d\mathbf{a} = \int_0^1 \int_0^1 2y\, dx\, dy = 1.$$

(vi) $$\int \mathbf{v} \cdot d\mathbf{a} = -\int_0^1 \int_0^1 0\, dx\, dy = 0.$$

So the total flux is:

$$\oint_{\mathcal{S}} \mathbf{v} \cdot d\mathbf{a} = \tfrac{1}{3} - \tfrac{1}{3} + \tfrac{4}{3} - \tfrac{1}{3} + 1 + 0 = 2,$$

as expected.

Problem 1.33 Test the divergence theorem for the function $\mathbf{v} = (xy)\,\hat{\mathbf{x}} + (2yz)\,\hat{\mathbf{y}} + (3zx)\,\hat{\mathbf{z}}$. Take as your volume the cube shown in Fig. 1.30, with sides of length 2.

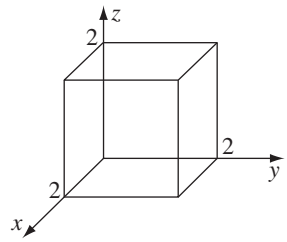


FIGURE 1.30

1.3.5 ■ The Fundamental Theorem for Curls

The fundamental theorem for curls, which goes by the special name of **Stokes' theorem**, states that

$$\int_{\mathcal{S}} (\nabla \times \mathbf{v}) \cdot d\mathbf{a} = \oint_{\mathcal{P}} \mathbf{v} \cdot d\mathbf{l}. \tag{1.57}$$

As always, the *integral* of a *derivative* (here, the *curl*) over a *region* (here, a patch of *surface*, $\mathcal{S}$) is equal to the value of the function at the *boundary* (here, the perimeter of the patch, $\mathcal{P}$). As in the case of the divergence theorem, the boundary term is itself an integral—specifically, a closed line integral.

Geometrical Interpretation: Recall that the curl measures the "twist" of the vectors $\mathbf{v}$; a region of high curl is a whirlpool—if you put a tiny paddle wheel there, it will rotate. Now, the integral of the curl over some surface (or, more precisely, the *flux* of the curl *through* that surface) represents the "total amount of swirl," and we can determine that just as well by going around the edge and finding how much the flow is following the boundary (Fig. 1.31). Indeed, $\oint \mathbf{v} \cdot d\mathbf{l}$ is sometimes called the **circulation** of $\mathbf{v}$.

You may have noticed an apparent ambiguity in Stokes' theorem: concerning the boundary line integral, which *way* are we supposed to go around (clockwise or counterclockwise)? If we go the "wrong" way, we'll pick up an overall sign error. The answer is that it doesn't matter which way you go as long as you are consistent, for there is a compensating sign ambiguity in the surface integral: Which way does $d\mathbf{a}$ point? For a *closed* surface (as in the divergence theorem), $d\mathbf{a}$ points in the direction of the *outward* normal; but for an *open* surface, which way is "out"? Consistency in Stokes' theorem (as in all such matters) is given by the right-hand rule: if your fingers point in the direction of the line integral, then your thumb fixes the direction of $d\mathbf{a}$ (Fig. 1.32).

Now, there are plenty of surfaces (infinitely many) that share any given boundary line. Twist a paper clip into a loop, and dip it in soapy water. The soap film constitutes a surface, with the wire loop as its boundary. If you blow on it, the soap film will expand, making a larger surface, with the same boundary. Ordinarily, a flux integral depends critically on what surface you integrate over, but evidently

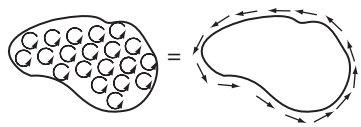


FIGURE 1.31

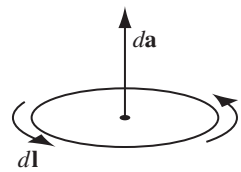


FIGURE 1.32

this is *not* the case with curls. For Stokes' theorem says that $\int (\nabla \times \mathbf{v}) \cdot d\mathbf{a}$ is equal to the line integral of $\mathbf{v}$ around the boundary, and the latter makes no reference to the specific surface you choose.

Corollary 1: $\int (\nabla \times \mathbf{v}) \cdot d\mathbf{a}$ depends only on the boundary line, not on the particular surface used.

Corollary 2: $\oint (\nabla \times \mathbf{v}) \cdot d\mathbf{a} = 0$ for any closed surface, since the boundary line, like the mouth of a balloon, shrinks down to a point, and hence the right side of Eq. 1.57 vanishes.

These corollaries are analogous to those for the gradient theorem. We will develop the parallel further in due course.

Example 1.11. Suppose $\mathbf{v} = (2xz + 3y^2)\hat{\mathbf{y}} + (4yz^2)\hat{\mathbf{z}}$. Check Stokes' theorem for the square surface shown in Fig. 1.33.

Solution
Here

$$\nabla \times \mathbf{v} = (4z^2 - 2x)\,\hat{\mathbf{x}} + 2z\,\hat{\mathbf{z}} \quad \text{and} \quad d\mathbf{a} = dy\,dz\,\hat{\mathbf{x}}.$$

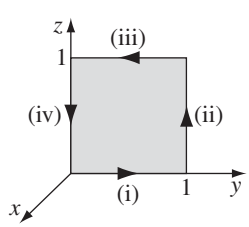


FIGURE 1.33

(In saying that $d\mathbf{a}$ points in the x direction, we are committing ourselves to a counterclockwise line integral. We could as well write $d\mathbf{a} = -dy\,dz\,\hat{\mathbf{x}}$, but then we would be obliged to go clockwise.) Since $x = 0$ for this surface,

$$\int (\nabla \times \mathbf{v}) \cdot d\mathbf{a} = \int_0^1 \int_0^1 4z^2\,dy\,dz = \frac{4}{3}.$$

Now, what about the line integral? We must break this up into four segments:

(i) $x = 0, \quad z = 0, \quad \mathbf{v} \cdot d\mathbf{l} = 3y^2\, dy, \quad \int \mathbf{v} \cdot d\mathbf{l} = \int_0^1 3y^2\, dy = 1,$

(ii) $x = 0, \quad y = 1, \quad \mathbf{v} \cdot d\mathbf{l} = 4z^2\, dz, \quad \int \mathbf{v} \cdot d\mathbf{l} = \int_0^1 4z^2\, dz = \frac{4}{3},$

(iii) $x = 0, \quad z = 1, \quad \mathbf{v} \cdot d\mathbf{l} = 3y^2\, dy, \quad \int \mathbf{v} \cdot d\mathbf{l} = \int_1^0 3y^2\, dy = -1,$

(iv) $x = 0, \quad y = 0, \quad \mathbf{v} \cdot d\mathbf{l} = 0, \quad \int \mathbf{v} \cdot d\mathbf{l} = \int_1^0 0\, dz = 0.$

So

$$\oint \mathbf{v} \cdot d\mathbf{l} = 1 + \frac{4}{3} - 1 + 0 = \frac{4}{3}.$$

It checks.

A point of strategy: notice how I handled step (iii). There is a temptation to write $d\mathbf{l} = -dy\, \hat{\mathbf{y}}$ here, since the path goes to the left. You can get away with this, if you absolutely insist, by running the integral from $0 \to 1$. But it is much safer to say $d\mathbf{l} = dx\, \hat{\mathbf{x}} + dy\, \hat{\mathbf{y}} + dz\, \hat{\mathbf{z}}$ *always* (never any minus signs) and let the limits of the integral take care of the direction.

Problem 1.34 Test Stokes' theorem for the function $\mathbf{v} = (xy)\, \hat{\mathbf{x}} + (2yz)\, \hat{\mathbf{y}} + (3zx)\, \hat{\mathbf{z}}$, using the triangular shaded area of Fig. 1.34.

Problem 1.35 Check Corollary 1 by using the same function and boundary line as in Ex. 1.11, but integrating over the five faces of the cube in Fig. 1.35. The back of the cube is open.

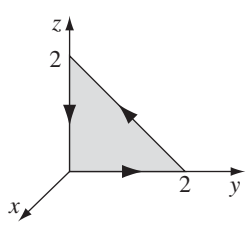


FIGURE 1.34

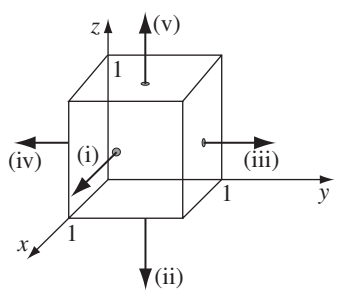


FIGURE 1.35

1.3.6 ■ Integration by Parts

The technique known (awkwardly) as **integration by parts** exploits the product rule for derivatives:

$$\frac{d}{dx}(fg) = f\left(\frac{dg}{dx}\right) + g\left(\frac{df}{dx}\right).$$

Integrating both sides, and invoking the fundamental theorem:

$$\int_a^b \frac{d}{dx}(fg)\,dx = fg\Big|_a^b = \int_a^b f\left(\frac{dg}{dx}\right)dx + \int_a^b g\left(\frac{df}{dx}\right)dx,$$

or

$$\int_a^b f\left(\frac{dg}{dx}\right)dx = -\int_a^b g\left(\frac{df}{dx}\right)dx + fg\Big|_a^b. \tag{1.58}$$

That's integration by parts. It applies to the situation in which you are called upon to integrate the product of one function (f) and the *derivative* of another (g); it says you can *transfer the derivative from g to f*, at the cost of a minus sign and a boundary term.

Example 1.12. Evaluate the integral

$$\int_0^\infty xe^{-x}\,dx.$$

Solution
The exponential can be expressed as a derivative:

$$e^{-x} = \frac{d}{dx}\left(-e^{-x}\right);$$

in this case, then, $f(x) = x$, $g(x) = -e^{-x}$, and $df/dx = 1$, so

$$\int_0^\infty xe^{-x}\,dx = \int_0^\infty e^{-x}\,dx - xe^{-x}\Big|_0^\infty = -e^{-x}\Big|_0^\infty = 1.$$

We can exploit the product rules of vector calculus, together with the appropriate fundamental theorems, in exactly the same way. For example, integrating

$$\boldsymbol{\nabla}\cdot(f\mathbf{A}) = f(\boldsymbol{\nabla}\cdot\mathbf{A}) + \mathbf{A}\cdot(\boldsymbol{\nabla}f)$$

over a volume, and invoking the divergence theorem, yields

$$\int \boldsymbol{\nabla}\cdot(f\mathbf{A})\,d\tau = \int f(\boldsymbol{\nabla}\cdot\mathbf{A})\,d\tau + \int \mathbf{A}\cdot(\boldsymbol{\nabla}f)\,d\tau = \oint f\mathbf{A}\cdot d\mathbf{a},$$

or

$$\int_{\mathcal{V}} f(\boldsymbol{\nabla}\cdot\mathbf{A})\,d\tau = -\int_{\mathcal{V}} \mathbf{A}\cdot(\boldsymbol{\nabla}f)\,d\tau + \oint_{\mathcal{S}} f\mathbf{A}\cdot d\mathbf{a}. \tag{1.59}$$

Here again the integrand is the product of one function (f) and the derivative (in this case the *divergence*) of another ($\mathbf{A}$), and integration by parts licenses us to

transfer the derivative from **A** to f (where it becomes a *gradient*), at the cost of a minus sign and a boundary term (in this case a surface integral).

You might wonder how often one is likely to encounter an integral involving the product of one function and the derivative of another; the answer is *surprisingly* often, and integration by parts turns out to be one of the most powerful tools in vector calculus.

Problem 1.36

(a) Show that

$$\int_{\mathcal{S}} f(\nabla \times \mathbf{A}) \cdot d\mathbf{a} = \int_{\mathcal{S}} [\mathbf{A} \times (\nabla f)] \cdot d\mathbf{a} + \oint_{\mathcal{P}} f\mathbf{A} \cdot d\mathbf{l}. \tag{1.60}$$

(b) Show that

$$\int_{\mathcal{V}} \mathbf{B} \cdot (\nabla \times \mathbf{A})\, d\tau = \int_{\mathcal{V}} \mathbf{A} \cdot (\nabla \times \mathbf{B})\, d\tau + \oint_{\mathcal{S}} (\mathbf{A} \times \mathbf{B}) \cdot d\mathbf{a}. \tag{1.61}$$

1.4 ■ CURVILINEAR COORDINATES

1.4.1 ■ Spherical Coordinates

You can label a point P by its Cartesian coordinates (x, y, z), but sometimes it is more convenient to use **spherical** coordinates (r, θ, ϕ); r is the distance from the origin (the magnitude of the position vector **r**), θ (the angle down from the z axis) is called the **polar angle**, and ϕ (the angle around from the x axis) is the **azimuthal angle**. Their relation to Cartesian coordinates can be read from Fig. 1.36:

$$x = r\sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta. \tag{1.62}$$

Figure 1.36 also shows three unit vectors, $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\phi}}$, pointing in the direction of increase of the corresponding coordinates. They constitute an orthogonal (mutually perpendicular) basis set (just like $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$), and any vector **A** can be expressed in terms of them, in the usual way:

$$\mathbf{A} = A_r\,\hat{\mathbf{r}} + A_\theta\,\hat{\boldsymbol{\theta}} + A_\phi\,\hat{\boldsymbol{\phi}}; \tag{1.63}$$

A_r, A_θ, and A_ϕ are the radial, polar, and azimuthal components of **A**. In terms of the Cartesian unit vectors,

$$\left.\begin{aligned} \hat{\mathbf{r}} &= \sin\theta\cos\phi\,\hat{\mathbf{x}} + \sin\theta\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}, \\ \hat{\boldsymbol{\theta}} &= \cos\theta\cos\phi\,\hat{\mathbf{x}} + \cos\theta\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}}, \\ \hat{\boldsymbol{\phi}} &= -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}}, \end{aligned}\right\} \tag{1.64}$$

as you can check for yourself (Prob. 1.38). I have put these formulas inside the back cover, for easy reference.

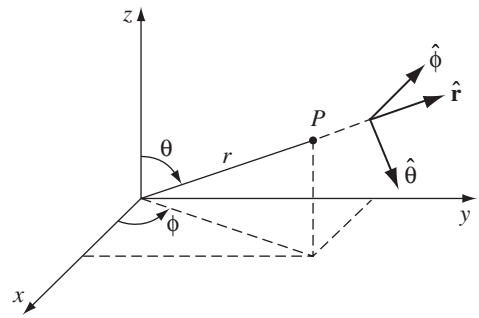


FIGURE 1.36

But there is a poisonous snake lurking here that I'd better warn you about: $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ are associated with a *particular point* P, and they *change direction* as P moves around. For example, $\hat{\mathbf{r}}$ always points radially outward, but "radially outward" can be the x direction, the y direction, or any other direction, depending on where you are. In Fig. 1.37, $\mathbf{A} = \hat{\mathbf{y}}$ and $\mathbf{B} = -\hat{\mathbf{y}}$, and yet *both* of them would be written as $\hat{\mathbf{r}}$ in spherical coordinates. One could take account of this by explicitly indicating the point of reference: $\hat{\mathbf{r}}(\theta, \phi)$, $\hat{\boldsymbol{\theta}}(\theta, \phi)$, $\hat{\boldsymbol{\phi}}(\theta, \phi)$, but this would be cumbersome, and as long as you are alert to the problem, I don't think it will cause difficulties.[9] In particular, do not naïvely combine the spherical components of vectors associated with different points (in Fig. 1.37, $\mathbf{A} + \mathbf{B} = \mathbf{0}$, not $2\hat{\mathbf{r}}$, and $\mathbf{A} \cdot \mathbf{B} = -1$, not $+1$). Beware of differentiating a vector that is expressed in spherical coordinates, since the unit vectors themselves are functions of position ($\partial\hat{\mathbf{r}}/\partial\theta = \hat{\boldsymbol{\theta}}$, for example). And do not take $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ outside an integral, as I did with $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ in Eq. 1.53. In general, if you're uncertain about the validity of an operation, rewrite the problem using Cartesian coordinates, for which this difficulty does not arise.

An infinitesimal displacement in the $\hat{\mathbf{r}}$ direction is simply dr (Fig. 1.38a), just as an infinitesimal element of length in the x direction is dx:

$$dl_r = dr. \tag{1.65}$$

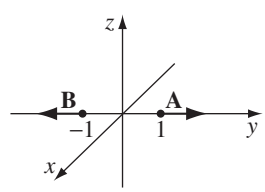


FIGURE 1.37

[9]I claimed back at the beginning that vectors have no location, and I'll stand by that. The vectors themselves live "out there," completely independent of our choice of coordinates. But the *notation* we use to represent them *does* depend on the point in question, in curvilinear coordinates.

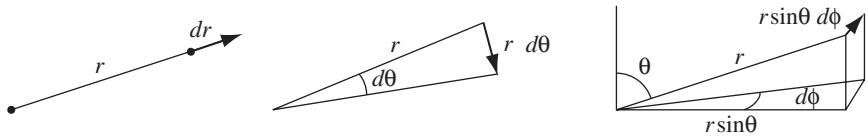


FIGURE 1.38

On the other hand, an infinitesimal element of length in the $\hat{\boldsymbol{\theta}}$ direction (Fig. 1.38b) is *not* just $d\theta$ (that's an *angle*—it doesn't even have the right *units* for a length); rather,

$$dl_\theta = r\,d\theta. \tag{1.66}$$

Similarly, an infinitesimal element of length in the $\hat{\boldsymbol{\phi}}$ direction (Fig. 1.38c) is

$$dl_\phi = r\sin\theta\,d\phi. \tag{1.67}$$

Thus the general infinitesimal displacement $d\mathbf{l}$ is

$$d\mathbf{l} = dr\,\hat{\mathbf{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}. \tag{1.68}$$

This plays the role (in line integrals, for example) that $d\mathbf{l} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}$ played in Cartesian coordinates.

The infinitesimal volume element $d\tau$, in spherical coordinates, is the product of the three infinitesimal displacements:

$$d\tau = dl_r\,dl_\theta\,dl_\phi = r^2\sin\theta\,dr\,d\theta\,d\phi. \tag{1.69}$$

I cannot give you a general expression for *surface* elements $d\mathbf{a}$, since these depend on the orientation of the surface. You simply have to analyze the geometry for any given case (this goes for Cartesian and curvilinear coordinates alike). If you are integrating over the surface of a sphere, for instance, then r is constant, whereas θ and ϕ change (Fig. 1.39), so

$$d\mathbf{a}_1 = dl_\theta\,dl_\phi\,\hat{\mathbf{r}} = r^2\sin\theta\,d\theta\,d\phi\,\hat{\mathbf{r}}.$$

On the other hand, if the surface lies in the xy plane, say, so that θ is constant (to wit: $\pi/2$) while r and ϕ vary, then

$$d\mathbf{a}_2 = dl_r\,dl_\phi\,\hat{\boldsymbol{\theta}} = r\,dr\,d\phi\,\hat{\boldsymbol{\theta}}.$$

Notice, finally, that r ranges from 0 to ∞, ϕ from 0 to 2π, and θ from 0 to π (*not* 2π—that would count every point twice).[10]

[10] Alternatively, you could run ϕ from 0 to π (the "eastern hemisphere") and cover the "western hemisphere" by extending θ from π up to 2π. But this is very bad notation, since, among other things, $\sin\theta$ will then run negative, and you'll have to put absolute value signs around that term in volume and surface elements (area and volume being intrinsically positive quantities).

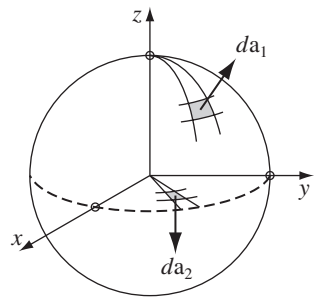


FIGURE 1.39

Example 1.13. Find the volume of a sphere of radius R.

Solution

$$V = \int d\tau = \int_{r=0}^{R}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} r^2 \sin\theta \, dr \, d\theta \, d\phi$$
$$= \left(\int_0^R r^2 \, dr\right)\left(\int_0^\pi \sin\theta \, d\theta\right)\left(\int_0^{2\pi} d\phi\right)$$
$$= \left(\frac{R^3}{3}\right)(2)(2\pi) = \frac{4}{3}\pi R^3$$

(not a big surprise).

So far we have talked only about the *geometry* of spherical coordinates. Now I would like to "translate" the vector derivatives (gradient, divergence, curl, and Laplacian) into r, θ, ϕ notation. In principle, this is entirely straightforward: in the case of the gradient,

$$\nabla T = \frac{\partial T}{\partial x}\hat{\mathbf{x}} + \frac{\partial T}{\partial y}\hat{\mathbf{y}} + \frac{\partial T}{\partial z}\hat{\mathbf{z}},$$

for instance, we would first use the chain rule to expand the partials:

$$\frac{\partial T}{\partial x} = \frac{\partial T}{\partial r}\left(\frac{\partial r}{\partial x}\right) + \frac{\partial T}{\partial \theta}\left(\frac{\partial \theta}{\partial x}\right) + \frac{\partial T}{\partial \phi}\left(\frac{\partial \phi}{\partial x}\right).$$

The terms in parentheses could be worked out from Eq. 1.62—or rather, the *inverse* of those equations (Prob. 1.37). Then we'd do the same for $\partial T/\partial y$ and $\partial T/\partial z$. Finally, we'd substitute in the formulas for $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ in terms of $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ (Prob. 1.38). It would take an hour to figure out the gradient in spherical coordinates by this brute-force method. I suppose this is how it was first done, but there is a much more efficient indirect approach, explained in Appendix A, which

has the extra advantage of treating all coordinate systems at once. I described the "straightforward" method only to show you that there is nothing subtle or mysterious about transforming to spherical coordinates: you're expressing the *same quantity* (gradient, divergence, or whatever) in different notation, that's all.

Here, then, are the vector derivatives in spherical coordinates:

Gradient:

$$\nabla T = \frac{\partial T}{\partial r}\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial T}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial T}{\partial \phi}\hat{\boldsymbol{\phi}}. \tag{1.70}$$

Divergence:

$$\nabla \cdot \mathbf{v} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta\, v_\theta) + \frac{1}{r\sin\theta}\frac{\partial v_\phi}{\partial \phi}. \tag{1.71}$$

Curl:

$$\nabla \times \mathbf{v} = \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(\sin\theta\, v_\phi) - \frac{\partial v_\theta}{\partial \phi}\right]\hat{\mathbf{r}} + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial v_r}{\partial \phi} - \frac{\partial}{\partial r}(r v_\phi)\right]\hat{\boldsymbol{\theta}} + \frac{1}{r}\left[\frac{\partial}{\partial r}(r v_\theta) - \frac{\partial v_r}{\partial \theta}\right]\hat{\boldsymbol{\phi}}. \tag{1.72}$$

Laplacian:

$$\nabla^2 T = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial T}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial T}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 T}{\partial \phi^2}. \tag{1.73}$$

For reference, these formulas are listed inside the front cover.

Problem 1.37 Find formulas for r, θ, ϕ in terms of x, y, z (the inverse, in other words, of Eq. 1.62).

• **Problem 1.38** Express the unit vectors $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ in terms of $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ (that is, derive Eq. 1.64). Check your answers several ways ($\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} \stackrel{?}{=} 1$, $\hat{\boldsymbol{\theta}} \cdot \hat{\boldsymbol{\phi}} \stackrel{?}{=} 0$, $\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} \stackrel{?}{=} \hat{\boldsymbol{\phi}}$, ...). Also work out the inverse formulas, giving $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ in terms of $\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ (and θ, ϕ).

• **Problem 1.39**

(a) Check the divergence theorem for the function $\mathbf{v}_1 = r^2\hat{\mathbf{r}}$, using as your volume the sphere of radius R, centered at the origin.

(b) Do the same for $\mathbf{v}_2 = (1/r^2)\hat{\mathbf{r}}$. (If the answer surprises you, look back at Prob. 1.16.)

Problem 1.40 Compute the divergence of the function

$$\mathbf{v} = (r\cos\theta)\,\hat{\mathbf{r}} + (r\sin\theta)\,\hat{\boldsymbol{\theta}} + (r\sin\theta\cos\phi)\,\hat{\boldsymbol{\phi}}.$$

Check the divergence theorem for this function, using as your volume the inverted hemispherical bowl of radius R, resting on the xy plane and centered at the origin (Fig. 1.40).

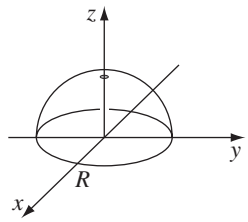


FIGURE 1.40

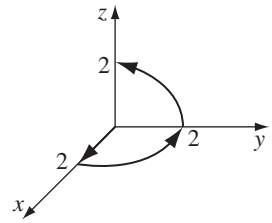


FIGURE 1.41

Problem 1.41 Compute the gradient and Laplacian of the function $T = r(\cos\theta + \sin\theta\cos\phi)$. Check the Laplacian by converting T to Cartesian coordinates and using Eq. 1.42. Test the gradient theorem for this function, using the path shown in Fig. 1.41, from (0, 0, 0) to (0, 0, 2).

1.4.2 ■ Cylindrical Coordinates

The cylindrical coordinates (s, ϕ, z) of a point P are defined in Fig. 1.42. Notice that ϕ has the same meaning as in spherical coordinates, and z is the same as Cartesian; s is the distance to P *from the z axis,* whereas the spherical coordinate r is the distance from the *origin*. The relation to Cartesian coordinates is

$$x = s\cos\phi, \qquad y = s\sin\phi, \qquad z = z. \tag{1.74}$$

The unit vectors (Prob. 1.42) are

$$\left.\begin{array}{rcl} \hat{\mathbf{s}} &=& \cos\phi\,\hat{\mathbf{x}} + \sin\phi\,\hat{\mathbf{y}}, \\ \hat{\boldsymbol{\phi}} &=& -\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}}, \\ \hat{\mathbf{z}} &=& \hat{\mathbf{z}}. \end{array}\right\} \tag{1.75}$$

The infinitesimal displacements are

$$dl_s = ds, \qquad dl_\phi = s\,d\phi, \qquad dl_z = dz, \tag{1.76}$$

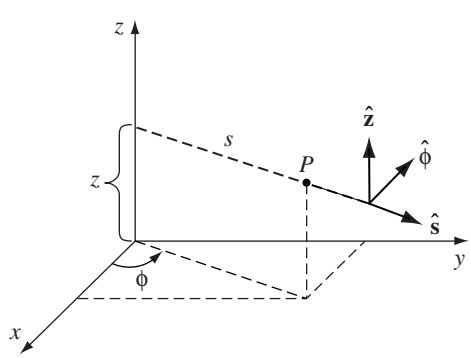


FIGURE 1.42

so

$$d\mathbf{l} = ds\,\hat{\mathbf{s}} + s\,d\phi\,\hat{\boldsymbol{\phi}} + dz\,\hat{\mathbf{z}}, \tag{1.77}$$

and the volume element is

$$d\tau = s\,ds\,d\phi\,dz. \tag{1.78}$$

The range of s is $0 \to \infty$, ϕ goes from $0 \to 2\pi$, and z from $-\infty$ to ∞.

The vector derivatives in cylindrical coordinates are:

Gradient:

$$\boldsymbol{\nabla} T = \frac{\partial T}{\partial s}\,\hat{\mathbf{s}} + \frac{1}{s}\frac{\partial T}{\partial \phi}\,\hat{\boldsymbol{\phi}} + \frac{\partial T}{\partial z}\,\hat{\mathbf{z}}. \tag{1.79}$$

Divergence:

$$\boldsymbol{\nabla} \cdot \mathbf{v} = \frac{1}{s}\frac{\partial}{\partial s}(s v_s) + \frac{1}{s}\frac{\partial v_\phi}{\partial \phi} + \frac{\partial v_z}{\partial z}. \tag{1.80}$$

Curl:

$$\boldsymbol{\nabla} \times \mathbf{v} = \left(\frac{1}{s}\frac{\partial v_z}{\partial \phi} - \frac{\partial v_\phi}{\partial z}\right)\hat{\mathbf{s}} + \left(\frac{\partial v_s}{\partial z} - \frac{\partial v_z}{\partial s}\right)\hat{\boldsymbol{\phi}} + \frac{1}{s}\left[\frac{\partial}{\partial s}(s v_\phi) - \frac{\partial v_s}{\partial \phi}\right]\hat{\mathbf{z}}. \tag{1.81}$$

Laplacian:

$$\nabla^2 T = \frac{1}{s}\frac{\partial}{\partial s}\left(s\frac{\partial T}{\partial s}\right) + \frac{1}{s^2}\frac{\partial^2 T}{\partial \phi^2} + \frac{\partial^2 T}{\partial z^2}. \tag{1.82}$$

These formulas are also listed inside the front cover.

Problem 1.42 Express the cylindrical unit vectors $\hat{\mathbf{s}}$, $\hat{\boldsymbol{\phi}}$, $\hat{\mathbf{z}}$ in terms of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ (that is, derive Eq. 1.75). "Invert" your formulas to get $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ in terms of $\hat{\mathbf{s}}$, $\hat{\boldsymbol{\phi}}$, $\hat{\mathbf{z}}$ (and ϕ).

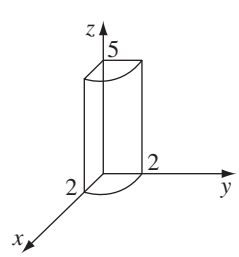


FIGURE 1.43

Problem 1.43

(a) Find the divergence of the function

$$\mathbf{v} = s(2 + \sin^2\phi)\,\hat{\mathbf{s}} + s\sin\phi\cos\phi\;\hat{\boldsymbol{\phi}} + 3z\;\hat{\mathbf{z}}.$$

(b) Test the divergence theorem for this function, using the quarter-cylinder (radius 2, height 5) shown in Fig. 1.43.

(c) Find the curl of $\mathbf{v}$.

1.5 ■ THE DIRAC DELTA FUNCTION

1.5.1 ■ The Divergence of $\hat{\mathbf{r}}/r^2$

Consider the vector function

$$\mathbf{v} = \frac{1}{r^2}\,\hat{\mathbf{r}}. \tag{1.83}$$

At every location, $\mathbf{v}$ is directed radially outward (Fig. 1.44); if ever there was a function that ought to have a large positive divergence, this is it. And yet, when you actually *calculate* the divergence (using Eq. 1.71), you get precisely *zero*:

$$\boldsymbol{\nabla}\cdot\mathbf{v} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{1}{r^2}\right) = \frac{1}{r^2}\frac{\partial}{\partial r}(1) = 0. \tag{1.84}$$

(You will have encountered this paradox already, if you worked Prob. 1.16.) The plot thickens when we apply the divergence theorem to this function. Suppose we integrate over a sphere of radius R, centered at the origin (Prob. 1.38b); the surface integral is

$$\oint \mathbf{v}\cdot d\mathbf{a} = \int \left(\frac{1}{R^2}\hat{\mathbf{r}}\right)\cdot(R^2\sin\theta\,d\theta\,d\phi\,\hat{\mathbf{r}})$$
$$= \left(\int_0^{\pi}\sin\theta\,d\theta\right)\left(\int_0^{2\pi}d\phi\right) = 4\pi. \tag{1.85}$$

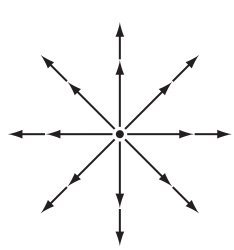


FIGURE 1.44

But the *volume* integral, $\int \boldsymbol{\nabla} \cdot \mathbf{v}\, d\tau$, is *zero*, if we are really to believe Eq. 1.84. Does this mean that the divergence theorem is false? What's going on here?

The source of the problem is the point $r = 0$, where $\mathbf{v}$ blows up (and where, in Eq. 1.84, we have unwittingly divided by zero). It is quite true that $\boldsymbol{\nabla} \cdot \mathbf{v} = 0$ everywhere *except* the origin, but right *at* the origin the situation is more complicated. Notice that the surface integral (Eq. 1.85) is *independent of* R; if the divergence theorem is right (and it *is*), we should get $\int (\boldsymbol{\nabla} \cdot \mathbf{v})\, d\tau = 4\pi$ for *any* sphere centered at the origin, no matter how small. Evidently the entire contribution must be coming from the point $r = 0$! Thus, $\boldsymbol{\nabla} \cdot \mathbf{v}$ has the bizarre property that it vanishes everywhere except at one point, and yet its *integral* (over any volume containing that point) is 4π. No ordinary function behaves like that. (On the other hand, a *physical* example *does* come to mind: the density (mass per unit volume) of a point particle. It's zero except at the exact location of the particle, and yet its *integral* is finite—namely, the mass of the particle.) What we have stumbled on is a mathematical object known to physicists as the **Dirac delta function**. It arises in many branches of theoretical physics. Moreover, the specific problem at hand (the divergence of the function $\hat{\mathbf{r}}/r^2$) is not just some arcane curiosity—it is, in fact, central to the whole theory of electrodynamics. So it is worthwhile to pause here and study the Dirac delta function with some care.

1.5.2 ■ The One-Dimensional Dirac Delta Function

The one-dimensional Dirac delta function, $\delta(x)$, can be pictured as an infinitely high, infinitesimally narrow "spike," with area 1 (Fig. 1.45). That is to say:

$$\delta(x) = \left\{ \begin{array}{ll} 0, & \text{if } x \neq 0 \\ \infty, & \text{if } x = 0 \end{array} \right\} \tag{1.86}$$

and[11]

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1. \tag{1.87}$$

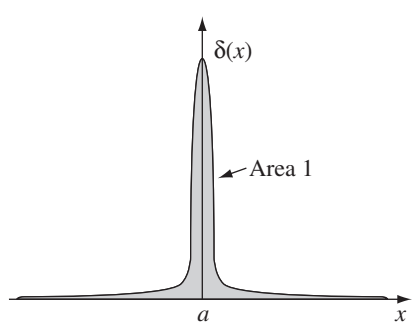


FIGURE 1.45

[11] Notice that the dimensions of $\delta(x)$ are one *over* the dimensions of its argument; if x is a length, $\delta(x)$ carries the units m^{-1}.

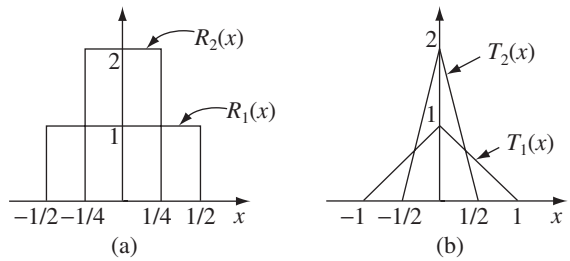


FIGURE 1.46

Technically, $\delta(x)$ is not a function at all, since its value is not finite at $x = 0$; in the mathematical literature it is known as a **generalized function**, or **distribution**. It is, if you like, the *limit* of a *sequence* of functions, such as rectangles $R_n(x)$, of height n and width $1/n$, or isosceles triangles $T_n(x)$, of height n and base $2/n$ (Fig. 1.46).

If $f(x)$ is some "ordinary" function (that is, *not* another delta function—in fact, just to be on the safe side, let's say that $f(x)$ is *continuous*), then the *product* $f(x)\delta(x)$ is zero everywhere except at $x = 0$. It follows that

$$f(x)\delta(x) = f(0)\delta(x). \tag{1.88}$$

(This is the most important fact about the delta function, so make sure you understand why it is true: since the product is zero anyway *except* at $x = 0$, we may as well replace $f(x)$ by the value it assumes at the origin.) In particular

$$\int_{-\infty}^{\infty} f(x)\delta(x)\,dx = f(0)\int_{-\infty}^{\infty} \delta(x)\,dx = f(0). \tag{1.89}$$

Under an integral, then, the delta function "picks out" the value of $f(x)$ at $x = 0$. (Here and below, the integral need not run from $-\infty$ to $+\infty$; it is sufficient that the domain extend across the delta function, and $-\epsilon$ to $+\epsilon$ would do as well.)

Of course, we can shift the spike from $x = 0$ to some other point, $x = a$ (Fig. 1.47):

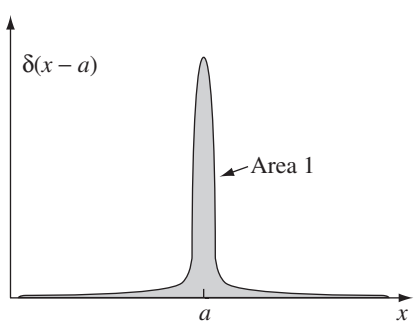


FIGURE 1.47

$$\delta(x-a) = \left\{ \begin{array}{ll} 0, & \text{if } x \neq a \\ \infty, & \text{if } x = a \end{array} \right\} \text{ with } \int_{-\infty}^{\infty} \delta(x-a)\,dx = 1. \tag{1.90}$$

Equation 1.88 becomes

$$f(x)\delta(x-a) = f(a)\delta(x-a), \tag{1.91}$$

and Eq. 1.89 generalizes to

$$\boxed{\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a).} \tag{1.92}$$

Example 1.14. Evaluate the integral

$$\int_0^3 x^3\delta(x-2)\,dx.$$

Solution

The delta function picks out the value of x^3 at the point $x = 2$, so the integral is $2^3 = 8$. Notice, however, that if the upper limit had been 1 (instead of 3), the answer would be 0, because the spike would then be outside the domain of integration.

Although δ itself is not a legitimate function, *integrals* over δ are perfectly acceptable. In fact, it's best to think of the delta function as something that is *always intended for use under an integral sign.* In particular, two expressions involving delta functions (say, $D_1(x)$ and $D_2(x)$) are considered equal if [12]

$$\int_{-\infty}^{\infty} f(x)D_1(x)\,dx = \int_{-\infty}^{\infty} f(x)D_2(x)\,dx, \tag{1.93}$$

for all ("ordinary") functions $f(x)$.

Example 1.15. Show that

$$\delta(kx) = \frac{1}{|k|}\delta(x), \tag{1.94}$$

where k is any (nonzero) constant. (In particular, $\delta(-x) = \delta(x)$.)

[12] I emphasize that the integrals must be equal for *any* $f(x)$. Suppose $D_1(x)$ and $D_2(x)$ actually *differed,* say, in the neighborhood of the point $x = 17$. Then we could pick a function $f(x)$ that was sharply peaked about $x = 17$, and the integrals would not be equal.

Solution
For an arbitrary test function $f(x)$, consider the integral

$$\int_{-\infty}^{\infty} f(x)\delta(kx)\,dx.$$

Changing variables, we let $y \equiv kx$, so that $x = y/k$, and $dx = 1/k\,dy$. If k is positive, the integration still runs from $-\infty$ to $+\infty$, but if k is *negative,* then $x = \infty$ implies $y = -\infty$, and vice versa, so the order of the limits is reversed. Restoring the "proper" order costs a minus sign. Thus

$$\int_{-\infty}^{\infty} f(x)\delta(kx)\,dx = \pm\int_{-\infty}^{\infty} f(y/k)\delta(y)\frac{dy}{k} = \pm\frac{1}{k}f(0) = \frac{1}{|k|}f(0).$$

(The lower signs apply when k is negative, and we account for this neatly by putting absolute value bars around the final k, as indicated.) Under the integral sign, then, $\delta(kx)$ serves the same purpose as $(1/|k|)\delta(x)$:

$$\int_{-\infty}^{\infty} f(x)\delta(kx)\,dx = \int_{-\infty}^{\infty} f(x)\left[\frac{1}{|k|}\delta(x)\right]dx.$$

According to the criterion Eq. 1.93, therefore, $\delta(kx)$ and $(1/|k|)\delta(x)$ are equal.

Problem 1.44 Evaluate the following integrals:

(a) $\int_2^6 (3x^2 - 2x - 1)\,\delta(x - 3)\,dx$.

(b) $\int_0^5 \cos x\,\delta(x - \pi)\,dx$.

(c) $\int_0^3 x^3\delta(x + 1)\,dx$.

(d) $\int_{-\infty}^{\infty} \ln(x + 3)\,\delta(x + 2)\,dx$.

Problem 1.45 Evaluate the following integrals:

(a) $\int_{-2}^{2} (2x + 3)\,\delta(3x)\,dx$.

(b) $\int_0^2 (x^3 + 3x + 2)\,\delta(1 - x)\,dx$.

(c) $\int_{-1}^{1} 9x^2\delta(3x + 1)\,dx$.

(d) $\int_{-\infty}^{a} \delta(x - b)\,dx$.

Problem 1.46

(a) Show that

$$x\frac{d}{dx}(\delta(x)) = -\delta(x).$$

[*Hint:* Use integration by parts.]

(b) Let $\theta(x)$ be the **step function**:

$$\theta(x) \equiv \left\{ \begin{array}{ll} 1, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{array} \right\}. \tag{1.95}$$

Show that $d\theta/dx = \delta(x)$.

1.5.3 ■ The Three-Dimensional Delta Function

It is easy to generalize the delta function to three dimensions:

$$\delta^3(\mathbf{r}) = \delta(x)\,\delta(y)\,\delta(z). \tag{1.96}$$

(As always, $\mathbf{r} \equiv x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$ is the position vector, extending from the origin to the point (x, y, z).) This three-dimensional delta function is zero everywhere except at (0, 0, 0), where it blows up. Its volume integral is 1:

$$\int_{\text{all space}} \delta^3(\mathbf{r})\,d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x)\,\delta(y)\,\delta(z)\,dx\,dy\,dz = 1. \tag{1.97}$$

And, generalizing Eq. 1.92,

$$\int_{\text{all space}} f(\mathbf{r})\delta^3(\mathbf{r} - \mathbf{a})\,d\tau = f(\mathbf{a}). \tag{1.98}$$

As in the one-dimensional case, integration with δ picks out the value of the function f at the location of the spike.

We are now in a position to resolve the paradox introduced in Sect. 1.5.1. As you will recall, we found that the divergence of $\hat{\mathbf{r}}/r^2$ is zero everywhere except at the origin, and yet its *integral* over any volume containing the origin is a constant (to wit: 4π). These are precisely the defining conditions for the Dirac delta function; evidently

$$\boldsymbol{\nabla} \cdot \left(\frac{\hat{\mathbf{r}}}{r^2}\right) = 4\pi\,\delta^3(\mathbf{r}). \tag{1.99}$$

More generally,

$$\boxed{\boldsymbol{\nabla} \cdot \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2}\right) = 4\pi\,\delta^3(\boldsymbol{\imath}),} \tag{1.100}$$

where, as always, $\boldsymbol{\imath}$ is the separation vector: $\boldsymbol{\imath} \equiv \mathbf{r} - \mathbf{r}'$. Note that differentiation here is with respect to $\mathbf{r}$, while $\mathbf{r}'$ is held constant. Incidentally, since

$$\boldsymbol{\nabla}\left(\frac{1}{\imath}\right) = -\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \tag{1.101}$$

(Prob. 1.13b), it follows that

$$\nabla^2 \frac{1}{\mathscr{r}} = -4\pi\delta^3(\boldsymbol{\mathscr{r}}). \tag{1.102}$$

Example 1.16. Evaluate the integral

$$J = \int_{\mathcal{V}} (r^2 + 2)\, \boldsymbol{\nabla} \cdot \left(\frac{\hat{\mathbf{r}}}{r^2}\right) d\tau,$$

where $\mathcal{V}$ is a sphere[13] of radius R centered at the origin.

Solution 1
Use Eq. 1.99 to rewrite the divergence, and Eq. 1.98 to do the integral:

$$J = \int_{\mathcal{V}} (r^2 + 2) 4\pi\delta^3(\mathbf{r})\, d\tau = 4\pi(0 + 2) = 8\pi.$$

This one-line solution demonstrates something of the power and beauty of the delta function, but I would like to show you a second method, which is much more cumbersome but serves to illustrate the method of integration by parts (Sect. 1.3.6).

Solution 2
Using Eq. 1.59, we transfer the derivative from $\hat{\mathbf{r}}/r^2$ to $(r^2 + 2)$:

$$J = -\int_{\mathcal{V}} \frac{\hat{\mathbf{r}}}{r^2} \cdot [\boldsymbol{\nabla}(r^2 + 2)]\, d\tau + \oint_{\mathcal{S}} (r^2 + 2) \frac{\hat{\mathbf{r}}}{r^2} \cdot d\mathbf{a}.$$

The gradient is

$$\boldsymbol{\nabla}(r^2 + 2) = 2r\hat{\mathbf{r}},$$

so the volume integral becomes

$$\int \frac{2}{r}\, d\tau = \int \frac{2}{r} r^2 \sin\theta\, dr\, d\theta\, d\phi = 8\pi \int_0^R r\, dr = 4\pi R^2.$$

Meanwhile, on the boundary of the sphere (where $r = R$),

$$d\mathbf{a} = R^2 \sin\theta\, d\theta\, d\phi\, \hat{\mathbf{r}},$$

so the surface integral is

$$\int (R^2 + 2) \sin\theta\, d\theta\, d\phi = 4\pi(R^2 + 2).$$

[13] In proper mathematical jargon, "sphere" denotes the *surface*, and "ball" the volume it encloses. But physicists are (as usual) sloppy about this sort of thing, and I use the word "sphere" for both the surface and the volume. Where the meaning is not clear from the context, I will write "spherical surface" or "spherical volume." The language police tell me that the former is redundant and the latter an oxymoron, but a poll of my physics colleagues reveals that this is (for us) the standard usage.

Putting it all together,

$$J = -4\pi R^2 + 4\pi(R^2 + 2) = 8\pi,$$

as before.

Problem 1.47

(a) Write an expression for the volume charge density $\rho(\mathbf{r})$ of a point charge q at $\mathbf{r}'$. Make sure that the volume integral of ρ equals q.

(b) What is the volume charge density of an electric dipole, consisting of a point charge $-q$ at the origin and a point charge $+q$ at $\mathbf{a}$?

(c) What is the volume charge density (in spherical coordinates) of a uniform, infinitesimally thin spherical shell of radius R and total charge Q, centered at the origin? [*Beware:* the integral over all space must equal Q.]

Problem 1.48 Evaluate the following integrals:

(a) $\int (r^2 + \mathbf{r}\cdot\mathbf{a} + a^2)\delta^3(\mathbf{r} - \mathbf{a})\, d\tau$, where $\mathbf{a}$ is a fixed vector, a is its magnitude, and the integral is over all space.

(b) $\int_{\mathcal{V}} |\mathbf{r} - \mathbf{b}|^2 \delta^3(5\mathbf{r})\, d\tau$, where $\mathcal{V}$ is a cube of side 2, centered on the origin, and $\mathbf{b} = 4\,\hat{\mathbf{y}} + 3\,\hat{\mathbf{z}}$.

(c) $\int_{\mathcal{V}} \left[r^4 + r^2(\mathbf{r}\cdot\mathbf{c}) + c^4\right] \delta^3(\mathbf{r} - \mathbf{c})\, d\tau$, where $\mathcal{V}$ is a sphere of radius 6 about the origin, $\mathbf{c} = 5\,\hat{\mathbf{x}} + 3\,\hat{\mathbf{y}} + 2\,\hat{\mathbf{z}}$, and c is its magnitude.

(d) $\int_{\mathcal{V}} \mathbf{r}\cdot(\mathbf{d} - \mathbf{r})\delta^3(\mathbf{e} - \mathbf{r})\, d\tau$, where $\mathbf{d} = (1, 2, 3)$, $\mathbf{e} = (3, 2, 1)$, and $\mathcal{V}$ is a sphere of radius 1.5 centered at (2, 2, 2).

Problem 1.49 Evaluate the integral

$$J = \int_{\mathcal{V}} e^{-r} \left(\boldsymbol{\nabla}\cdot\frac{\hat{\mathbf{r}}}{r^2}\right) d\tau$$

(where $\mathcal{V}$ is a sphere of radius R, centered at the origin) by two different methods, as in Ex. 1.16.

1.6 ■ THE THEORY OF VECTOR FIELDS

1.6.1 ■ The Helmholtz Theorem

Ever since Faraday, the laws of electricity and magnetism have been expressed in terms of **electric** and **magnetic fields**, **E** and **B**. Like many physical laws,

these are most compactly expressed as differential equations. Since **E** and **B** are *vectors*, the differential equations naturally involve vector derivatives: divergence and curl. Indeed, Maxwell reduced the entire theory to four equations, specifying respectively the divergence and the curl of **E** and **B**.

Maxwell's formulation raises an important mathematical question: To what extent is a vector function determined by its divergence and curl? In other words, if I tell you that the *divergence* of **F** (which stands for **E** or **B**, as the case may be) is a specified (scalar) function D,

$$\nabla \cdot \mathbf{F} = D,$$

and the curl of **F** is a specified (vector) function **C**,

$$\nabla \times \mathbf{F} = \mathbf{C},$$

(for consistency, **C** must be divergenceless,

$$\nabla \cdot \mathbf{C} = 0,$$

because the divergence of a curl is always zero), can you then determine the function **F**?

Well. . . not quite. For example, as you may have discovered in Prob. 1.20, there are many functions whose divergence and curl are both zero everywhere—the trivial case $\mathbf{F} = \mathbf{0}$, of course, but also $\mathbf{F} = yz\,\hat{\mathbf{x}} + zx\,\hat{\mathbf{y}} + xy\,\hat{\mathbf{z}}$, $\mathbf{F} = \sin x \cosh y\,\hat{\mathbf{x}} - \cos x \sinh y\,\hat{\mathbf{y}}$, etc. To solve a differential equation you must also be supplied with appropriate **boundary conditions**. In electrodynamics we typically require that the fields go to zero "at infinity" (far away from all charges).[14] With that extra information, the **Helmholtz theorem** guarantees that the field is uniquely determined by its divergence and curl. (The Helmholtz theorem is discussed in Appendix B.)

1.6.2 ■ Potentials

If the curl of a vector field (**F**) vanishes (everywhere), then **F** can be written as the gradient of a **scalar potential** (V):

$$\nabla \times \mathbf{F} = \mathbf{0} \Longleftrightarrow \mathbf{F} = -\nabla V. \tag{1.103}$$

(The minus sign is purely conventional.) That's the essential burden of the following theorem:

Theorem 1
Curl-less (or "**irrotational**") **fields**. The following conditions are equivalent (that is, **F** satisfies one if and only if it satisfies all the others):

[14] In some textbook problems the charge itself extends to infinity (we speak, for instance, of the electric field of an infinite plane, or the magnetic field of an infinite wire). In such cases the normal boundary conditions do not apply, and one must invoke symmetry arguments to determine the fields uniquely.

(a) $\nabla \times \mathbf{F} = \mathbf{0}$ everywhere.

(b) $\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{F} \cdot d\mathbf{l}$ is independent of path, for any given end points.

(c) $\oint \mathbf{F} \cdot d\mathbf{l} = 0$ for any closed loop.

(d) $\mathbf{F}$ is the gradient of some scalar function: $\mathbf{F} = -\nabla V$.

The potential is not unique—any constant can be added to V with impunity, since this will not affect its gradient.

If the divergence of a vector field ($\mathbf{F}$) vanishes (everywhere), then $\mathbf{F}$ can be expressed as the curl of a **vector potential** ($\mathbf{A}$):

$$\nabla \cdot \mathbf{F} = 0 \Longleftrightarrow \mathbf{F} = \nabla \times \mathbf{A}. \tag{1.104}$$

That's the main conclusion of the following theorem:

Theorem 2
Divergence-less (or "**solenoidal**") **fields**. The following conditions are equivalent:

(a) $\nabla \cdot \mathbf{F} = 0$ everywhere.

(b) $\int \mathbf{F} \cdot d\mathbf{a}$ is independent of surface, for any given boundary line.

(c) $\oint \mathbf{F} \cdot d\mathbf{a} = 0$ for any closed surface.

(d) $\mathbf{F}$ is the curl of some vector function: $\mathbf{F} = \nabla \times \mathbf{A}$.

The vector potential is not unique—the gradient of any scalar function can be added to $\mathbf{A}$ without affecting the curl, since the curl of a gradient is zero.

You should by now be able to prove all the connections in these theorems, save for the ones that say (a), (b), or (c) implies (d). Those are more subtle, and will come later. Incidentally, in *all* cases (*whatever* its curl and divergence may be) a vector field $\mathbf{F}$ can be written as the gradient of a scalar plus the curl of a vector:[15]

$$\mathbf{F} = -\nabla V + \nabla \times \mathbf{A} \qquad \text{(always)}. \tag{1.105}$$

Problem 1.50

(a) Let $\mathbf{F}_1 = x^2\,\hat{\mathbf{z}}$ and $\mathbf{F}_2 = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$. Calculate the divergence and curl of $\mathbf{F}_1$ and $\mathbf{F}_2$. Which one can be written as the gradient of a scalar? Find a scalar potential that does the job. Which one can be written as the curl of a vector? Find a suitable vector potential.

[15] In physics, the word **field** denotes generically any function of position (x, y, z) and time (t). But in electrodynamics two particular fields ($\mathbf{E}$ and $\mathbf{B}$) are of such paramount importance as to preempt the term. Thus technically the potentials are also "fields," but we never call them that.

(b) Show that $\mathbf{F}_3 = yz\,\hat{\mathbf{x}} + zx\,\hat{\mathbf{y}} + xy\,\hat{\mathbf{z}}$ can be written both as the gradient of a scalar and as the curl of a vector. Find scalar and vector potentials for this function.

Problem 1.51 For Theorem 1, show that (d) $\Rightarrow$ (a), (a) $\Rightarrow$ (c), (c) $\Rightarrow$ (b), (b) $\Rightarrow$ (c), and (c) $\Rightarrow$ (a).

Problem 1.52 For Theorem 2, show that (d) $\Rightarrow$ (a), (a) $\Rightarrow$ (c), (c) $\Rightarrow$ (b), (b) $\Rightarrow$ (c), and (c) $\Rightarrow$ (a).

Problem 1.53

(a) Which of the vectors in Problem 1.15 can be expressed as the gradient of a scalar? Find a scalar function that does the job.

(b) Which can be expressed as the curl of a vector? Find such a vector.

More Problems on Chapter 1

Problem 1.54 Check the divergence theorem for the function

$$\mathbf{v} = r^2 \cos\theta\,\hat{\mathbf{r}} + r^2 \cos\phi\,\hat{\boldsymbol{\theta}} - r^2 \cos\theta \sin\phi\,\hat{\boldsymbol{\phi}},$$

using as your volume one octant of the sphere of radius R (Fig. 1.48). Make sure you include the *entire* surface. [*Answer:* $\pi R^4/4$]

Problem 1.55 Check Stokes' theorem using the function $\mathbf{v} = ay\,\hat{\mathbf{x}} + bx\,\hat{\mathbf{y}}$ (a and b are constants) and the circular path of radius R, centered at the origin in the xy plane. [*Answer:* $\pi R^2(b - a)$]

Problem 1.56 Compute the line integral of

$$\mathbf{v} = 6\,\hat{\mathbf{x}} + yz^2\,\hat{\mathbf{y}} + (3y + z)\,\hat{\mathbf{z}}$$

along the triangular path shown in Fig. 1.49. Check your answer using Stokes' theorem. [*Answer:* 8/3]

Problem 1.57 Compute the line integral of

$$\mathbf{v} = (r\cos^2\theta)\,\hat{\mathbf{r}} - (r\cos\theta\sin\theta)\,\hat{\boldsymbol{\theta}} + 3r\,\hat{\boldsymbol{\phi}}$$

around the path shown in Fig. 1.50 (the points are labeled by their Cartesian coordinates). Do it either in cylindrical or in spherical coordinates. Check your answer, using Stokes' theorem. [*Answer:* $3\pi/2$]

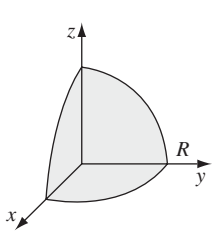


FIGURE 1.48

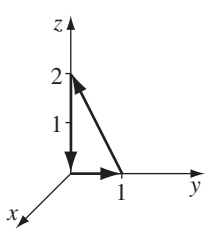


FIGURE 1.49

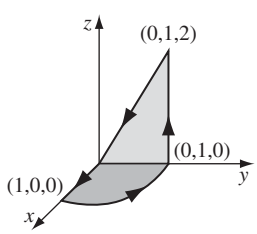


FIGURE 1.50

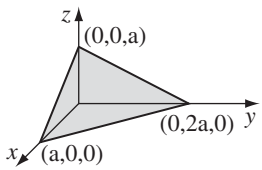


FIGURE 1.51

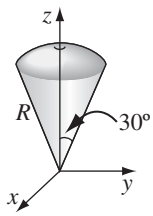


FIGURE 1.52

Problem 1.58 Check Stokes' theorem for the function $\mathbf{v} = y\,\hat{\mathbf{z}}$, using the triangular surface shown in Fig. 1.51. [*Answer:* a^2]

Problem 1.59 Check the divergence theorem for the function

$$\mathbf{v} = r^2 \sin\theta\,\hat{\mathbf{r}} + 4r^2 \cos\theta\,\hat{\boldsymbol{\theta}} + r^2 \tan\theta\,\hat{\boldsymbol{\phi}},$$

using the volume of the "ice-cream cone" shown in Fig. 1.52 (the top surface is spherical, with radius R and centered at the origin). [*Answer:* $(\pi R^4/12)(2\pi + 3\sqrt{3})$]

Problem 1.60 Here are two cute checks of the fundamental theorems:

(a) Combine Corollary 2 to the gradient theorem with Stokes' theorem ($\mathbf{v} = \boldsymbol{\nabla} T$, in this case). Show that the result is consistent with what you already knew about second derivatives.

(b) Combine Corollary 2 to Stokes' theorem with the divergence theorem. Show that the result is consistent with what you already knew.

• **Problem 1.61** Although the gradient, divergence, and curl theorems are the fundamental integral theorems of vector calculus, it is possible to derive a number of corollaries from them. Show that:

(a) $\int_{\mathcal{V}} (\boldsymbol{\nabla} T)\,d\tau = \oint_{\mathcal{S}} T\,d\mathbf{a}$. [*Hint:* Let $\mathbf{v} = \mathbf{c}T$, where $\mathbf{c}$ is a constant, in the divergence theorem; use the product rules.]

(b) $\int_{\mathcal{V}} (\boldsymbol{\nabla} \times \mathbf{v})\,d\tau = -\oint_{\mathcal{S}} \mathbf{v} \times d\mathbf{a}$. [*Hint:* Replace $\mathbf{v}$ by $(\mathbf{v} \times \mathbf{c})$ in the divergence theorem.]

(c) $\int_{\mathcal{V}} [T\nabla^2 U + (\boldsymbol{\nabla} T) \cdot (\boldsymbol{\nabla} U)]\,d\tau = \oint_{\mathcal{S}} (T\boldsymbol{\nabla} U) \cdot d\mathbf{a}$. [*Hint:* Let $\mathbf{v} = T\boldsymbol{\nabla} U$ in the divergence theorem.]

(d) $\int_{\mathcal{V}} (T\nabla^2 U - U\nabla^2 T)\,d\tau = \oint_{\mathcal{S}} (T\boldsymbol{\nabla} U - U\boldsymbol{\nabla} T) \cdot d\mathbf{a}$. [*Comment:* This is sometimes called **Green's second identity**; it follows from (c), which is known as **Green's identity**.]

(e) $\int_{\mathcal{S}} \boldsymbol{\nabla} T \times d\mathbf{a} = -\oint_{\mathcal{P}} T\,d\mathbf{l}$. [*Hint:* Let $\mathbf{v} = \mathbf{c}T$ in Stokes' theorem.]

• **Problem 1.62** The integral

$$\mathbf{a} \equiv \int_{\mathcal{S}} d\mathbf{a} \tag{1.106}$$

is sometimes called the **vector area** of the surface $\mathcal{S}$. If $\mathcal{S}$ happens to be *flat*, then $|\mathbf{a}|$ is the *ordinary* (scalar) area, obviously.

(a) Find the vector area of a hemispherical bowl of radius R.

(b) Show that $\mathbf{a} = \mathbf{0}$ for any *closed* surface. [*Hint:* Use Prob. 1.61a.]

(c) Show that $\mathbf{a}$ is the same for all surfaces sharing the same boundary.

(d) Show that

$$\mathbf{a} = \tfrac{1}{2} \oint \mathbf{r} \times d\mathbf{l}, \tag{1.107}$$

where the integral is around the boundary line. [*Hint:* One way to do it is to draw the cone subtended by the loop at the origin. Divide the conical surface up into infinitesimal triangular wedges, each with vertex at the origin and opposite side $d\mathbf{l}$, and exploit the geometrical interpretation of the cross product (Fig. 1.8).]

(e) Show that

$$\oint (\mathbf{c} \cdot \mathbf{r})\, d\mathbf{l} = \mathbf{a} \times \mathbf{c}, \tag{1.108}$$

for any constant vector $\mathbf{c}$. [*Hint:* Let $T = \mathbf{c} \cdot \mathbf{r}$ in Prob. 1.61e.]

• **Problem 1.63**

(a) Find the divergence of the function

$$\mathbf{v} = \frac{\hat{\mathbf{r}}}{r}.$$

First compute it directly, as in Eq. 1.84. Test your result using the divergence theorem, as in Eq. 1.85. Is there a delta function at the origin, as there was for $\hat{\mathbf{r}}/r^2$? What is the general formula for the divergence of $r^n\hat{\mathbf{r}}$? [*Answer:* $\boldsymbol{\nabla} \cdot (r^n\hat{\mathbf{r}}) = (n+2)r^{n-1}$, unless $n = -2$, in which case it is $4\pi\delta^3(\mathbf{r})$; for $n < -2$, the divergence is ill-defined at the origin.]

(b) Find the *curl* of $r^n\hat{\mathbf{r}}$. Test your conclusion using Prob. 1.61b. [*Answer:* $\boldsymbol{\nabla} \times (r^n\hat{\mathbf{r}}) = \mathbf{0}$]

Problem 1.64 In case you're not persuaded that $\nabla^2(1/r) = -4\pi\delta^3(\mathbf{r})$ (Eq. 1.102 with $\mathbf{r}' = \mathbf{0}$ for simplicity), try replacing r by $\sqrt{r^2+\epsilon^2}$, and watching what happens as $\epsilon \to 0$.[16] Specifically, let

$$D(r,\epsilon) \equiv -\frac{1}{4\pi}\nabla^2 \frac{1}{\sqrt{r^2+\epsilon^2}}.$$

[16]This problem was suggested by Frederick Strauch.

To demonstrate that this goes to $\delta^3(\mathbf{r})$ as $\epsilon \to 0$:

(a) Show that $D(r, \epsilon) = (3\epsilon^2/4\pi)(r^2 + \epsilon^2)^{-5/2}$.

(b) Check that $D(0, \epsilon) \to \infty$, as $\epsilon \to 0$.

(c) Check that $D(r, \epsilon) \to 0$, as $\epsilon \to 0$, for all $r \neq 0$.

(d) Check that the integral of $D(r, \epsilon)$ over all space is 1.

CHAPTER

2 Electrostatics

2.1 ■ THE ELECTRIC FIELD

2.1.1 ■ Introduction

The fundamental problem electrodynamics hopes to solve is this (Fig. 2.1): We have some electric charges, $q_1, q_2, q_3, \ldots$ (call them **source charges**); what force do they exert on another charge, Q (call it the **test charge**)? The positions of the source charges are *given* (as functions of time); the trajectory of the test particle is *to be calculated.* In general, both the source charges and the test charge are in motion.

The solution to this problem is facilitated by the **principle of superposition**, which states that the interaction between any two charges is completely unaffected by the presence of others. This means that to determine the force on Q, we can first compute the force $\mathbf{F}_1$, due to q_1 alone (ignoring all the others); then we compute the force $\mathbf{F}_2$, due to q_2 alone; and so on. Finally, we take the vector sum of all these individual forces: $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \ldots$ Thus, if we can find the force on Q due to a *single* source charge q, we are, in principle, done (the rest is just a question of repeating the same operation over and over, and adding it all up).[1]

Well, at first sight this looks very easy: Why don't I just write down the formula for the force on Q due to q, and be done with it? I *could,* and in Chapter 10 I shall, but you would be shocked to see it at this stage, for not only does the force on Q depend on the separation distance $\mathcal{r}$ between the charges (Fig. 2.2), it also

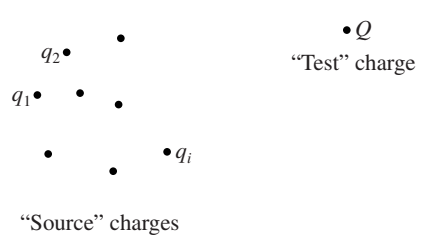


FIGURE 2.1

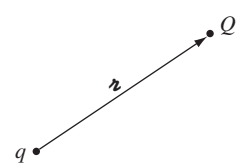


FIGURE 2.2

[1]The principle of superposition may seem "obvious" to you, but it did not have to be so simple: if the electromagnetic force were proportional to the *square* of the total source charge, for instance, the principle of superposition would not hold, since $(q_1 + q_2)^2 \neq q_1^2 + q_2^2$ (there would be "cross terms" to consider). Superposition is not a logical necessity, but an experimental fact.

depends on *both* their velocities and on the *acceleration* of q. Moreover, it is not the position, velocity, and acceleration of q *right now* that matter: electromagnetic "news" travels at the speed of light, so what concerns Q is the position, velocity, and acceleration q *had* at some earlier time, when the message left.

Therefore, in spite of the fact that the basic question ("What is the force on Q due to q?") is easy to state, it does not pay to confront it head on; rather, we shall go at it by stages. In the meantime, the theory we develop will allow for the solution of more subtle electromagnetic problems that do not present themselves in quite this simple format. To begin with, we shall consider the special case of **electrostatics** in which all the *source charges are stationary* (though the test charge may be moving).

2.1.2 ■ Coulomb's Law

What is the force on a test charge Q due to a single point charge q, that is at *rest* a distance $\mathcal{r}$ away? The answer (based on experiments) is given by **Coulomb's law**:

$$\boxed{\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{qQ}{\mathcal{r}^2}\hat{\boldsymbol{\mathcal{r}}}.} \tag{2.1}$$

The constant ϵ_0 is called (ludicrously) the **permittivity of free space.** In SI units, where force is in newtons (N), distance in meters (m), and charge in coulombs (C),

$$\epsilon_0 = 8.85 \times 10^{-12}\,\frac{\mathrm{C}^2}{\mathrm{N}\cdot\mathrm{m}^2}.$$

In words, the force is proportional to the product of the charges and inversely proportional to the square of the separation distance. As always (Sect. 1.1.4), $\boldsymbol{\mathcal{r}}$ is the separation vector from $\mathbf{r}'$ (the location of q) to $\mathbf{r}$ (the location of Q):

$$\boldsymbol{\mathcal{r}} = \mathbf{r} - \mathbf{r}'; \tag{2.2}$$

$\mathcal{r}$ is its magnitude, and $\hat{\boldsymbol{\mathcal{r}}}$ is its direction. The force points along the line from q to Q; it is repulsive if q and Q have the same sign, and attractive if their signs are opposite.

Coulomb's law and the principle of superposition constitute the physical input for electrostatics—the rest, except for some special properties of matter, is mathematical elaboration of these fundamental rules.

Problem 2.1

(a) Twelve equal charges, q, are situated at the corners of a regular 12-sided polygon (for instance, one on each numeral of a clock face). What is the net force on a test charge Q at the center?

(b) Suppose *one* of the 12 q's is removed (the one at "6 o'clock"). What is the force on Q? Explain your reasoning carefully.

(c) Now 13 equal charges, q, are placed at the corners of a regular 13-sided polygon. What is the force on a test charge Q at the center?

(d) If one of the 13 q's is removed, what is the force on Q? Explain your reasoning.

2.1.3 ■ The Electric Field

If we have *several* point charges $q_1, q_2, \ldots, q_n$, at distances $\imath_1, \imath_2, \ldots, \imath_n$ from Q, the total force on Q is evidently

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \ldots = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1 Q}{\imath_1^2}\hat{\boldsymbol{\imath}}_1 + \frac{q_2 Q}{\imath_2^2}\hat{\boldsymbol{\imath}}_2 + \ldots\right)$$
$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{q_1}{\imath_1^2}\hat{\boldsymbol{\imath}}_1 + \frac{q_2}{\imath_2^2}\hat{\boldsymbol{\imath}}_2 + \frac{q_3}{\imath_3^2}\hat{\boldsymbol{\imath}}_3 + \ldots\right),$$

or

$$\boxed{\mathbf{F} = Q\mathbf{E},} \tag{2.3}$$

where

$$\mathbf{E}(\mathbf{r}) \equiv \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{n}\frac{q_i}{\imath_i^2}\hat{\boldsymbol{\imath}}_i. \tag{2.4}$$

E is called the **electric field** of the source charges. Notice that it is a function of position (**r**), because the separation vectors $\boldsymbol{\imath}_i$ depend on the location of the **field point** P (Fig. 2.3). But it makes no reference to the test charge Q. The electric field is a vector quantity that varies from point to point and is determined by the configuration of source charges; physically, **E**(**r**) is the force per unit charge that would be exerted on a test charge, if you were to place one at P.

What exactly *is* an electric field? I have deliberately begun with what you might call the "minimal" interpretation of **E**, as an intermediate step in the calculation of electric forces. But I encourage you to think of the field as a "real" physical

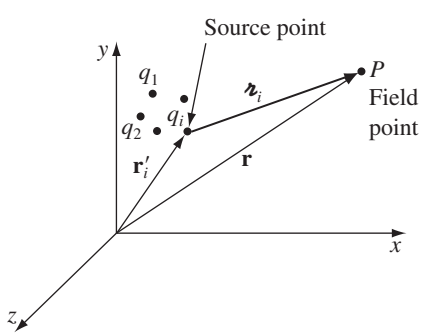


FIGURE 2.3

entity, filling the space around electric charges. Maxwell himself came to believe that electric and magnetic fields are stresses and strains in an invisible primordial jellylike "ether." Special relativity has forced us to abandon the notion of ether, and with it Maxwell's mechanical interpretation of electromagnetic fields. (It is even possible, though cumbersome, to formulate classical electrodynamics as an "action-at-a-distance" theory, and dispense with the field concept altogether.) I can't tell you, then, what a field *is*—only how to calculate it and what it can do for you once you've got it.

Example 2.1. Find the electric field a distance z above the midpoint between two equal charges (q), a distance d apart (Fig. 2.4a).

Solution

Let $\mathbf{E}_1$ be the field of the left charge alone, and $\mathbf{E}_2$ that of the right charge alone (Fig. 2.4b). Adding them (vectorially), the horizontal components cancel and the vertical components conspire:

$$E_z = 2\frac{1}{4\pi\epsilon_0}\frac{q}{\imath^2}\cos\theta.$$

Here $\imath = \sqrt{z^2 + (d/2)^2}$ and $\cos\theta = z/\imath$, so

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{2qz}{\left[z^2 + (d/2)^2\right]^{3/2}}\,\hat{\mathbf{z}}.$$

Check: When $z \gg d$ you're so far away that it just looks like a single charge $2q$, so the field should reduce to $\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{2q}{z^2}\,\hat{\mathbf{z}}$. And it *does* (just set $d \to 0$ in the formula).

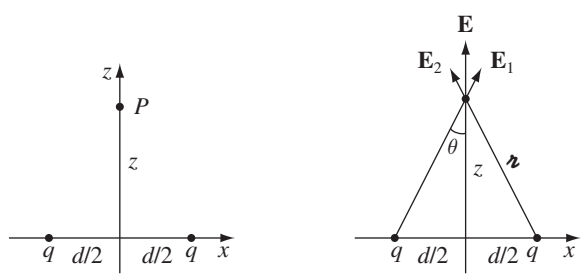
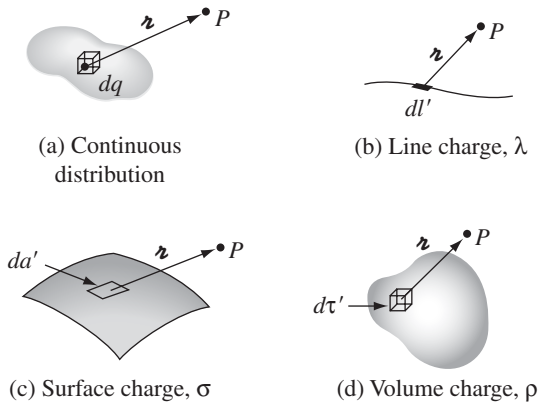


FIGURE 2.4

Problem 2.2 Find the electric field (magnitude and direction) a distance z above the midpoint between equal and opposite charges ($\pm q$), a distance d apart (same as Example 2.1, except that the charge at $x = +d/2$ is $-q$).



(a) Continuous distribution

(b) Line charge, λ

(c) Surface charge, σ

(d) Volume charge, ρ

FIGURE 2.5

2.1.4 ■ Continuous Charge Distributions

Our definition of the electric field (Eq. 2.4) assumes that the source of the field is a collection of discrete point charges q_i. If, instead, the charge is distributed continuously over some region, the sum becomes an integral (Fig. 2.5a):

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{\imath^2}\hat{\boldsymbol{\imath}}\, dq. \tag{2.5}$$

If the charge is spread out along a *line* (Fig. 2.5b), with charge-per-unit-length λ, then $dq = \lambda\, dl'$ (where dl' is an element of length along the line); if the charge is smeared out over a *surface* (Fig. 2.5c), with charge-per-unit-area σ, then $dq = \sigma\, da'$ (where da' is an element of area on the surface); and if the charge fills a *volume* (Fig. 2.5d), with charge-per-unit-volume ρ, then $dq = \rho\, d\tau'$ (where $d\tau'$ is an element of volume):

$$dq \to \lambda\, dl' \sim \sigma\, da' \sim \rho\, d\tau'.$$

Thus the electric field of a line charge is

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\lambda(\mathbf{r}')}{\imath^2}\hat{\boldsymbol{\imath}}\, dl'; \tag{2.6}$$

for a surface charge,

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\sigma(\mathbf{r}')}{\imath^2}\hat{\boldsymbol{\imath}}\, da'; \tag{2.7}$$

and for a volume charge,

$$\boxed{\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{\imath^2}\hat{\boldsymbol{\imath}}\, d\tau'.} \tag{2.8}$$

Equation 2.8 itself is often referred to as "Coulomb's law," because it is such a short step from the original (2.1), and because a volume charge is in a sense the most general and realistic case. Please note carefully the meaning of $\boldsymbol{\imath}$ in these formulas. Originally, in Eq. 2.4, $\boldsymbol{\imath}_i$ stood for the vector from the source charge q_i to the field point $\mathbf{r}$. Correspondingly, in Eqs. 2.5–2.8, $\boldsymbol{\imath}$ is the vector from dq (therefore from dl', da', or $d\tau'$) to the field point $\mathbf{r}$.[2]

Example 2.2. Find the electric field a distance z above the midpoint of a straight line segment of length $2L$ that carries a uniform line charge λ (Fig. 2.6).

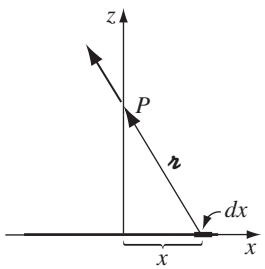


FIGURE 2.6

Solution

The simplest method is to chop the line into symmetrically placed pairs (at $\pm x$), quote the result of Ex. 2.1 (with $d/2 \to x$, $q \to \lambda\, dx$), and integrate ($x : 0 \to L$). But here's a more general approach:[3]

$$\mathbf{r} = z\,\hat{\mathbf{z}}, \quad \mathbf{r}' = x\,\hat{\mathbf{x}}, \quad dl' = dx;$$

$$\boldsymbol{\imath} = \mathbf{r} - \mathbf{r}' = z\,\hat{\mathbf{z}} - x\,\hat{\mathbf{x}}, \quad \imath = \sqrt{z^2 + x^2}, \quad \hat{\boldsymbol{\imath}} = \frac{\boldsymbol{\imath}}{\imath} = \frac{z\,\hat{\mathbf{z}} - x\,\hat{\mathbf{x}}}{\sqrt{z^2 + x^2}}.$$

$$\begin{aligned}
\mathbf{E} &= \frac{1}{4\pi\epsilon_0} \int_{-L}^{L} \frac{\lambda}{z^2 + x^2} \frac{z\,\hat{\mathbf{z}} - x\,\hat{\mathbf{x}}}{\sqrt{z^2 + x^2}}\, dx \\
&= \frac{\lambda}{4\pi\epsilon_0} \left[z\,\hat{\mathbf{z}} \int_{-L}^{L} \frac{1}{(z^2 + x^2)^{3/2}}\, dx - \hat{\mathbf{x}} \int_{-L}^{L} \frac{x}{(z^2 + x^2)^{3/2}}\, dx \right] \\
&= \frac{\lambda}{4\pi\epsilon_0} \left[z\,\hat{\mathbf{z}} \left(\frac{x}{z^2\sqrt{z^2 + x^2}} \right) \Bigg|_{-L}^{L} - \hat{\mathbf{x}} \left(-\frac{1}{\sqrt{z^2 + x^2}} \right) \Bigg|_{-L}^{L} \right] \\
&= \frac{1}{4\pi\epsilon_0} \frac{2\lambda L}{z\sqrt{z^2 + L^2}}\, \hat{\mathbf{z}}.
\end{aligned}$$

[2] *Warning:* The unit vector $\hat{\boldsymbol{\imath}}$ is *not* constant; its *direction* depends on the source point $\mathbf{r}'$, and hence it *cannot be taken outside the integrals* (Eqs. 2.5–2.8). In practice, *you must work with Cartesian components* ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ *are* constant, and *do* come out), even if you use curvilinear coordinates to perform the integration.

[3] Ordinarily I'll put a prime on the source coordinates, but where no confusion can arise I'll remove the prime to simplify the notation.

For points far from the line ($z \gg L$),

$$E \cong \frac{1}{4\pi\epsilon_0}\frac{2\lambda L}{z^2}.$$

This makes sense: From far away the line looks like a point charge $q = 2\lambda L$. In the limit $L \to \infty$, on the other hand, we obtain the field of an infinite straight wire:

$$E = \frac{1}{4\pi\epsilon_0}\frac{2\lambda}{z}. \tag{2.9}$$

Problem 2.3 Find the electric field a distance z above one end of a straight line segment of length L (Fig. 2.7) that carries a uniform line charge λ. Check that your formula is consistent with what you would expect for the case $z \gg L$.

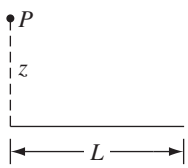


FIGURE 2.7

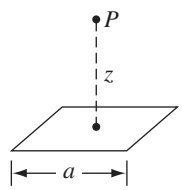


FIGURE 2.8

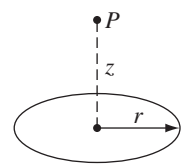


FIGURE 2.9

Problem 2.4 Find the electric field a distance z above the center of a square loop (side a) carrying uniform line charge λ (Fig. 2.8). [*Hint:* Use the result of Ex. 2.2.]

Problem 2.5 Find the electric field a distance z above the center of a circular loop of radius r (Fig. 2.9) that carries a uniform line charge λ.

Problem 2.6 Find the electric field a distance z above the center of a flat circular disk of radius R (Fig. 2.10) that carries a uniform surface charge σ. What does your formula give in the limit $R \to \infty$? Also check the case $z \gg R$.

! **Problem 2.7** Find the electric field a distance z from the center of a spherical surface of radius R (Fig. 2.11) that carries a uniform charge density σ. Treat the case $z < R$ (inside) as well as $z > R$ (outside). Express your answers in terms of the total charge q on the sphere. [*Hint:* Use the law of cosines to write $\imath$ in terms of R and θ. Be sure to take the *positive* square root: $\sqrt{R^2 + z^2 - 2Rz} = (R - z)$ if $R > z$, but it's $(z - R)$ if $R < z$.]

Problem 2.8 Use your result in Prob. 2.7 to find the field inside and outside a solid sphere of radius R that carries a uniform volume charge density ρ. Express your answers in terms of the total charge of the sphere, q. Draw a graph of $|\mathbf{E}|$ as a function of the distance from the center.

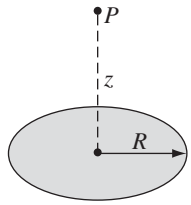


FIGURE 2.10

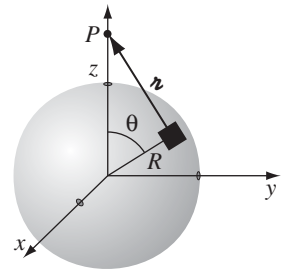


FIGURE 2.11

2.2 ■ DIVERGENCE AND CURL OF ELECTROSTATIC FIELDS

2.2.1 ■ Field Lines, Flux, and Gauss's Law

In principle, we are *done* with the subject of electrostatics. Equation 2.8 tells us how to compute the field of a charge distribution, and Eq. 2.3 tells us what the force on a charge Q placed in this field will be. Unfortunately, as you may have discovered in working Prob. 2.7, the integrals involved in computing $\mathbf{E}$ can be formidable, even for reasonably simple charge distributions. Much of the rest of electrostatics is devoted to assembling a bag of tools and tricks for avoiding these integrals. It all begins with the divergence and curl of $\mathbf{E}$. I shall calculate the divergence of $\mathbf{E}$ directly from Eq. 2.8, in Sect. 2.2.2, but first I want to show you a more qualitative, and perhaps more illuminating, intuitive approach.

Let's begin with the simplest possible case: a single point charge q, situated at the origin:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}}. \tag{2.10}$$

To get a "feel" for this field, I might sketch a few representative vectors, as in Fig. 2.12a. Because the field falls off like $1/r^2$, the vectors get shorter as you go farther away from the origin; they always point radially outward. But there is a

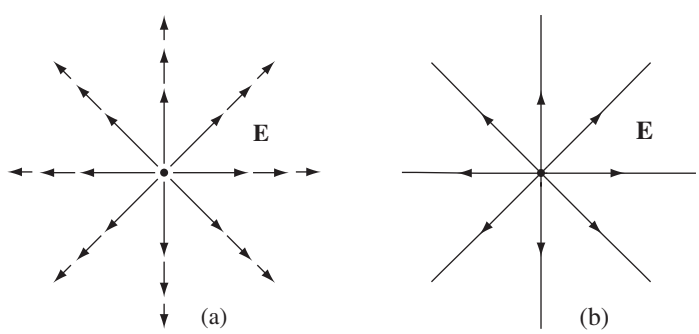


FIGURE 2.12

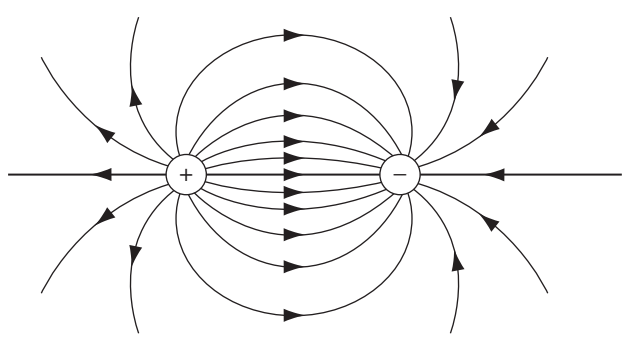
nicer way to represent this field, and that's to connect up the arrows, to form **field lines** (Fig. 2.12b). You might think that I have thereby thrown away information about the *strength* of the field, which was contained in the length of the arrows. But actually I have not. The magnitude of the field is indicated by the *density* of the field lines: it's strong near the center where the field lines are close together, and weak farther out, where they are relatively far apart.

In truth, the field-line diagram is deceptive, when I draw it on a two-dimensional surface, for the density of lines passing through a circle of radius r is the total number divided by the circumference ($n/2\pi r$), which goes like ($1/r$), not ($1/r^2$). But if you imagine the model in three dimensions (a pincushion with needles sticking out in all directions), then the density of lines is the total number divided by the area of the sphere ($n/4\pi r^2$), which *does* go like ($1/r^2$).

Such diagrams are also convenient for representing more complicated fields. Of course, the number of lines you draw depends on how lazy you are (and how sharp your pencil is), though you ought to include enough to get an accurate sense of the field, and you must be consistent: If q gets 8 lines, then $2q$ deserves 16. And you must space them fairly—they emanate from a point charge symmetrically in all directions. Field lines begin on positive charges and end on negative ones; they cannot simply terminate in midair,[4] though they may extend out to infinity. Moreover, field lines can never cross—at the intersection, the field would have two different directions at once! With all this in mind, it is easy to sketch the field of any simple configuration of point charges: Begin by drawing the lines in the neighborhood of each charge, and then connect them up or extend them to infinity (Figs. 2.13 and 2.14).

In this model, the *flux* of **E** through a surface $\mathcal{S}$,

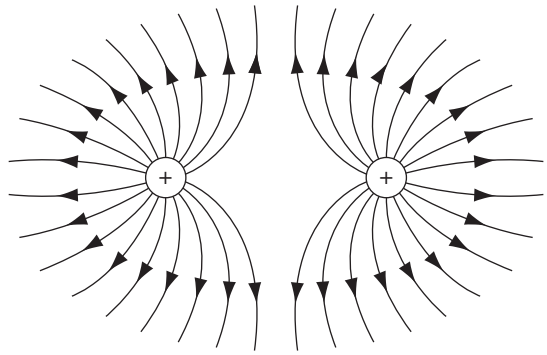
$$\Phi_E \equiv \int_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{a}, \tag{2.11}$$



Opposite charges

FIGURE 2.13

[4]If they *did*, the divergence of **E** would not be zero, and (as we shall soon see) that cannot happen in empty space.



Equal charges

FIGURE 2.14

is a measure of the "number of field lines" passing through $\mathcal{S}$. I put this in quotes because of course we can only draw a representative *sample* of the field lines—the *total* number would be infinite. But *for a given sampling rate* the flux is *proportional* to the number of lines drawn, because the field strength, remember, is proportional to the density of field lines (the number per unit area), and hence $\mathbf{E} \cdot d\mathbf{a}$ is proportional to the number of lines passing through the infinitesimal area $d\mathbf{a}$. (The dot product picks out the component of $d\mathbf{a}$ along the direction of $\mathbf{E}$, as indicated in Fig. 2.15. It is the area *in the plane perpendicular to* $\mathbf{E}$ that we have in mind when we say that the density of field lines is the number per unit area.)

This suggests that the flux through any *closed* surface is a measure of the total charge inside. For the field lines that originate on a positive charge must either pass out through the surface or else terminate on a negative charge inside (Fig. 2.16a). On the other hand, a charge *outside* the surface will contribute nothing to the total flux, since its field lines pass in one side and out the other (Fig. 2.16b). This is the *essence* of **Gauss's law**. Now let's make it quantitative.

In the case of a point charge q at the origin, the flux of $\mathbf{E}$ through a spherical surface of radius r is

$$\oint \mathbf{E} \cdot d\mathbf{a} = \int \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r^2}\hat{\mathbf{r}}\right) \cdot (r^2 \sin\theta \, d\theta \, d\phi \, \hat{\mathbf{r}}) = \frac{1}{\epsilon_0} q. \tag{2.12}$$

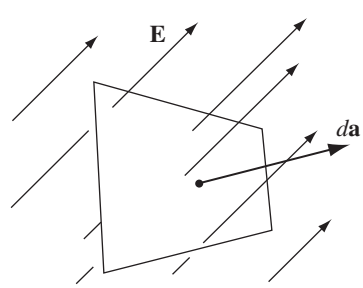


FIGURE 2.15

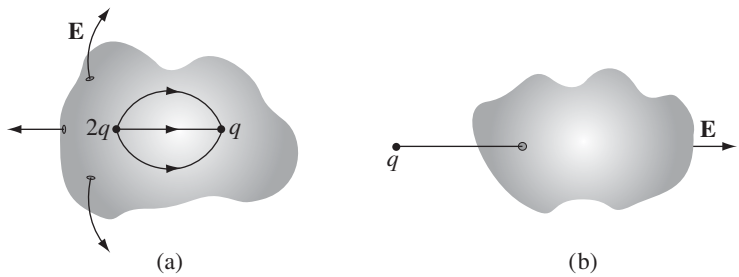


FIGURE 2.16

Notice that the radius of the sphere cancels out, for while the surface area goes *up* as r^2, the field goes *down* as $1/r^2$, so the product is constant. In terms of the field-line picture, this makes good sense, since the same number of field lines pass through any sphere centered at the origin, regardless of its size. In fact, it didn't have to be a sphere—*any* closed surface, whatever its shape, would be pierced by the same number of field lines. Evidently *the flux through any surface enclosing the charge is* q/ϵ_0.

Now suppose that instead of a single charge at the origin, we have a bunch of charges scattered about. According to the principle of superposition, the total field is the (vector) sum of all the individual fields:

$$\mathbf{E} = \sum_{i=1}^{n} \mathbf{E}_i.$$

The flux through a surface that encloses them all is

$$\oint \mathbf{E} \cdot d\mathbf{a} = \sum_{i=1}^{n} \left(\oint \mathbf{E}_i \cdot d\mathbf{a} \right) = \sum_{i=1}^{n} \left(\frac{1}{\epsilon_0} q_i \right)$$

For any closed surface, then,

$$\boxed{\oint \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{enc}},} \tag{2.13}$$

where Q_{enc} is the total charge enclosed within the surface. This is the quantitative statement of Gauss's law. Although it contains no information that was not already present in Coulomb's law plus the principle of superposition, it is of almost magical power, as you will see in Sect. 2.2.3. Notice that it all hinges on the $1/r^2$ character of Coulomb's law; without that the crucial cancellation of the r's in Eq. 2.12 would not take place, and the total flux of $\mathbf{E}$ would depend on the surface chosen, not merely on the total charge enclosed. Other $1/r^2$ forces (I am thinking particularly of Newton's law of universal gravitation) will obey "Gauss's laws" of their own, and the applications we develop here carry over directly.

As it stands, Gauss's law is an *integral* equation, but we can easily turn it into a *differential* one, by applying the divergence theorem:

$$\oint_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{a} = \int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \mathbf{E})\, d\tau.$$

Rewriting Q_{enc} in terms of the charge density ρ, we have

$$Q_{\text{enc}} = \int_{\mathcal{V}} \rho\, d\tau.$$

So Gauss's law becomes

$$\int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \mathbf{E})\, d\tau = \int_{\mathcal{V}} \left(\frac{\rho}{\epsilon_0}\right) d\tau.$$

And since this holds for *any* volume, the integrands must be equal:

$$\boxed{\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho.} \tag{2.14}$$

Equation 2.14 carries the same message as Eq. 2.13; it is **Gauss's law in differential form.** The differential version is tidier, but the integral form has the advantage in that it accommodates point, line, and surface charges more naturally.

Problem 2.9 Suppose the electric field in some region is found to be $\mathbf{E} = kr^3\hat{\mathbf{r}}$, in spherical coordinates (k is some constant).

(a) Find the charge density ρ.

(b) Find the total charge contained in a sphere of radius R, centered at the origin. (Do it two different ways.)

Problem 2.10 A charge q sits at the back corner of a cube, as shown in Fig. 2.17. What is the flux of $\mathbf{E}$ through the shaded side?

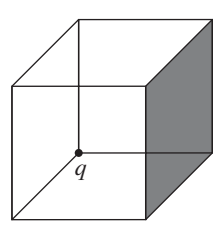


FIGURE 2.17

2.2.2 ■ The Divergence of E

Let's go back, now, and calculate the divergence of **E** directly from Eq. 2.8:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_{\text{all space}} \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \rho(\mathbf{r}')\, d\tau'. \tag{2.15}$$

(Originally the integration was over the volume occupied by the charge, but I may as well extend it to all space, since $\rho = 0$ in the exterior region anyway.) Noting that the **r**-dependence is contained in $\boldsymbol{\imath} = \mathbf{r} - \mathbf{r}'$, we have

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{4\pi\epsilon_0} \int \boldsymbol{\nabla} \cdot \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) \rho(\mathbf{r}')\, d\tau'.$$

This is precisely the divergence we calculated in Eq. 1.100:

$$\boldsymbol{\nabla} \cdot \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = 4\pi\delta^3(\boldsymbol{\imath}).$$

Thus

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{1}{4\pi\epsilon_0} \int 4\pi\delta^3(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}')\, d\tau' = \frac{1}{\epsilon_0}\rho(\mathbf{r}), \tag{2.16}$$

which is Gauss's law in differential form (Eq. 2.14). To recover the integral form (Eq. 2.13), we run the previous argument in reverse—integrate over a volume and apply the divergence theorem:

$$\int_{\mathcal{V}} \boldsymbol{\nabla} \cdot \mathbf{E}\, d\tau = \oint_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} \int_{\mathcal{V}} \rho\, d\tau = \frac{1}{\epsilon_0} Q_{\text{enc}}.$$

2.2.3 ■ Applications of Gauss's Law

I must interrupt the theoretical development at this point to show you the extraordinary power of Gauss's law, in integral form. When symmetry permits, it affords *by far* the quickest and easiest way of computing electric fields. I'll illustrate the method with a series of examples.

Example 2.3. Find the field outside a uniformly charged solid sphere of radius R and total charge q.

Solution

Imagine a spherical surface at radius $r > R$ (Fig. 2.18); this is called a **Gaussian surface** in the trade. Gauss's law says that

$$\oint_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{enc}},$$

and in this case $Q_{\text{enc}} = q$. At first glance this doesn't seem to get us very far, because the quantity we want (**E**) is buried inside the surface integral. Luckily, symmetry allows us to extract **E** from under the integral sign: **E** certainly points radially outward,[5] as does $d\mathbf{a}$, so we can drop the dot product,

$$\int_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{a} = \int_{\mathcal{S}} |\mathbf{E}|\, da,$$

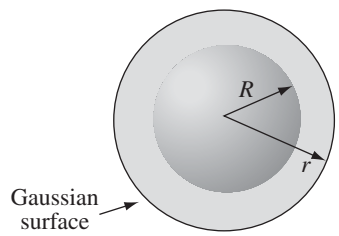


FIGURE 2.18

and the *magnitude* of **E** is constant over the Gaussian surface, so it comes outside the integral:

$$\int_{\mathcal{S}} |\mathbf{E}|\, da = |\mathbf{E}| \int_{\mathcal{S}} da = |\mathbf{E}|\, 4\pi r^2.$$

Thus

$$|\mathbf{E}|\, 4\pi r^2 = \frac{1}{\epsilon_0} q,$$

or

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2} \hat{\mathbf{r}}.$$

Notice a remarkable feature of this result: The field outside the sphere is exactly *the same as it would have been if all the charge had been concentrated at the center.*

Gauss's law is always *true*, but it is not always *useful.* If ρ had not been uniform (or, at any rate, not spherically symmetrical), or if I had chosen some other shape for my Gaussian surface, it would still have been true that the flux of **E** is q/ϵ_0, but **E** would not have pointed in the same direction as $d\mathbf{a}$, and its magnitude would not have been constant over the surface, and without that I cannot get $|\mathbf{E}|$ outside

[5] If you doubt that **E** is radial, consider the alternative. Suppose, say, that it points due east, at the "equator." But the orientation of the equator is perfectly arbitrary—nothing is spinning here, so there is no natural "north-south" axis—any argument purporting to show that **E** points east could just as well be used to show it points west, or north, or any other direction. The only *unique* direction on a sphere is radial.

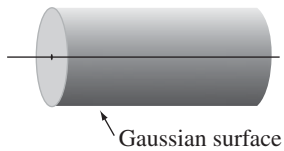


FIGURE 2.19

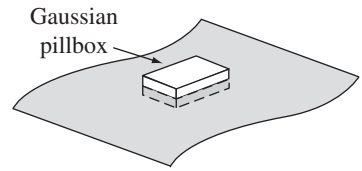


FIGURE 2.20

of the integral. *Symmetry is crucial* to this application of Gauss's law. As far as I know, there are only three kinds of symmetry that work:

1. *Spherical symmetry.* Make your Gaussian surface a concentric sphere.
2. *Cylindrical symmetry.* Make your Gaussian surface a coaxial cylinder (Fig. 2.19).
3. *Plane symmetry.* Use a Gaussian "pillbox" that straddles the surface (Fig. 2.20).

Although (2) and (3) technically require infinitely long cylinders, and planes extending to infinity, we shall often use them to get approximate answers for "long" cylinders or "large" planes, at points far from the edges.

Example 2.4. A long cylinder (Fig. 2.21) carries a charge density that is proportional to the distance from the axis: $\rho = ks$, for some constant k. Find the electric field inside this cylinder.

Solution

Draw a Gaussian cylinder of length l and radius s. For this surface, Gauss's law states:

$$\oint_S \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{enc}}.$$

The enclosed charge is

$$Q_{\text{enc}} = \int \rho \, d\tau = \int (ks')(s' \, ds' \, d\phi \, dz) = 2\pi kl \int_0^s s'^2 \, ds' = \tfrac{2}{3}\pi kls^3.$$

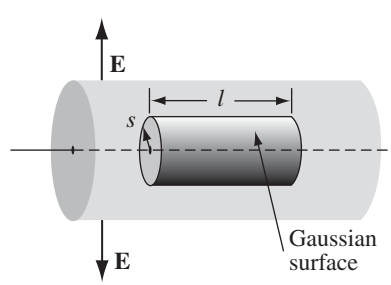


FIGURE 2.21

(I used the volume element appropriate to cylindrical coordinates, Eq. 1.78, and integrated ϕ from 0 to 2π, dz from 0 to l. I put a prime on the integration variable s', to distinguish it from the radius s of the Gaussian surface.)

Now, symmetry dictates that **E** must point radially outward, so for the curved portion of the Gaussian cylinder we have:

$$\int \mathbf{E} \cdot d\mathbf{a} = \int |\mathbf{E}|\, da = |\mathbf{E}| \int da = |\mathbf{E}|\, 2\pi s l,$$

while the two ends contribute nothing (here **E** is perpendicular to $d\mathbf{a}$). Thus,

$$|\mathbf{E}|\, 2\pi s l = \frac{1}{\epsilon_0} \frac{2}{3} \pi k l s^3,$$

or, finally,

$$\mathbf{E} = \frac{1}{3\epsilon_0} k s^2 \hat{\mathbf{s}}.$$

Example 2.5. An infinite plane carries a uniform surface charge σ. Find its electric field.

Solution

Draw a "Gaussian pillbox," extending equal distances above and below the plane (Fig. 2.22). Apply Gauss's law to this surface:

$$\oint \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{enc}}.$$

In this case, $Q_{\text{enc}} = \sigma A$, where A is the area of the lid of the pillbox. By symmetry, **E** points away from the plane (upward for points above, downward for points below). So the top and bottom surfaces yield

$$\int \mathbf{E} \cdot d\mathbf{a} = 2A|\mathbf{E}|,$$

whereas the sides contribute nothing. Thus

$$2A\,|\mathbf{E}| = \frac{1}{\epsilon_0} \sigma A,$$

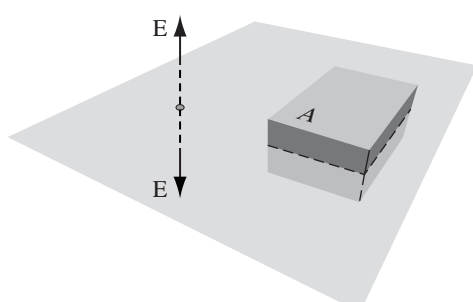


FIGURE 2.22

or

$$\mathbf{E} = \frac{\sigma}{2\epsilon_0}\hat{\mathbf{n}}, \tag{2.17}$$

where $\hat{\mathbf{n}}$ is a unit vector pointing away from the surface. In Prob. 2.6, you obtained this same result by a much more laborious method.

It seems surprising, at first, that the field of an infinite plane is *independent of how far away you are.* What about the $1/r^2$ in Coulomb's law? The point is that as you move farther and farther away from the plane, more and more charge comes into your "field of view" (a cone shape extending out from your eye), and this compensates for the diminishing influence of any particular piece. The electric field of a sphere falls off like $1/r^2$; the electric field of an infinite line falls off like $1/r$; and the electric field of an infinite plane does not fall off at all (you cannot escape from an infinite plane).

Although the direct use of Gauss's law to compute electric fields is limited to cases of spherical, cylindrical, and planar symmetry, we can put together *combinations* of objects possessing such symmetry, even though the arrangement as a whole is not symmetrical. For example, invoking the principle of superposition, we could find the field in the vicinity of two uniformly charged parallel cylinders, or a sphere near an infinite charged plane.

Example 2.6. Two infinite parallel planes carry equal but opposite uniform charge densities $\pm\sigma$ (Fig. 2.23). Find the field in each of the three regions: (i) to the left of both, (ii) between them, (iii) to the right of both.

Solution

The left plate produces a field $(1/2\epsilon_0)\sigma$, which points away from it (Fig. 2.24)—to the left in region (i) and to the right in regions (ii) and (iii). The right plate, being negatively charged, produces a field $(1/2\epsilon_0)\sigma$, which points *toward* it—to the right in regions (i) and (ii) and to the left in region (iii). The two fields cancel in regions (i) and (iii); they conspire in region (ii). *Conclusion:* The field between the plates is σ/ϵ_0, and points to the right; elsewhere it is zero.

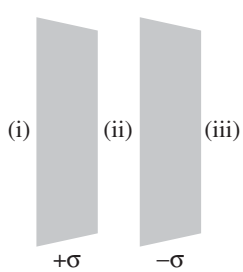


FIGURE 2.23

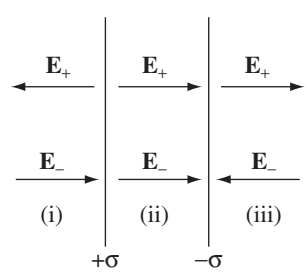


FIGURE 2.24

Problem 2.11 Use Gauss's law to find the electric field inside and outside a spherical shell of radius R that carries a uniform surface charge density σ. Compare your answer to Prob. 2.7.

Problem 2.12 Use Gauss's law to find the electric field inside a uniformly charged solid sphere (charge density ρ). Compare your answer to Prob. 2.8.

Problem 2.13 Find the electric field a distance s from an infinitely long straight wire that carries a uniform line charge λ. Compare Eq. 2.9.

Problem 2.14 Find the electric field inside a sphere that carries a charge density proportional to the distance from the origin, $\rho = kr$, for some constant k. [*Hint:* This charge density is *not* uniform, and you must *integrate* to get the enclosed charge.]

Problem 2.15 A thick spherical shell carries charge density

$$\rho = \frac{k}{r^2} \quad (a \leq r \leq b)$$

(Fig. 2.25). Find the electric field in the three regions: (i) $r < a$, (ii) $a < r < b$, (iii) $r > b$. Plot $|\mathbf{E}|$ as a function of r, for the case $b = 2a$.

Problem 2.16 A long coaxial cable (Fig. 2.26) carries a uniform *volume* charge density ρ on the inner cylinder (radius a), and a uniform *surface* charge density on the outer cylindrical shell (radius b). This surface charge is negative and is of just the right magnitude that the cable as a whole is electrically neutral. Find the electric field in each of the three regions: (i) inside the inner cylinder ($s < a$), (ii) between the cylinders ($a < s < b$), (iii) outside the cable ($s > b$). Plot $|\mathbf{E}|$ as a function of s.

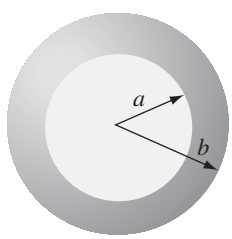


FIGURE 2.25

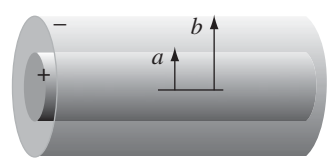


FIGURE 2.26

Problem 2.17 An infinite plane slab, of thickness $2d$, carries a uniform volume charge density ρ (Fig. 2.27). Find the electric field, as a function of y, where $y = 0$ at the center. Plot E versus y, calling E positive when it points in the $+y$ direction and negative when it points in the $-y$ direction.

• **Problem 2.18** Two spheres, each of radius R and carrying uniform volume charge densities $+\rho$ and $-\rho$, respectively, are placed so that they partially overlap (Fig. 2.28). Call the vector from the positive center to the negative center **d**. Show that the field in the region of overlap is constant, and find its value. [*Hint:* Use the answer to Prob. 2.12.]

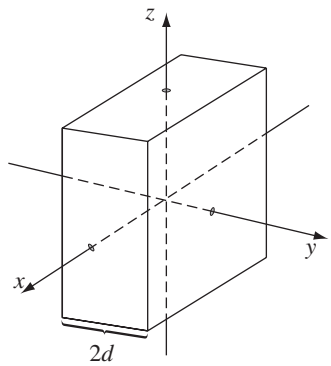


FIGURE 2.27

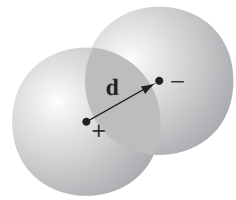


FIGURE 2.28

2.2.4 ■ The Curl of E

I'll calculate the curl of **E**, as I did the divergence in Sect. 2.2.1, by studying first the simplest possible configuration: a point charge at the origin. In this case

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}}.$$

Now, a glance at Fig. 2.12 should convince you that the curl of this field has to be zero, but I suppose we ought to come up with something a little more rigorous than that. What if we calculate the line integral of this field from some point **a** to some other point **b** (Fig. 2.29):

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l}.$$

In spherical coordinates, $d\mathbf{l} = dr\,\hat{\mathbf{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}$, so

$$\mathbf{E} \cdot d\mathbf{l} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}dr.$$

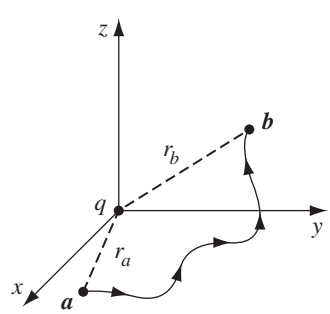


FIGURE 2.29

Therefore,

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l} = \frac{1}{4\pi\epsilon_0} \int_{\mathbf{a}}^{\mathbf{b}} \frac{q}{r^2} dr = \frac{-1}{4\pi\epsilon_0} \frac{q}{r} \Big|_{r_a}^{r_b} = \frac{1}{4\pi\epsilon_0} \left(\frac{q}{r_a} - \frac{q}{r_b} \right), \tag{2.18}$$

where r_a is the distance from the origin to the point **a** and r_b is the distance to **b**. The integral around a *closed* path is evidently zero (for then $r_a = r_b$):

$$\boxed{\oint \mathbf{E} \cdot d\mathbf{l} = 0,} \tag{2.19}$$

and hence, applying Stokes' theorem,

$$\boxed{\nabla \times \mathbf{E} = \mathbf{0}.} \tag{2.20}$$

Now, I proved Eqs. 2.19 and 2.20 only for the field of a single point charge at the origin, but these results make no reference to what is, after all, a perfectly arbitrary choice of coordinates; they hold no matter *where* the charge is located. Moreover, if we have many charges, the principle of superposition states that the total field is a vector sum of their individual fields:

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \ldots,$$

so

$$\nabla \times \mathbf{E} = \nabla \times (\mathbf{E}_1 + \mathbf{E}_2 + \ldots) = (\nabla \times \mathbf{E}_1) + (\nabla \times \mathbf{E}_2) + \ldots = \mathbf{0}.$$

Thus, Eqs. 2.19 and 2.20 hold for *any static charge distribution whatever.*

Problem 2.19 Calculate $\nabla \times \mathbf{E}$ directly from Eq. 2.8, by the method of Sect. 2.2.2. Refer to Prob. 1.63 if you get stuck.

2.3 ■ ELECTRIC POTENTIAL

2.3.1 ■ Introduction to Potential

The electric field **E** is not just *any* old vector function. It is a very special *kind* of vector function: one whose curl is zero. $\mathbf{E} = y\hat{\mathbf{x}}$, for example, could not possibly be an electrostatic field; *no* set of charges, regardless of their sizes and positions, could ever produce such a field. We're going to exploit this special property of electric fields to reduce a *vector* problem (finding **E**) to a much simpler *scalar* problem. The first theorem in Sect. 1.6.2 asserts that any vector whose curl is zero is equal to the gradient of some scalar. What I'm going to do now amounts to a proof of that claim, in the context of electrostatics.

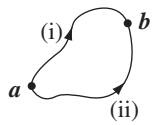


FIGURE 2.30

Because $\nabla \times \mathbf{E} = \mathbf{0}$, the line integral of $\mathbf{E}$ around any closed loop is zero (that follows from Stokes' theorem). Because $\oint \mathbf{E} \cdot d\mathbf{l} = 0$, the line integral of $\mathbf{E}$ from point **a** to point **b** is the same for all paths (otherwise you could go *out* along path (i) and return along path (ii)—Fig. 2.30—and obtain $\oint \mathbf{E} \cdot d\mathbf{l} \neq 0$). Because the line integral is independent of path, we can define a function[6]

$$\boxed{V(\mathbf{r}) \equiv -\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E} \cdot d\mathbf{l}.} \tag{2.21}$$

Here $\mathcal{O}$ is some standard reference point on which we have agreed beforehand; V then depends only on the point $\mathbf{r}$. It is called the **electric potential.**

The potential *difference* between two points **a** and **b** is

$$\begin{aligned} V(\mathbf{b}) - V(\mathbf{a}) &= -\int_{\mathcal{O}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l} + \int_{\mathcal{O}}^{\mathbf{a}} \mathbf{E} \cdot d\mathbf{l} \\ &= -\int_{\mathcal{O}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l} - \int_{\mathbf{a}}^{\mathcal{O}} \mathbf{E} \cdot d\mathbf{l} = -\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l}. \end{aligned} \tag{2.22}$$

Now, the fundamental theorem for gradients states that

$$V(\mathbf{b}) - V(\mathbf{a}) = \int_{\mathbf{a}}^{\mathbf{b}} (\nabla V) \cdot d\mathbf{l},$$

so

$$\int_{\mathbf{a}}^{\mathbf{b}} (\nabla V) \cdot d\mathbf{l} = -\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l}.$$

Since, finally, this is true for *any* points **a** and **b**, the integrands must be equal:

$$\boxed{\mathbf{E} = -\nabla V.} \tag{2.23}$$

[6]To avoid any possible ambiguity, I should perhaps put a prime on the integration variable:

$$V(\mathbf{r}) = -\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E}(\mathbf{r}') \cdot d\mathbf{l}'.$$

But this makes for cumbersome notation, and I prefer whenever possible to reserve the primes for source points. However, when (as in Ex. 2.7) we calculate such integrals explicitly, I will put in the primes.

Equation 2.23 is the differential version of Eq. 2.21; it says that the electric field is the gradient of a scalar potential, which is what we set out to prove.

Notice the subtle but crucial role played by path independence (or, equivalently, the fact that $\nabla \times \mathbf{E} = \mathbf{0}$) in this argument. If the line integral of $\mathbf{E}$ depended on the path taken, then the "definition" of V, Eq. 2.21, would be nonsense. It simply would not define a function, since changing the path would alter the value of $V(\mathbf{r})$. By the way, don't let the minus sign in Eq. 2.23 distract you; it carries over from Eq. 2.21 and is largely a matter of convention.

Problem 2.20 One of these is an impossible electrostatic field. Which one?

(a) $\mathbf{E} = k[xy\,\hat{\mathbf{x}} + 2yz\,\hat{\mathbf{y}} + 3xz\,\hat{\mathbf{z}}]$;

(b) $\mathbf{E} = k[y^2\,\hat{\mathbf{x}} + (2xy + z^2)\,\hat{\mathbf{y}} + 2yz\,\hat{\mathbf{z}}]$.

Here k is a constant with the appropriate units. For the *possible* one, find the potential, using the *origin* as your reference point. Check your answer by computing ∇V. [*Hint:* You must select a specific path to integrate along. It doesn't matter *what* path you choose, since the answer is path-independent, but you simply cannot integrate unless you have a definite path in mind.]

2.3.2 ■ Comments on Potential

(i) The name. The word "potential" is a hideous misnomer because it inevitably reminds you of potential *energy.* This is particularly insidious, because there *is* a connection between "potential" and "potential energy," as you will see in Sect. 2.4. I'm sorry that it is impossible to escape this word. The best I can do is to insist once and for all that "potential" and "potential energy" are completely different terms and should, by all rights, have different names. Incidentally, a surface over which the potential is constant is called an **equipotential**.

(ii) Advantage of the potential formulation. If you know V, you can easily get $\mathbf{E}$—just take the gradient: $\mathbf{E} = -\nabla V$. This is quite extraordinary when you stop to think about it, for $\mathbf{E}$ is a *vector* quantity (three components), but V is a *scalar* (one component). How can *one* function possibly contain all the information that *three* independent functions carry? The answer is that the three components of $\mathbf{E}$ are not really as independent as they look; in fact, they are explicitly interrelated by the very condition we started with, $\nabla \times \mathbf{E} = \mathbf{0}$. In terms of components,

$$\frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x}, \qquad \frac{\partial E_z}{\partial y} = \frac{\partial E_y}{\partial z}, \qquad \frac{\partial E_x}{\partial z} = \frac{\partial E_z}{\partial x}.$$

This brings us back to my observation at the beginning of Sect. 2.3.1: $\mathbf{E}$ *is a very special kind of vector.* What the potential formulation does is to exploit this feature to maximum advantage, reducing a vector problem to a scalar one, in which there is no need to fuss with components.

(iii) The reference point $\mathcal{O}$. There is an essential ambiguity in the definition of potential, since the choice of reference point $\mathcal{O}$ was arbitrary. Changing reference points amounts to adding a constant K to the potential:

$$V'(\mathbf{r}) = -\int_{\mathcal{O}'}^{\mathbf{r}} \mathbf{E}\cdot d\mathbf{l} = -\int_{\mathcal{O}'}^{\mathcal{O}} \mathbf{E}\cdot d\mathbf{l} - \int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E}\cdot d\mathbf{l} = K + V(\mathbf{r}),$$

where K is the line integral of $\mathbf{E}$ from the old reference point $\mathcal{O}$ to the new one $\mathcal{O}'$. Of course, adding a constant to V will not affect the potential *difference* between two points:

$$V'(\mathbf{b}) - V'(\mathbf{a}) = V(\mathbf{b}) - V(\mathbf{a}),$$

since the K's cancel out. (Actually, it was already clear from Eq. 2.22 that the potential difference is independent of $\mathcal{O}$, because it can be written as the line integral of $\mathbf{E}$ from $\mathbf{a}$ to $\mathbf{b}$, with no reference to $\mathcal{O}$.) Nor does the ambiguity affect the gradient of V:

$$\boldsymbol{\nabla} V' = \boldsymbol{\nabla} V,$$

since the derivative of a constant is zero. That's why all such V's, differing only in their choice of reference point, correspond to the same field $\mathbf{E}$.

Potential as such carries no real physical significance, for at any given point we can adjust its value at will by a suitable relocation of $\mathcal{O}$. In this sense, it is rather like altitude: If I ask you how high Denver is, you will probably tell me its height above sea level, because that is a convenient and traditional reference point. But we could as well agree to measure altitude above Washington, D.C., or Greenwich, or wherever. That would add (or, rather, subtract) a fixed amount from all our sea-level readings, but it wouldn't change anything about the real world. The only quantity of intrinsic interest is the *difference* in altitude between two points, and *that* is the same *whatever* your reference level.

Having said this, however, there *is* a "natural" spot to use for $\mathcal{O}$ in electrostatics—analogous to sea level for altitude—and that is a point infinitely far from the charge. Ordinarily, then, we "set the zero of potential at infinity." (Since $V(\mathcal{O}) = 0$, choosing a reference point is equivalent to selecting a place where V is to be zero.) But I must warn you that there is one special circumstance in which this convention fails: when the charge distribution itself extends to infinity. The symptom of trouble, in such cases, is that the potential blows up. For instance, the field of a uniformly charged plane is $(\sigma/2\epsilon_0)\hat{\mathbf{n}}$, as we found in Ex. 2.5; if we naïvely put $\mathcal{O} = \infty$, then the potential at height z above the plane becomes

$$V(z) = -\int_{\infty}^{z} \frac{1}{2\epsilon_0}\sigma\, dz = -\frac{1}{2\epsilon_0}\sigma(z - \infty).$$

The remedy is simply to choose some other reference point (in this example you might use a point on the plane). Notice that the difficulty occurs only in textbook problems; in "real life" there is no such thing as a charge distribution that goes on forever, and we can *always* use infinity as our reference point.

(iv) Potential obeys the superposition principle. The original superposition principle pertains to the force on a test charge Q. It says that the total force on Q is the vector sum of the forces attributable to the source charges individually:

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \dots$$

Dividing through by Q, we see that the electric field, too, obeys the superposition principle:

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \dots$$

Integrating from the common reference point to $\mathbf{r}$, it follows that the potential also satisfies such a principle:

$$V = V_1 + V_2 + \dots$$

That is, the potential at any given point is the sum of the potentials due to all the source charges separately. Only this time it is an *ordinary* sum, not a *vector* sum, which makes it a lot easier to work with.

(v) Units of Potential. In our units, force is measured in newtons and charge in coulombs, so electric fields are in newtons per coulomb. Accordingly, potential is newton-meters per coulomb, or joules per coulomb. A joule per coulomb is a **volt**.

Example 2.7. Find the potential inside and outside a spherical shell of radius R (Fig. 2.31) that carries a uniform surface charge. Set the reference point at infinity.

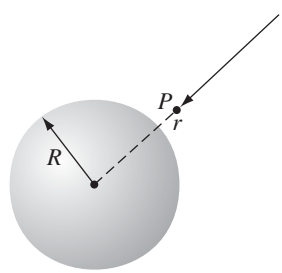


FIGURE 2.31

Solution

From Gauss's law, the field outside is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}},$$

where q is the total charge on the sphere. The field inside is zero. For points outside the sphere ($r > R$),

$$V(r) = -\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E}\cdot d\mathbf{l} = \frac{-1}{4\pi\epsilon_0}\int_{\infty}^{r}\frac{q}{r'^2}\,dr' = \frac{1}{4\pi\epsilon_0}\frac{q}{r'}\Big|_{\infty}^{r} = \frac{1}{4\pi\epsilon_0}\frac{q}{r}.$$

To find the potential inside the sphere ($r < R$), we must break the integral into two pieces, using in each region the field that prevails there:

$$V(r) = \frac{-1}{4\pi\epsilon_0}\int_\infty^R \frac{q}{r'^2}\,dr' - \int_R^r (0)\,dr' = \frac{1}{4\pi\epsilon_0}\frac{q}{r'}\bigg|_\infty^R + 0 = \frac{1}{4\pi\epsilon_0}\frac{q}{R}.$$

Notice that the potential is *not* zero inside the shell, even though the field is. V is a *constant* in this region, to be sure, so that $\boldsymbol{\nabla} V = \mathbf{0}$—that's what matters. In problems of this type, you must always *work your way in from the reference point;* that's where the potential is "nailed down." It is tempting to suppose that you could figure out the potential inside the sphere on the basis of the field there alone, but this is false: The potential inside the sphere is sensitive to what's going on outside the sphere as well. If I placed a second uniformly charged shell out at radius $R' > R$, the potential inside R would change, even though the field would still be zero. Gauss's law guarantees that charge exterior to a given point (that is, at larger r) produces no net *field* at that point, provided it is spherically or cylindrically symmetric, but there is no such rule for *potential,* when infinity is used as the reference point.

Problem 2.21 Find the potential inside and outside a uniformly charged solid sphere whose radius is R and whose total charge is q. Use infinity as your reference point. Compute the gradient of V in each region, and check that it yields the correct field. Sketch $V(r)$.

Problem 2.22 Find the potential a distance s from an infinitely long straight wire that carries a uniform line charge λ. Compute the gradient of your potential, and check that it yields the correct field.

Problem 2.23 For the charge configuration of Prob. 2.15, find the potential at the center, using infinity as your reference point.

Problem 2.24 For the configuration of Prob. 2.16, find the potential difference between a point on the axis and a point on the outer cylinder. Note that it is not necessary to commit yourself to a particular reference point, if you use Eq. 2.22.

2.3.3 ■ Poisson's Equation and Laplace's Equation

We found in Sect. 2.3.1 that the electric field can be written as the gradient of a scalar potential.

$$\mathbf{E} = -\boldsymbol{\nabla} V.$$

The question arises: What do the divergence and curl of $\mathbf{E}$,

$$\boldsymbol{\nabla}\cdot\mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \text{and} \qquad \boldsymbol{\nabla}\times\mathbf{E} = \mathbf{0},$$

look like, in terms of V? Well, $\boldsymbol{\nabla} \cdot \mathbf{E} = \boldsymbol{\nabla} \cdot (-\boldsymbol{\nabla} V) = -\nabla^2 V$, so, apart from that persistent minus sign, the divergence of $\mathbf{E}$ is the Laplacian of V. Gauss's law, then, says

$$\boxed{\nabla^2 V = -\frac{\rho}{\epsilon_0}.} \tag{2.24}$$

This is known as **Poisson's equation**. In regions where there is no charge, so $\rho = 0$, Poisson's equation reduces to **Laplace's equation**,

$$\nabla^2 V = 0. \tag{2.25}$$

We'll explore this equation more fully in Chapter 3.

So much for Gauss's law. What about the curl law? This says that

$$\boldsymbol{\nabla} \times \mathbf{E} = \boldsymbol{\nabla} \times (-\boldsymbol{\nabla} V) = \mathbf{0}.$$

But that's no condition on V—curl of gradient is *always* zero. Of course, we *used* the curl law to show that $\mathbf{E}$ could be expressed as the gradient of a scalar, so it's not really surprising that this works out: $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$ *permits* $\mathbf{E} = -\boldsymbol{\nabla} V$; in return, $\mathbf{E} = -\boldsymbol{\nabla} V$ *guarantees* $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$. It takes only *one* differential equation (Poisson's) to determine V, because V is a scalar; for $\mathbf{E}$ we needed *two,* the divergence and the curl.

2.3.4 ■ The Potential of a Localized Charge Distribution

I defined V in terms of $\mathbf{E}$ (Eq. 2.21). Ordinarily, though, it's $\mathbf{E}$ that we're looking for (if we already knew $\mathbf{E}$, there wouldn't be much point in calculating V). The idea is that it might be easier to get V first, and then calculate $\mathbf{E}$ by taking the gradient. Typically, then, we know where the charge is (that is, we know ρ), and we want to find V. Now, Poisson's equation relates V and ρ, but unfortunately it's "the wrong way around": it would give us ρ, if we knew V, whereas we want V, knowing ρ. What we must do, then, is "invert" Poisson's equation. That's the program for this section, although I shall do it by roundabout means, beginning, as always, with a point charge at the origin.

The electric field is $\mathbf{E} = (1/4\pi\epsilon_0)(1/r^2)\,\hat{\mathbf{r}}$, and $d\mathbf{l} = dr\,\hat{\mathbf{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}$ (Eq. 1.68), so

$$\mathbf{E} \cdot d\mathbf{l} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\,dr.$$

Setting the reference point at infinity, the potential of a point charge q at the origin is

$$V(r) = -\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{E} \cdot d\mathbf{l} = \frac{-1}{4\pi\epsilon_0}\int_{\infty}^{r} \frac{q}{r'^2}\,dr' = \frac{1}{4\pi\epsilon_0}\frac{q}{r'}\bigg|_{\infty}^{r} = \frac{1}{4\pi\epsilon_0}\frac{q}{r}.$$

(You see here the advantage of using infinity for the reference point: it kills the lower limit on the integral.) Notice the sign of V; presumably the conventional

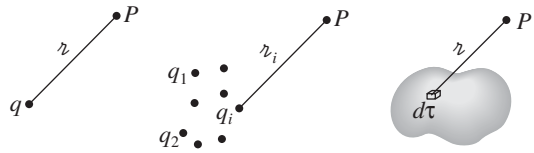


FIGURE 2.32

minus sign in the definition (Eq. 2.21) was chosen in order to make the potential of a positive charge come out positive. It is useful to remember that regions of positive charge are potential "hills," regions of negative charge are potential "valleys," and the electric field points "downhill," from plus toward minus.

In general, the potential of a point charge q is

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q}{\imath}, \tag{2.26}$$

where $\imath$, as always, is the distance from q to $\mathbf{r}$ (Fig. 2.32). Invoking the superposition principle, then, the potential of a collection of charges is

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{n}\frac{q_i}{\imath_i}, \tag{2.27}$$

or, for a continuous distribution,

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{\imath}dq. \tag{2.28}$$

In particular, for a volume charge, it's

$$\boxed{V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{\imath}\,d\tau'.} \tag{2.29}$$

This is the equation we were looking for, telling us how to compute V when we know ρ; it is, if you like, the "solution" to Poisson's equation, for a localized charge distribution.[7] Compare Eq. 2.29 with the corresponding formula for the electric *field* in terms of ρ (Eq. 2.8):

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{\imath^2}\hat{\boldsymbol{\imath}}\,d\tau'.$$

The main point to notice is that the pesky unit vector $\hat{\boldsymbol{\imath}}$ is gone, so there is no need to fuss with components. The potentials of line and surface charges are

$$V = \frac{1}{4\pi\epsilon_0}\int \frac{\lambda(\mathbf{r}')}{\imath}\,dl' \quad \text{and} \quad V = \frac{1}{4\pi\epsilon_0}\int \frac{\sigma(\mathbf{r}')}{\imath}\,da'. \tag{2.30}$$

I should warn you that everything in this section is predicated on the assumption that the reference point is at infinity. This is hardly apparent in Eq. 2.29, but

[7]Equation 2.29 is an example of the Helmholtz theorem (Appendix B).

remember that we *got* that equation from the potential of a point charge at the origin, $(1/4\pi\epsilon_0)(q/r)$, which is valid only when $\mathcal{O} = \infty$. If you try to apply these formulas to one of those artificial problems in which the charge itself extends to infinity, the integral will diverge.

Example 2.8. Find the potential of a uniformly charged spherical shell of radius R (Fig. 2.33).

Solution

This is the same problem we solved in Ex. 2.7, but this time let's do it using Eq. 2.30:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\sigma}{\imath}\, da'.$$

We might as well set the point P on the z axis and use the law of cosines to express $\imath$:

$$\imath^2 = R^2 + z^2 - 2Rz\cos\theta'.$$

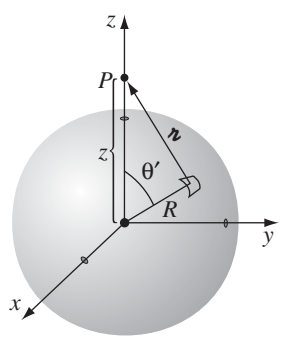


FIGURE 2.33

An element of surface area on the sphere is $R^2\sin\theta'\,d\theta'\,d\phi'$, so

$$\begin{aligned}
4\pi\epsilon_0 V(z) &= \sigma\int \frac{R^2\sin\theta'\,d\theta'\,d\phi'}{\sqrt{R^2+z^2-2Rz\cos\theta'}} \\
&= 2\pi R^2\sigma\int_0^\pi \frac{\sin\theta'}{\sqrt{R^2+z^2-2Rz\cos\theta'}}\,d\theta' \\
&= 2\pi R^2\sigma\left(\frac{1}{Rz}\sqrt{R^2+z^2-2Rz\cos\theta'}\right)\Bigg|_0^\pi \\
&= \frac{2\pi R\sigma}{z}\left(\sqrt{R^2+z^2+2Rz}-\sqrt{R^2+z^2-2Rz}\right) \\
&= \frac{2\pi R\sigma}{z}\left[\sqrt{(R+z)^2}-\sqrt{(R-z)^2}\right].
\end{aligned}$$

At this stage, we must be very careful to take the *positive* root. For points *outside* the sphere, z is greater than R, and hence $\sqrt{(R-z)^2} = z - R$; for points *inside* the sphere, $\sqrt{(R-z)^2} = R - z$. Thus,

$$V(z) = \frac{R\sigma}{2\epsilon_0 z}[(R+z) - (z-R)] = \frac{R^2\sigma}{\epsilon_0 z}, \qquad \text{outside;}$$

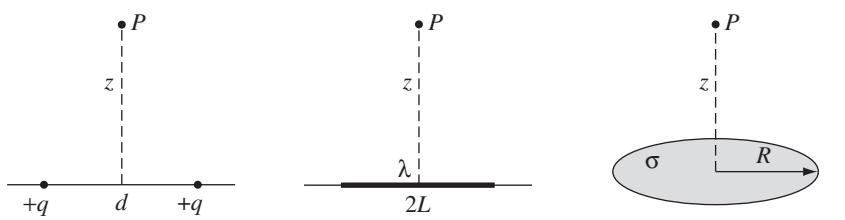
$$V(z) = \frac{R\sigma}{2\epsilon_0 z}[(R+z) - (R-z)] = \frac{R\sigma}{\epsilon_0}, \qquad \text{inside.}$$

In terms of r and the total charge on the shell, $q = 4\pi R^2\sigma$,

$$V(r) = \begin{cases} \dfrac{1}{4\pi\epsilon_0}\dfrac{q}{r} & (r \geq R), \\ \dfrac{1}{4\pi\epsilon_0}\dfrac{q}{R} & (r \leq R). \end{cases}$$

Of course, in this particular case, it was easier to get V by using Eq. 2.21 than Eq. 2.30, because Gauss's law gave us **E** with so little effort. But if you compare Ex. 2.8 with Prob. 2.7, you will appreciate the power of the potential formulation.

Problem 2.25 Using Eqs. 2.27 and 2.30, find the potential at a distance z above the center of the charge distributions in Fig. 2.34. In each case, compute $\mathbf{E} = -\nabla V$, and compare your answers with Ex. 2.1, Ex. 2.2, and Prob. 2.6, respectively. Suppose that we changed the right-hand charge in Fig. 2.34a to $-q$; what then is the potential at P? What field does that suggest? Compare your answer to Prob. 2.2, and explain carefully any discrepancy.



(a) Two point charges (b) Uniform line charge (c) Uniform surface charge

FIGURE 2.34

Problem 2.26 A conical surface (an empty ice-cream cone) carries a uniform surface charge σ. The height of the cone is h, as is the radius of the top. Find the potential difference between points **a** (the vertex) and **b** (the center of the top).

Problem 2.27 Find the potential on the axis of a uniformly charged solid cylinder, a distance z from the center. The length of the cylinder is L, its radius is R, and the charge density is ρ. Use your result to calculate the electric field at this point. (Assume that $z > L/2$.)

Problem 2.28 Use Eq. 2.29 to calculate the potential inside a uniformly charged solid sphere of radius R and total charge q. Compare your answer to Prob. 2.21.

Problem 2.29 Check that Eq. 2.29 satisfies Poisson's equation, by applying the Laplacian and using Eq. 1.102.

2.3.5 ■ Boundary Conditions

In the typical electrostatic problem you are given a source charge distribution ρ, and you want to find the electric field $\mathbf{E}$ it produces. Unless the symmetry of the problem allows a solution by Gauss's law, it is generally to your advantage to calculate the potential first, as an intermediate step. These are the three fundamental quantities of electrostatics: ρ, $\mathbf{E}$, and V. We have, in the course of our discussion, derived all six formulas interrelating them. These equations are neatly summarized in Fig. 2.35. We began with just two experimental observations: (1) the principle of superposition—a broad general rule applying to *all* electromagnetic forces, and (2) Coulomb's law—the fundamental law of electrostatics. From these, all else followed.

You may have noticed, in studying Exs. 2.5 and 2.6, or working problems such as 2.7, 2.11, and 2.16, that the electric field always undergoes a discontinuity when you cross a surface charge σ. In fact, it is a simple matter to find the *amount* by which $\mathbf{E}$ changes at such a boundary. Suppose we draw a wafer-thin Gaussian pillbox, extending just barely over the edge in each direction (Fig. 2.36). Gauss's law says that

$$\oint_S \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{enc}} = \frac{1}{\epsilon_0} \sigma A,$$

where A is the area of the pillbox lid. (If σ varies from point to point or the surface is curved, we must pick A to be extremely small.) Now, the *sides* of the pillbox

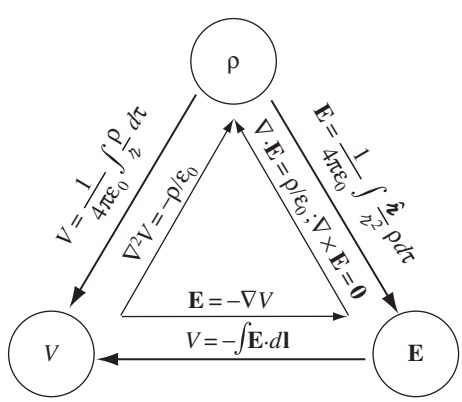


FIGURE 2.35

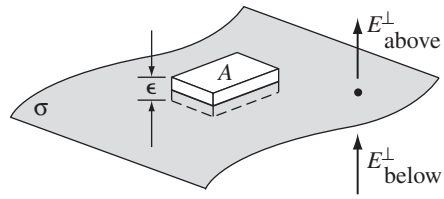


FIGURE 2.36

contribute nothing to the flux, in the limit as the thickness ϵ goes to zero, so we are left with

$$E^{\perp}_{\text{above}} - E^{\perp}_{\text{below}} = \frac{1}{\epsilon_0}\sigma, \tag{2.31}$$

where $E^{\perp}_{\text{above}}$ denotes the component of **E** that is perpendicular to the surface immediately above, and $E^{\perp}_{\text{below}}$ is the same, only just below the surface. For consistency, we let "upward" be the positive direction for both. *Conclusion: The normal component of* **E** *is discontinuous by an amount* σ/ϵ_0 *at any boundary.* In particular, where there is *no* surface charge, $E^{\perp}$ is continuous, as for instance at the surface of a uniformly charged solid sphere.

The *tangential* component of **E**, by contrast, is *always* continuous. For if we apply Eq. 2.19,

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0,$$

to the thin rectangular loop of Fig. 2.37, the ends give nothing (as $\epsilon \to 0$), and the sides give $(E^{\parallel}_{\text{above}} l - E^{\parallel}_{\text{below}} l)$, so

$$\mathbf{E}^{\parallel}_{\text{above}} = \mathbf{E}^{\parallel}_{\text{below}}, \tag{2.32}$$

where $\mathbf{E}^{\parallel}$ stands for the components of **E** *parallel* to the surface. The boundary conditions on **E** (Eqs. 2.31 and 2.32) can be combined into a single formula:

$$\mathbf{E}_{\text{above}} - \mathbf{E}_{\text{below}} = \frac{\sigma}{\epsilon_0}\hat{\mathbf{n}}, \tag{2.33}$$

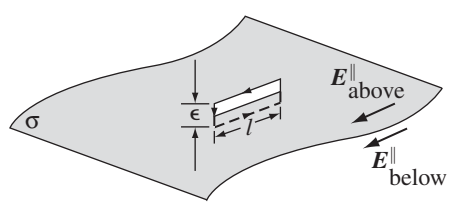


FIGURE 2.37

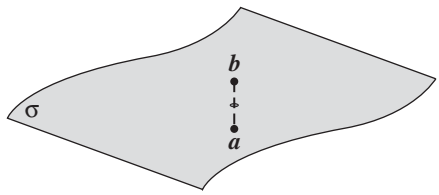


FIGURE 2.38

where $\hat{\mathbf{n}}$ is a unit vector perpendicular to the surface, pointing from "below" to "above."[8]

The potential, meanwhile, is continuous across *any* boundary (Fig. 2.38), since

$$V_{\text{above}} - V_{\text{below}} = -\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l};$$

as the path length shrinks to zero, so too does the integral:

$$V_{\text{above}} = V_{\text{below}}. \tag{2.34}$$

However, the *gradient* of V inherits the discontinuity in $\mathbf{E}$; since $\mathbf{E} = -\boldsymbol{\nabla} V$, Eq. 2.33 implies that

$$\boldsymbol{\nabla} V_{\text{above}} - \boldsymbol{\nabla} V_{\text{below}} = -\frac{1}{\epsilon_0}\sigma \hat{\mathbf{n}}, \tag{2.35}$$

or, more conveniently,

$$\frac{\partial V_{\text{above}}}{\partial n} - \frac{\partial V_{\text{below}}}{\partial n} = -\frac{1}{\epsilon_0}\sigma, \tag{2.36}$$

where

$$\frac{\partial V}{\partial n} = \boldsymbol{\nabla} V \cdot \hat{\mathbf{n}} \tag{2.37}$$

denotes the **normal derivative** of V (that is, the rate of change in the direction perpendicular to the surface).

Please note that these boundary conditions relate the fields and potentials *just* above and *just* below the surface. For example, the derivatives in Eq. 2.36 are the *limiting* values as we approach the surface from either side.

[8]Notice that it doesn't matter which side you call "above" and which "below," since reversal would switch the direction of $\hat{\mathbf{n}}$. Incidentally, if you're only interested in the field due to the (essentially flat) local patch of surface charge itself, the answer is $(\sigma/2\epsilon_0)\hat{\mathbf{n}}$ immediately above the surface, and $-(\sigma/2\epsilon_0)\hat{\mathbf{n}}$ immediately below. This follows from Ex. 2.5, for if you are close enough to the patch it "looks" like an infinite plane. Evidently the entire discontinuity in $\mathbf{E}$ is attributable to this local patch of surface charge.

Problem 2.30

(a) Check that the results of Exs. 2.5 and 2.6, and Prob. 2.11, are consistent with Eq. 2.33.

(b) Use Gauss's law to find the field inside and outside a long hollow cylindrical tube, which carries a uniform surface charge σ. Check that your result is consistent with Eq. 2.33.

(c) Check that the result of Ex. 2.8 is consistent with boundary conditions 2.34 and 2.36.

2.4 ■ WORK AND ENERGY IN ELECTROSTATICS

2.4.1 ■ The Work It Takes to Move a Charge

Suppose you have a stationary configuration of source charges, and you want to move a test charge Q from point **a** to point **b** (Fig. 2.39). *Question:* How much work will you have to do? At any point along the path, the electric force on Q is $\mathbf{F} = Q\mathbf{E}$; the force *you* must exert, in opposition to this electrical force, is $-Q\mathbf{E}$. (If the sign bothers you, think about lifting a brick: gravity exerts a force mg *downward,* but *you* exert a force mg *upward.* Of course, you *could* apply an even greater force—then the brick would accelerate, and part of your effort would be "wasted" generating kinetic energy. What we're interested in here is the *minimum* force you must exert to do the job.) The work you do is therefore

$$W = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{F} \cdot d\mathbf{l} = -Q \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l} = Q[V(\mathbf{b}) - V(\mathbf{a})].$$

Notice that the answer is independent of the path you take from **a** to **b**; in mechanics, then, we would call the electrostatic force "conservative." Dividing through by Q, we have

$$V(\mathbf{b}) - V(\mathbf{a}) = \frac{W}{Q}. \tag{2.38}$$

In words, *the potential difference between points* **a** *and* **b** *is equal to the work per unit charge required to carry a particle from* **a** *to* **b**. In particular, if you want to bring Q in from far away and stick it at point **r**, the work you must do is

$$W = Q[V(\mathbf{r}) - V(\infty)],$$

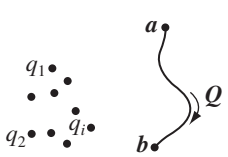


FIGURE 2.39

so, if you have set the reference point at infinity,

$$W = QV(\mathbf{r}). \tag{2.39}$$

In this sense, *potential* is potential *energy* (the work it takes to create the system) *per unit charge* (just as the *field* is the *force* per unit charge).

2.4.2 ■ The Energy of a Point Charge Distribution

How much work would it take to assemble an entire *collection* of point charges? Imagine bringing in the charges, one by one, from far away (Fig. 2.40). The first charge, q_1, takes *no* work, since there is no field yet to fight against. Now bring in q_2. According to Eq. 2.39, this will cost you $q_2 V_1(\mathbf{r}_2)$, where V_1 is the potential due to q_1, and $\mathbf{r}_2$ is the place we're putting q_2:

$$W_2 = \frac{1}{4\pi\epsilon_0} q_2 \left(\frac{q_1}{r_{12}}\right)$$

(r_{12} is the distance between q_1 and q_2 once they are in position). As you bring in each charge, nail it down in its final location, so it doesn't move when you bring in the next charge. Now bring in q_3; this requires work $q_3 V_{1,2}(\mathbf{r}_3)$, where $V_{1,2}$ is the potential due to charges q_1 and q_2, namely, $(1/4\pi\epsilon_0)(q_1/r_{13} + q_2/r_{23})$. Thus

$$W_3 = \frac{1}{4\pi\epsilon_0} q_3 \left(\frac{q_1}{r_{13}} + \frac{q_2}{r_{23}}\right).$$

Similarly, the extra work to bring in q_4 will be

$$W_4 = \frac{1}{4\pi\epsilon_0} q_4 \left(\frac{q_1}{r_{14}} + \frac{q_2}{r_{24}} + \frac{q_3}{r_{34}}\right).$$

The *total* work necessary to assemble the first four charges, then, is

$$W = \frac{1}{4\pi\epsilon_0} \left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_1 q_4}{r_{14}} + \frac{q_2 q_3}{r_{23}} + \frac{q_2 q_4}{r_{24}} + \frac{q_3 q_4}{r_{34}}\right).$$

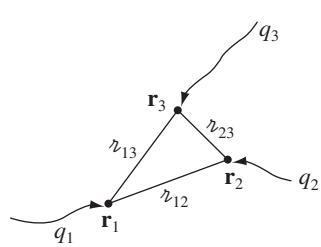


FIGURE 2.40

You see the general rule: Take the product of each pair of charges, divide by their separation distance, and add it all up:

$$W = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{q_i q_j}{r_{ij}}. \tag{2.40}$$

The stipulation $j > i$ is to remind you not to count the same pair twice. A nicer way to accomplish this is *intentionally* to count each pair twice, and then divide by 2:

$$W = \frac{1}{8\pi\epsilon_0} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{q_i q_j}{r_{ij}} \tag{2.41}$$

(we must still avoid $i = j$, of course). Notice that in this form the answer plainly does not depend on the *order* in which you assemble the charges, since every pair occurs in the sum.

Finally, let's pull out the factor q_i:

$$W = \frac{1}{2} \sum_{i=1}^{n} q_i \left(\sum_{j\neq i}^{n} \frac{1}{4\pi\epsilon_0} \frac{q_j}{r_{ij}} \right).$$

The term in parentheses is the potential at point $\mathbf{r}_i$ (the position of q_i) due to all the *other* charges—all of them, now, not just the ones that were present at some stage during the assembly. Thus,

$$W = \frac{1}{2} \sum_{i=1}^{n} q_i V(\mathbf{r}_i). \tag{2.42}$$

That's how much work it takes to assemble a configuration of point charges; it's also the amount of work you'd get back if you dismantled the system. In the meantime, it represents energy stored in the configuration ("potential" energy, if you insist, though for obvious reasons I prefer to avoid that word in this context).

Problem 2.31

(a) Three charges are situated at the corners of a square (side a), as shown in Fig. 2.41. How much work does it take to bring in another charge, $+q$, from far away and place it in the fourth corner?

(b) How much work does it take to assemble the whole configuration of four charges?

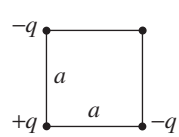


FIGURE 2.41

Problem 2.32 Two positive point charges, q_A and q_B (masses m_A and m_B) are at rest, held together by a massless string of length a. Now the string is cut, and the particles fly off in opposite directions. How fast is each one going, when they are far apart?

Problem 2.33 Consider an infinite chain of point charges, $\pm q$ (with alternating signs), strung out along the x axis, each a distance a from its nearest neighbors. Find the work per particle required to assemble this system. [*Partial Answer:* $-\alpha q^2/(4\pi\epsilon_0 a)$, for some dimensionless number α; your problem is to determine α. It is known as the **Madelung constant**. Calculating the Madelung constant for 2- and 3-dimensional arrays is much more subtle and difficult.]

2.4.3 ■ The Energy of a Continuous Charge Distribution

For a volume charge density ρ, Eq. 2.42 becomes

$$W = \frac{1}{2}\int \rho V \, d\tau. \tag{2.43}$$

(The corresponding integrals for line and surface charges would be $\int \lambda V \, dl$ and $\int \sigma V \, da$.) There is a lovely way to rewrite this result, in which ρ and V are eliminated in favor of $\mathbf{E}$. First use Gauss's law to express ρ in terms of $\mathbf{E}$:

$$\rho = \epsilon_0 \nabla \cdot \mathbf{E}, \quad \text{so} \quad W = \frac{\epsilon_0}{2}\int (\nabla \cdot \mathbf{E}) V \, d\tau.$$

Now use integration by parts (Eq. 1.59) to transfer the derivative from $\mathbf{E}$ to V:

$$W = \frac{\epsilon_0}{2}\left[-\int \mathbf{E} \cdot (\nabla V)\, d\tau + \oint V\mathbf{E} \cdot d\mathbf{a}\right].$$

But $\nabla V = -\mathbf{E}$, so

$$W = \frac{\epsilon_0}{2}\left(\int_{\mathcal{V}} E^2 \, d\tau + \oint_{\mathcal{S}} V\mathbf{E} \cdot d\mathbf{a}\right). \tag{2.44}$$

But what volume *is* this we're integrating over? Let's go back to the formula we started with, Eq. 2.43. From its derivation, it is clear that we should integrate over the region where the charge is located. But actually, any *larger* volume would do just as well: The "extra" territory we throw in will contribute nothing to the integral, since $\rho = 0$ out there. With this in mind, we return to Eq. 2.44. What happens *here,* as we enlarge the volume beyond the minimum necessary to trap all the charge? Well, the integral of E^2 can only increase (the integrand being positive); evidently the surface integral must decrease correspondingly to leave the sum intact. (In fact, at large distances from the charge, E goes like $1/r^2$ and V like $1/r$, while the surface area grows like r^2; roughly speaking, then, the surface integral goes down like $1/r$.) Please understand: Eq. 2.44 gives you the correct

energy W, *whatever* volume you use (as long as it encloses all the charge), but the contribution from the volume integral goes up, and that of the surface integral goes down, as you take larger and larger volumes. In particular, why not integrate over *all* space? Then the surface integral goes to zero, and we are left with

$$\boxed{W = \frac{\epsilon_0}{2}\int E^2\, d\tau} \quad \text{(all space)}. \tag{2.45}$$

Example 2.9. Find the energy of a uniformly charged spherical shell of total charge q and radius R.

Solution 1
Use Eq. 2.43, in the version appropriate to surface charges:

$$W = \frac{1}{2}\int \sigma V\, da.$$

Now, the potential at the surface of this sphere is $(1/4\pi\epsilon_0)q/R$ (a constant—Ex. 2.7), so

$$W = \frac{1}{8\pi\epsilon_0}\frac{q}{R}\int \sigma\, da = \frac{1}{8\pi\epsilon_0}\frac{q^2}{R}.$$

Solution 2
Use Eq. 2.45. Inside the sphere, $\mathbf{E} = \mathbf{0}$; outside,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}}, \quad \text{so} \quad E^2 = \frac{q^2}{(4\pi\epsilon_0)^2 r^4}.$$

Therefore,

$$W_{\text{tot}} = \frac{\epsilon_0}{2(4\pi\epsilon_0)^2}\int_{\text{outside}} \left(\frac{q^2}{r^4}\right)(r^2\sin\theta\, dr\, d\theta\, d\phi)$$

$$= \frac{1}{32\pi^2\epsilon_0}q^2 4\pi\int_R^\infty \frac{1}{r^2}\, dr = \frac{1}{8\pi\epsilon_0}\frac{q^2}{R}.$$

Problem 2.34 Find the energy stored in a uniformly charged solid sphere of radius R and charge q. Do it three different ways:

(a) Use Eq. 2.43. You found the potential in Prob. 2.21.

(b) Use Eq. 2.45. Don't forget to integrate over *all space.*

(c) Use Eq. 2.44. Take a spherical volume of radius a. What happens as $a \to \infty$?

Problem 2.35 Here is a fourth way of computing the energy of a uniformly charged solid sphere: Assemble it like a snowball, layer by layer, each time bringing in an infinitesimal charge dq from far away and smearing it uniformly over the surface, thereby increasing the radius. How much work dW does it take to build up the radius by an amount dr? Integrate this to find the work necessary to create the entire sphere of radius R and total charge q.

2.4.4 ■ Comments on Electrostatic Energy

(i) A perplexing "inconsistency." Equation 2.45 clearly implies that the energy of a stationary charge distribution is always *positive.* On the other hand, Eq. 2.42 (from which 2.45 was in fact derived), can be positive or negative. For instance, according to Eq. 2.42, the energy of two equal but opposite charges a distance $\imath$ apart is $-(1/4\pi\epsilon_0)(q^2/\imath)$. What's gone wrong? Which equation is correct?

The answer is that *both* are correct, but they speak to slightly different questions. Equation 2.42 does not take into account the work necessary to *make* the point charges in the first place; we *started* with point charges and simply found the work required to bring them together. This is wise strategy, since Eq. 2.45 indicates that the energy of a point charge is in fact *infinite*:

$$W = \frac{\epsilon_0}{2(4\pi\epsilon_0)^2}\int\left(\frac{q^2}{r^4}\right)(r^2\sin\theta\, dr\, d\theta\, d\phi) = \frac{q^2}{8\pi\epsilon_0}\int_0^\infty \frac{1}{r^2}\, dr = \infty.$$

Equation 2.45 is more *complete,* in the sense that it tells you the *total* energy stored in a charge configuration, but Eq. 2.42 is more appropriate when you're dealing with point charges, because we prefer (for good reason!) to leave out that portion of the total energy that is attributable to the fabrication of the point charges themselves. In practice, after all, the point charges (electrons, say) are *given* to us ready-made; all *we* do is move them around. Since we did not put them together, and we cannot take them apart, it is immaterial how much work the process would involve. (Still, the infinite energy of a point charge is a recurring source of embarrassment for electromagnetic theory, afflicting the quantum version as well as the classical. We shall return to the problem in Chapter 11.)

Now, you may wonder where the inconsistency crept into an apparently watertight derivation. The "flaw" lies between Eqs. 2.42 and 2.43: in the former, $V(\mathbf{r}_i)$ represents the potential due to all the *other* charges *but not* q_i, whereas in the latter, $V(\mathbf{r})$ is the *full* potential. For a *continuous* distribution, there is no distinction, since the amount of charge *right at the point* $\mathbf{r}$ is vanishingly small, and its contribution to the potential is zero. But in the presence of point charges you'd better stick with Eq. 2.42.

(ii) Where is the energy stored? Equations 2.43 and 2.45 offer two different ways of calculating the same thing. The first is an integral over the charge distribution; the second is an integral over the field. These can involve completely different regions. For instance, in the case of the spherical shell (Ex. 2.9) the charge is confined to the surface, whereas the electric field is everywhere *outside*

this surface. Where *is* the energy, then? Is it stored in the field, as Eq. 2.45 seems to suggest, or is it stored in the charge, as Eq. 2.43 implies? At the present stage this is simply an unanswerable question: I can tell you what the total energy is, and I can provide you with several different ways to compute it, but it is impertinent to worry about *where* the energy is located. In the context of radiation theory (Chapter 11) it is useful (and in general relativity it is *essential*) to regard the energy as stored in the field, with a density

$$\frac{\epsilon_0}{2}E^2 = \text{ energy per unit volume.} \tag{2.46}$$

But in electrostatics one could just as well say it is stored in the charge, with a density $\frac{1}{2}\rho V$. The difference is purely a matter of bookkeeping.

(iii) The superposition principle. Because electrostatic energy is *quadratic* in the fields, it does *not* obey a superposition principle. The energy of a compound system is *not* the sum of the energies of its parts considered separately—there are also "cross terms":

$$\begin{aligned} W_{\text{tot}} &= \frac{\epsilon_0}{2}\int E^2\,d\tau = \frac{\epsilon_0}{2}\int (\mathbf{E}_1+\mathbf{E}_2)^2\,d\tau \\ &= \frac{\epsilon_0}{2}\int \left(E_1^2+E_2^2+2\mathbf{E}_1\cdot\mathbf{E}_2\right)\,d\tau \\ &= W_1+W_2+\epsilon_0\int \mathbf{E}_1\cdot\mathbf{E}_2\,d\tau. \end{aligned} \tag{2.47}$$

For example, if you double the charge everywhere, you *quadruple* the total energy.

Problem 2.36 Consider two concentric spherical shells, of radii a and b. Suppose the inner one carries a charge q, and the outer one a charge $-q$ (both of them uniformly distributed over the surface). Calculate the energy of this configuration, (a) using Eq. 2.45, and (b) using Eq. 2.47 and the results of Ex. 2.9.

Problem 2.37 Find the interaction energy ($\epsilon_0\int \mathbf{E}_1\cdot\mathbf{E}_2\,d\tau$ in Eq. 2.47) for two point charges, q_1 and q_2, a distance a apart. [*Hint:* Put q_1 at the origin and q_2 on the z axis; use spherical coordinates, and do the r integral first.]

2.5 ■ CONDUCTORS

2.5.1 ■ Basic Properties

In an **insulator,** such as glass or rubber, each electron is on a short leash, attached to a particular atom. In a metallic **conductor,** by contrast, one or more electrons per atom are free to roam. (In liquid conductors such as salt water, it is ions that do the moving.) A *perfect* conductor would contain an *unlimited* supply of free charges. In real life there are no perfect conductors, but metals come pretty close, for most purposes.

From this definition, the basic electrostatic properties of ideal conductors immediately follow:

(i) E = 0 inside a conductor. Why? Because if there *were* any field, those free charges would move, and it wouldn't be electro*statics* any more. Hmm . . . that's hardly a satisfactory explanation; maybe all it proves is that you can't have electrostatics when conductors are present. We had better examine what happens when you put a conductor into an external electric field $\mathbf{E}_0$ (Fig. 2.42). Initially, the field will drive any free positive charges to the right, and negative ones to the left. (In practice, it's the negative charges—electrons—that do the moving, but when they depart, the right side is left with a net positive charge—the stationary nuclei—so it doesn't really matter which charges move; the effect is the same.) When they come to the edge of the material, the charges pile up: plus on the right side, minus on the left. Now, these **induced charges** produce a field of their own, $\mathbf{E}_1$, which, as you can see from the figure, is in the *opposite direction* to $\mathbf{E}_0$. That's the crucial point, for it means that the field of the induced charges *tends to cancel the original field.* Charge will continue to flow until this cancellation is complete, and the resultant field inside the conductor is precisely zero.[9] The whole process is practically instantaneous.

(ii) ρ = 0 inside a conductor. This follows from Gauss's law: $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$. If $\mathbf{E}$ is zero, so also is ρ. There is still charge around, but exactly as much plus as minus, so the *net* charge density in the interior is zero.

(iii) Any net charge resides on the surface. That's the only place left.

(iv) A conductor is an equipotential. For if **a** and **b** are any two points within (or at the surface of) a given conductor, $V(\mathbf{b}) - V(\mathbf{a}) = -\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{E} \cdot d\mathbf{l} = 0$, and hence $V(\mathbf{a}) = V(\mathbf{b})$.

(v) E is perpendicular to the surface, just outside a conductor. Otherwise, as in **(i),** charge will immediately flow around the surface until it kills off the tangential component (Fig. 2.43). (*Perpendicular* to the surface, charge cannot flow, of course, since it is confined to the conducting object.)

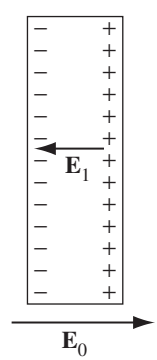


FIGURE 2.42

[9] *Outside* the conductor the field is *not* zero, for here $\mathbf{E}_0$ and $\mathbf{E}_1$ do *not* tend to cancel.

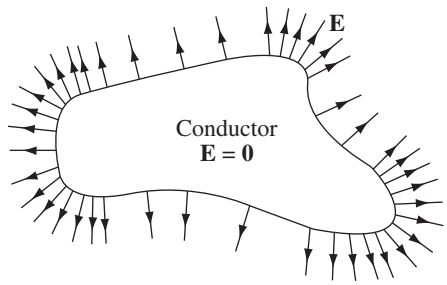


FIGURE 2.43

I think it is astonishing that the charge on a conductor flows to the surface. Because of their mutual repulsion, the charges naturally spread out as much as possible, but for *all* of them to go to the surface seems like a waste of the interior space. Surely we could do better, from the point of view of making each charge as far as possible from its neighbors, to sprinkle *some* of them throughout the volume ... Well, it simply is not so. You do best to put *all* the charge on the surface, and this is true regardless of the size or shape of the conductor.[10]

The problem can also be phrased in terms of energy. Like any other free dynamical system, the charge on a conductor will seek the configuration that minimizes its potential energy. What property **(iii)** asserts is that the electrostatic energy of a solid object (with specified shape and total charge) is a minimum when that charge is spread over the surface. For instance, the energy of a sphere is $(1/8\pi\epsilon_0)(q^2/R)$ if the charge is uniformly distributed over the surface, as we found in Ex. 2.9, but it is greater, $(3/20\pi\epsilon_0)(q^2/R)$, if the charge is uniformly distributed throughout the volume (Prob. 2.34).

2.5.2 ■ Induced Charges

If you hold a charge $+q$ near an uncharged conductor (Fig. 2.44), the two will attract one another. The reason for this is that q will pull minus charges over to the near side and repel plus charges to the far side. (Another way to think of it is that the charge moves around in such a way as to kill off the field of q for points inside the conductor, where the total field must be zero.) Since the negative induced charge is closer to q, there is a net force of attraction. (In Chapter 3 we shall calculate this force explicitly, for the case of a spherical conductor.)

When I speak of the field, charge, or potential "inside" a conductor, I mean in the "meat" of the conductor; if there is some hollow *cavity* in the conductor, and

[10]By the way, the one- and two-dimensional analogs are quite different: The charge on a conducting *disk* does *not* all go to the perimeter (R. Friedberg, *Am. J. Phys.* **61**, 1084 (1993)), nor does the charge on a conducting needle go to the ends (D. J. Griffiths and Y. Li, *Am. J. Phys.* **64**, 706 (1996))—see Prob. 2.57. Moreover, if the exponent of r in Coulomb's law were not precisely 2, the charge on a solid conductor would not all go to the surface—see D. J. Griffiths and D. Z. Uvanovic, *Am. J. Phys.* **69**, 435 (2001), and Prob. 2.54g.

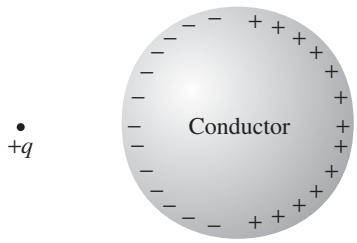


FIGURE 2.44

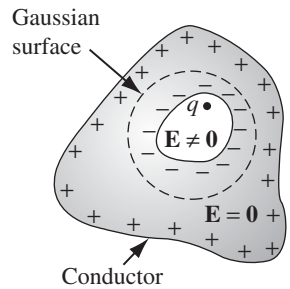


FIGURE 2.45

within that cavity you put some charge, then the field *in the cavity* will *not* be zero. But in a remarkable way the cavity and its contents are electrically isolated from the outside world by the surrounding conductor (Fig. 2.45). No external fields penetrate the conductor; they are canceled at the outer surface by the induced charge there. Similarly, the field due to charges within the cavity is canceled, for all exterior points, by the induced charge on the inner surface. However, the compensating charge left over on the *outer* surface of the conductor effectively "communicates" the presence of q to the outside world. The total charge induced on the cavity wall is equal and opposite to the charge inside, for if we surround the cavity with a Gaussian surface, all points of which are in the conductor (Fig. 2.45), $\oint \mathbf{E} \cdot d\mathbf{a} = 0$, and hence (by Gauss's law) the net enclosed charge must be zero. But $Q_{\text{enc}} = q + q_{\text{induced}}$, so $q_{\text{induced}} = -q$. Then if the conductor as a whole is electrically neutral, there must be a charge $+q$ on its outer surface.

Example 2.10. An uncharged spherical conductor centered at the origin has a cavity of some weird shape carved out of it (Fig. 2.46). Somewhere within the cavity is a charge q. *Question:* What is the field outside the sphere?

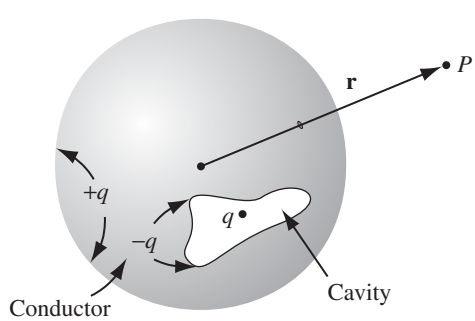


FIGURE 2.46

Solution

At first glance, it would appear that the answer depends on the shape of the cavity and the location of the charge. But that's wrong: the answer is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}}$$

regardless. The conductor conceals from us all information concerning the nature of the cavity, revealing only the total charge it contains. How can this be? Well, the charge $+q$ induces an opposite charge $-q$ on the wall of the cavity, which distributes itself in such a way that its field cancels that of q, for all points exterior to the cavity. Since the conductor carries no net charge, this leaves $+q$ to distribute itself uniformly over the surface of the sphere. (It's *uniform* because the asymmetrical influence of the point charge $+q$ is negated by that of the induced charge $-q$ on the inner surface.) For points outside the sphere, then, the only thing that survives is the field of the leftover $+q$, uniformly distributed over the outer surface.

It may occur to you that in one respect this argument is open to challenge: There are actually *three* fields at work here: $\mathbf{E}_q$, $\mathbf{E}_{\text{induced}}$, and $\mathbf{E}_{\text{leftover}}$. All we know for certain is that the sum of the three is zero inside the conductor, yet I claimed that the first two *alone* cancel, while the third is separately zero there. Moreover, even if the first two cancel within the conductor, who is to say they still cancel for points outside? They do not, after all, cancel for points *inside* the cavity. I cannot give you a completely satisfactory answer at the moment, but this much at least is true: There *exists* a way of distributing $-q$ over the inner surface so as to cancel the field of q at all exterior points. For that same cavity could have been carved out of a *huge* spherical conductor with a radius of 27 miles or light years or whatever. In that case, the leftover $+q$ on the outer surface is simply too far away to produce a significant field, and the other two fields would *have* to accomplish the cancellation by themselves. So we know they *can* do it . . . but are we sure they *choose* to? Perhaps for small spheres nature prefers some complicated three-way cancellation. Nope: As we'll see in the uniqueness theorems of Chapter 3, electrostatics is very stingy with its options; there is always precisely one way—no more—of distributing the charge on a conductor so as to make the field inside zero. Having found a *possible* way, we are guaranteed that no alternative exists, even in principle.

If a cavity surrounded by conducting material is itself empty of charge, then the field within the cavity is zero. For any field line would have to begin and end on the cavity wall, going from a plus charge to a minus charge (Fig. 2.47). Letting that field line be part of a closed loop, the rest of which is entirely inside the conductor (where $\mathbf{E} = \mathbf{0}$), the integral $\oint \mathbf{E} \cdot d\mathbf{l}$ is distinctly *positive,* in violation of Eq. 2.19. It follows that $\mathbf{E} = \mathbf{0}$ within an *empty* cavity, and there is in fact *no* charge on the surface of the cavity. (This is why you are relatively safe inside a metal car during a thunderstorm—you may get *cooked,* if lightning strikes, but you will not be *electrocuted.* The same principle applies to the placement of sensitive apparatus

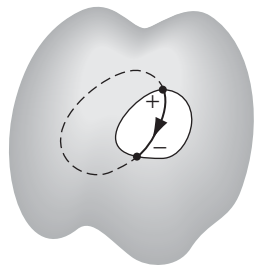


FIGURE 2.47

inside a grounded **Faraday cage,** to shield out stray electric fields. In practice, the enclosure doesn't even have to be solid conductor—chicken wire will often suffice.)

Problem 2.38 A metal sphere of radius R, carrying charge q, is surrounded by a thick concentric metal shell (inner radius a, outer radius b, as in Fig. 2.48). The shell carries no net charge.

(a) Find the surface charge density σ at R, at a, and at b.

(b) Find the potential at the center, using infinity as the reference point.

(c) Now the outer surface is touched to a grounding wire, which drains off charge and lowers its potential to zero (same as at infinity). How do your answers to (a) and (b) change?

Problem 2.39 Two spherical cavities, of radii a and b, are hollowed out from the interior of a (neutral) conducting sphere of radius R (Fig. 2.49). At the center of each cavity a point charge is placed—call these charges q_a and q_b.

(a) Find the surface charge densities σ_a, σ_b, and σ_R.

(b) What is the field outside the conductor?

(c) What is the field within each cavity?

(d) What is the force on q_a and q_b?

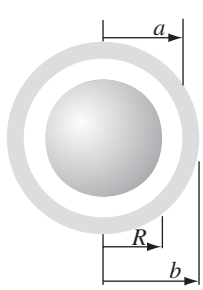


FIGURE 2.48

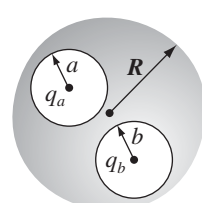


FIGURE 2.49

(e) Which of these answers would change if a third charge, q_c, were brought near the conductor?

Problem 2.40

(a) A point charge q is inside a cavity in an uncharged conductor (Fig. 2.45). Is the force on q necessarily zero?[11]

(b) Is the force between a point charge and a nearby uncharged conductor always attractive?[12]

2.5.3 ■ Surface Charge and the Force on a Conductor

Because the field inside a conductor is zero, boundary condition 2.33 requires that the field immediately *outside* is

$$\mathbf{E} = \frac{\sigma}{\epsilon_0}\hat{\mathbf{n}}, \tag{2.48}$$

consistent with our earlier conclusion that the field is normal to the surface. In terms of potential, Eq. 2.36 yields

$$\sigma = -\epsilon_0 \frac{\partial V}{\partial n}. \tag{2.49}$$

These equations enable you to calculate the surface charge on a conductor, if you can determine $\mathbf{E}$ or V; we shall use them frequently in the next chapter.

In the presence of an electric field, a surface charge will experience a force; the force per unit area, $\mathbf{f}$, is $\sigma\mathbf{E}$. But there's a problem here, for the electric field is *discontinuous* at a surface charge, so what are we supposed to use: $\mathbf{E}_{\text{above}}$, $\mathbf{E}_{\text{below}}$, or something in between? The answer is that we should use the *average* of the two:

$$\mathbf{f} = \sigma\mathbf{E}_{\text{average}} = \frac{1}{2}\sigma(\mathbf{E}_{\text{above}} + \mathbf{E}_{\text{below}}). \tag{2.50}$$

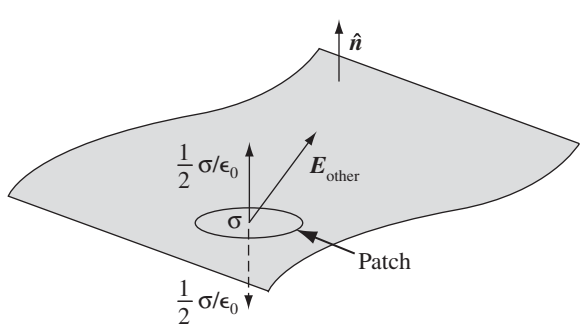


FIGURE 2.50

[11]This problem was suggested by Nelson Christensen.
[12]See M. Levin and S. G. Johnson, *Am. J. Phys.* **79**, 843 (2011).

Why the average? The reason is very simple, though the telling makes it sound complicated: Let's focus our attention on a tiny patch of surface surrounding the point in question (Fig. 2.50). (Make it small enough so it is essentially flat and the surface charge on it is essentially constant.) The *total* field consists of two parts—that attributable to the patch itself, and that due to everything else (other regions of the surface, as well as any external sources that may be present):

$$\mathbf{E} = \mathbf{E}_{\text{patch}} + \mathbf{E}_{\text{other}}.$$

Now, the patch cannot exert a force on itself, any more than you can lift yourself by standing in a basket and pulling up on the handles. The force on the patch, then, is due exclusively to $\mathbf{E}_{\text{other}}$, and *this* suffers *no* discontinuity (if we removed the patch, the field in the "hole" would be perfectly smooth). The discontinuity is due entirely to the charge on the patch, which puts out a field $(\sigma/2\epsilon_0)$ on either side, pointing away from the surface. Thus,

$$\mathbf{E}_{\text{above}} = \mathbf{E}_{\text{other}} + \frac{\sigma}{2\epsilon_0}\hat{\mathbf{n}},$$

$$\mathbf{E}_{\text{below}} = \mathbf{E}_{\text{other}} - \frac{\sigma}{2\epsilon_0}\hat{\mathbf{n}},$$

and hence

$$\mathbf{E}_{\text{other}} = \frac{1}{2}(\mathbf{E}_{\text{above}} + \mathbf{E}_{\text{below}}) = \mathbf{E}_{\text{average}}.$$

Averaging is really just a device for removing the contribution of the patch itself.

That argument applies to *any* surface charge; in the particular case of a conductor, the field is zero inside and $(\sigma/\epsilon_0)\hat{\mathbf{n}}$ outside (Eq. 2.48), so the average is $(\sigma/2\epsilon_0)\hat{\mathbf{n}}$, and the force per unit area is

$$\mathbf{f} = \frac{1}{2\epsilon_0}\sigma^2\hat{\mathbf{n}}. \tag{2.51}$$

This amounts to an outward **electrostatic pressure** on the surface, tending to draw the conductor into the field, regardless of the sign of σ. Expressing the pressure in terms of the field just outside the surface,

$$P = \frac{\epsilon_0}{2}E^2. \tag{2.52}$$

Problem 2.41 Two large metal plates (each of area A) are held a small distance d apart. Suppose we put a charge Q on each plate; what is the electrostatic pressure on the plates?

Problem 2.42 A metal sphere of radius R carries a total charge Q. What is the force of repulsion between the "northern" hemisphere and the "southern" hemisphere?

FIGURE 2.51

2.5.4 ■ Capacitors

Suppose we have *two* conductors, and we put charge $+Q$ on one and $-Q$ on the other (Fig. 2.51). Since V is constant over a conductor, we can speak unambiguously of the potential difference between them:

$$V = V_+ - V_- = -\int_{(-)}^{(+)} \mathbf{E} \cdot d\mathbf{l}.$$

We don't know how the charge distributes itself over the two conductors, and calculating the field would be a nightmare, if their shapes are complicated, but this much we *do* know: $\mathbf{E}$ is *proportional* to Q. For $\mathbf{E}$ is given by Coulomb's law:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \int \frac{\rho}{\imath^2} \hat{\boldsymbol{\imath}}\, d\tau,$$

so if you double ρ, you double $\mathbf{E}$. [Wait a minute! How do we know that doubling Q (and also $-Q$) simply doubles ρ? Maybe the charge *moves around* into a completely different configuration, quadrupling ρ in some places and halving it in others, just so the *total* charge on each conductor is doubled. The *fact* is that this concern is unwarranted—doubling Q *does* double ρ everywhere; it *doesn't* shift the charge around. The proof of this will come in Chapter 3; for now you'll just have to trust me.]

Since $\mathbf{E}$ is proportional to Q, so also is V. The constant of proportionality is called the **capacitance** of the arrangement:

$$C \equiv \frac{Q}{V}. \tag{2.53}$$

Capacitance is a purely geometrical quantity, determined by the sizes, shapes, and separation of the two conductors. In SI units, C is measured in **farads** (F); a farad is a coulomb-per-volt. Actually, this turns out to be inconveniently large; more practical units are the microfarad (10^{-6} F) and the picofarad (10^{-12} F).

Notice that V is, by definition, the potential of the *positive* conductor less that of the negative one; likewise, Q is the charge of the *positive* conductor. Accordingly, capacitance is an intrinsically positive quantity. (By the way, you will occasionally hear someone speak of the capacitance of a *single* conductor. In this case the "second conductor," with the negative charge, is an imaginary spherical shell of infinite radius surrounding the one conductor. It contributes nothing to the field, so the capacitance is given by Eq. 2.53, where V is the potential with infinity as the reference point.)

Example 2.11. Find the capacitance of a **parallel-plate capacitor** consisting of two metal surfaces of area A held a distance d apart (Fig. 2.52).

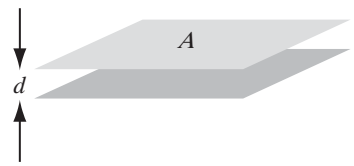


FIGURE 2.52

Solution
If we put $+Q$ on the top and $-Q$ on the bottom, they will spread out uniformly over the two surfaces, provided the area is reasonably large and the separation small.[13] The surface charge density, then, is $\sigma = Q/A$ on the top plate, and so the field, according to Ex. 2.6, is $(1/\epsilon_0)Q/A$. The potential difference between the plates is therefore

$$V = \frac{Q}{A\epsilon_0}d,$$

and hence

$$C = \frac{A\epsilon_0}{d}. \tag{2.54}$$

If, for instance, the plates are square with sides 1 cm long, and they are held 1 mm apart, then the capacitance is 9×10^{-13} F.

Example 2.12. Find the capacitance of two concentric spherical metal shells, with radii a and b.

Solution
Place charge $+Q$ on the inner sphere, and $-Q$ on the outer one. The field between the spheres is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}},$$

so the potential difference between them is

$$V = -\int_b^a \mathbf{E}\cdot d\mathbf{l} = -\frac{Q}{4\pi\epsilon_0}\int_b^a \frac{1}{r^2}\,dr = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{a}-\frac{1}{b}\right).$$

[13]The *exact* solution is not easy—even for the simpler case of circular plates. See G. T. Carlson and B. L. Illman, *Am. J. Phys.* **62**, 1099 (1994).

As promised, V is proportional to Q; the capacitance is

$$C = \frac{Q}{V} = 4\pi\epsilon_0 \frac{ab}{(b-a)}.$$

To "charge up" a capacitor, you have to remove electrons from the positive plate and carry them to the negative plate. In doing so, you fight against the electric field, which is pulling them back toward the positive conductor and pushing them away from the negative one. How much work does it take, then, to charge the capacitor up to a final amount Q? Suppose that at some intermediate stage in the process the charge on the positive plate is q, so that the potential difference is q/C. According to Eq. 2.38, the work you must do to transport the next piece of charge, dq, is

$$dW = \left(\frac{q}{C}\right) dq.$$

The total work necessary, then, to go from $q = 0$ to $q = Q$, is

$$W = \int_0^Q \left(\frac{q}{C}\right) dq = \frac{1}{2}\frac{Q^2}{C},$$

or, since $Q = CV$,

$$W = \frac{1}{2}CV^2, \tag{2.55}$$

where V is the final potential of the capacitor.

Problem 2.43 Find the capacitance per unit length of two coaxial metal cylindrical tubes, of radii a and b (Fig. 2.53).

FIGURE 2.53

Problem 2.44 Suppose the plates of a parallel-plate capacitor move closer together by an infinitesimal distance ϵ, as a result of their mutual attraction.

(a) Use Eq. 2.52 to express the work done by electrostatic forces, in terms of the field E, and the area of the plates, A.

(b) Use Eq. 2.46 to express the energy lost by the field in this process.

(This problem is supposed to be easy, but it contains the embryo of an alternative derivation of Eq. 2.52, using conservation of energy.)

More Problems on Chapter 2

Problem 2.45 Find the electric field at a height z above the center of a square sheet (side a) carrying a uniform surface charge σ. Check your result for the limiting cases $a \to \infty$ and $z \gg a$.

$\left[Answer: (\sigma/2\epsilon_0)\left\{(4/\pi)\tan^{-1}\sqrt{1+(a^2/2z^2)}-1\right\}\right]$

Problem 2.46 If the electric field in some region is given (in spherical coordinates) by the expression

$$\mathbf{E}(\mathbf{r}) = \frac{k}{r}\left[3\,\hat{\mathbf{r}} + 2\sin\theta\cos\theta\sin\phi\,\hat{\boldsymbol{\theta}} + \sin\theta\cos\phi\,\hat{\boldsymbol{\phi}}\right],$$

for some constant k, what is the charge density? [*Answer:* $3k\epsilon_0(1+\cos 2\theta\sin\phi)/r^2$]

Problem 2.47 Find the net force that the southern hemisphere of a uniformly charged solid sphere exerts on the northern hemisphere. Express your answer in terms of the radius R and the total charge Q. [*Answer:* $(1/4\pi\epsilon_0)(3Q^2/16R^2)$]

Problem 2.48 An inverted hemispherical bowl of radius R carries a uniform surface charge density σ. Find the potential difference between the "north pole" and the center. [*Answer:* $(R\sigma/2\epsilon_0)(\sqrt{2}-1)$]

Problem 2.49 A sphere of radius R carries a charge density $\rho(r) = kr$ (where k is a constant). Find the energy of the configuration. Check your answer by calculating it in at least two different ways. [*Answer:* $\pi k^2 R^7/7\epsilon_0$]

Problem 2.50 The electric potential of some configuration is given by the expression

$$V(\mathbf{r}) = A\frac{e^{-\lambda r}}{r},$$

where A and λ are constants. Find the electric field $\mathbf{E}(\mathbf{r})$, the charge density $\rho(r)$, and the total charge Q. [*Answer:* $\rho = \epsilon_0 A(4\pi\delta^3(\mathbf{r}) - \lambda^2 e^{-\lambda r}/r)$]

Problem 2.51 Find the potential on the rim of a uniformly charged disk (radius R, charge density σ). [*Hint:* First show that $V = k(\sigma R/\pi\epsilon_0)$, for some dimensionless number k, which you can express as an integral. Then evaluate k analytically, if you can, or by computer.]

! **Problem 2.52** Two infinitely long wires running parallel to the x axis carry uniform charge densities $+\lambda$ and $-\lambda$ (Fig. 2.54).

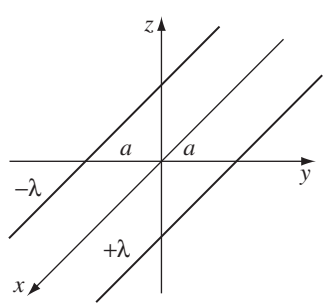


FIGURE 2.54

(a) Find the potential at any point (x, y, z), using the origin as your reference.

(b) Show that the equipotential surfaces are circular cylinders, and locate the axis and radius of the cylinder corresponding to a given potential V_0.

! **Problem 2.53** In a vacuum diode, electrons are "boiled" off a hot **cathode**, at potential zero, and accelerated across a gap to the **anode**, which is held at positive potential V_0. The cloud of moving electrons within the gap (called **space charge**) quickly builds up to the point where it reduces the field at the surface of the cathode to zero. From then on, a steady current I flows between the plates.

Suppose the plates are large relative to the separation ($A \gg d^2$ in Fig. 2.55), so that edge effects can be neglected. Then V, ρ, and v (the speed of the electrons) are all functions of x alone.

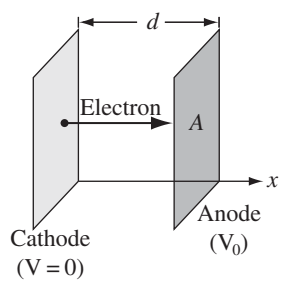


FIGURE 2.55

(a) Write Poisson's equation for the region between the plates.

(b) Assuming the electrons start from rest at the cathode, what is their speed at point x, where the potential is $V(x)$?

(c) In the steady state, I is independent of x. What, then, is the relation between ρ and v?

(d) Use these three results to obtain a differential equation for V, by eliminating ρ and v.

(e) Solve this equation for V as a function of x, V_0, and d. Plot $V(x)$, and compare it to the potential *without* space-charge. Also, find ρ and v as functions of x.

(f) Show that

$$I = K V_0^{3/2}, \tag{2.56}$$

and find the constant K. (Equation 2.56 is called the **Child-Langmuir law**. It holds for other geometries as well, whenever space-charge limits the current. Notice that the space-charge limited diode is *nonlinear*—it does not obey Ohm's law.)

! **Problem 2.54** Imagine that new and extraordinarily precise measurements have revealed an error in Coulomb's law. The *actual* force of interaction between two point charges is found to be

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{\imath^2}\left(1 + \frac{\imath}{\lambda}\right)e^{-(\imath/\lambda)}\hat{\boldsymbol{\imath}},$$

where λ is a new constant of nature (it has dimensions of length, obviously, and is a huge number—say half the radius of the known universe—so that the correction is small, which is why no one ever noticed the discrepancy before). You are charged with the task of reformulating electrostatics to accommodate the new discovery. Assume the principle of superposition still holds.

(a) What is the electric field of a charge distribution ρ (replacing Eq. 2.8)?

(b) Does this electric field admit a scalar potential? Explain briefly how you reached your conclusion. (No formal proof necessary—just a persuasive argument.)

(c) Find the potential of a point charge q—the analog to Eq. 2.26. (If your answer to (b) was "no," better go back and change it!) Use ∞ as your reference point.

(d) For a point charge q at the origin, show that

$$\oint_{\mathcal{S}} \mathbf{E}\cdot d\mathbf{a} + \frac{1}{\lambda^2}\int_{\mathcal{V}} V\, d\tau = \frac{1}{\epsilon_0}q,$$

where $\mathcal{S}$ is the surface, $\mathcal{V}$ the volume, of any sphere centered at q.

(e) Show that this result generalizes:

$$\oint_{\mathcal{S}} \mathbf{E}\cdot d\mathbf{a} + \frac{1}{\lambda^2}\int_{\mathcal{V}} V\, d\tau = \frac{1}{\epsilon_0}Q_{\text{enc}},$$

for *any* charge distribution. (This is the next best thing to Gauss's Law, in the new "electrostatics.")

(f) Draw the triangle diagram (like Fig. 2.35) for this world, putting in all the appropriate formulas. (Think of Poisson's equation as the formula for ρ in terms of V, and Gauss's law (differential form) as an equation for ρ in terms of $\mathbf{E}$.)

(g) Show that *some* of the charge on a conductor distributes itself (uniformly!) over the volume, with the remainder on the surface. [*Hint:* $\mathbf{E}$ is still zero, inside a conductor.]

Problem 2.55 Suppose an electric field $\mathbf{E}(x, y, z)$ has the form

$$E_x = ax, \qquad E_y = 0, \qquad E_z = 0$$

where a is a constant. What is the charge density? How do you account for the fact that the field points in a particular direction, when the charge density is uniform? [This is a more subtle problem than it looks, and worthy of careful thought.]

Problem 2.56 All of electrostatics follows from the $1/r^2$ character of Coulomb's law, together with the principle of superposition. An analogous theory can therefore be constructed for Newton's law of universal gravitation. What is the gravitational energy of a sphere, of mass M and radius R, assuming the density is uniform? Use your result to estimate the gravitational energy of the sun (look up the relevant numbers). Note that the energy is *negative*—masses *attract*, whereas (like) electric charges *repel*. As the matter "falls in," to create the sun, its energy is converted into other forms (typically thermal), and it is subsequently released in the form of radiation. The sun radiates at a rate of 3.86×10^{26} W; if all this came from gravitational energy, how long would the sun last? [The sun is in fact much older than that, so evidently this is *not* the source of its power.[14]]

! **Problem 2.57** We know that the charge on a conductor goes to the surface, but just how it distributes itself there is not easy to determine. One famous example in which the surface charge density can be calculated explicitly is the ellipsoid:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

In this case[15]

$$\sigma = \frac{Q}{4\pi abc}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2}, \tag{2.57}$$

where Q is the total charge. By choosing appropriate values for a, b, and c, obtain (from Eq. 2.57): (a) the net (both sides) surface charge density $\sigma(r)$ on a circular disk of radius R; (b) the net surface charge density $\sigma(x)$ on an infinite conducting "ribbon" in the xy plane, which straddles the y axis from $x = -a$ to $x = a$ (let Λ be the total charge per unit length of ribbon); (c) the net charge per unit length $\lambda(x)$ on a conducting "needle," running from $x = -a$ to $x = a$. In each case, sketch the graph of your result.

Problem 2.58

(a) Consider an equilateral triangle, inscribed in a circle of radius a, with a point charge q at each vertex. The electric field is zero (obviously) at the center, but (surprisingly) there are three *other* points inside the triangle where the field is zero. Where are they? [*Answer:* $r = 0.285\,a$—you'll probably need a computer to get it.]

(b) For a regular n-sided polygon there are n points (in addition to the center) where the field is zero.[16] Find their distance from the center for $n = 4$ and $n = 5$. What do you suppose happens as $n \to \infty$?

[14]Lord Kelvin used this argument to counter Darwin's theory of evolution, which called for a much older Earth. Of course, we now know that the source of the Sun's energy is nuclear fusion, not gravity.
[15]For the derivation (which is a real *tour de force*), see W. R. Smythe, *Static and Dynamic Electricity,* 3rd ed. (New York: Hemisphere, 1989), Sect. 5.02.
[16]S. D. Baker, *Am. J. Phys.* **52**, 165 (1984); D. Kiang and D. A. Tindall, *Am. J. Phys.* **53**, 593 (1985).

Problem 2.59 Prove or disprove (with a counterexample) the following

> **Theorem:** Suppose a conductor carrying a net charge Q, when placed in an external electric field $\mathbf{E}_e$, experiences a force $\mathbf{F}$; if the external field is now reversed ($\mathbf{E}_e \rightarrow -\mathbf{E}_e$), the force also reverses ($\mathbf{F} \rightarrow -\mathbf{F}$).

What if we stipulate that the external field is *uniform*?

Problem 2.60 A point charge q is at the center of an uncharged spherical conducting shell, of inner radius a and outer radius b. *Question:* How much work would it take to move the charge out to infinity (through a tiny hole drilled in the shell)? [*Answer:* $(q^2/8\pi\epsilon_0)(1/a - 1/b)$.]

Problem 2.61 What is the minimum-energy configuration for a system of N equal point charges placed on or inside a circle of radius R?[17] Because the charge on a conductor goes to the surface, you might think the N charges would arrange themselves (uniformly) around the circumference. Show (to the contrary) that for $N = 12$ it is better to place 11 on the circumference and one at the center. How about for $N = 11$ (is the energy lower if you put all 11 around the circumference, or if you put 10 on the circumference and one at the center)? [*Hint:* Do it numerically—you'll need at least 4 significant digits. Express all energies as multiples of $q^2/4\pi\epsilon_0 R$]

[17] M. G. Calkin, D. Kiang, and D. A. Tindall, *Am. H. Phys.* **55**, 157 (1987).

CHAPTER

3 Potentials

3.1 ■ LAPLACE'S EQUATION

3.1.1 ■ Introduction

The primary task of electrostatics is to find the electric field of a given stationary charge distribution. In principle, this purpose is accomplished by Coulomb's law, in the form of Eq. 2.8:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\hat{\boldsymbol{\imath}}}{\imath^2}\rho(\mathbf{r}')\,d\tau'. \tag{3.1}$$

Unfortunately, integrals of this type can be difficult to calculate for any but the simplest charge configurations. Occasionally we can get around this by exploiting symmetry and using Gauss's law, but ordinarily the best strategy is first to calculate the *potential,* V, which is given by the somewhat more tractable Eq. 2.29:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{\imath}\rho(\mathbf{r}')\,d\tau'. \tag{3.2}$$

Still, even *this* integral is often too tough to handle analytically. Moreover, in problems involving conductors ρ itself may not be known in advance; since charge is free to move around, the only thing we control directly is the *total* charge (or perhaps the potential) of each conductor.

In such cases, it is fruitful to recast the problem in differential form, using Poisson's equation (2.24),

$$\nabla^2 V = -\frac{1}{\epsilon_0}\rho, \tag{3.3}$$

which, together with appropriate boundary conditions, is equivalent to Eq. 3.2. Very often, in fact, we are interested in finding the potential in a region where $\rho = 0$. (If $\rho = 0$ *everywhere,* of course, then $V = 0$, and there is nothing further to say—that's not what I mean. There may be plenty of charge *elsewhere,* but we're confining our attention to places where there is no charge.) In this case, Poisson's equation reduces to Laplace's equation:

$$\nabla^2 V = 0, \tag{3.4}$$

or, written out in Cartesian coordinates,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0. \tag{3.5}$$

This formula is so fundamental to the subject that one might almost say electrostatics *is* the study of Laplace's equation. At the same time, it is a ubiquitous equation, appearing in such diverse branches of physics as gravitation and magnetism, the theory of heat, and the study of soap bubbles. In mathematics, it plays a major role in analytic function theory. To get a feel for Laplace's equation and its solutions (which are called **harmonic functions**), we shall begin with the one- and two-dimensional versions, which are easier to picture, and illustrate all the essential properties of the three-dimensional case.

3.1.2 ■ Laplace's Equation in One Dimension

Suppose V depends on only one variable, x. Then Laplace's equation becomes

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

The general solution is

$$V(x) = mx + b, \tag{3.6}$$

the equation for a straight line. It contains two undetermined constants (m and b), as is appropriate for a second-order (ordinary) differential equation. They are fixed, in any particular case, by the boundary conditions of that problem. For instance, it might be specified that $V = 4$ at $x = 1$, and $V = 0$ at $x = 5$. In that case, $m = -1$ and $b = 5$, so $V = -x + 5$ (see Fig. 3.1).

I want to call your attention to two features of this result; they may seem silly and obvious in one dimension, where I can write down the general solution explicitly, but the analogs in two and three dimensions are powerful and by no means obvious:

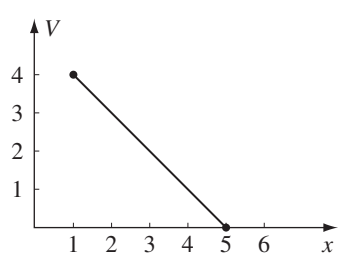


FIGURE 3.1

1. $V(x)$ is the *average* of $V(x+a)$ and $V(x-a)$, for any a:

$$V(x) = \frac{1}{2}[V(x+a) + V(x-a)].$$

Laplace's equation is a kind of averaging instruction; it tells you to assign to the point x the average of the values to the left and to the right of x. Solutions to Laplace's equation are, in this sense, *as boring as they could possibly be,* and yet fit the end points properly.

2. Laplace's equation tolerates *no local maxima or minima;* extreme values of V must occur at the end points. Actually, this is a consequence of (1), for if there *were* a local maximum, V would be greater at that point than on either side, and therefore could not be the average. (Ordinarily, you expect the second derivative to be negative at a maximum and positive at a minimum. Since Laplace's equation requires, on the contrary, that the second derivative is zero, it seems reasonable that solutions should exhibit no extrema. However, this is not a *proof,* since there exist functions that have maxima and minima at points where the second derivative vanishes: x^4, for example, has such a minimum at the point $x = 0$.)

3.1.3 ■ Laplace's Equation in Two Dimensions

If V depends on two variables, Laplace's equation becomes

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0.$$

This is no longer an *ordinary* differential equation (that is, one involving ordinary derivatives only); it is a *partial* differential equation. As a consequence, some of the simple rules you may be familiar with do not apply. For instance, the general solution to this equation doesn't contain just two arbitrary constants—or, for that matter, *any* finite number—despite the fact that it's a second-order equation. Indeed, one cannot write down a "general solution" (at least, not in a closed form like Eq. 3.6). Nevertheless, it is possible to deduce certain properties common to all solutions.

It may help to have a physical example in mind. Picture a thin rubber sheet (or a soap film) stretched over some support. For definiteness, suppose you take a cardboard box, cut a wavy line all the way around, and remove the top part (Fig. 3.2). Now glue a tightly stretched rubber membrane over the box, so that it fits like a drum head (it won't be a *flat* drumhead, of course, unless you chose to cut the edges off straight). Now, if you lay out coordinates (x, y) on the bottom of the box, the height $V(x, y)$ of the sheet above the point (x, y) will satisfy Laplace's

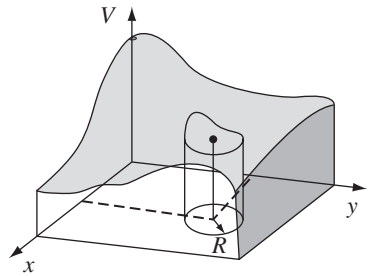


FIGURE 3.2

equation.[1] (The one-dimensional analog would be a rubber band stretched between two points. Of course, it would form a straight line.)

Harmonic functions in two dimensions have the same properties we noted in one dimension:

1. The value of V at a point (x, y) is the average of those *around* the point. More precisely, if you draw a circle of any radius R about the point (x, y), the average value of V on the circle is equal to the value at the center:

$$V(x, y) = \frac{1}{2\pi R} \oint_{\text{circle}} V \, dl.$$

(This, incidentally, suggests the **method of relaxation**, on which computer solutions to Laplace's equation are based: Starting with specified values for V at the boundary, and reasonable guesses for V on a grid of interior points, the first pass reassigns to each point the average of its nearest neighbors. The second pass repeats the process, using the corrected values, and so on. After a few iterations, the numbers begin to settle down, so that subsequent passes produce negligible changes, and a numerical solution to Laplace's equation, with the given boundary values, has been achieved.)[2]

2. V has no local maxima or minima; all extrema occur at the boundaries. (As before, this follows from (1).) Again, Laplace's equation picks the most featureless function possible, consistent with the boundary conditions: no hills, no valleys, just the smoothest conceivable surface. For instance, if you put a ping-pong ball on the stretched rubber sheet of Fig. 3.2, it will

[1] Actually, the equation satisfied by a rubber sheet is

$$\frac{\partial}{\partial x}\left(g\frac{\partial V}{\partial x}\right) + \frac{\partial}{\partial y}\left(g\frac{\partial V}{\partial y}\right) = 0, \quad \text{where } g = \left[1 + \left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2\right]^{-1/2};$$

it reduces (approximately) to Laplace's equation as long as the surface does not deviate too radically from a plane.

[2] See, for example, E. M. Purcell, *Electricity and Magnetism,* 2nd ed. (New York: McGraw-Hill, 1985), problem 3.30.

roll over to one side and fall off—it will not find a "pocket" somewhere to settle into, for Laplace's equation allows no such dents in the surface. From a geometrical point of view, just as a straight line is the shortest distance between two points, so a harmonic function in two dimensions minimizes the surface area spanning the given boundary line.

3.1.4 ■ Laplace's Equation in Three Dimensions

In three dimensions I can neither provide you with an explicit solution (as in one dimension) nor offer a suggestive physical example to guide your intuition (as I did in two dimensions). Nevertheless, the same two properties remain true, and this time I will sketch a proof.[3]

1. The value of V at point $\mathbf{r}$ is the average value of V over a spherical surface of radius R centered at $\mathbf{r}$:
$$V(\mathbf{r}) = \frac{1}{4\pi R^2} \oint_{\text{sphere}} V \, da.$$

2. As a consequence, V can have no local maxima or minima; the extreme values of V must occur at the boundaries. (For if V had a local maximum at $\mathbf{r}$, then by the very nature of maximum I could draw a sphere around $\mathbf{r}$ over which all values of V—and *a fortiori* the average—would be less than at $\mathbf{r}$.)

Proof. Let's begin by calculating the average potential over a spherical surface of radius R due to a *single* point charge q located outside the sphere. We may as well center the sphere at the origin and choose coordinates so that q lies on the z-axis (Fig. 3.3). The potential at a point on the surface is

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{\mathscr{r}},$$

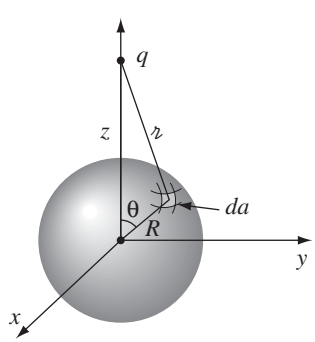


FIGURE 3.3

[3]For a proof that does not rely on Coulomb's law (only on Laplace's equation), see Prob. 3.37.

where

$$\imath^2 = z^2 + R^2 - 2zR\cos\theta,$$

so

$$\begin{aligned} V_{\text{ave}} &= \frac{1}{4\pi R^2}\frac{q}{4\pi\epsilon_0}\int [z^2 + R^2 - 2zR\cos\theta]^{-1/2} R^2 \,\sin\theta\, d\theta\, d\phi \\ &= \frac{q}{4\pi\epsilon_0}\frac{1}{2zR}\sqrt{z^2 + R^2 - 2zR\cos\theta}\,\Big|_0^\pi \\ &= \frac{q}{4\pi\epsilon_0}\frac{1}{2zR}[(z+R) - (z-R)] = \frac{1}{4\pi\epsilon_0}\frac{q}{z}. \end{aligned}$$

But this is precisely the potential due to q at the *center* of the sphere! By the superposition principle, the same goes for any *collection* of charges outside the sphere: their average potential over the sphere is equal to the net potential they produce at the center. □

Problem 3.1 Find the average potential over a spherical surface of radius R due to a point charge q located *inside* (same as above, in other words, only with $z < R$). (In this case, of course, Laplace's equation does not hold within the sphere.) Show that, in general,

$$V_{\text{ave}} = V_{\text{center}} + \frac{Q_{\text{enc}}}{4\pi\epsilon_0 R},$$

where V_{center} is the potential at the center due to all the *external* charges, and Q_{enc} is the total enclosed charge.

Problem 3.2 In one sentence, justify **Earnshaw's Theorem:** *A charged particle cannot be held in a stable equilibrium by electrostatic forces alone.* As an example, consider the cubical arrangement of fixed charges in Fig. 3.4. It *looks,* off hand, as though a positive charge at the center would be suspended in midair, since it is repelled away from each corner. Where is the leak in this "electrostatic bottle"? [To harness nuclear fusion as a practical energy source it is necessary to heat a plasma (soup of charged particles) to fantastic temperatures—so hot that contact would vaporize any ordinary pot. Earnshaw's theorem says that electrostatic containment is also out of the question. Fortunately, it *is* possible to confine a hot plasma magnetically.]

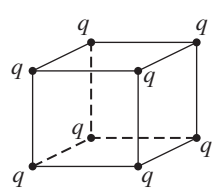


FIGURE 3.4

Problem 3.3 Find the general solution to Laplace's equation in spherical coordinates, for the case where V depends only on r. Do the same for cylindrical coordinates, assuming V depends only on s.

Problem 3.4

(a) Show that the average electric *field* over a spherical surface, due to charges outside the sphere, is the same as the field at the center.

(b) What is the average due to charges *inside* the sphere?

3.1.5 ■ Boundary Conditions and Uniqueness Theorems

Laplace's equation does not by itself determine V; in addition, suitable boundary conditions must be supplied. This raises a delicate question: What are appropriate boundary conditions, sufficient to determine the answer and yet not so strong as to generate inconsistencies? The one-dimensional case is easy, for here the general solution $V = mx + b$ contains two arbitrary constants, and we therefore require two boundary conditions. We might, for instance, specify the value of the function at each end, or we might give the value of the function and its derivative at one end, or the value at one end and the derivative at the other, and so on. But we cannot get away with *just* the value or *just* the derivative at *one* end—this is insufficient information. Nor would it do to specify the derivatives at both ends—this would either be redundant (if the two are equal) or inconsistent (if they are not).

In two or three dimensions we are confronted by a *partial* differential equation, and it is not so obvious what would constitute acceptable boundary conditions. Is the shape of a taut rubber membrane, for instance, uniquely determined by the frame over which it is stretched, or, like a canning jar lid, can it snap from one stable configuration to another? The answer, as I think your intuition would suggest, is that V *is* uniquely determined by its value at the boundary (canning jars evidently do not obey Laplace's equation). However, other boundary conditions can also be used (see Prob. 3.5). The *proof* that a proposed set of boundary conditions will suffice is usually presented in the form of a **uniqueness theorem.** There are many such theorems for electrostatics, all sharing the same basic format—I'll show you the two most useful ones.[4]

> **First uniqueness theorem:** The solution to Laplace's equation in some volume $\mathcal{V}$ is uniquely determined if V is specified on the boundary surface $\mathcal{S}$.

Proof. In Fig. 3.5 I have drawn such a region and its boundary. (There could also be "islands" inside, so long as V is given on all their surfaces; also, the outer

[4] I do not intend to prove the *existence* of solutions here—that's a much more difficult job. In context, the existence is generally clear on physical grounds.

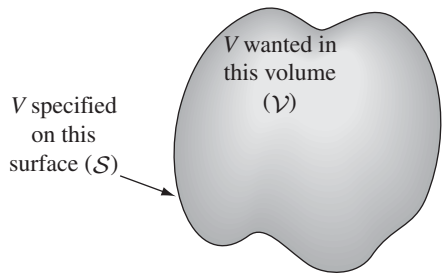


FIGURE 3.5

boundary could be at infinity, where V is ordinarily taken to be zero.) Suppose there were *two* solutions to Laplace's equation:

$$\nabla^2 V_1 = 0 \quad \text{and} \quad \nabla^2 V_2 = 0,$$

both of which assume the specified value on the surface. I want to prove that they must be equal. The trick is look at their *difference:*

$$V_3 \equiv V_1 - V_2.$$

This obeys Laplace's equation,

$$\nabla^2 V_3 = \nabla^2 V_1 - \nabla^2 V_2 = 0,$$

and it takes the value *zero* on all boundaries (since V_1 and V_2 are equal there). But Laplace's equation allows no local maxima or minima—all extrema occur on the boundaries. So the maximum and minimum of V_3 are both zero. Therefore V_3 must be zero everywhere, and hence

$$V_1 = V_2. \qquad \square$$

Example 3.1. Show that the potential is *constant* inside an enclosure completely surrounded by conducting material, provided there is no charge within the enclosure.

Solution

The potential on the cavity wall is some constant, V_0 (that's item (iv), in Sect. 2.5.1), so the potential inside is a function that satisfies Laplace's equation and has the constant value V_0 at the boundary. It doesn't take a genius to think of *one* solution to this problem: $V = V_0$ everywhere. The uniqueness theorem guarantees that this is the *only* solution. (It follows that the *field* inside an empty cavity is zero—the same result we found in Sect. 2.5.2 on rather different grounds.)

The uniqueness theorem is a license to your imagination. It doesn't matter *how* you come by your solution; if (a) it satisfies Laplace's equation and (b) it has

the correct value on the boundaries, then it's *right.* You'll see the power of this argument when we come to the method of images.

Incidentally, it is easy to improve on the first uniqueness theorem: I assumed there was no charge inside the region in question, so the potential obeyed Laplace's equation, but we may as well throw in some charge (in which case V obeys Poisson's equation). The argument is the same, only this time

$$\nabla^2 V_1 = -\frac{1}{\epsilon_0}\rho, \qquad \nabla^2 V_2 = -\frac{1}{\epsilon_0}\rho,$$

so

$$\nabla^2 V_3 = \nabla^2 V_1 - \nabla^2 V_2 = -\frac{1}{\epsilon_0}\rho + \frac{1}{\epsilon_0}\rho = 0.$$

Once again the *difference* ($V_3 \equiv V_1 - V_2$) satisfies Laplace's equation and has the value zero on all boundaries, so $V_3 = 0$ and hence $V_1 = V_2$.

Corollary: The potential in a volume $\mathcal{V}$ is uniquely determined if (a) the charge density throughout the region, and (b) the value of V on all boundaries, are specified.

3.1.6 ■ Conductors and the Second Uniqueness Theorem

The *simplest* way to set the boundary conditions for an electrostatic problem is to specify the value of V on all surfaces surrounding the region of interest. And this situation often occurs in practice: In the laboratory, we have conductors connected to batteries, which maintain a given potential, or to **ground**, which is the experimentalist's word for $V = 0$. However, there are other circumstances in which we do not know the *potential* at the boundary, but rather the *charges* on various conducting surfaces. Suppose I put charge Q_a on the first conductor, Q_b on the second, and so on—I'm not telling you how the charge distributes itself over each conducting surface, because as soon as I put it on, it moves around in a way I do not control. And for good measure, let's say there is some specified charge density ρ in the region between the conductors. Is the electric field now uniquely determined? Or are there perhaps a number of different ways the charges could arrange themselves on their respective conductors, each leading to a different field?

Second uniqueness theorem: In a volume $\mathcal{V}$ surrounded by conductors and containing a specified charge density ρ, the electric field is uniquely determined if the *total charge* on each conductor is given (Fig. 3.6). (The region as a whole can be bounded by another conductor, or else unbounded.)

Proof. Suppose there are *two* fields satisfying the conditions of the problem. Both obey Gauss's law in differential form in the space between the conductors:

$$\boldsymbol{\nabla} \cdot \mathbf{E}_1 = \frac{1}{\epsilon_0}\rho, \qquad \boldsymbol{\nabla} \cdot \mathbf{E}_2 = \frac{1}{\epsilon_0}\rho.$$

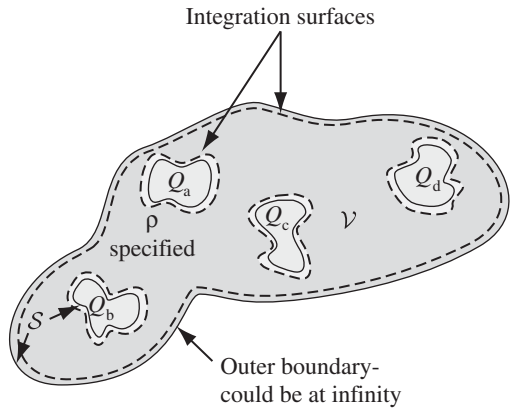


FIGURE 3.6

And both obey Gauss's law in integral form for a Gaussian surface enclosing each conductor:

$$\oint_{\substack{i\text{th conducting}\\ \text{surface}}} \mathbf{E}_1 \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_i, \qquad \oint_{\substack{i\text{th conducting}\\ \text{surface}}} \mathbf{E}_2 \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_i.$$

Likewise, for the outer boundary (whether this is just inside an enclosing conductor or at infinity),

$$\oint_{\substack{\text{outer}\\ \text{boundary}}} \mathbf{E}_1 \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{tot}}, \qquad \oint_{\substack{\text{outer}\\ \text{boundary}}} \mathbf{E}_2 \cdot d\mathbf{a} = \frac{1}{\epsilon_0} Q_{\text{tot}}.$$

As before, we examine the difference

$$\mathbf{E}_3 \equiv \mathbf{E}_1 - \mathbf{E}_2,$$

which obeys

$$\boldsymbol{\nabla} \cdot \mathbf{E}_3 = 0 \tag{3.7}$$

in the region between the conductors, and

$$\oint \mathbf{E}_3 \cdot d\mathbf{a} = 0 \tag{3.8}$$

over each boundary surface.

Now there is one final piece of information we must exploit: Although we do not know how the charge Q_i distributes itself over the ith conductor, we *do* know that each conductor is an equipotential, and hence V_3 is a *constant* (not

necessarily the *same* constant) over each conducting surface. (It need not be *zero,* for the potentials V_1 and V_2 may not be equal—all we know for sure is that *both* are *constant* over any given conductor.) Next comes a trick. Invoking product rule number 5 (inside front cover), we find that

$$\nabla \cdot (V_3 \mathbf{E}_3) = V_3(\nabla \cdot \mathbf{E}_3) + \mathbf{E}_3 \cdot (\nabla V_3) = -(E_3)^2.$$

Here I have used Eq. 3.7, and $\mathbf{E}_3 = -\nabla V_3$. Integrating this over $\mathcal{V}$, and applying the divergence theorem to the left side:

$$\int_{\mathcal{V}} \nabla \cdot (V_3 \mathbf{E}_3)\, d\tau = \oint_{\mathcal{S}} V_3 \mathbf{E}_3 \cdot d\mathbf{a} = -\int_{\mathcal{V}} (E_3)^2\, d\tau.$$

The surface integral covers all boundaries of the region in question—the conductors and outer boundary. Now V_3 is a constant over each surface (if the outer boundary is infinity, $V_3 = 0$ there), so it comes outside each integral, and what remains is zero, according to Eq. 3.8. Therefore,

$$\int_{\mathcal{V}} (E_3)^2\, d\tau = 0.$$

But this integrand is never negative; the only way the integral can vanish is if $E_3 = 0$ everywhere. Consequently, $\mathbf{E}_1 = \mathbf{E}_2$, and the theorem is proved. □

This proof was not easy, and there is a real danger that the theorem itself will seem more plausible to you than the proof. In case you think the second uniqueness theorem is "obvious," consider this example of Purcell's: Figure 3.7 shows a simple electrostatic configuration, consisting of four conductors with charges $\pm Q$, situated so that the plusses are near the minuses. It all looks very comfortable. Now, what happens if we join them in pairs, by tiny wires, as indicated in Fig. 3.8? Since the positive charges are very near negative charges (which is where they *like* to be) you might well guess that *nothing* will happen—the configuration looks stable.

Well, that sounds reasonable, but it's wrong. The configuration in Fig. 3.8 is *impossible.* For there are now effectively *two* conductors, and the total charge on each is *zero*. *One* possible way to distribute zero charge over these conductors is to have no accumulation of charge anywhere, and hence zero field

FIGURE 3.7

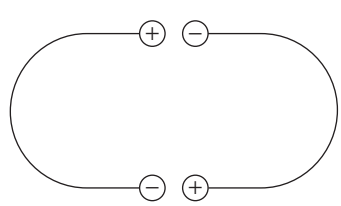


FIGURE 3.8

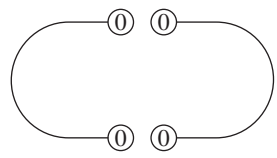


FIGURE 3.9

everywhere (Fig. 3.9). By the second uniqueness theorem, this must be *the* solution: The charge will flow down the tiny wires, canceling itself off.

Problem 3.5 Prove that the field is uniquely determined when the charge density ρ is given and *either* V *or* the normal derivative $\partial V/\partial n$ is specified on each boundary surface. Do not assume the boundaries are conductors, or that V is constant over any given surface.

Problem 3.6 A more elegant proof of the second uniqueness theorem uses Green's identity (Prob. 1.61c), with $T = U = V_3$. Supply the details.

3.2 ■ THE METHOD OF IMAGES

3.2.1 ■ The Classic Image Problem

Suppose a point charge q is held a distance d above an infinite grounded conducting plane (Fig. 3.10). *Question:* What is the potential in the region above the plane? It's not just $(1/4\pi\epsilon_0)q/\mathscr{r}$, for q will induce a certain amount of negative charge on the nearby surface of the conductor; the total potential is due in part to q directly, and in part to this induced charge. But how can we possibly calculate the potential, when we don't know how much charge is induced or how it is distributed?

From a mathematical point of view, our problem is to solve Poisson's equation in the region $z > 0$, with a single point charge q at $(0, 0, d)$, subject to the boundary conditions:

1. $V = 0$ when $z = 0$ (since the conducting plane is grounded), and
2. $V \to 0$ far from the charge $\big($that is, for $x^2 + y^2 + z^2 \gg d^2\big)$.

The first uniqueness theorem (actually, its corollary) guarantees that there is only one function that meets these requirements. If by trick or clever guess we can discover such a function, it's got to be the answer.

Trick: Forget about the actual problem; we're going to study a *completely different* situation. This new configuration consists of *two* point charges, $+q$ at

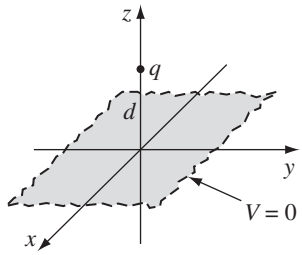


FIGURE 3.10

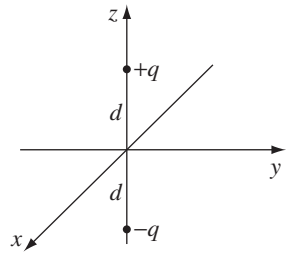


FIGURE 3.11

$(0, 0, d)$ and $-q$ at $(0, 0, -d)$, and *no* conducting plane (Fig. 3.11). For this configuration, I can easily write down the potential:

$$V(x, y, z) = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{\sqrt{x^2 + y^2 + (z-d)^2}} - \frac{q}{\sqrt{x^2 + y^2 + (z+d)^2}}\right]. \quad (3.9)$$

(The denominators represent the distances from (x, y, z) to the charges $+q$ and $-q$, respectively.) It follows that

1. $V = 0$ when $z = 0$,
2. $V \to 0$ for $x^2 + y^2 + z^2 \gg d^2$,

and the only charge in the region $z > 0$ is the point charge $+q$ at $(0, 0, d)$. But these are precisely the conditions of the original problem! Evidently the second configuration happens to produce exactly the same potential as the first configuration, in the "upper" region $z \geq 0$. (The "lower" region, $z < 0$, is completely different, but who cares? The upper part is all we need.) *Conclusion:* The potential of a point charge above an infinite grounded conductor is given by Eq. 3.9, for $z \geq 0$.

Notice the crucial role played by the uniqueness theorem in this argument: without it, no one would believe this solution, since it was obtained for a completely different charge distribution. But the uniqueness theorem certifies it: If it satisfies Poisson's equation in the region of interest, and assumes the correct value at the boundaries, then it must be right.

3.2.2 ■ Induced Surface Charge

Now that we know the potential, it is a straightforward matter to compute the surface charge σ induced on the conductor. According to Eq. 2.49,

$$\sigma = -\epsilon_0 \frac{\partial V}{\partial n},$$

where $\partial V/\partial n$ is the normal derivative of V at the surface. In this case the normal direction is the z direction, so

$$\sigma = -\epsilon_0 \frac{\partial V}{\partial z}\bigg|_{z=0}.$$

From Eq. 3.9,

$$\frac{\partial V}{\partial z} = \frac{1}{4\pi\epsilon_0}\left\{\frac{-q(z-d)}{[x^2+y^2+(z-d)^2]^{3/2}} + \frac{q(z+d)}{[x^2+y^2+(z+d)^2]^{3/2}}\right\},$$

so[5]

$$\sigma(x, y) = \frac{-qd}{2\pi(x^2+y^2+d^2)^{3/2}}. \tag{3.10}$$

As expected, the induced charge is negative (assuming q is positive) and greatest at $x = y = 0$.

While we're at it, let's compute the *total* induced charge

$$Q = \int \sigma \, da.$$

This integral, over the xy plane, could be done in Cartesian coordinates, with $da = dx\,dy$, but it's a little easier to use polar coordinates (r, ϕ), with $r^2 = x^2+y^2$ and $da = r\,dr\,d\phi$. Then

$$\sigma(r) = \frac{-qd}{2\pi(r^2+d^2)^{3/2}},$$

and

$$Q = \int_0^{2\pi}\int_0^{\infty} \frac{-qd}{2\pi(r^2+d^2)^{3/2}} r\,dr\,d\phi = \frac{qd}{\sqrt{r^2+d^2}}\bigg|_0^{\infty} = -q. \tag{3.11}$$

The total charge induced on the plane is $-q$, as (with benefit of hindsight) you can perhaps convince yourself it *had* to be.

3.2.3 ■ Force and Energy

The charge q is attracted toward the plane, because of the negative induced charge. Let's calculate the force of attraction. Since the potential in the vicinity of q is the same as in the analog problem (the one with $+q$ and $-q$ but no conductor), so also is the field and, therefore, the force:

$$\mathbf{F} = -\frac{1}{4\pi\epsilon_0}\frac{q^2}{(2d)^2}\hat{\mathbf{z}}. \tag{3.12}$$

[5]For an entirely different derivation of this result, see Prob. 3.38.

Beware: It is easy to get carried away, and assume that *everything* is the same in the two problems. Energy, however, is *not* the same. With the two point charges and no conductor, Eq. 2.42 gives

$$W = -\frac{1}{4\pi\epsilon_0}\frac{q^2}{2d}. \tag{3.13}$$

But for a single charge and conducting plane, the energy is *half* of this:

$$W = -\frac{1}{4\pi\epsilon_0}\frac{q^2}{4d}. \tag{3.14}$$

Why half? Think of the energy stored in the fields (Eq. 2.45):

$$W = \frac{\epsilon_0}{2}\int E^2\, d\tau.$$

In the first case, both the upper region ($z > 0$) and the lower region ($z < 0$) contribute—and by symmetry they contribute equally. But in the second case, only the upper region contains a nonzero field, and hence the energy is half as great.[6]

Of course, one could also determine the energy by calculating the work required to bring q in from infinity. The force required (to oppose the electrical force in Eq. 3.12) is $(1/4\pi\epsilon_0)(q^2/4z^2)\hat{\mathbf{z}}$, so

$$W = \int_\infty^d \mathbf{F}\cdot d\mathbf{l} = \frac{1}{4\pi\epsilon_0}\int_\infty^d \frac{q^2}{4z^2}\, dz$$
$$= \frac{1}{4\pi\epsilon_0}\left(-\frac{q^2}{4z}\right)\Bigg|_\infty^d = -\frac{1}{4\pi\epsilon_0}\frac{q^2}{4d}.$$

As I move q toward the conductor, I do work *only on* q. It is true that induced charge is moving in over the conductor, but this costs me nothing, since the whole conductor is at potential zero. By contrast, if I simultaneously bring in *two* point charges (with no conductor), I do work on *both* of them, and the total is (again) twice as great.

3.2.4 ■ Other Image Problems

The method just described is not limited to a single point charge; *any* stationary charge distribution near a grounded conducting plane can be treated in the same way, by introducing its mirror image—hence the name **method of images.** (Remember that the image charges have the *opposite sign;* this is what guarantees that the xy plane will be at potential zero.) There are also some exotic problems that can be handled in similar fashion; the nicest of these is the following.

[6]For a generalization of this result, see M. M. Taddei, T. N. C. Mendes, and C. Farina, *Eur. J. Phys.* **30**, 965 (2009), and Prob. 3.41b.

Example 3.2. A point charge q is situated a distance a from the center of a grounded conducting sphere of radius R (Fig. 3.12). Find the potential outside the sphere.

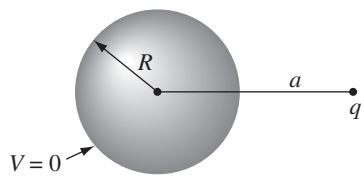


FIGURE 3.12

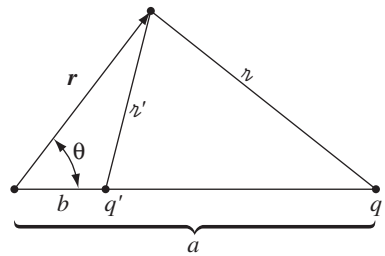


FIGURE 3.13

Solution

Examine the *completely different* configuration, consisting of the point charge q together with another point charge

$$q' = -\frac{R}{a}q, \tag{3.15}$$

placed a distance

$$b = \frac{R^2}{a} \tag{3.16}$$

to the right of the center of the sphere (Fig. 3.13). No conductor, now—just the two point charges. The potential of this configuration is

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r} + \frac{q'}{r'}\right), \tag{3.17}$$

where r and r' are the distances from q and q', respectively. Now, it happens (see Prob. 3.8) that this potential vanishes at all points on the sphere, and therefore fits the boundary conditions for our original problem, in the exterior region.[7]

Conclusion: Eq. 3.17 is the potential of a point charge near a grounded conducting sphere. (Notice that b is less than R, so the "image" charge q' is safely inside the sphere—*you cannot put image charges in the region where you are calculating V*; that would change ρ, and you'd be solving Poisson's equation with

[7]This solution is due to William Thomson (later Lord Kelvin), who published it in 1848, when he was just 24. It was apparently inspired by a theorem of Apollonius (200 BC) that says the locus of points with a fixed ratio of distances from two given points is a sphere. See J. C. Maxwell, "Treatise on Electricity and Magnetism, Vol. I," Dover, New York, p. 245. I thank Gabriel Karl for this interesting history.

the wrong source.) In particular, the force of attraction between the charge and the sphere is

$$F = \frac{1}{4\pi\epsilon_0}\frac{qq'}{(a-b)^2} = -\frac{1}{4\pi\epsilon_0}\frac{q^2Ra}{(a^2-R^2)^2}. \tag{3.18}$$

The method of images is delightfully simple ... when it works. But it is as much an art as a science, for you must somehow think up just the right "auxiliary" configuration, and for most shapes this is forbiddingly complicated, if not impossible.

Problem 3.7 Find the force on the charge $+q$ in Fig. 3.14. (The xy plane is a grounded conductor.)

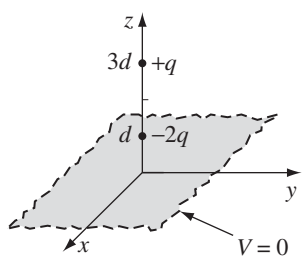


FIGURE 3.14

Problem 3.8

(a) Using the law of cosines, show that Eq. 3.17 can be written as follows:

$$V(r,\theta) = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{\sqrt{r^2+a^2-2ra\,\cos\theta}} - \frac{q}{\sqrt{R^2+(ra/R)^2-2ra\,\cos\theta}}\right], \tag{3.19}$$

where r and θ are the usual spherical polar coordinates, with the z axis along the line through q. In this form, it is obvious that $V = 0$ on the sphere, $r = R$.

(b) Find the induced surface charge on the sphere, as a function of θ. Integrate this to get the total induced charge. (What *should* it be?)

(c) Calculate the energy of this configuration.

Problem 3.9 In Ex. 3.2 we assumed that the conducting sphere was grounded ($V = 0$). But with the addition of a second image charge, the same basic model will handle the case of a sphere at *any* potential V_0 (relative, of course, to infinity). What charge should you use, and where should you put it? Find the force of attraction between a point charge q and a *neutral* conducting sphere.

! **Problem 3.10** A uniform line charge λ is placed on an infinite straight wire, a distance d above a grounded conducting plane. (Let's say the wire runs parallel to the x-axis and directly above it, and the conducting plane is the xy plane.)

(a) Find the potential in the region above the plane. [*Hint:* Refer to Prob. 2.52.]

(b) Find the charge density σ induced on the conducting plane.

Problem 3.11 Two semi-infinite grounded conducting planes meet at right angles. In the region between them, there is a point charge q, situated as shown in Fig. 3.15. Set up the image configuration, and calculate the potential in this region. What charges do you need, and where should they be located? What is the force on q? How much work did it take to bring q in from infinity? Suppose the planes met at some angle other than 90°; would you still be able to solve the problem by the method of images? If not, for what particular angles *does* the method work?

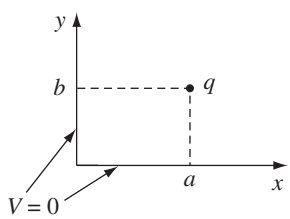


FIGURE 3.15

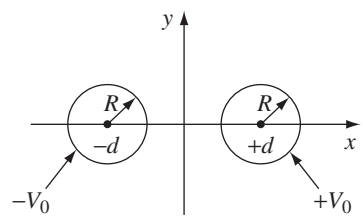


FIGURE 3.16

! **Problem 3.12** Two long, straight copper pipes, each of radius R, are held a distance $2d$ apart. One is at potential V_0, the other at $-V_0$ (Fig. 3.16). Find the potential everywhere. [*Hint:* Exploit the result of Prob. 2.52.]

3.3 ■ SEPARATION OF VARIABLES

In this section we shall attack Laplace's equation directly, using the method of **separation of variables**, which is the physicist's favorite tool for solving partial differential equations. The method is applicable in circumstances where the potential (V) or the charge density (σ) is specified on the boundaries of some region, and we are asked to find the potential in the interior. The basic strategy is very simple: *We look for solutions that are* products *of functions, each of which depends on only* one *of the coordinates*. The algebraic details, however, can be formidable, so I'm going to develop the method through a sequence of examples. We'll start with Cartesian coordinates and then do spherical coordinates (I'll leave the cylindrical case for you to tackle on your own, in Prob. 3.24).

3.3.1 ■ Cartesian Coordinates

Example 3.3. Two infinite grounded metal plates lie parallel to the xz plane, one at $y = 0$, the other at $y = a$ (Fig. 3.17). The left end, at $x = 0$, is closed off with an infinite strip insulated from the two plates, and maintained at a specific potential $V_0(y)$. Find the potential inside this "slot."

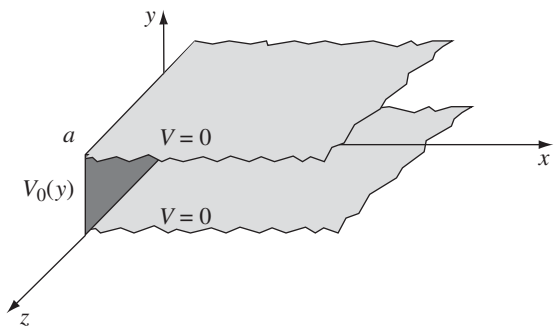


FIGURE 3.17

Solution

The configuration is independent of z, so this is really a *two*-dimensional problem. In mathematical terms, we must solve Laplace's equation,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0, \tag{3.20}$$

subject to the boundary conditions

$$\left.\begin{array}{ll} \text{(i)} & V = 0 \text{ when } y = 0, \\ \text{(ii)} & V = 0 \text{ when } y = a, \\ \text{(iii)} & V = V_0(y) \text{ when } x = 0, \\ \text{(iv)} & V \to 0 \text{ as } x \to \infty. \end{array}\right\} \tag{3.21}$$

(The latter, although not explicitly stated in the problem, is necessary on physical grounds: as you get farther and farther away from the "hot" strip at $x = 0$, the potential should drop to zero.) Since the potential is specified on all boundaries, the answer is uniquely determined.

The first step is to look for solutions in the form of products:

$$V(x, y) = X(x)Y(y). \tag{3.22}$$

On the face of it, this is an absurd restriction—the overwhelming majority of solutions to Laplace's equation do *not* have such a form. For example, $V(x, y) =$

$(5x + 6y)$ satisfies Eq. 3.20, but you can't express it as the product of a function x times a function y. Obviously, we're only going to get a tiny subset of all possible solutions by this means, and it would be a *miracle* if one of them happened to fit the boundary conditions of our problem ... But hang on, because the solutions we *do* get are very special, and it turns out that by pasting them together we can construct the general solution.

Anyway, putting Eq. 3.22 into Eq. 3.20, we obtain

$$Y\frac{d^2X}{dx^2} + X\frac{d^2Y}{dy^2} = 0.$$

The next step is to "separate the variables" (that is, collect all the x-dependence into one term and all the y-dependence into another). Typically, this is accomplished by dividing through by V:

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} = 0. \tag{3.23}$$

Here the first term depends only on x and the second only on y; in other words, we have an equation of the form

$$f(x) + g(y) = 0. \tag{3.24}$$

Now, there's only one way this could possibly be true: *f and g must both be constant.* For what if $f(x)$ *changed,* as you vary x—then if we held y fixed and fiddled with x, the sum $f(x) + g(y)$ would *change,* in violation of Eq. 3.24, which says it's always zero. (That's a simple but somehow rather elusive argument; don't accept it without due thought, because the whole method rides on it.)

It follows from Eq. 3.23, then, that

$$\frac{1}{X}\frac{d^2X}{dx^2} = C_1 \quad \text{and} \quad \frac{1}{Y}\frac{d^2Y}{dy^2} = C_2, \quad \text{with} \quad C_1 + C_2 = 0. \tag{3.25}$$

One of these constants is positive, the other negative (or perhaps both are zero). In general, one must investigate all the possibilities; however, in our particular problem we need C_1 positive and C_2 negative, for reasons that will appear in a moment. Thus

$$\frac{d^2X}{dx^2} = k^2X, \qquad \frac{d^2Y}{dy^2} = -k^2Y. \tag{3.26}$$

Notice what has happened: A *partial* differential equation (3.20) has been converted into two *ordinary* differential equations (3.26). The advantage of this is obvious—ordinary differential equations are a lot easier to solve. Indeed:

$$X(x) = Ae^{kx} + Be^{-kx}, \qquad Y(y) = C\sin ky + D\cos ky,$$

so

$$V(x, y) = (Ae^{kx} + Be^{-kx})(C\sin ky + D\cos ky). \tag{3.27}$$

This is the appropriate separable solution to Laplace's equation; it remains to impose the boundary conditions, and see what they tell us about the constants. To begin at the end, condition (iv) requires that A equal zero.[8] Absorbing B into C and D, we are left with

$$V(x, y) = e^{-kx}(C \sin ky + D \cos ky).$$

Condition (i) now demands that D equal zero, so

$$V(x, y) = Ce^{-kx} \sin ky. \tag{3.28}$$

Meanwhile (ii) yields $\sin ka = 0$, from which it follows that

$$k = \frac{n\pi}{a}, \qquad (n = 1, 2, 3, \ldots). \tag{3.29}$$

(At this point you can see why I chose C_1 positive and C_2 negative: If X were sinusoidal, we could never arrange for it to go to zero at infinity, and if Y were exponential we could not make it vanish at both 0 and a. Incidentally, $n = 0$ is no good, for in that case the potential vanishes *everywhere*. And we have already excluded negative n's.)

That's as far as we can go, using separable solutions, and unless $V_0(y)$ just happens to have the form $\sin(n\pi y/a)$ for some integer n, we simply *can't fit* the final boundary condition at $x = 0$. But now comes the crucial step that redeems the method: Separation of variables has given us an *infinite family* of solutions (one for each n), and whereas none of them *by itself* satisfies the final boundary condition, it is possible to combine them in a way that *does.* Laplace's equation is *linear,* in the sense that if V_1, V_2, V_3, ... satisfy it, so does any **linear combination**, $V = \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 + \ldots$, where $\alpha_1, \alpha_2, \ldots$ are arbitrary constants. For

$$\nabla^2 V = \alpha_1 \nabla^2 V_1 + \alpha_2 \nabla^2 V_2 + \ldots = 0\alpha_1 + 0\alpha_2 + \ldots = 0.$$

Exploiting this fact, we can patch together the separable solutions (Eq. 3.28) to construct a much more general solution:

$$V(x, y) = \sum_{n=1}^{\infty} C_n e^{-n\pi x/a} \sin(n\pi y/a). \tag{3.30}$$

This still satisfies three of the boundary conditions; the question is, can we (by astute choice of the coefficients C_n) fit the final boundary condition (iii)?

$$V(0, y) = \sum_{n=1}^{\infty} C_n \sin(n\pi y/a) = V_0(y). \tag{3.31}$$

[8] I'm assuming k is positive, but this involves no loss of generality—negative k gives the same solution (Eq. 3.27), only with the constants shuffled ($A \leftrightarrow B, C \rightarrow -C$). Occasionally (though not in this example) $k = 0$ must also be included (see Prob. 3.54).

Well, you may recognize this sum—it's a **Fourier sine series**. And Dirichlet's theorem[9] guarantees that virtually *any* function $V_0(y)$—it can even have a finite number of discontinuities—can be expanded in such a series.

But how do we actually *determine* the coefficients C_n, buried as they are in that infinite sum? The device for accomplishing this is so lovely it deserves a name—I call it **Fourier's trick**, though it seems Euler had used essentially the same idea somewhat earlier. Here's how it goes: Multiply Eq. 3.31 by $\sin(n'\pi y/a)$ (where n' is a positive integer), and integrate from 0 to a:

$$\sum_{n=1}^{\infty} C_n \int_0^a \sin(n\pi y/a)\, \sin(n'\pi y/a)\, dy = \int_0^a V_0(y) \sin(n'\pi y/a)\, dy. \quad (3.32)$$

You can work out the integral on the left for yourself; the answer is

$$\int_0^a \sin(n\pi y/a)\, \sin(n'\pi y/a)\, dy = \begin{cases} 0, & \text{if } n' \neq n, \\ \dfrac{a}{2}, & \text{if } n' = n. \end{cases} \quad (3.33)$$

Thus all the terms in the series drop out, save only the one where $n = n'$, and the left side of Eq. 3.32, reduces to $(a/2)C_{n'}$. *Conclusion:*[10]

$$C_n = \frac{2}{a} \int_0^a V_0(y) \sin(n\pi y/a)\, dy. \quad (3.34)$$

That *does* it: Eq. 3.30 is the solution, with coefficients given by Eq. 3.34. As a concrete example, suppose the strip at $x = 0$ is a metal plate with constant potential V_0 (remember, it's insulated from the grounded plates at $y = 0$ and $y = a$). Then

$$C_n = \frac{2V_0}{a} \int_0^a \sin(n\pi y/a)\, dy = \frac{2V_0}{n\pi}(1 - \cos n\pi) = \begin{cases} 0, & \text{if } n \text{ is even}, \\ \dfrac{4V_0}{n\pi}, & \text{if } n \text{ is odd}. \end{cases} \quad (3.35)$$

Thus

$$V(x, y) = \frac{4V_0}{\pi} \sum_{n=1,3,5\ldots} \frac{1}{n} e^{-n\pi x/a} \sin(n\pi y/a). \quad (3.36)$$

Figure 3.18 is a plot of this potential; Fig. 3.19 shows how the first few terms in the Fourier series combine to make a better and better approximation to the constant V_0: (a) is $n = 1$ only, (b) includes n up to 5, (c) is the sum of the first 10 terms, and (d) is the sum of the first 100 terms.

[9] Boas, M., *Mathematical Methods in the Physical Sciences,* 2nd ed. (New York: John Wiley, 1983).

[10] For aesthetic reasons I've dropped the prime; Eq. 3.34 holds for $n = 1, 2, 3, \ldots$, and it doesn't matter (obviously) what letter you use for the "dummy" index.

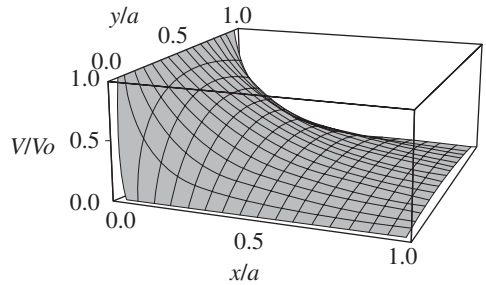


FIGURE 3.18

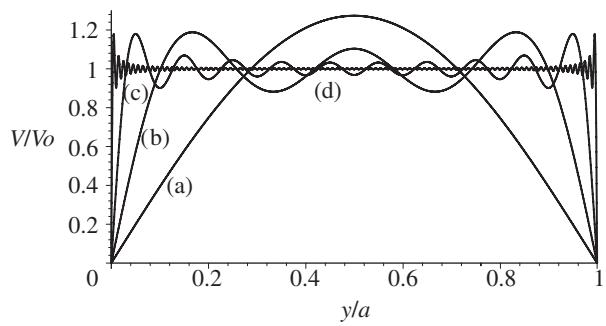


FIGURE 3.19

Incidentally, the infinite series in Eq. 3.36 can be summed explicitly (try your hand at it, if you like); the result is

$$V(x, y) = \frac{2V_0}{\pi} \tan^{-1}\left(\frac{\sin(\pi y/a)}{\sinh(\pi x/a)}\right). \tag{3.37}$$

In this form, it is easy to check that Laplace's equation is obeyed and the four boundary conditions (Eq. 3.21) are satisfied.

The success of this method hinged on two extraordinary properties of the separable solutions (Eqs. 3.28 and 3.29): **completeness** and **orthogonality**. A set of functions $f_n(y)$ is said to be **complete** if any other function $f(y)$ can be expressed as a linear combination of them:

$$f(y) = \sum_{n=1}^{\infty} C_n f_n(y). \tag{3.38}$$

The functions $\sin(n\pi y/a)$ are complete on the interval $0 \leq y \leq a$. It was this fact, guaranteed by Dirichlet's theorem, that assured us Eq. 3.31 could be satisfied, given the proper choice of the coefficients C_n. (The *proof* of completeness, for a particular set of functions, is an extremely difficult business, and I'm afraid

physicists tend to *assume* it's true and leave the checking to others.) A set of functions is **orthogonal** if the integral of the product of any two different members of the set is zero:

$$\int_0^a f_n(y)\, f_{n'}(y)\, dy = 0 \qquad \text{for } n' \neq n. \tag{3.39}$$

The sine functions are orthogonal (Eq. 3.33); this is the property on which Fourier's trick is based, allowing us to kill off all terms but one in the infinite series and thereby solve for the coefficients C_n. (Proof of orthogonality is generally quite simple, either by direct integration or by analysis of the differential equation from which the functions came.)

Example 3.4. Two infinitely-long grounded metal plates, again at $y = 0$ and $y = a$, are connected at $x = \pm b$ by metal strips maintained at a constant potential V_0, as shown in Fig. 3.20 (a thin layer of insulation at each corner prevents them from shorting out). Find the potential inside the resulting rectangular pipe.

Solution

Once again, the configuration is independent of z. Our problem is to solve Laplace's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0,$$

subject to the boundary conditions

$$\left.\begin{array}{ll} \text{(i)} & V = 0 \text{ when } y = 0, \\ \text{(ii)} & V = 0 \text{ when } y = a, \\ \text{(iii)} & V = V_0 \text{ when } x = b, \\ \text{(iv)} & V = V_0 \text{ when } x = -b. \end{array}\right\} \tag{3.40}$$

The argument runs as before, up to Eq. 3.27:

$$V(x, y) = (Ae^{kx} + Be^{-kx})(C \sin ky + D \cos ky).$$

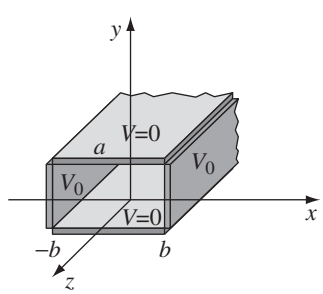


FIGURE 3.20

This time, however, we cannot set $A = 0$; the region in question does not extend to $x = \infty$, so e^{kx} is perfectly acceptable. On the other hand, the situation is *symmetric* with respect to x, so $V(-x, y) = V(x, y)$, and it follows that $A = B$. Using

$$e^{kx} + e^{-kx} = 2\cosh kx,$$

and absorbing $2A$ into C and D, we have

$$V(x, y) = \cosh kx\,(C \sin ky + D \cos ky).$$

Boundary conditions (i) and (ii) require, as before, that $D = 0$ and $k = n\pi/a$, so

$$V(x, y) = C \cosh(n\pi x/a)\, \sin(n\pi y/a). \tag{3.41}$$

Because $V(x, y)$ is even in x, it will automatically meet condition (iv) if it fits (iii). It remains, therefore, to construct the general linear combination,

$$V(x, y) = \sum_{n=1}^{\infty} C_n \cosh(n\pi x/a)\, \sin(n\pi y/a),$$

and pick the coefficients C_n in such a way as to satisfy condition (iii):

$$V(b, y) = \sum_{n=1}^{\infty} C_n \cosh(n\pi b/a)\, \sin(n\pi y/a) = V_0.$$

This is the same problem in Fourier analysis that we faced before; I quote the result from Eq. 3.35:

$$C_n \cosh(n\pi b/a) = \begin{cases} 0, & \text{if } n \text{ is even} \\ \dfrac{4V_0}{n\pi}, & \text{if } n \text{ is odd} \end{cases}$$

Conclusion: The potential in this case is given by

$$V(x, y) = \frac{4V_0}{\pi} \sum_{n=1,3,5\ldots} \frac{1}{n} \frac{\cosh(n\pi x/a)}{\cosh(n\pi b/a)}\, \sin(n\pi y/a). \tag{3.42}$$

This function is shown in Fig. 3.21.

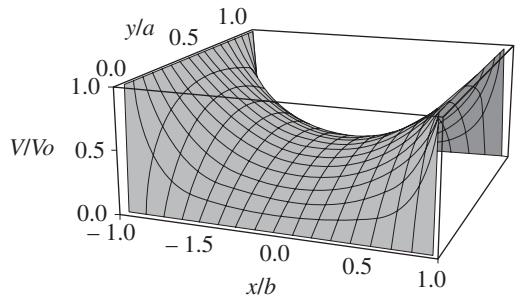


FIGURE 3.21

Example 3.5. An infinitely long rectangular metal pipe (sides a and b) is grounded, but one end, at $x = 0$, is maintained at a specified potential $V_0(y, z)$, as indicated in Fig. 3.22. Find the potential inside the pipe.

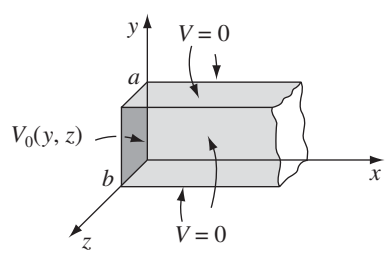


FIGURE 3.22

Solution

This is a genuinely three-dimensional problem,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0, \tag{3.43}$$

subject to the boundary conditions

$$\left.\begin{array}{ll} \text{(i)} & V = 0 \text{ when } y = 0, \\ \text{(ii)} & V = 0 \text{ when } y = a, \\ \text{(iii)} & V = 0 \text{ when } z = 0, \\ \text{(iv)} & V = 0 \text{ when } z = b, \\ \text{(v)} & V \to 0 \text{ as } x \to \infty, \\ \text{(vi)} & V = V_0(y, z) \text{ when } x = 0. \end{array}\right\} \tag{3.44}$$

As always, we look for solutions that are products:

$$V(x, y, z) = X(x)Y(y)Z(z). \tag{3.45}$$

Putting this into Eq. 3.43, and dividing by V, we find

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = 0.$$

It follows that

$$\frac{1}{X}\frac{d^2X}{dx^2} = C_1, \quad \frac{1}{Y}\frac{d^2Y}{dy^2} = C_2, \quad \frac{1}{Z}\frac{d^2Z}{dz^2} = C_3, \quad \text{with } C_1 + C_2 + C_3 = 0.$$

Our previous experience (Ex. 3.3) suggests that C_1 must be positive, C_2 and C_3 negative. Setting $C_2 = -k^2$ and $C_3 = -l^2$, we have $C_1 = k^2 + l^2$, and hence

$$\frac{d^2X}{dx^2} = (k^2 + l^2)X, \quad \frac{d^2Y}{dy^2} = -k^2Y, \quad \frac{d^2Z}{dz^2} = -l^2Z. \tag{3.46}$$

Once again, separation of variables has turned a *partial* differential equation into *ordinary* differential equations. The solutions are

$$\begin{aligned} X(x) &= Ae^{\sqrt{k^2+l^2}\,x} + Be^{-\sqrt{k^2+l^2}\,x}, \\ Y(y) &= C\sin ky + D\cos ky, \\ Z(z) &= E\sin lz + F\cos lz. \end{aligned}$$

Boundary condition (v) implies $A = 0$, (i) gives $D = 0$, and (iii) yields $F = 0$, whereas (ii) and (iv) require that $k = n\pi/a$ and $l = m\pi/b$, where n and m are positive integers. Combining the remaining constants, we are left with

$$V(x, y, z) = Ce^{-\pi\sqrt{(n/a)^2+(m/b)^2}\,x} \sin(n\pi y/a)\, \sin(m\pi z/b). \tag{3.47}$$

This solution meets all the boundary conditions except (vi). It contains *two* unspecified integers (n and m), and the most general linear combination is a *double* sum:

$$V(x, y, z) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m}e^{-\pi\sqrt{(n/a)^2+(m/b)^2}\,x} \sin(n\pi y/a)\, \sin(m\pi z/b). \tag{3.48}$$

We hope to fit the remaining boundary condition,

$$V(0, y, z) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m} \sin(n\pi y/a)\, \sin(m\pi z/b) = V_0(y, z), \tag{3.49}$$

by appropriate choice of the coefficients $C_{n,m}$. To determine these constants, we multiply by $\sin(n'\pi y/a)\,\sin(m'\pi z/b)$, where n' and m' are arbitrary positive integers, and integrate:

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m}\int_0^a \sin(n\pi y/a)\,\sin(n'\pi y/a)\,dy\int_0^b \sin(m\pi z/b)\,\sin(m'\pi z/b)\,dz$$
$$= \int_0^a\int_0^b V_0(y,z)\sin(n'\pi y/a)\,\sin(m'\pi z/b)\,dy\,dz.$$

Quoting Eq. 3.33, the left side is $(ab/4)C_{n',m'}$, so

$$C_{n,m} = \frac{4}{ab}\int_0^a\int_0^b V_0(y,z)\sin(n\pi y/a)\,\sin(m\pi z/b)\,dy\,dz. \tag{3.50}$$

Equation 3.48, with the coefficients given by Eq. 3.50, is the solution to our problem.

For instance, if the end of the tube is a conductor at *constant* potential V_0,

$$C_{n,m} = \frac{4V_0}{ab}\int_0^a \sin(n\pi y/a)\,dy\int_0^b \sin(m\pi z/b)\,dz$$
$$= \begin{cases} 0, & \text{if } n \text{ or } m \text{ is even,} \\ \dfrac{16V_0}{\pi^2 nm}, & \text{if } n \text{ and } m \text{ are odd.} \end{cases} \tag{3.51}$$

In this case

$$V(x,y,z) = \frac{16V_0}{\pi^2}\sum_{n,m=1,3,5\ldots}^{\infty}\frac{1}{nm}\,e^{-\pi\sqrt{(n/a)^2+(m/b)^2}\,x}\sin(n\pi y/a)\,\sin(m\pi z/b). \tag{3.52}$$

Notice that the successive terms decrease rapidly; a reasonable approximation would be obtained by keeping only the first few.

Problem 3.13 Find the potential in the infinite slot of Ex. 3.3 if the boundary at $x = 0$ consists of two metal strips: one, from $y = 0$ to $y = a/2$, is held at a constant potential V_0, and the other, from $y = a/2$ to $y = a$, is at potential $-V_0$.

Problem 3.14 For the infinite slot (Ex. 3.3), determine the charge density $\sigma(y)$ on the strip at $x = 0$, assuming it is a conductor at constant potential V_0.

Problem 3.15 A rectangular pipe, running parallel to the z-axis (from $-\infty$ to $+\infty$), has three grounded metal sides, at $y = 0$, $y = a$, and $x = 0$. The fourth side, at $x = b$, is maintained at a specified potential $V_0(y)$.

(a) Develop a general formula for the potential inside the pipe.

(b) Find the potential explicitly, for the case $V_0(y) = V_0$ (a constant).

Problem 3.16 A cubical box (sides of length a) consists of five metal plates, which are welded together and grounded (Fig. 3.23). The top is made of a separate sheet of metal, insulated from the others, and held at a constant potential V_0. Find the potential inside the box. [What should the potential at the center $(a/2, a/2, a/2)$ be? Check numerically that your formula is consistent with this value.][11]

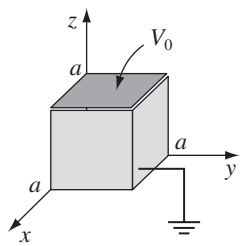


FIGURE 3.23

3.3.2 ■ Spherical Coordinates

In the examples considered so far, Cartesian coordinates were clearly appropriate, since the boundaries were planes. For round objects, spherical coordinates are more natural. In the spherical system, Laplace's equation reads:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 V}{\partial\phi^2}=0. \tag{3.53}$$

I shall assume the problem has **azimuthal symmetry**, so that V *is independent of* ϕ;[12] in that case, Eq. 3.53 reduces to

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial V}{\partial\theta}\right)=0. \tag{3.54}$$

As before, we look for solutions that are products:

$$V(r,\theta)=R(r)\Theta(\theta). \tag{3.55}$$

Putting this into Eq. 3.54, and dividing by V,

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{\Theta\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)=0. \tag{3.56}$$

[11]This cute test was suggested by J. Castro.

[12]The general case, for ϕ-dependent potentials, is treated in all the graduate texts. See, for instance, J. D. Jackson's *Classical Electrodynamics*, 3rd ed. (New York: John Wiley, 1999), Chapter 3.

Since the first term depends only on r, and the second only on θ, it follows that each must be a constant:

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = l(l+1), \quad \frac{1}{\Theta \sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) = -l(l+1). \tag{3.57}$$

Here $l(l+1)$ is just a fancy way of writing the separation constant—you'll see in a minute why this is convenient.

As always, separation of variables has converted a *partial* differential equation (3.54) into *ordinary* differential equations (3.57). The radial equation,

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = l(l+1)R, \tag{3.58}$$

has the general solution

$$R(r) = Ar^l + \frac{B}{r^{l+1}}, \tag{3.59}$$

as you can easily check; A and B are the two arbitrary constants to be expected in the solution of a second-order differential equation. But the angular equation,

$$\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) = -l(l+1)\sin\theta\,\Theta, \tag{3.60}$$

is not so simple. The solutions are **Legendre polynomials** in the variable $\cos\theta$:

$$\Theta(\theta) = P_l(\cos\theta). \tag{3.61}$$

$P_l(x)$ is most conveniently defined by the **Rodrigues formula:**

$$P_l(x) \equiv \frac{1}{2^l l!}\left(\frac{d}{dx}\right)^l (x^2-1)^l. \tag{3.62}$$

The first few Legendre polynomials are listed in Table 3.1.

$P_0(x)$	=	1
$P_1(x)$	=	x
$P_2(x)$	=	$(3x^2-1)/2$
$P_3(x)$	=	$(5x^3-3x)/2$
$P_4(x)$	=	$(35x^4-30x^2+3)/8$
$P_5(x)$	=	$(63x^5-70x^3+15x)/8$

TABLE 3.1 Legendre Polynomials.

Notice that $P_l(x)$ is (as the name suggests) an lth-order *polynomial* in x; it contains only *even* powers, if l is even, and *odd* powers, if l is odd. The factor in front $(1/2^l l!)$ was chosen in order that

$$P_l(1) = 1. \tag{3.63}$$

The Rodrigues formula obviously works only for nonnegative integer values of l. Moreover, it provides us with only *one* solution. But Eq. 3.60 is *second*-order, and it should possess *two* independent solutions, for *every* value of l. It turns out that these "other solutions" blow up at $\theta = 0$ and/or $\theta = \pi$, and are therefore unacceptable on physical grounds.[13] For instance, the second solution for $l = 0$ is

$$\Theta(\theta) = \ln\left(\tan\frac{\theta}{2}\right). \tag{3.64}$$

You might want to check for yourself that this satisfies Eq. 3.60.

In the case of azimuthal symmetry, then, the most general separable solution to Laplace's equation, consistent with minimal physical requirements, is

$$V(r,\theta) = \left(Ar^l + \frac{B}{r^{l+1}}\right) P_l(\cos\theta).$$

(There was no need to include an overall constant in Eq. 3.61 because it can be absorbed into A and B at this stage.) As before, separation of variables yields an infinite set of solutions, one for each l. The *general* solution is the linear combination of separable solutions:

$$\boxed{V(r,\theta) = \sum_{l=0}^{\infty}\left(A_l r^l + \frac{B_l}{r^{l+1}}\right) P_l(\cos\theta).} \tag{3.65}$$

The following examples illustrate the power of this important result.

Example 3.6. The potential $V_0(\theta)$ is specified on the surface of a hollow sphere, of radius R. Find the potential inside the sphere.

Solution

In this case, $B_l = 0$ for all l—otherwise the potential would blow up at the origin. Thus,

$$V(r,\theta) = \sum_{l=0}^{\infty} A_l r^l P_l(\cos\theta). \tag{3.66}$$

[13] In rare cases where the z axis is excluded, these "other solutions" do have to be considered.

At $r = R$ this must match the specified function $V_0(\theta)$:

$$V(R, \theta) = \sum_{l=0}^{\infty} A_l R^l P_l(\cos\theta) = V_0(\theta). \tag{3.67}$$

Can this equation be satisfied, for an appropriate choice of coefficients A_l? *Yes:* The Legendre polynomials (like the sines) constitute a complete set of functions, on the interval $-1 \leq x \leq 1$ ($0 \leq \theta \leq \pi$). How do we determine the constants? Again, by Fourier's trick, for the Legendre polynomials (like the sines) are *orthogonal* functions:[14]

$$\begin{aligned} \int_{-1}^{1} P_l(x) P_{l'}(x)\, dx &= \int_0^{\pi} P_l(\cos\theta) P_{l'}(\cos\theta) \sin\theta\, d\theta \\ &= \begin{cases} 0, & \text{if } l' \neq l, \\ \dfrac{2}{2l+1}, & \text{if } l' = l. \end{cases} \end{aligned} \tag{3.68}$$

Thus, multiplying Eq. 3.67 by $P_{l'}(\cos\theta)\sin\theta$ and integrating, we have

$$A_{l'} R^{l'} \frac{2}{2l'+1} = \int_0^{\pi} V_0(\theta) P_{l'}(\cos\theta) \sin\theta\, d\theta,$$

or

$$A_l = \frac{2l+1}{2R^l} \int_0^{\pi} V_0(\theta) P_l(\cos\theta) \sin\theta\, d\theta. \tag{3.69}$$

Equation 3.66 is the solution to our problem, with the coefficients given by Eq. 3.69.

It can be difficult to evaluate integrals of the form 3.69 analytically, and in practice it is often easier to solve Eq. 3.67 "by eyeball."[15] For instance, suppose we are told that the potential on the sphere is

$$V_0(\theta) = k \sin^2(\theta/2), \tag{3.70}$$

where k is a constant. Using the half-angle formula, we rewrite this as

$$V_0(\theta) = \frac{k}{2}(1 - \cos\theta) = \frac{k}{2}[P_0(\cos\theta) - P_1(\cos\theta)].$$

[14] M. Boas, *Mathematical Methods in the Physical Sciences,* 2nd ed. (New York: John Wiley, 1983), Section 12.7.

[15] This is certainly true whenever $V_0(\theta)$ can be expressed as a polynomial in $\cos\theta$. The degree of the polynomial tells us the highest l we require, and the leading coefficient determines the corresponding A_l. Subtracting off $A_l R^l P_l(\cos\theta)$ and repeating the process, we systematically work our way down to A_0. Notice that if V_0 is an *even* function of $\cos\theta$, then only even terms will occur in the sum (and likewise for odd functions).

Putting this into Eq. 3.67, we read off immediately that $A_0 = k/2$, $A_1 = -k/(2R)$, and all other A_l's vanish. Therefore,

$$V(r,\theta) = \frac{k}{2}\left[r^0 P_0(\cos\theta) - \frac{r^1}{R}P_1(\cos\theta)\right] = \frac{k}{2}\left(1 - \frac{r}{R}\cos\theta\right). \tag{3.71}$$

Example 3.7. The potential $V_0(\theta)$ is again specified on the surface of a sphere of radius R, but this time we are asked to find the potential *outside*, assuming there is no charge there.

Solution

In this case it's the A_l's that must be zero (or else V would not go to zero at ∞), so

$$V(r,\theta) = \sum_{l=0}^{\infty} \frac{B_l}{r^{l+1}} P_l(\cos\theta). \tag{3.72}$$

At the surface of the sphere, we require that

$$V(R,\theta) = \sum_{l=0}^{\infty} \frac{B_l}{R^{l+1}} P_l(\cos\theta) = V_0(\theta).$$

Multiplying by $P_{l'}(\cos\theta)\sin\theta$ and integrating—exploiting, again, the orthogonality relation 3.68—we have

$$\frac{B_{l'}}{R^{l'+1}}\frac{2}{2l'+1} = \int_0^\pi V_0(\theta) P_{l'}(\cos\theta)\sin\theta\, d\theta,$$

or

$$B_l = \frac{2l+1}{2} R^{l+1} \int_0^\pi V_0(\theta) P_l(\cos\theta)\sin\theta\, d\theta. \tag{3.73}$$

Equation 3.72, with the coefficients given by Eq. 3.73, is the solution to our problem.

Example 3.8. An uncharged metal sphere of radius R is placed in an otherwise uniform electric field $\mathbf{E} = E_0\hat{\mathbf{z}}$. The field will push positive charge to the "northern" surface of the sphere, and—symmetrically—negative charge to the "southern" surface (Fig. 3.24). This induced charge, in turn, distorts the field in the neighborhood of the sphere. Find the potential in the region outside the sphere.

Solution

The sphere is an equipotential—we may as well set it to zero. Then by symmetry the entire xy plane is at potential zero. This time, however, V does *not* go to zero at large z. In fact, far from the sphere the field is $E_0\hat{\mathbf{z}}$, and hence

$$V \to -E_0 z + C.$$

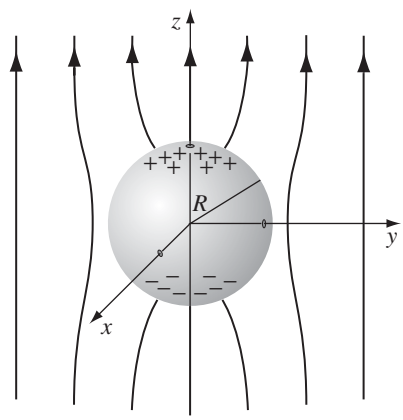


FIGURE 3.24

Since $V = 0$ in the equatorial plane, the constant C must be zero. Accordingly, the boundary conditions for this problem are

$$\left.\begin{array}{ll} \text{(i)} & V = 0 \qquad \text{when } r = R, \\ \text{(ii)} & V \to -E_0 r\cos\theta \qquad \text{for } r \gg R. \end{array}\right\} \tag{3.74}$$

We must fit these boundary conditions with a function of the form 3.65.

The first condition yields

$$A_l R^l + \frac{B_l}{R^{l+1}} = 0,$$

or

$$B_l = -A_l R^{2l+1}, \tag{3.75}$$

so

$$V(r,\theta) = \sum_{l=0}^{\infty} A_l \left(r^l - \frac{R^{2l+1}}{r^{l+1}}\right) P_l(\cos\theta).$$

For $r \gg R$, the second term in parentheses is negligible, and therefore condition (ii) requires that

$$\sum_{l=0}^{\infty} A_l r^l P_l(\cos\theta) = -E_0 r\cos\theta.$$

Evidently only one term is present: $l = 1$. In fact, since $P_1(\cos\theta) = \cos\theta$, we can read off immediately

$$A_1 = -E_0, \quad \text{all other } A_l\text{'s zero.}$$

Conclusion:

$$V(r,\theta) = -E_0\left(r - \frac{R^3}{r^2}\right)\cos\theta. \tag{3.76}$$

The first term $(-E_0 r\cos\theta)$ is due to the external field; the contribution attributable to the induced charge is

$$E_0\frac{R^3}{r^2}\cos\theta.$$

If you want to know the induced charge density, it can be calculated in the usual way:

$$\sigma(\theta) = -\epsilon_0\frac{\partial V}{\partial r}\bigg|_{r=R} = \epsilon_0 E_0\left(1 + 2\frac{R^3}{r^3}\right)\cos\theta\bigg|_{r=R} = 3\epsilon_0 E_0\cos\theta. \tag{3.77}$$

As expected, it is positive in the "northern" hemisphere ($0 \le \theta \le \pi/2$) and negative in the "southern" ($\pi/2 \le \theta \le \pi$).

Example 3.9. A specified charge density $\sigma_0(\theta)$ is glued over the surface of a spherical shell of radius R. Find the resulting potential inside and outside the sphere.

Solution

You could, of course, do this by direct integration:

$$V = \frac{1}{4\pi\epsilon_0}\int \frac{\sigma_0}{\mathscr{r}}\,da,$$

but separation of variables is often easier. For the interior region, we have

$$V(r,\theta) = \sum_{l=0}^{\infty} A_l r^l P_l(\cos\theta) \qquad (r \le R) \tag{3.78}$$

(no B_l terms—they blow up at the origin); in the exterior region

$$V(r,\theta) = \sum_{l=0}^{\infty} \frac{B_l}{r^{l+1}} P_l(\cos\theta) \qquad (r \ge R) \tag{3.79}$$

(no A_l terms—they don't go to zero at infinity). These two functions must be joined together by the appropriate boundary conditions at the surface itself. First, the potential is *continuous* at $r = R$ (Eq. 2.34):

$$\sum_{l=0}^{\infty} A_l R^l P_l(\cos\theta) = \sum_{l=0}^{\infty} \frac{B_l}{R^{l+1}} P_l(\cos\theta). \tag{3.80}$$

It follows that the coefficients of like Legendre polynomials are equal:

$$B_l = A_l R^{2l+1}. \tag{3.81}$$

(To prove that formally, multiply both sides of Eq. 3.80 by $P_{l'}(\cos\theta)\sin\theta$ and integrate from 0 to π, using the orthogonality relation 3.68.) Second, the radial derivative of V suffers a discontinuity at the surface (Eq. 2.36):

$$\left(\frac{\partial V_{\text{out}}}{\partial r} - \frac{\partial V_{\text{in}}}{\partial r}\right)\bigg|_{r=R} = -\frac{1}{\epsilon_0}\sigma_0(\theta). \tag{3.82}$$

Thus

$$-\sum_{l=0}^{\infty}(l+1)\frac{B_l}{R^{l+2}}P_l(\cos\theta) - \sum_{l=0}^{\infty} l A_l R^{l-1} P_l(\cos\theta) = -\frac{1}{\epsilon_0}\sigma_0(\theta),$$

or, using Eq. 3.81,

$$\sum_{l=0}^{\infty}(2l+1)A_l R^{l-1} P_l(\cos\theta) = \frac{1}{\epsilon_0}\sigma_0(\theta). \tag{3.83}$$

From here, the coefficients can be determined using Fourier's trick:

$$A_l = \frac{1}{2\epsilon_0 R^{l-1}} \int_0^{\pi} \sigma_0(\theta) P_l(\cos\theta)\sin\theta\, d\theta. \tag{3.84}$$

Equations 3.78 and 3.79 constitute the solution to our problem, with the coefficients given by Eqs. 3.81 and 3.84.

For instance, if

$$\sigma_0(\theta) = k\cos\theta = k P_1(\cos\theta), \tag{3.85}$$

for some constant k, then all the A_l's are zero except for $l = 1$, and

$$A_1 = \frac{k}{2\epsilon_0}\int_0^{\pi} [P_1(\cos\theta)]^2 \sin\theta\, d\theta = \frac{k}{3\epsilon_0}.$$

The potential inside the sphere is therefore

$$V(r,\theta) = \frac{k}{3\epsilon_0} r\cos\theta \qquad (r \le R), \tag{3.86}$$

whereas outside the sphere

$$V(r,\theta) = \frac{kR^3}{3\epsilon_0}\frac{1}{r^2}\cos\theta \qquad (r \geq R). \tag{3.87}$$

In particular, if $\sigma_0(\theta)$ is the induced charge on a metal sphere in an external field $E_0\hat{\mathbf{z}}$, so that $k = 3\epsilon_0 E_0$ (Eq. 3.77), then the potential inside is $E_0 r\cos\theta = E_0 z$, and the field is $-E_0\hat{\mathbf{z}}$—exactly right to cancel off the external field, as of course it *should* be. Outside the sphere the potential due to this surface charge is

$$E_0\frac{R^3}{r^2}\cos\theta,$$

consistent with our conclusion in Ex. 3.8.

Problem 3.17 Derive $P_3(x)$ from the Rodrigues formula, and check that $P_3(\cos\theta)$ satisfies the angular equation (3.60) for $l = 3$. Check that P_3 and P_1 are orthogonal by explicit integration.

Problem 3.18

(a) Suppose the potential is a *constant* V_0 over the surface of the sphere. Use the results of Ex. 3.6 and Ex. 3.7 to find the potential inside and outside the sphere. (Of course, you know the answers in advance—this is just a consistency check on the method.)

(b) Find the potential inside and outside a spherical shell that carries a uniform surface charge σ_0, using the results of Ex. 3.9.

Problem 3.19 The potential at the surface of a sphere (radius R) is given by

$$V_0 = k\cos 3\theta,$$

where k is a constant. Find the potential inside and outside the sphere, as well as the surface charge density $\sigma(\theta)$ on the sphere. (Assume there's no charge inside or outside the sphere.)

Problem 3.20 Suppose the potential $V_0(\theta)$ at the surface of a sphere is specified, and there is no charge inside or outside the sphere. Show that the charge density on the sphere is given by

$$\sigma(\theta) = \frac{\epsilon_0}{2R}\sum_{l=0}^{\infty}(2l+1)^2 C_l P_l(\cos\theta), \tag{3.88}$$

where

$$C_l = \int_0^\pi V_0(\theta)P_l(\cos\theta)\sin\theta\, d\theta. \tag{3.89}$$

Problem 3.21 Find the potential outside a *charged* metal sphere (charge Q, radius R) placed in an otherwise uniform electric field $\mathbf{E}_0$. Explain clearly where you are setting the zero of potential.

Problem 3.22 In Prob. 2.25, you found the potential on the axis of a uniformly charged disk:

$$V(r, 0) = \frac{\sigma}{2\epsilon_0}\left(\sqrt{r^2 + R^2} - r\right).$$

(a) Use this, together with the fact that $P_l(1) = 1$, to evaluate the first three terms in the expansion (Eq. 3.72) for the potential of the disk at points *off* the axis, assuming $r > R$.

(b) Find the potential for $r < R$ by the same method, using Eq. 3.66. [*Note:* You must break the interior region up into two hemispheres, above and below the disk. Do *not* assume the coefficients A_l are the same in both hemispheres.]

Problem 3.23 A spherical shell of radius R carries a uniform surface charge σ_0 on the "northern" hemisphere and a uniform surface charge $-\sigma_0$ on the "southern" hemisphere. Find the potential inside and outside the sphere, calculating the coefficients explicitly up to A_6 and B_6.

• **Problem 3.24** Solve Laplace's equation by separation of variables in *cylindrical* coordinates, assuming there is no dependence on z (cylindrical symmetry). [Make sure you find *all* solutions to the radial equation; in particular, your result must accommodate the case of an infinite line charge, for which (of course) we already know the answer.]

Problem 3.25 Find the potential outside an infinitely long metal pipe, of radius R, placed at right angles to an otherwise uniform electric field $\mathbf{E}_0$. Find the surface charge induced on the pipe. [Use your result from Prob. 3.24.]

Problem 3.26 Charge density

$$\sigma(\phi) = a \sin 5\phi$$

(where a is a constant) is glued over the surface of an infinite cylinder of radius R (Fig. 3.25). Find the potential inside and outside the cylinder. [Use your result from Prob. 3.24.]

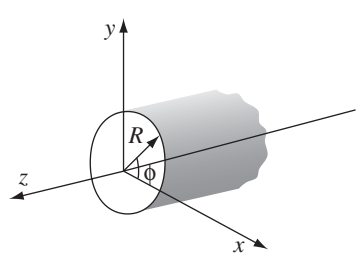


FIGURE 3.25

3.4 ■ MULTIPOLE EXPANSION

3.4.1 ■ Approximate Potentials at Large Distances

If you are very far away from a localized charge distribution, it "looks" like a point charge, and the potential is—to good approximation—$(1/4\pi\epsilon_0)Q/r$, where Q is the total charge. We have often used this as a check on formulas for V. But what if Q is *zero?* You might reply that the potential is then approximately zero, and of course, you're right, in a sense (indeed, the potential at large r is *pretty small* even if Q is *not* zero). But we're looking for something a bit more informative than that.

Example 3.10. A (physical) **electric dipole** consists of two equal and opposite charges ($\pm q$) separated by a distance d. Find the approximate potential at points far from the dipole.

Solution

Let $\mathscr{r}_-$ be the distance from $-q$ and $\mathscr{r}_+$ the distance from $+q$ (Fig. 3.26). Then

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{\mathscr{r}_+} - \frac{q}{\mathscr{r}_-}\right),$$

and (from the law of cosines)

$$\mathscr{r}_\pm^2 = r^2 + (d/2)^2 \mp rd\cos\theta = r^2\left(1 \mp \frac{d}{r}\cos\theta + \frac{d^2}{4r^2}\right).$$

We're interested in the régime $r \gg d$, so the third term is negligible, and the binomial expansion yields

$$\frac{1}{\mathscr{r}_\pm} \cong \frac{1}{r}\left(1 \mp \frac{d}{r}\cos\theta\right)^{-1/2} \cong \frac{1}{r}\left(1 \pm \frac{d}{2r}\cos\theta\right).$$

Thus

$$\frac{1}{\mathscr{r}_+} - \frac{1}{\mathscr{r}_-} \cong \frac{d}{r^2}\cos\theta,$$

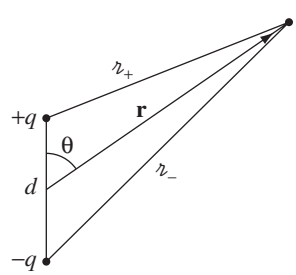


FIGURE 3.26

and hence

$$V(\mathbf{r}) \cong \frac{1}{4\pi\epsilon_0}\frac{qd\cos\theta}{r^2}. \tag{3.90}$$

The potential of a dipole goes like $1/r^2$ at large r; as we might have anticipated, it falls off more rapidly than the potential of a point charge. If we put together a pair of equal and opposite *dipoles* to make a **quadrupole**, the potential goes like $1/r^3$; for back-to-back *quadrupoles* (an **octopole**), it goes like $1/r^4$; and so on. Figure 3.27 summarizes this hierarchy; for completeness I have included the electric **monopole** (point charge), whose potential, of course, goes like $1/r$.

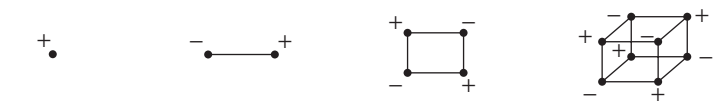


FIGURE 3.27

Example 3.10 pertains to a very special charge configuration. I propose now to develop a systematic expansion for the potential of *any* localized charge distribution, in powers of $1/r$. Figure 3.28 defines the relevant variables; the potential at $\mathbf{r}$ is given by

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{\mathscr{r}}\rho(\mathbf{r}')\,d\tau'. \tag{3.91}$$

Using the law of cosines,

$$\mathscr{r}^2 = r^2 + (r')^2 - 2rr'\cos\alpha = r^2\left[1 + \left(\frac{r'}{r}\right)^2 - 2\left(\frac{r'}{r}\right)\cos\alpha\right],$$

where α is the angle between $\mathbf{r}$ and $\mathbf{r}'$. Thus

$$\mathscr{r} = r\sqrt{1+\epsilon}, \tag{3.92}$$

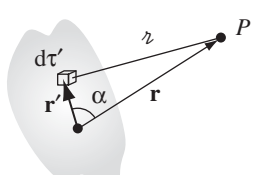


FIGURE 3.28

with

$$\epsilon \equiv \left(\frac{r'}{r}\right)\left(\frac{r'}{r} - 2\cos\alpha\right).$$

For points well outside the charge distribution, ϵ is much less than 1, and this invites a binomial expansion:

$$\frac{1}{\mathscr{r}} = \frac{1}{r}(1+\epsilon)^{-1/2} = \frac{1}{r}\left(1 - \frac{1}{2}\epsilon + \frac{3}{8}\epsilon^2 - \frac{5}{16}\epsilon^3 + \dots\right), \tag{3.93}$$

or, in terms of r, r', and α:

$$\begin{aligned}\frac{1}{\mathscr{r}} &= \frac{1}{r}\left[1 - \frac{1}{2}\left(\frac{r'}{r}\right)\left(\frac{r'}{r} - 2\cos\alpha\right) + \frac{3}{8}\left(\frac{r'}{r}\right)^2\left(\frac{r'}{r} - 2\cos\alpha\right)^2\right. \\ &\qquad \left. -\frac{5}{16}\left(\frac{r'}{r}\right)^3\left(\frac{r'}{r} - 2\cos\alpha\right)^3 + \dots\right] \\ &= \frac{1}{r}\left[1 + \left(\frac{r'}{r}\right)(\cos\alpha) + \left(\frac{r'}{r}\right)^2\left(\frac{3\cos^2\alpha - 1}{2}\right)\right. \\ &\qquad \left. + \left(\frac{r'}{r}\right)^3\left(\frac{5\cos^3\alpha - 3\cos\alpha}{2}\right) + \dots\right].\end{aligned}$$

In the last step, I have collected together like powers of (r'/r); surprisingly, their coefficients (the terms in parentheses) are Legendre polynomials! The remarkable result[16] is that

$$\frac{1}{\mathscr{r}} = \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\alpha). \tag{3.94}$$

Substituting this back into Eq. 3.91, and noting that r is a constant, as far as the integration is concerned, I conclude that

$$\boxed{V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{n=0}^{\infty}\frac{1}{r^{(n+1)}}\int (r')^n P_n(\cos\alpha)\rho(\mathbf{r}')\,d\tau'}, \tag{3.95}$$

or, more explicitly,

$$\begin{aligned}V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left[\frac{1}{r}\int \rho(\mathbf{r}')\,d\tau' + \frac{1}{r^2}\int r'\cos\alpha\,\rho(\mathbf{r}')\,d\tau'\right. \\ \left. + \frac{1}{r^3}\int (r')^2\left(\frac{3}{2}\cos^2\alpha - \frac{1}{2}\right)\rho(\mathbf{r}')\,d\tau' + \dots\right].\end{aligned} \tag{3.96}$$

[16]This suggests a second way of defining the Legendre polynomials (the first being Rodrigues' formula); $1/\mathscr{r}$ is called the **generating function** for Legendre polynomials.

This is the desired result—the **multipole expansion** of V in powers of $1/r$. The first term ($n = 0$) is the monopole contribution (it goes like $1/r$); the second ($n = 1$) is the dipole (it goes like $1/r^2$); the third is quadrupole; the fourth octopole; and so on. Remember that α is the angle between $\mathbf{r}$ and $\mathbf{r}'$, so the integrals depend on the direction to the field point. If you are interested in the potential along the z' axis (or—putting it the other way around—if you orient your $\mathbf{r}'$ coordinates so the z' axis lies along $\mathbf{r}$), then α is the usual polar angle θ'.

As it stands, Eq. 3.95 is *exact,* but it is *useful* primarily as an approximation scheme: the lowest nonzero term in the expansion provides the approximate potential at large r, and the successive terms tell us how to improve the approximation if greater precision is required.

Problem 3.27 A sphere of radius R, centered at the origin, carries charge density

$$\rho(r, \theta) = k\frac{R}{r^2}(R - 2r)\sin\theta,$$

where k is a constant, and r, θ are the usual spherical coordinates. Find the approximate potential for points on the z axis, far from the sphere.

Problem 3.28 A circular ring in the xy plane (radius R, centered at the origin) carries a uniform line charge λ. Find the first three terms ($n = 0, 1, 2$) in the multipole expansion for $V(r, \theta)$.

3.4.2 ■ The Monopole and Dipole Terms

Ordinarily, the multipole expansion is dominated (at large r) by the monopole term:

$$V_{\text{mon}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}, \tag{3.97}$$

where $Q = \int \rho\, d\tau$ is the total charge of the configuration. This is just what we expect for the approximate potential at large distances from the charge. For a *point* charge *at the origin,* V_{mon} is the *exact* potential, not merely a first approximation at large r; in this case, all the higher multipoles vanish.

If the total charge is zero, the dominant term in the potential will be the dipole (unless, of course, it *also* vanishes):

$$V_{\text{dip}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^2}\int r' \cos\alpha\, \rho(\mathbf{r}')\, d\tau'.$$

Since α is the angle between $\mathbf{r}'$ and $\mathbf{r}$ (Fig. 3.28),

$$r' \cos\alpha = \hat{\mathbf{r}} \cdot \mathbf{r}',$$

and the dipole potential can be written more succinctly:

$$V_{\text{dip}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^2}\hat{\mathbf{r}} \cdot \int \mathbf{r}'\rho(\mathbf{r}')\, d\tau'.$$

This integral (which does not depend on $\mathbf{r}$) is called the **dipole moment** of the distribution:

$$\boxed{\mathbf{p} \equiv \int \mathbf{r}' \rho(\mathbf{r}')\, d\tau'}, \tag{3.98}$$

and the dipole contribution to the potential simplifies to

$$\boxed{V_{\text{dip}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2}}. \tag{3.99}$$

The dipole moment is determined by the geometry (size, shape, and density) of the charge distribution. Equation 3.98 translates in the usual way (Sect. 2.1.4) for point, line, and surface charges. Thus, the dipole moment of a collection of *point* charges is

$$\mathbf{p} = \sum_{i=1}^{n} q_i \mathbf{r}'_i. \tag{3.100}$$

For a **physical dipole** (equal and opposite charges, $\pm q$),

$$\mathbf{p} = q\mathbf{r}'_+ - q\mathbf{r}'_- = q(\mathbf{r}'_+ - \mathbf{r}'_-) = q\mathbf{d}, \tag{3.101}$$

where $\mathbf{d}$ is the vector from the negative charge to the positive one (Fig. 3.29).

Is this consistent with what we got in Ex. 3.10? Yes: If you put Eq. 3.101 into Eq. 3.99, you recover Eq. 3.90. Notice, however, that this is only the *approximate* potential of the physical dipole—evidently there are higher multipole contributions. Of course, as you go farther and farther away, V_{dip} becomes a better and better approximation, since the higher terms die off more rapidly with increasing r. By the same token, at a fixed r the dipole approximation improves as you shrink the separation d. To construct a **perfect** (point) **dipole** whose potential is given *exactly* by Eq. 3.99, you'd have to let d approach zero. Unfortunately, you then lose the dipole term *too,* unless you simultaneously arrange for q to go to infinity! A *physical* dipole becomes a *pure* dipole, then, in the rather artificial limit $d \to 0$, $q \to \infty$, with the product $qd = p$ held fixed. When someone uses the word "dipole," you can't always tell whether they mean a *physical* dipole (with

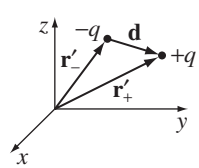


FIGURE 3.29

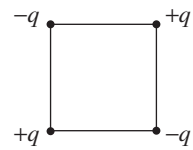


FIGURE 3.30

finite separation between the charges) or an *ideal* (point) dipole. If in doubt, assume that d is small enough (compared to r) that you can safely apply Eq. 3.99.

Dipole moments are *vectors,* and they add accordingly: if you have two dipoles, $\mathbf{p}_1$ and $\mathbf{p}_2$, the total dipole moment is $\mathbf{p}_1 + \mathbf{p}_2$. For instance, with four charges at the corners of a square, as shown in Fig. 3.30, the net dipole moment is zero. You can see this by combining the charges in pairs (vertically, $\downarrow + \uparrow = 0$, or horizontally, $\rightarrow + \leftarrow = 0$) or by adding up the four contributions individually, using Eq. 3.100. This is a *quadrupole,* as I indicated earlier, and its potential is dominated by the quadrupole term in the multipole expansion.

Problem 3.29 Four particles (one of charge q, one of charge $3q$, and two of charge $-2q$) are placed as shown in Fig. 3.31, each a distance a from the origin. Find a simple approximate formula for the potential, valid at points far from the origin. (Express your answer in spherical coordinates.)

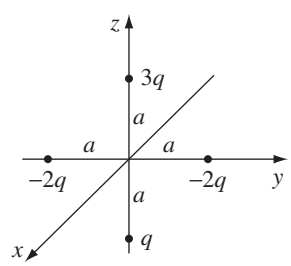


FIGURE 3.31

Problem 3.30 In Ex. 3.9, we derived the exact potential for a spherical shell of radius R, which carries a surface charge $\sigma = k\cos\theta$.

(a) Calculate the dipole moment of this charge distribution.

(b) Find the approximate potential, at points far from the sphere, and compare the exact answer (Eq. 3.87). What can you conclude about the higher multipoles?

Problem 3.31 For the dipole in Ex. 3.10, expand $1/\mathscr{r}_{\pm}$ to order $(d/r)^3$, and use this to determine the quadrupole and octopole terms in the potential.

3.4.3 ■ Origin of Coordinates in Multipole Expansions

I mentioned earlier that a point charge at the origin constitutes a "pure" monopole. If it is *not* at the origin, it's no longer a pure monopole. For instance, the charge in Fig. 3.32 has a dipole moment $\mathbf{p} = qd\hat{\mathbf{y}}$, and a corresponding dipole term in its potential. The monopole potential $(1/4\pi\epsilon_0)q/r$ is not quite correct for this configuration; rather, the exact potential is $(1/4\pi\epsilon_0)q/\imath$. The multipole expansion is, remember, a series in inverse powers of r (the distance to the *origin*), and when we expand $1/\imath$, we get *all* powers, not just the first.

So moving the origin (or, what amounts to the same thing, moving the *charge*) can radically alter a multipole expansion. The **monopole moment** Q does not change, since the total charge is obviously independent of the coordinate system. (In Fig. 3.32, the monopole term was unaffected when we moved q away from the origin—it's just that it was no longer the whole story: a dipole term—and for that matter all higher poles—appeared as well.) Ordinarily, the dipole moment *does* change when you shift the origin, but there is an important exception: *If the total charge is zero, then the dipole moment is independent of the choice of origin.* For suppose we displace the origin by an amount $\mathbf{a}$ (Fig. 3.33). The new dipole moment is then

$$\bar{\mathbf{p}} = \int \bar{\mathbf{r}}'\rho(\mathbf{r}')\,d\tau' = \int (\mathbf{r}' - \mathbf{a})\rho(\mathbf{r}')\,d\tau'$$
$$= \int \mathbf{r}'\rho(\mathbf{r}')\,d\tau' - \mathbf{a}\int \rho(\mathbf{r}')\,d\tau' = \mathbf{p} - Q\mathbf{a}.$$

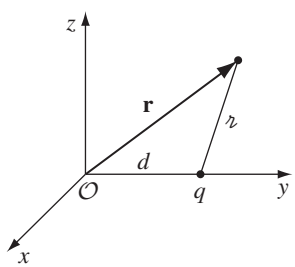


FIGURE 3.32

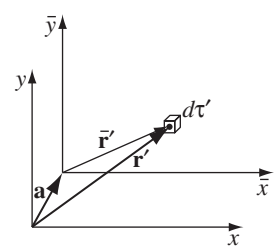
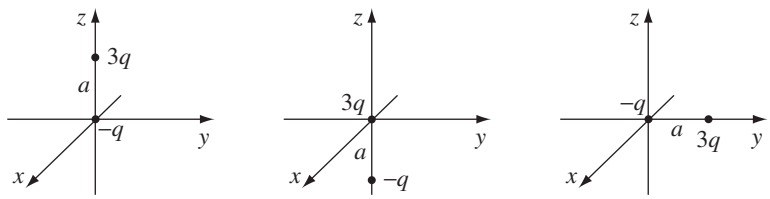


FIGURE 3.33

In particular, if $Q = 0$, then $\bar{\mathbf{p}} = \mathbf{p}$. So if someone asks for the dipole moment in Fig. 3.34(a), you can answer with confidence "$q\mathbf{d}$," but if you're asked for the dipole moment in Fig. 3.34(b), the appropriate response would be "With respect to *what origin*?"

FIGURE 3.34

Problem 3.32 Two point charges, $3q$ and $-q$, are separated by a distance a. For each of the arrangements in Fig. 3.35, find (i) the monopole moment, (ii) the dipole moment, and (iii) the approximate potential (in spherical coordinates) at large r (include both the monopole and dipole contributions).



(a) (b) (c)

FIGURE 3.35

3.4.4 ■ The Electric Field of a Dipole

So far we have worked only with *potentials.* Now I would like to calculate the electric *field* of a (perfect) dipole. If we choose coordinates so that **p** is at the origin and points in the z direction (Fig. 3.36), then the potential at r, θ is (Eq. 3.99):

$$V_{\text{dip}}(r, \theta) = \frac{\hat{\mathbf{r}} \cdot \mathbf{p}}{4\pi\epsilon_0 r^2} = \frac{p \cos\theta}{4\pi\epsilon_0 r^2}. \tag{3.102}$$

To get the field, we take the negative gradient of V:

$$E_r = -\frac{\partial V}{\partial r} = \frac{2p\cos\theta}{4\pi\epsilon_0 r^3},$$
$$E_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta} = \frac{p\sin\theta}{4\pi\epsilon_0 r^3},$$
$$E_\phi = -\frac{1}{r\sin\theta}\frac{\partial V}{\partial \phi} = 0.$$

Thus

$$\boxed{\mathbf{E}_{\text{dip}}(r, \theta) = \frac{p}{4\pi\epsilon_0 r^3}(2\cos\theta\,\hat{\mathbf{r}} + \sin\theta\,\hat{\boldsymbol{\theta}}).} \tag{3.103}$$

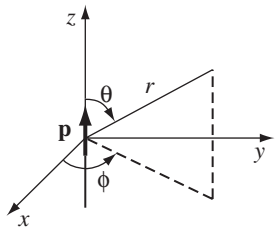
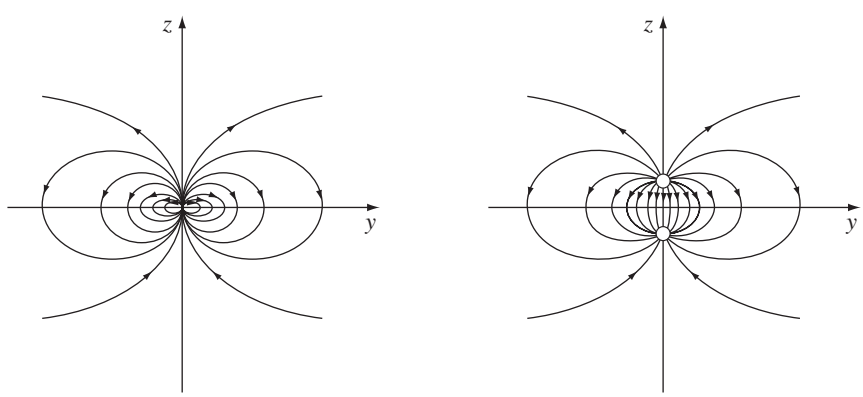


FIGURE 3.36

This formula makes explicit reference to a particular coordinate system (spherical) and assumes a particular orientation for $\mathbf{p}$ (along z). It can be recast in a coordinate-free form, analogous to the potential in Eq. 3.99—see Prob. 3.36.

Notice that the dipole field falls off as the inverse *cube* of r; the *monopole* field $(Q/4\pi\epsilon_0 r^2)\hat{\mathbf{r}}$ goes as the inverse *square,* of course. Quadrupole fields go like $1/r^4$, octopole like $1/r^5$, and so on. (This merely reflects the fact that monopole *potentials* fall off like $1/r$, dipole like $1/r^2$, quadrupole like $1/r^3$, and so on—the gradient introduces another factor of $1/r$.)

Figure 3.37(a) shows the field lines of a "pure" dipole (Eq. 3.103). For comparison, I have also sketched the field lines for a "physical" dipole, in Fig. 3.37(b). Notice how similar the two pictures become if you blot out the central region; up close, however, they are entirely different. Only for points $r \gg d$ does Eq. 3.103 represent a valid approximation to the field of a physical dipole. As I mentioned earlier, this régime can be reached either by going to large r or by squeezing the charges very close together.[17]



(a) Field of a "pure" dipole (b) Field of a "physical" dipole

FIGURE 3.37

[17]Even in the limit, there remains an infinitesimal region at the origin where the field of a physical dipole points in the "wrong" direction, as you can see by "walking" down the z axis in Fig. 3.35(b). If you want to explore this subtle and important point, work Prob. 3.48.

Problem 3.33 A "pure" dipole p is situated at the origin, pointing in the z direction.

(a) What is the force on a point charge q at $(a, 0, 0)$ (Cartesian coordinates)?

(b) What is the force on q at $(0, 0, a)$?

(c) How much work does it take to move q from $(a, 0, 0)$ to $(0, 0, a)$?

Problem 3.34 Three point charges are located as shown in Fig. 3.38, each a distance a from the origin. Find the approximate electric field at points far from the origin. Express your answer in spherical coordinates, and include the two lowest orders in the multipole expansion.

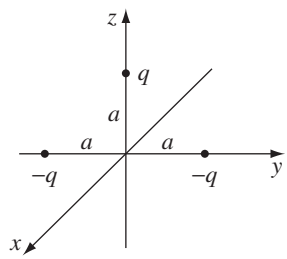


FIGURE 3.38

Problem 3.35 A solid sphere, radius R, is centered at the origin. The "northern" hemisphere carries a uniform charge density ρ_0, and the "southern" hemisphere a uniform charge density $-\rho_0$. Find the approximate field $\mathbf{E}(r, \theta)$ for points far from the sphere ($r \gg R$).

• **Problem 3.36** Show that the electric field of a (perfect) dipole (Eq. 3.103) can be written in the coordinate-free form

$$\boxed{\mathbf{E}_{\text{dip}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}\left[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}\right].} \tag{3.104}$$

More Problems on Chapter 3

Problem 3.37 In Section 3.1.4, I proved that the electrostatic potential at any point P in a charge-free region is equal to its average value over any spherical surface (radius R) centered at P. Here's an alternative argument that does not rely on Coulomb's law, only on Laplace's equation. We might as well set the origin at P. Let $V_{\text{ave}}(R)$ be the average; first show that

$$\frac{dV_{\text{ave}}}{dR} = \frac{1}{4\pi R^2}\oint \nabla V \cdot d\mathbf{a}$$

(note that the R^2 in $d\mathbf{a}$ cancels the $1/R^2$ out front, so the only dependence on R is in V itself). Now use the divergence theorem, and conclude that if V satisfies Laplace's equation, then $V_{\text{ave}}(R) = V_{\text{ave}}(0) = V(P)$, for all R.[18]

[18] I thank Ted Jacobson for suggesting this proof.

Problem 3.38 Here's an alternative derivation of Eq. 3.10 (the surface charge density induced on a grounded conducted plane by a point charge q a distance d above the plane). This approach[19] (which generalizes to many other problems) does not rely on the method of images. The total field is due in part to q, and in part to the induced surface charge. Write down the z components of these fields—in terms of q and the as-yet-unknown $\sigma(x, y)$—just below the surface. The sum must be zero, of course, because this is inside a conductor. Use that to determine σ.

Problem 3.39 Two infinite parallel grounded conducting planes are held a distance a apart. A point charge q is placed in the region between them, a distance x from one plate. Find the force on q.[20] Check that your answer is correct for the special cases $a \to \infty$ and $x = a/2$.

Problem 3.40 Two long straight wires, carrying opposite uniform line charges $\pm\lambda$, are situated on either side of a long conducting cylinder (Fig. 3.39). The cylinder (which carries no net charge) has radius R, and the wires are a distance a from the axis. Find the potential.

$$\left[\textit{Answer: } V(s,\phi) = \frac{\lambda}{4\pi\epsilon_0} \ln\left\{\frac{(s^2 + a^2 + 2sa\cos\phi)[(sa/R)^2 + R^2 - 2sa\cos\phi]}{(s^2 + a^2 - 2sa\cos\phi)[(sa/R)^2 + R^2 + 2sa\cos\phi]}\right\}\right]$$

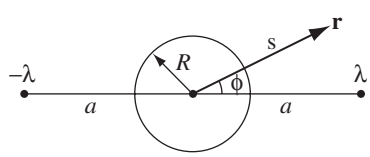


FIGURE 3.39

Problem 3.41 Buckminsterfullerine is a molecule of 60 carbon atoms arranged like the stitching on a soccer-ball. It may be approximated as a conducting spherical shell of radius $R = 3.5$ Å. A nearby electron would be *attracted*, according to Prob. 3.9, so it is not surprising that the ion C_{60}^- exists. (Imagine that the electron—on average—smears itself out uniformly over the surface.) But how about a *second* electron? At large distances it would be *repelled* by the ion, obviously, but at a certain distance r (from the center), the net force is zero, and closer than this it would be attracted. So an electron with enough energy to get in that close should bind.

(a) Find r, in Å. [You'll have to do it numerically.]

(b) How much energy (in electron volts) would it take to push an electron in (from infinity) to the point r?

[Incidentally, the C_{60}^{--} ion has been observed.][21]

[19] See J. L. R. Marrero, *Am. J. Phys.* **78**, 639 (2010).

[20] Obtaining the induced surface charge is not so easy. See B. G. Dick, *Am. J. Phys.* **41**, 1289 (1973), M. Zahn, *Am. J. Phys.* **44**, 1132 (1976), J. Pleines and S. Mahajan, *Am. J. Phys.* **45**, 868 (1977), and Prob. 3.51 below.

[21] Richard Mawhorter suggested this problem.

Problem 3.42 You can use the superposition principle to combine solutions obtained by separation of variables. For example, in Prob. 3.16 you found the potential inside a cubical box, if five faces are grounded and the sixth is at a constant potential V_0; by a six-fold superposition of the result, you could obtain the potential inside a cube with the faces maintained at specified constant voltages V_1, V_2, $\ldots V_6$. In this way, using Ex. 3.4 and Prob. 3.15, find the potential inside a rectangular pipe with two facing sides ($x = \pm b$) at potential V_0, a third ($y = a$) at V_1, and the last (at $y = 0$) grounded.

Problem 3.43 A conducting sphere of radius a, at potential V_0, is surrounded by a thin concentric spherical shell of radius b, over which someone has glued a surface charge

$$\sigma(\theta) = k\cos\theta,$$

where k is a constant and θ is the usual spherical coordinate.

(a) Find the potential in each region: (i) $r > b$, and (ii) $a < r < b$.

(b) Find the induced surface charge $\sigma_i(\theta)$ on the conductor.

(c) What is the total charge of this system? Check that your answer is consistent with the behavior of V at large r.

$$\left[\text{Answer: } V(r,\theta) = \begin{cases} aV_0/r + (b^3 - a^3)k\cos\theta/3r^2\epsilon_0, & r \geq b \\ aV_0/r + (r^3 - a^3)k\cos\theta/3r^2\epsilon_0, & r \leq b \end{cases}\right]$$

Problem 3.44 A charge $+Q$ is distributed uniformly along the z axis from $z = -a$ to $z = +a$. Show that the electric potential at a point $\mathbf{r}$ is given by

$$V(r,\theta) = \frac{Q}{4\pi\epsilon_0}\frac{1}{r}\left[1 + \frac{1}{3}\left(\frac{a}{r}\right)^2 P_2(\cos\theta) + \frac{1}{5}\left(\frac{a}{r}\right)^4 P_4(\cos\theta) + \ldots\right],$$

for $r > a$.

Problem 3.45 A long cylindrical shell of radius R carries a uniform surface charge σ_0 on the upper half and an opposite charge $-\sigma_0$ on the lower half (Fig. 3.40). Find the electric potential inside and outside the cylinder.

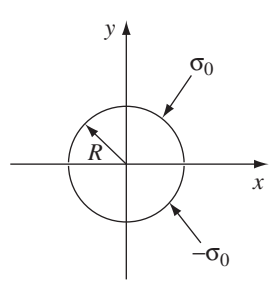


FIGURE 3.40

Problem 3.46 A thin insulating rod, running from $z = -a$ to $z = +a$, carries the indicated line charges. In each case, find the leading term in the multipole expansion of the potential: (a) $\lambda = k\cos(\pi z/2a)$, (b) $\lambda = k\sin(\pi z/a)$, (c) $\lambda = k\cos(\pi z/a)$, where k is a constant.

• **Problem 3.47** Show that the average field inside a sphere of radius R, due to all the charge within the sphere, is

$$\mathbf{E}_{\text{ave}} = -\frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}}{R^3}, \tag{3.105}$$

where $\mathbf{p}$ is the total dipole moment. There are several ways to prove this delightfully simple result. Here's one method:[22]

(a) Show that the average field due to a single charge q at point $\mathbf{r}$ inside the sphere is the same as the field at $\mathbf{r}$ due to a uniformly charged sphere with $\rho = -q/(\frac{4}{3}\pi R^3)$, namely

$$\frac{1}{4\pi\epsilon_0}\frac{1}{(\frac{4}{3}\pi R^3)}\int \frac{q}{\mathscr{r}^2}\,\hat{\boldsymbol{\mathscr{r}}}\,d\tau',$$

where $\boldsymbol{\mathscr{r}}$ is the vector from $\mathbf{r}$ to $d\tau'$.

(b) The latter can be found from Gauss's law (see Prob. 2.12). Express the answer in terms of the dipole moment of q.

(c) Use the superposition principle to generalize to an arbitrary charge distribution.

(d) While you're at it, show that the average field over the volume of a sphere, due to all the charges *outside*, is the same as the field they produce at the center.

Problem 3.48

(a) Using Eq. 3.103, calculate the average electric field of a dipole, over a spherical volume of radius R, centered at the origin. Do the angular integrals first. [*Note:* You *must* express $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ in terms of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ (see back cover) before integrating. If you don't understand why, reread the discussion in Sect. 1.4.1.] Compare your answer with the general theorem (Eq. 3.105). The discrepancy here is related to the fact that the field of a dipole blows up at $r = 0$. The angular integral is zero, but the radial integral is infinite, so we really don't know *what* to make of the answer. To resolve this dilemma, let's say that Eq. 3.103 applies *outside a tiny sphere of radius* ϵ—its contribution to E_{ave} is then *unambiguously* zero, and the whole answer has to come from the field *inside* the ϵ-sphere.

(b) What must the field *inside* the ϵ-sphere be, in order for the general theorem (Eq. 3.105) to hold? [*Hint:* since ϵ is arbitrarily small, we're talking about something that is infinite at $r = 0$ and whose integral over an infinitesimal volume is finite.] [*Answer:* $-(\mathbf{p}/3\epsilon_0)\delta^3(\mathbf{r})$]

Evidently, the *true* field of a dipole is

$$\mathbf{E}_{\text{dip}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}\left[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}\right] - \frac{1}{3\epsilon_0}\mathbf{p}\,\delta^3(\mathbf{r}). \tag{3.106}$$

[22] Another method exploits the result of Prob. 3.4. See B. Y.-K. Hu, *Eur. J. Phys.* **30**, L29 (2009).

You may wonder how we missed the delta-function term[23] when we calculated the field back in Sect. 3.4.4. The answer is that the differentiation leading to Eq. 3.103 is valid *except* at $r = 0$, but we should have known (from our experience in Sect. 1.5.1) that the point $r = 0$ would be problematic.[24]

Problem 3.49 In Ex. 3.9, we obtained the potential of a spherical shell with surface charge $\sigma(\theta) = k\cos\theta$. In Prob. 3.30, you found that the field is pure dipole outside; it's *uniform* inside (Eq. 3.86). Show that the limit $R \to 0$ reproduces the delta function term in Eq. 3.106.

Problem 3.50

(a) Suppose a charge distribution $\rho_1(\mathbf{r})$ produces a potential $V_1(\mathbf{r})$, and some other charge distribution $\rho_2(\mathbf{r})$ produces a potential $V_2(\mathbf{r})$. [The two situations may have nothing in common, for all I care—perhaps number 1 is a uniformly charged sphere and number 2 is a parallel-plate capacitor. Please understand that ρ_1 and ρ_2 are not present *at the same time;* we are talking about two *different problems,* one in which only ρ_1 is present, and another in which only ρ_2 is present.] Prove **Green's reciprocity theorem:**[25]

$$\int_{\text{all space}} \rho_1 V_2 \, d\tau = \int_{\text{all space}} \rho_2 V_1 \, d\tau.$$

[*Hint:* Evaluate $\int \mathbf{E}_1 \cdot \mathbf{E}_2 \, d\tau$ two ways, first writing $\mathbf{E}_1 = -\boldsymbol{\nabla} V_1$ and using integration by parts to transfer the derivative to $\mathbf{E}_2$, then writing $\mathbf{E}_2 = -\boldsymbol{\nabla} V_2$ and transferring the derivative to $\mathbf{E}_1$.]

(b) Suppose now that you have two separated conductors (Fig. 3.41). If you charge up conductor a by amount Q (leaving b uncharged), the resulting potential of b is, say, V_{ab}. On the other hand, if you put that same charge Q on conductor b (leaving a uncharged), the potential of a would be V_{ba}. Use Green's reciprocity theorem to show that $V_{ab} = V_{ba}$ (an astonishing result, since we assumed nothing about the shapes or placement of the conductors).

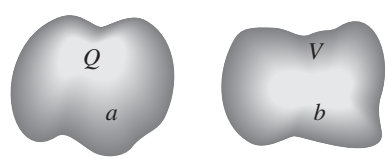


FIGURE 3.41

[23]There are other ways of getting the delta-function term in the field of a dipole—my own favorite is Prob. 3.49. Note that unless you are right on *top* of the dipole, Eq. 3.104 is perfectly adequate.
[24]See C. P. Frahm, *Am. J. Phys.* **51**, 826 (1983). For applications, see D. J. Griffiths, *Am. J. Phys.* **50,** 698 (1982). There are other (perhaps preferable) ways of expressing the **contact** (delta-function) **term** in Eq. 3.106; see A. Gsponer, *Eur. J. Phys.* **28**, 267 (2007), J. Franklin, *Am. J. Phys.* **78**, 1225 (2010), and V. Hnizdo, *Eur. J. Phys.* **32**, 287 (2011).
[25]For interesting commentary, see B. Y.-K. Hu, *Am. J. Phys.* **69**, 1280 (2001).

Problem 3.51 Use Green's reciprocity theorem (Prob. 3.50) to solve the following two problems. [*Hint:* for distribution 1, use the actual situation; for distribution 2, remove q, and set one of the conductors at potential V_0.]

(a) Both plates of a parallel-plate capacitor are grounded, and a point charge q is placed between them at a distance x from plate 1. The plate separation is d. Find the induced charge on each plate. [*Answer:* $Q_1 = q(x/d - 1)$; $Q_2 = -qx/d$]

(b) Two concentric spherical conducting shells (radii a and b) are grounded, and a point charge q is placed between them (at radius r). Find the induced charge on each sphere.

Problem 3.52

(a) Show that the quadrupole term in the multipole expansion can be written

$$V_{\text{quad}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}\sum_{i,j=1}^{3}\hat{r}_i\hat{r}_j Q_{ij}$$

(in the notation of Eq. 1.31), where

$$Q_{ij} \equiv \frac{1}{2}\int [3r'_i r'_j - (r')^2\delta_{ij}]\rho(\mathbf{r}')\,d\tau'.$$

Here

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

is the **Kronecker delta**, and Q_{ij} is the **quadrupole moment** of the charge distribution. Notice the hierarchy:

$$V_{\text{mon}} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}; \quad V_{\text{dip}} = \frac{1}{4\pi\epsilon_0}\frac{\sum \hat{r}_i p_i}{r^2}; \quad V_{\text{quad}} = \frac{1}{4\pi\epsilon_0}\frac{\sum \hat{r}_i\hat{r}_j Q_{ij}}{r^3}; \ \ldots$$

The monopole moment (Q) is a scalar, the dipole moment ($\mathbf{p}$) is a vector, the quadrupole moment (Q_{ij}) is a second-rank tensor, and so on.

(b) Find all nine components of Q_{ij} for the configuration in Fig. 3.30 (assume the square has side a and lies in the xy plane, centered at the origin).

(c) Show that the quadrupole moment is independent of origin if the monopole and dipole moments both vanish. (This works all the way up the hierarchy—the lowest nonzero multipole moment is always independent of origin.)

(d) How would you define the **octopole moment**? Express the octopole term in the multipole expansion in terms of the octopole moment.

Problem 3.53 In Ex. 3.8 we determined the electric field outside a spherical conductor (radius R) placed in a uniform external field $\mathbf{E}_0$. Solve the problem now using the method of images, and check that your answer agrees with Eq. 3.76. [*Hint:* Use Ex. 3.2, but put another charge, $-q$, diametrically opposite q. Let $a \to \infty$, with $(1/4\pi\epsilon_0)(2q/a^2) = -E_0$ held constant.]

! **Problem 3.54** For the infinite rectangular pipe in Ex. 3.4, suppose the potential on the bottom ($y = 0$) and the two sides ($x = \pm b$) is zero, but the potential on the top ($y = a$) is a nonzero constant V_0. Find the potential inside the pipe. [*Note:* This is a rotated version of Prob. 3.15(b), but set it up as in Ex. 3.4, using sinusoidal functions in y and hyperbolics in x. It is an unusual case in which $k = 0$ must be included. Begin by finding the general solution to Eq. 3.26 when $k = 0$.][26]

$\left[\textit{Answer: } V_0 \left(\frac{y}{a} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \frac{\cosh(n\pi x/a)}{\cosh(n\pi b/a)} \sin(n\pi y/a) \right)\right.$. Alternatively, using sinusoidal functions of x and hyperbolics in y, $\left. -\frac{2V_0}{b} \sum_{n=1}^{\infty} \frac{(-1)^n \sinh(\alpha_n y)}{\alpha_n \sinh(\alpha_n a)} \cos(\alpha_n x), \text{ where } \alpha_n \equiv (2n-1)\pi/2b \right]$

! **Problem 3.55**

(a) A long metal pipe of square cross-section (side a) is grounded on three sides, while the fourth (which is insulated from the rest) is maintained at constant potential V_0. Find the net charge per unit length on the side *opposite* to V_0. [*Hint:* Use your answer to Prob. 3.15 or Prob. 3.54.]

(b) A long metal pipe of circular cross-section (radius R) is divided (lengthwise) into four equal sections, three of them grounded and the fourth maintained at constant potential V_0. Find the net charge per unit length on the section opposite to V_0. [*Answer to both (a) and (b):* $\lambda = -(\epsilon_0 V_0/\pi) \ln 2$][27]

Problem 3.56 An ideal electric dipole is situated at the origin, and points in the z direction, as in Fig. 3.36. An electric charge is released from rest at a point in the xy plane. Show that it swings back and forth in a semi-circular arc, as though it were a pendulum supported at the origin.[28]

Problem 3.57 A stationary electric dipole $\mathbf{p} = p\,\hat{z}$ is situated at the origin. A positive point charge q (mass m) executes circular motion (radius s) at constant speed in the field of the dipole. Characterize the plane of the orbit. Find the speed, angular momentum and total energy of the charge.[29] $\left[\textit{Answer: } L = \sqrt{qpm/3\sqrt{3}\pi\epsilon_0}\right]$

Problem 3.58 Find the charge density $\sigma(\theta)$ on the surface of a sphere (radius R) that produces the same electric field, for points exterior to the sphere, as a charge q at the point $a < R$ on the z axis. $\left[\textit{Answer: } \frac{q}{4\pi R}(R^2 - a^2)(R^2 + a^2 - 2Ra\cos\theta)^{-3/2}\right]$

[26] For further discussion, see S. Hassani, *Am. J. Phys.* **59**, 470 (1991).

[27] These are special cases of the **Thompson-Lampard theorem**; see J. D. Jackson, *Am. J. Phys.* **67**, 107 (1999).

[28] This charming result is due to R. S. Jones, *Am. J. Phys.* **63**, 1042 (1995).

[29] G. P. Sastry, V. Srinivas, and A. V. Madhav, *Eur. J. Phys.* **17**, 275 (1996).

CHAPTER

4 Electric Fields in Matter

4.1 ■ POLARIZATION

4.1.1 ■ Dielectrics

In this chapter, we shall study electric fields in matter. Matter, of course, comes in many varieties—solids, liquids, gases, metals, woods, glasses—and these substances do not all respond in the same way to electrostatic fields. Nevertheless, *most* everyday objects belong (at least, in good approximation) to one of two large classes: **conductors** and **insulators** (or **dielectrics**). We have already talked about conductors; these are substances that contain an "unlimited" supply of charges that are free to move about through the material. In practice, what this ordinarily means is that many of the electrons (one or two per atom, in a typical metal) are not associated with any particular nucleus, but roam around at will. In dielectrics, by contrast, *all charges are attached to specific atoms or molecules*—they're on a tight leash, and all they can do is move a bit *within* the atom or molecule. Such microscopic displacements are not as dramatic as the wholesale rearrangement of charge in a conductor, but their cumulative effects account for the characteristic behavior of dielectric materials. There are actually *two* principal mechanisms by which electric fields can distort the charge distribution of a dielectric atom or molecule: *stretching* and *rotating*. In the next two sections I'll discuss these processes.

4.1.2 ■ Induced Dipoles

What happens to a neutral atom when it is placed in an electric field **E**? Your first guess might well be: "Absolutely nothing—since the atom is not charged, the field has no effect on it." But that is incorrect. Although the atom as a whole is electrically neutral, there *is* a positively charged core (the nucleus) and a negatively charged electron cloud surrounding it. These two regions of charge within the atom are influenced by the field: the nucleus is pushed in the direction of the field, and the electrons the opposite way. In principle, if the field is large enough, it can pull the atom apart completely, "ionizing" it (the substance then becomes a conductor). With less extreme fields, however, an equilibrium is soon established, for if the center of the electron cloud does not coincide with the nucleus, these positive and negative charges attract one another, and that holds the atom together. The two opposing forces—**E** pulling the electrons and nucleus apart, their mutual attraction drawing them back together—reach a balance, leaving the

H	He	Li	Be	C	Ne	Na	Ar	K	Cs
0.667	0.205	24.3	5.60	1.67	0.396	24.1	1.64	43.4	59.4

TABLE 4.1 Atomic Polarizabilities ($\alpha/4\pi\epsilon_0$, in units of 10^{-30} m^3). *Data from: Handbook of Chemistry and Physics,* 91st ed. (Boca Raton: CRC Press, 2010).

atom **polarized,** with plus charge shifted slightly one way, and minus the other. The atom now has a tiny dipole moment **p**, which points in the *same direction as* **E**. Typically, this induced dipole moment is approximately proportional to the field (as long as the latter is not too strong):

$$\mathbf{p} = \alpha \mathbf{E}. \tag{4.1}$$

The constant of proportionality α is called **atomic polarizability.** Its value depends on the detailed structure of the atom in question. Table 4.1 lists some experimentally determined atomic polarizabilities.

Example 4.1. A primitive model for an atom consists of a point nucleus $(+q)$ surrounded by a uniformly charged spherical cloud $(-q)$ of radius a (Fig. 4.1). Calculate the atomic polarizability of such an atom.

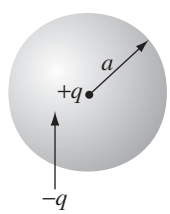


FIGURE 4.1

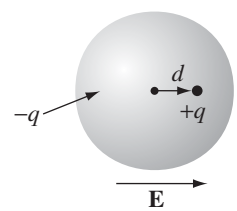


FIGURE 4.2

Solution

In the presence of an external field **E**, the nucleus will be shifted slightly to the right and the electron cloud to the left, as shown in Fig. 4.2. (Because the actual displacements involved are extremely small, as you'll see in Prob. 4.1, it is reasonable to assume that the electron cloud retains its spherical shape.) Say that equilibrium occurs when the nucleus is displaced a distance d from the center of the sphere. At that point, the external field pushing the nucleus to the right exactly balances the internal field pulling it to the left: $E = E_e$, where E_e is the field produced by the electron cloud. Now the field at a distance d from the center of a uniformly charged sphere is

$$E_e = \frac{1}{4\pi\epsilon_0}\frac{qd}{a^3}$$

(Prob. 2.12). At equilibrium, then,

$$E = \frac{1}{4\pi\epsilon_0}\frac{qd}{a^3}, \quad \text{or} \quad p = qd = (4\pi\epsilon_0 a^3)E.$$

The atomic polarizability is therefore

$$\alpha = 4\pi\epsilon_0 a^3 = 3\epsilon_0 v, \tag{4.2}$$

where v is the volume of the atom. Although this atomic model is extremely crude, the result (Eq. 4.2) is not too bad—it's accurate to within a factor of four or so for many simple atoms.

For molecules the situation is not quite so simple, because frequently they polarize more readily in some directions than in others. Carbon dioxide (Fig. 4.3), for instance, has a polarizability of 4.5×10^{-40} C^2·m/N when you apply the field along the axis of the molecule, but only 2×10^{-40} for fields perpendicular to this direction. When the field is at some *angle* to the axis, you must resolve it into parallel and perpendicular components, and multiply each by the pertinent polarizability:

$$\mathbf{p} = \alpha_\perp \mathbf{E}_\perp + \alpha_\parallel \mathbf{E}_\parallel.$$

In this case, the induced dipole moment may not even be in the same *direction* as **E**. And CO_2 is relatively simple, as molecules go, since the atoms at least arrange themselves in a straight line; for a completely asymmetrical molecule, Eq. 4.1 is replaced by the most general linear relation between **E** and **p**:

$$\left.\begin{aligned} p_x &= \alpha_{xx}E_x + \alpha_{xy}E_y + \alpha_{xz}E_z \\ p_y &= \alpha_{yx}E_x + \alpha_{yy}E_y + \alpha_{yz}E_z \\ p_z &= \alpha_{zx}E_x + \alpha_{zy}E_y + \alpha_{zz}E_z \end{aligned}\right\} \tag{4.3}$$

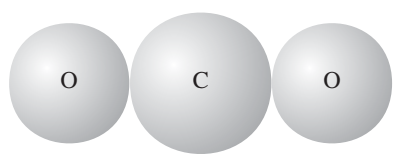


FIGURE 4.3

The set of nine constants α_{ij} constitute the **polarizability tensor** for the molecule. Their values depend on the orientation of the axes you use, though it is always possible to choose "principal" axes such that all the off-diagonal terms (α_{xy}, α_{zx}, etc.) vanish, leaving just three nonzero polarizabilities: α_{xx}, α_{yy}, and α_{zz}.

Problem 4.1 A hydrogen atom (with the Bohr radius of half an angstrom) is situated between two metal plates 1 mm apart, which are connected to opposite terminals of a 500 V battery. What fraction of the atomic radius does the separation distance d amount to, roughly? Estimate the voltage you would need with this apparatus to ionize the atom. [Use the value of α in Table 4.1. *Moral:* The displacements we're talking about are *minute,* even on an atomic scale.]

Problem 4.2 According to quantum mechanics, the electron cloud for a hydrogen atom in the ground state has a charge density

$$\rho(r) = \frac{q}{\pi a^3} e^{-2r/a},$$

where q is the charge of the electron and a is the Bohr radius. Find the atomic polarizability of such an atom. [*Hint:* First calculate the electric field of the electron cloud, $E_e(r)$; then expand the exponential, assuming $r \ll a$.[1]

Problem 4.3 According to Eq. 4.1, the induced dipole moment of an atom is proportional to the external field. This is a "rule of thumb," not a fundamental law, and it is easy to concoct exceptions—in theory. Suppose, for example, the charge density of the electron cloud were proportional to the distance from the center, out to a radius R. To what power of E would p be proportional in that case? Find the condition on $\rho(r)$ such that Eq. 4.1 will hold in the weak-field limit.

Problem 4.4 A point charge q is situated a large distance r from a neutral atom of polarizability α. Find the force of attraction between them.

4.1.3 ■ Alignment of Polar Molecules

The neutral atom discussed in Sect. 4.1.2 had no dipole moment to start with—**p** was *induced* by the applied field. Some molecules have built-in, permanent dipole moments. In the water molecule, for example, the electrons tend to cluster around the oxygen atom (Fig. 4.4), and since the molecule is bent at 105°, this leaves a negative charge at the vertex and a net positive charge on the opposite side. (The dipole moment of water is unusually large: 6.1×10^{-30} C·m; in fact, this is what accounts for its effectiveness as a solvent.) What happens when such molecules (called **polar molecules**) are placed in an electric field?

If the field is uniform, the *force* on the positive end, $\mathbf{F}_+ = q\mathbf{E}$, exactly cancels the force on the negative end, $\mathbf{F}_- = -q\mathbf{E}$ (Fig. 4.5). However, there will be a *torque:*

$$\begin{aligned}\mathbf{N} &= (\mathbf{r}_+ \times \mathbf{F}_+) + (\mathbf{r}_- \times \mathbf{F}_-) \\ &= \left[(\mathbf{d}/2) \times (q\mathbf{E})\right] + \left[(-\mathbf{d}/2) \times (-q\mathbf{E})\right] = q\mathbf{d} \times \mathbf{E}.\end{aligned}$$

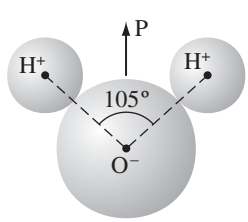


FIGURE 4.4

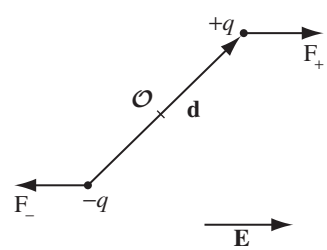


FIGURE 4.5

[1]For a more sophisticated approach, see W. A. Bowers, *Am. J. Phys.* **54,** 347 (1986).

Thus a dipole $\mathbf{p} = q\mathbf{d}$ in a uniform field $\mathbf{E}$ experiences a torque

$$\boxed{\mathbf{N} = \mathbf{p} \times \mathbf{E}.} \tag{4.4}$$

Notice that $\mathbf{N}$ is in such a direction as to line $\mathbf{p}$ up *parallel* to $\mathbf{E}$; a polar molecule that is free to rotate will swing around until it points in the direction of the applied field.

If the field is *non*uniform, so that $\mathbf{F}_+$ does not exactly balance $\mathbf{F}_-$, there will be a net *force* on the dipole, in addition to the torque. Of course, $\mathbf{E}$ must change rather abruptly for there to be significant variation in the space of one molecule, so this is not ordinarily a major consideration in discussing the behavior of dielectrics. Nevertheless, the formula for the force on a dipole in a nonuniform field is of some interest:

$$\mathbf{F} = \mathbf{F}_+ + \mathbf{F}_- = q(\mathbf{E}_+ - \mathbf{E}_-) = q(\Delta\mathbf{E}),$$

where $\Delta\mathbf{E}$ represents the difference between the field at the plus end and the field at the minus end. Assuming the dipole is very short, we may use Eq. 1.35 to approximate the small change in E_x:

$$\Delta E_x \equiv (\nabla E_x) \cdot \mathbf{d},$$

with corresponding formulas for E_y and E_z. More compactly,

$$\Delta\mathbf{E} = (\mathbf{d} \cdot \nabla)\mathbf{E},$$

and therefore[2]

$$\boxed{\mathbf{F} = (\mathbf{p} \cdot \nabla)\mathbf{E}.} \tag{4.5}$$

For a "perfect" dipole of infinitesimal length, Eq. 4.4 gives the torque *about the center of the dipole* even in a *non*uniform field; about any *other* point $\mathbf{N} = (\mathbf{p} \times \mathbf{E}) + (\mathbf{r} \times \mathbf{F})$.

Problem 4.5 In Fig. 4.6, $\mathbf{p}_1$ and $\mathbf{p}_2$ are (perfect) dipoles a distance r apart. What is the torque on $\mathbf{p}_1$ due to $\mathbf{p}_2$? What is the torque on $\mathbf{p}_2$ due to $\mathbf{p}_1$? [In each case, I want the torque on the dipole *about its own center*. If it bothers you that the answers are not equal and opposite, see Prob. 4.29.]

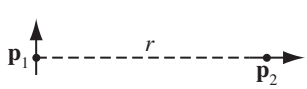


FIGURE 4.6

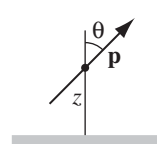


FIGURE 4.7

[2]In the present context, Eq. 4.5 could be written more conveniently as $\mathbf{F} = \nabla(\mathbf{p} \cdot \mathbf{E})$. However, it is safer to stick with $(\mathbf{p} \cdot \nabla)\mathbf{E}$, because we will be applying the formula to materials in which the dipole moment (per unit volume) is itself a function of position and this second expression would imply (incorrectly) that $\mathbf{p}$ *too* is to be differentiated.

Problem 4.6 A (perfect) dipole **p** is situated a distance z above an infinite grounded conducting plane (Fig. 4.7). The dipole makes an angle θ with the perpendicular to the plane. Find the torque on **p**. If the dipole is free to rotate, in what orientation will it come to rest?

Problem 4.7 Show that the energy of an ideal dipole **p** in an electric field **E** is given by

$$\boxed{U = -\mathbf{p} \cdot \mathbf{E}.} \tag{4.6}$$

Problem 4.8 Show that the interaction energy of two dipoles separated by a displacement **r** is

$$U = \frac{1}{4\pi\epsilon_0}\frac{1}{r^3}[\mathbf{p}_1 \cdot \mathbf{p}_2 - 3(\mathbf{p}_1 \cdot \hat{\mathbf{r}})(\mathbf{p}_2 \cdot \hat{\mathbf{r}})]. \tag{4.7}$$

[*Hint:* Use Prob. 4.7 and Eq. 3.104.]

Problem 4.9 A dipole **p** is a distance r from a point charge q, and oriented so that **p** makes an angle θ with the vector **r** from q to **p**.

(a) What is the force on **p**?

(b) What is the force on q?

4.1.4 ■ Polarization

In the previous two sections, we have considered the effect of an external electric field on an individual atom or molecule. We are now in a position to answer (qualitatively) the original question: What happens to a piece of dielectric material when it is placed in an electric field? If the substance consists of neutral atoms (or nonpolar molecules), the field will induce in each a tiny dipole moment, pointing in the same direction as the field.[3] If the material is made up of polar molecules, each permanent dipole will experience a torque, tending to line it up along the field direction. (Random thermal motions compete with this process, so the alignment is never complete, especially at higher temperatures, and disappears almost at once when the field is removed.)

Notice that these two mechanisms produce the same basic result: *a lot of little dipoles pointing along the direction of the field*—the material becomes **polarized**. A convenient measure of this effect is

$$\mathbf{P} \equiv \textit{ dipole moment per unit volume,}$$

which is called the **polarization**. From now on we shall not worry much about how the polarization *got* there. Actually, the two mechanisms I described are not as clear-cut as I tried to pretend. Even in polar molecules there will be

[3]In asymmetric molecules, the induced dipole moment may not be parallel to the field, but if the molecules are randomly oriented, the perpendicular contributions will *average* to zero. Within a single crystal, the orientations are certainly *not* random, and we would have to treat this case separately.

some polarization by displacement (though generally it is a lot easier to rotate a molecule than to stretch it, so the second mechanism dominates). It's even possible in some materials to "freeze in" polarization, so that it persists after the field is removed. But let's forget for a moment about the *cause* of the polarization, and let's study the field that a chunk of polarized material *itself* produces. Then, in Sect. 4.3, we'll put it all together: the original field, which was *responsible* for **P**, plus the new field, which is *due* to **P**.

4.2 ■ THE FIELD OF A POLARIZED OBJECT

4.2.1 ■ Bound Charges

Suppose we have a piece of polarized material—that is, an object containing a lot of microscopic dipoles lined up. The dipole moment per unit volume **P** is given. *Question:* What is the field produced by this object (not the field that may have *caused* the polarization, but the field the polarization *itself* causes)? Well, we know what the field of an individual dipole looks like, so why not chop the material up into infinitesimal dipoles and integrate to get the total? As usual, it's easier to work with the potential. For a single dipole **p** (Eq. 3.99),

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\boldsymbol{\imath}}}{\imath^2}, \tag{4.8}$$

where $\boldsymbol{\imath}$ is the vector from the dipole to the point at which we are evaluating the potential (Fig. 4.8). In the present context, we have a dipole moment $\mathbf{p} = \mathbf{P}\,d\tau'$ in each volume element $d\tau'$, so the total potential is

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int_{\mathcal{V}} \frac{\mathbf{P}(\mathbf{r}')\cdot\hat{\boldsymbol{\imath}}}{\imath^2}\,d\tau'. \tag{4.9}$$

That *does* it, in principle. But a little sleight-of-hand casts this integral into a much more illuminating form. Observing that

$$\boldsymbol{\nabla}'\left(\frac{1}{\imath}\right) = \frac{\hat{\boldsymbol{\imath}}}{\imath^2},$$

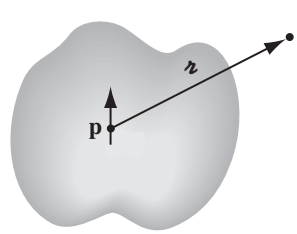


FIGURE 4.8

where (unlike Prob. 1.13) the differentiation is with respect to the *source* coordinates ($\mathbf{r}'$), we have

$$V = \frac{1}{4\pi\epsilon_0} \int_{\mathcal{V}} \mathbf{P} \cdot \boldsymbol{\nabla}' \left(\frac{1}{\imath}\right) d\tau'.$$

Integrating by parts, using product rule number 5 (in the front cover), gives

$$V = \frac{1}{4\pi\epsilon_0} \left[\int_{\mathcal{V}} \boldsymbol{\nabla}' \cdot \left(\frac{\mathbf{P}}{\imath}\right) d\tau' - \int_{\mathcal{V}} \frac{1}{\imath} (\boldsymbol{\nabla}' \cdot \mathbf{P})\, d\tau' \right],$$

or, invoking the divergence theorem,

$$V = \frac{1}{4\pi\epsilon_0} \oint_{\mathcal{S}} \frac{1}{\imath} \mathbf{P} \cdot d\mathbf{a}' - \frac{1}{4\pi\epsilon_0} \int_{\mathcal{V}} \frac{1}{\imath} (\boldsymbol{\nabla}' \cdot \mathbf{P})\, d\tau'. \tag{4.10}$$

The first term looks like the potential of a surface charge

$$\boxed{\sigma_b \equiv \mathbf{P} \cdot \hat{\mathbf{n}}} \tag{4.11}$$

(where $\hat{\mathbf{n}}$ is the normal unit vector), while the second term looks like the potential of a volume charge

$$\boxed{\rho_b \equiv -\boldsymbol{\nabla} \cdot \mathbf{P}.} \tag{4.12}$$

With these definitions, Eq. 4.10 becomes

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \oint_{\mathcal{S}} \frac{\sigma_b}{\imath}\, da' + \frac{1}{4\pi\epsilon_0} \int_{\mathcal{V}} \frac{\rho_b}{\imath}\, d\tau'. \tag{4.13}$$

What this means is that the potential (and hence also the field) of a polarized object is the same as that produced by a volume charge density $\rho_b = -\boldsymbol{\nabla} \cdot \mathbf{P}$ plus a surface charge density $\sigma_b = \mathbf{P} \cdot \hat{\mathbf{n}}$. Instead of integrating the contributions of all the infinitesimal dipoles, as in Eq. 4.9, we could first find those **bound charges**, and then calculate the fields *they* produce, in the same way we calculate the field of any other volume and surface charges (for example, using Gauss's law).

Example 4.2. Find the electric field produced by a uniformly polarized sphere of radius R.

Solution

We may as well choose the z axis to coincide with the direction of polarization (Fig. 4.9). The volume bound charge density ρ_b is zero, since $\mathbf{P}$ is uniform, but

$$\sigma_b = \mathbf{P} \cdot \hat{\mathbf{n}} = P \cos\theta,$$

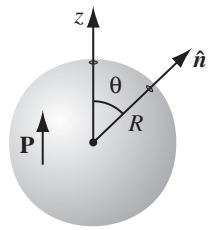


FIGURE 4.9

where θ is the usual spherical coordinate. What we want, then, is the field produced by a charge density $P\cos\theta$ plastered over the surface of a sphere. But we already computed the potential of such a configuration, in Ex. 3.9:

$$V(r,\theta) = \begin{cases} \dfrac{P}{3\epsilon_0} r\cos\theta, & \text{for} \quad r \le R, \\ \dfrac{P}{3\epsilon_0}\dfrac{R^3}{r^2}\cos\theta, & \text{for} \quad r \ge R. \end{cases}$$

Since $r\cos\theta = z$, the *field* inside the sphere is *uniform*:

$$\mathbf{E} = -\boldsymbol{\nabla} V = -\frac{P}{3\epsilon_0}\hat{\mathbf{z}} = -\frac{1}{3\epsilon_0}\mathbf{P}, \quad \text{for} \quad r < R. \tag{4.14}$$

This remarkable result will be very useful in what follows. Outside the sphere the potential is identical to that of a perfect dipole at the origin,

$$V = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2}, \quad \text{for} \quad r \ge R, \tag{4.15}$$

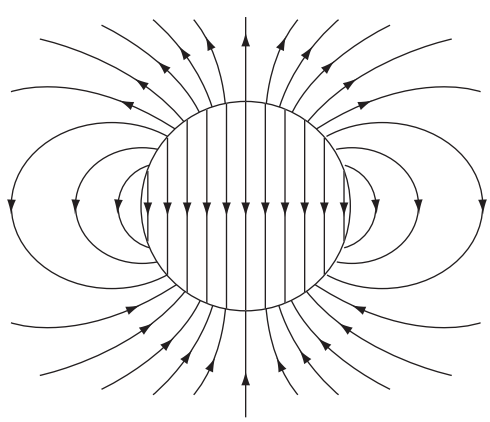


FIGURE 4.10

whose dipole moment is, not surprisingly, equal to the total dipole moment of the sphere:

$$\mathbf{p} = \frac{4}{3}\pi R^3 \mathbf{P}. \tag{4.16}$$

The field of the uniformly polarized sphere is shown in Fig. 4.10.

Problem 4.10 A sphere of radius R carries a polarization

$$\mathbf{P}(\mathbf{r}) = k\mathbf{r},$$

where k is a constant and $\mathbf{r}$ is the vector from the center.

(a) Calculate the bound charges σ_b and ρ_b.

(b) Find the field inside and outside the sphere.

Problem 4.11 A short cylinder, of radius a and length L, carries a "frozen-in" uniform polarization $\mathbf{P}$, parallel to its axis. Find the bound charge, and sketch the electric field (i) for $L \gg a$, (ii) for $L \ll a$, and (iii) for $L \approx a$. [This is known as a **bar electret**; it is the electrical analog to a bar magnet. In practice, only very special materials—barium titanate is the most "familiar" example—will hold a permanent electric polarization. That's why you can't buy electrets at the toy store.]

Problem 4.12 Calculate the potential of a uniformly polarized sphere (Ex. 4.2) directly from Eq. 4.9.

4.2.2 ■ Physical Interpretation of Bound Charges

In the last section we found that the field of a polarized object is identical to the field that would be produced by a certain distribution of "bound charges," σ_b and ρ_b. But this conclusion emerged in the course of abstract manipulations on the integral in Eq. 4.9, and left us with no clue as to the physical meaning of these bound charges. Indeed, some authors give you the impression that bound charges are in some sense "fictitious"—mere bookkeeping devices used to facilitate the calculation of fields. Nothing could be further from the truth: ρ_b and σ_b represent *perfectly genuine accumulations of charge.* In this section I'll explain how polarization leads to these charge distributions.

The basic idea is very simple: Suppose we have a long string of dipoles, as shown in Fig. 4.11. Along the line, the head of one effectively cancels the tail of its neighbor, but at the ends there are two charges left over: plus at the right end and minus at the left. It is as if we had peeled off an electron at one end and carried it all the way down to the other end, though in fact no single electron made the whole trip—a lot of tiny displacements add up to one large one. We call the net charge at the ends a *bound* charge to remind ourselves that it cannot be removed;

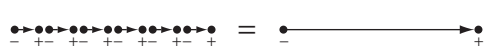


FIGURE 4.11

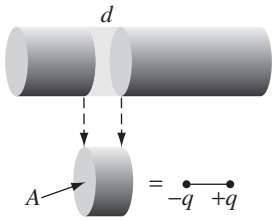


FIGURE 4.12

FIGURE 4.13

in a dielectric every electron is attached to a specific atom or molecule. But apart from that, bound charge is no different from any other kind.

To calculate the actual *amount* of bound charge resulting from a given polarization, examine a "tube" of dielectric parallel to **P**. The dipole moment of the tiny chunk shown in Fig. 4.12 is $P(Ad)$, where A is the cross-sectional area of the tube and d is the length of the chunk. In terms of the charge (q) at the end, this same dipole moment can be written qd. The bound charge that piles up at the right end of the tube is therefore

$$q = PA.$$

If the ends have been sliced off perpendicularly, the surface charge density is

$$\sigma_b = \frac{q}{A} = P.$$

For an oblique cut (Fig. 4.13), the *charge* is still the same, but $A = A_{\text{end}} \cos\theta$, so

$$\sigma_b = \frac{q}{A_{\text{end}}} = P\cos\theta = \mathbf{P}\cdot\hat{\mathbf{n}}.$$

The effect of the polarization, then, is to paint a bound charge $\sigma_b = \mathbf{P}\cdot\hat{\mathbf{n}}$ over the surface of the material. This is exactly what we found by more rigorous means in Sect. 4.2.1. But now we know where the bound charge *comes* from.

If the polarization is nonuniform, we get accumulations of bound charge *within* the material, as well as on the surface. A glance at Fig. 4.14 suggests that a diverging **P** results in a pileup of negative charge. Indeed, the net bound charge $\int \rho_b\, d\tau$

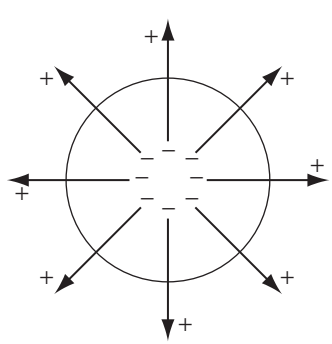


FIGURE 4.14

in a given volume is equal and opposite to the amount that has been pushed out through the surface. The latter (by the same reasoning we used before) is $\mathbf{P} \cdot \hat{\mathbf{n}}$ per unit area, so

$$\int_{\mathcal{V}} \rho_b \, d\tau = -\oint_{\mathcal{S}} \mathbf{P} \cdot d\mathbf{a} = -\int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \mathbf{P}) \, d\tau.$$

Since this is true for *any* volume, we have

$$\rho_b = -\boldsymbol{\nabla} \cdot \mathbf{P},$$

confirming, again, the more rigorous conclusion of Sect. 4.2.1.

Example 4.3. There is another way of analyzing the uniformly polarized sphere (Ex. 4.2), which nicely illustrates the idea of a bound charge. What we have, really, is *two* spheres of charge: a positive sphere and a negative sphere. Without polarization the two are superimposed and cancel completely. But when the material is uniformly polarized, all the plus charges move slightly *upward* (the z direction), and all the minus charges move slightly *downward* (Fig. 4.15). The two spheres no longer overlap perfectly: at the top there's a "cap" of leftover positive charge and at the bottom a cap of negative charge. This "leftover" charge is precisely the bound surface charge σ_b.

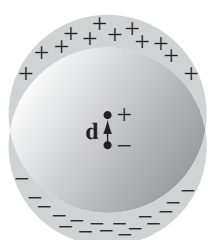


FIGURE 4.15

In Prob. 2.18, you calculated the field in the region of overlap between two uniformly charged spheres; the answer was

$$\mathbf{E} = -\frac{1}{4\pi\epsilon_0} \frac{q\mathbf{d}}{R^3},$$

where q is the total charge of the positive sphere, $\mathbf{d}$ is the vector from the negative center to the positive center, and R is the radius of the sphere. We can express this in terms of the polarization of the sphere, $\mathbf{p} = q\mathbf{d} = (\frac{4}{3}\pi R^3)\mathbf{P}$, as

$$\mathbf{E} = -\frac{1}{3\epsilon_0}\mathbf{P}.$$

Meanwhile, for points *outside,* it is as though all the charge on each sphere were concentrated at the respective center. We have, then, a dipole, with potential

$$V = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2}.$$

(Remember that **d** is some small fraction of an atomic radius; Fig. 4.15 is grossly exaggerated.) These answers agree, of course, with the results of Ex. 4.2.

Problem 4.13 A very long cylinder, of radius a, carries a uniform polarization **P** perpendicular to its axis. Find the electric field inside the cylinder. Show that the field *outside* the cylinder can be expressed in the form

$$\mathbf{E}(\mathbf{r}) = \frac{a^2}{2\epsilon_0 s^2}[2(\mathbf{P}\cdot\hat{\mathbf{s}})\hat{\mathbf{s}} - \mathbf{P}].$$

[*Careful:* I said "uniform," not "radial"!]

Problem 4.14 When you polarize a neutral dielectric, the charge moves a bit, but the *total* remains zero. This fact should be reflected in the bound charges σ_b and ρ_b. Prove from Eqs. 4.11 and 4.12 that the total bound charge vanishes.

4.2.3 ■ The Field Inside a Dielectric[4]

I have been sloppy about the distinction between "pure" dipoles and "physical" dipoles. In developing the theory of bound charges, I assumed we were working with the pure kind—indeed, I started with Eq. 4.8, the formula for the potential of a perfect dipole. And yet, an actual polarized dielectric consists of *physical* dipoles, albeit extremely tiny ones. What is more, I presumed to represent discrete molecular dipoles by a continuous density function **P**. How can I justify this method? *Outside* the dielectric there is no real problem: here we are far away from the molecules ($\mathcal{r}$ is many times greater than the separation distance between plus and minus charges), so the dipole potential dominates overwhelmingly and the detailed "graininess" of the source is blurred by distance. *Inside* the dielectric, however, we can hardly pretend to be far from all the dipoles, and the procedure I used in Sect. 4.2.1 is open to serious challenge.

In fact, when you stop to think about it, the electric field inside matter must be fantastically complicated, on the microscopic level. If you happen to be very near an electron, the field is gigantic, whereas a short distance away it may be small or may point in a totally different direction. Moreover, an instant later, as the atoms move about, the field will have altered entirely. This true **microscopic** field would be utterly impossible to calculate, nor would it be of much interest if you could. Just as, for macroscopic purposes, we regard water as a continuous fluid, ignoring its molecular structure, so also we can ignore the microscopic

[4]This section can be skipped without loss of continuity.

bumps and wrinkles in the electric field inside matter, and concentrate on the **macroscopic** field. This is defined as the *average* field over regions large enough to contain many thousands of atoms (so that the uninteresting microscopic fluctuations are smoothed over), and yet small enough to ensure that we do not wash out any significant large-scale variations in the field. (In practice, this means we must average over regions much smaller than the dimensions of the object itself.) Ordinarily, the macroscopic field is what people *mean* when they speak of "the" field inside matter.[5]

It remains to show that the macroscopic field is what we actually obtain when we use the methods of Sect. 4.2.1. The argument is subtle, so hang on. Suppose I want to calculate the macroscopic field at some point $\mathbf{r}$ within a dielectric (Fig. 4.16). I know I must average the true (microscopic) field over an appropriate volume, so let me draw a small sphere about $\mathbf{r}$, of radius, say, a thousand times the size of a molecule. The macroscopic field at $\mathbf{r}$, then, consists of two parts: the average field over the sphere due to all charges *outside*, plus the average due to all charges *inside:*

$$\mathbf{E} = \mathbf{E}_{\text{out}} + \mathbf{E}_{\text{in}}.$$

You proved in Prob. 3.47(d) that the average field (over a sphere), produced by charges *outside*, is equal to the field they produce at the center, so $\mathbf{E}_{\text{out}}$ is the field at $\mathbf{r}$ due to the dipoles exterior to the sphere. These are far enough away that we can safely use Eq. 4.9:

$$V_{\text{out}} = \frac{1}{4\pi\epsilon_0} \int_{\text{outside}} \frac{\mathbf{P}(\mathbf{r}') \cdot \hat{\boldsymbol{\imath}}}{\imath^2} \, d\tau'. \tag{4.17}$$

The dipoles *inside* the sphere are too close to treat in this fashion. But fortunately all we need is their *average* field, and that, according to Eq. 3.105, is

$$\mathbf{E}_{\text{in}} = -\frac{1}{4\pi\epsilon_0} \frac{\mathbf{p}}{R^3},$$

regardless of the details of the charge distribution within the sphere. The only relevant quantity is the total dipole moment, $\mathbf{p} = (\frac{4}{3}\pi R^3)\,\mathbf{P}$:

$$\mathbf{E}_{\text{in}} = -\frac{1}{3\epsilon_0}\mathbf{P}. \tag{4.18}$$

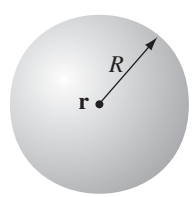


FIGURE 4.16

[5]In case the notion of macroscopic fields sounds suspicious to you, let me point out that you do *exactly* the same averaging whenever you speak of the *density* of a material.

Now, by assumption, the sphere is small enough that $\mathbf{P}$ does not vary significantly over its volume, so the term *left out* of the integral in Eq. 4.17 corresponds to the field at the center of a *uniformly* polarized sphere, to wit: $-(1/3\epsilon_0)\mathbf{P}$ (Eq. 4.14). But this is precisely what $\mathbf{E}_{\text{in}}$ (Eq. 4.18) puts back in! The macroscopic field, then, is given by the potential

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\mathbf{P}(\mathbf{r}')\cdot\hat{\boldsymbol{\imath}}}{\imath^2}\,d\tau', \tag{4.19}$$

where the integral runs over the *entire* volume of the dielectric. This is, of course, what we used in Sect. 4.2.1; without realizing it, we were correctly calculating the averaged, macroscopic field, for points inside the dielectric.

You may have to reread the last couple of paragraphs for the argument to sink in. Notice that it all revolves around the curious fact that the average field over *any* sphere (due to the charge inside) is the same as the field at the center of a *uniformly polarized* sphere with the same total dipole moment. This means that no matter how crazy the actual microscopic charge configuration, we can replace it by a nice smooth distribution of perfect dipoles, if all we want is the macroscopic (average) field. Incidentally, while the argument ostensibly relies on the spherical shape I chose to average over, the macroscopic field is certainly independent of the geometry of the averaging region, and this is reflected in the final answer, Eq. 4.19. Presumably one could reproduce the same argument for a cube or an ellipsoid or whatever—the calculation might be more difficult, but the conclusion would be the same.

4.3 ■ THE ELECTRIC DISPLACEMENT

4.3.1 ■ Gauss's Law in the Presence of Dielectrics

In Sect. 4.2 we found that the effect of polarization is to produce accumulations of (bound) charge, $\rho_b = -\boldsymbol{\nabla}\cdot\mathbf{P}$ within the dielectric and $\sigma_b = \mathbf{P}\cdot\hat{\mathbf{n}}$ on the surface. The field due to polarization of the medium is just the field of this bound charge. We are now ready to put it all together: the field attributable to bound charge plus the field due to everything *else* (which, for want of a better term, we call **free charge**, ρ_f). The free charge might consist of electrons on a conductor or ions embedded in the dielectric material or whatever; any charge, in other words, that is *not* a result of polarization. Within the dielectric, the total charge density can be written:

$$\rho = \rho_b + \rho_f, \tag{4.20}$$

and Gauss's law reads

$$\epsilon_0\boldsymbol{\nabla}\cdot\mathbf{E} = \rho = \rho_b + \rho_f = -\boldsymbol{\nabla}\cdot\mathbf{P} + \rho_f,$$

where $\mathbf{E}$ is now the *total* field, not just that portion generated by polarization.

It is convenient to combine the two divergence terms:

$$\nabla \cdot (\epsilon_0 \mathbf{E} + \mathbf{P}) = \rho_f.$$

The expression in parentheses, designated by the letter **D**,

$$\boxed{\mathbf{D} \equiv \epsilon_0 \mathbf{E} + \mathbf{P},} \tag{4.21}$$

is known as the **electric displacement**. In terms of **D**, Gauss's law reads

$$\boxed{\nabla \cdot \mathbf{D} = \rho_f,} \tag{4.22}$$

or, in integral form,

$$\oint \mathbf{D} \cdot d\mathbf{a} = Q_{f_{\text{enc}}}, \tag{4.23}$$

where $Q_{f_{\text{enc}}}$ denotes the total free charge enclosed in the volume. This is a particularly useful way to express Gauss's law, in the context of dielectrics, because *it makes reference only to free charges,* and free charge is the stuff we control. Bound charge comes along for the ride: when we put the free charge in place, a certain polarization automatically ensues, by the mechanisms of Sect. 4.1, and this polarization produces the bound charge. In a typical problem, therefore, we know ρ_f, but we do not (initially) know ρ_b; Eq. 4.23 lets us go right to work with the information at hand. In particular, whenever the requisite symmetry is present, we can immediately calculate **D** by the standard Gauss's law methods.

Example 4.4. A long straight wire, carrying uniform line charge λ, is surrounded by rubber insulation out to a radius a (Fig. 4.17). Find the electric displacement.

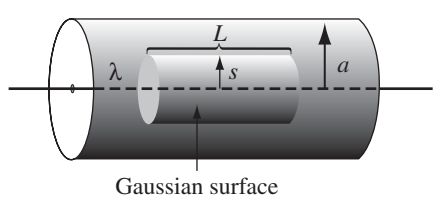


FIGURE 4.17

Solution

Drawing a cylindrical Gaussian surface, of radius s and length L, and applying Eq. 4.23, we find

$$D(2\pi s L) = \lambda L.$$

Therefore,

$$\mathbf{D} = \frac{\lambda}{2\pi s}\hat{\mathbf{s}}. \tag{4.24}$$

Notice that this formula holds both within the insulation and outside it. In the latter region, $\mathbf{P} = 0$, so

$$\mathbf{E} = \frac{1}{\epsilon_0}\mathbf{D} = \frac{\lambda}{2\pi\epsilon_0 s}\hat{\mathbf{s}}, \qquad \text{for } s > a.$$

Inside the rubber, the electric field cannot be determined, since we do not know **P**.

It may appear to you that I left out the surface bound charge σ_b in deriving Eq. 4.22, and in a sense that is true. We cannot apply Gauss's law precisely *at* the surface of a dielectric, for here ρ_b blows up,[6] taking the divergence of **E** with it. But everywhere *else* the logic is sound, and in fact if we picture the edge of the dielectric as having some finite thickness, within which the polarization tapers off to zero (probably a more realistic model than an abrupt cut-off anyway), then there *is* no surface bound charge; ρ_b varies rapidly but smoothly within this "skin," and Gauss's law can be safely applied *everywhere.* At any rate, the integral form (Eq. 4.23) is free from this "defect."

Problem 4.15 A thick spherical shell (inner radius a, outer radius b) is made of dielectric material with a "frozen-in" polarization

$$\mathbf{P}(\mathbf{r}) = \frac{k}{r}\hat{\mathbf{r}},$$

where k is a constant and r is the distance from the center (Fig. 4.18). (There is no *free* charge in the problem.) Find the electric field in all three regions by two different methods:

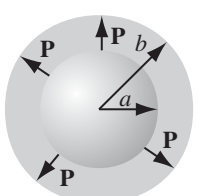


FIGURE 4.18

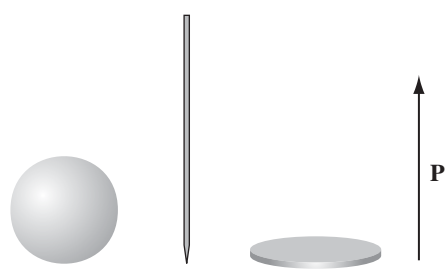


FIGURE 4.19

[6]The polarization drops abruptly to zero outside the material, so its *derivative* is a delta function (see Prob. 1.46). The surface bound charge *is* precisely this term—in this sense it is actually *included* in ρ_b, but we ordinarily prefer to handle it separately as σ_b.

(a) Locate all the bound charge, and use Gauss's law (Eq. 2.13) to calculate the field it produces.

(b) Use Eq. 4.23 to find **D**, and then get **E** from Eq. 4.21. [Notice that the second method is much faster, and it avoids any explicit reference to the bound charges.]

Problem 4.16 Suppose the field inside a large piece of dielectric is $\mathbf{E}_0$, so that the electric displacement is $\mathbf{D}_0 = \epsilon_0\mathbf{E}_0 + \mathbf{P}$.

(a) Now a small spherical cavity (Fig. 4.19a) is hollowed out of the material. Find the field at the center of the cavity in terms of $\mathbf{E}_0$ and **P**. Also find the displacement at the center of the cavity in terms of $\mathbf{D}_0$ and **P**. Assume the polarization is "frozen in," so it doesn't change when the cavity is excavated.

(b) Do the same for a long needle-shaped cavity running parallel to **P** (Fig. 4.19b).

(c) Do the same for a thin wafer-shaped cavity perpendicular to **P** (Fig. 4.19c).

Assume the cavities are small enough that **P**, $\mathbf{E}_0$, and $\mathbf{D}_0$ are essentially uniform. [*Hint:* Carving out a cavity is the same as superimposing an object of the same shape but opposite polarization.]

4.3.2 ■ A Deceptive Parallel

Equation 4.22 looks just like Gauss's law, only the *total* charge density ρ is replaced by the *free* charge density ρ_f, and **D** is substituted for $\epsilon_0\mathbf{E}$. For this reason, you may be tempted to conclude that **D** is "just like" **E** (apart from the factor ϵ_0), except that its source is ρ_f instead of ρ: "To solve problems involving dielectrics, you just forget all about the bound charge—calculate the field as you ordinarily would, only call the answer **D** instead of **E**." This reasoning is seductive, but the conclusion is false; in particular, there is no "Coulomb's law" for **D**:

$$\mathbf{D}(\mathbf{r}) \neq \frac{1}{4\pi}\int \frac{\hat{\boldsymbol{\imath}}}{\imath^2}\rho_f(\mathbf{r}')\,d\tau'.$$

The parallel between **E** and **D** is more subtle than that.

For the divergence alone is insufficient to determine a vector field; you need to know the curl as well. One tends to forget this in the case of electrostatic fields because the curl of **E** is always zero. But the curl of **D** is *not* always zero.

$$\nabla \times \mathbf{D} = \epsilon_0(\nabla \times \mathbf{E}) + (\nabla \times \mathbf{P}) = \nabla \times \mathbf{P}, \tag{4.25}$$

and there is no reason, in general, to suppose that the curl of **P** vanishes. Sometimes it does, as in Ex. 4.4 and Prob. 4.15, but more often it does not. The bar electret of Prob. 4.11 is a case in point: here there is no free charge anywhere, so if you really believe that the only source of **D** is ρ_f, you will be forced to conclude that $\mathbf{D} = \mathbf{0}$ everywhere, and hence that $\mathbf{E} = (-1/\epsilon_0)\mathbf{P}$ inside and $\mathbf{E} = \mathbf{0}$ outside the electret, which is obviously wrong. (I leave it for you to find the place where $\nabla \times \mathbf{P} \neq \mathbf{0}$ in this problem.) Because $\nabla \times \mathbf{D} \neq \mathbf{0}$, moreover, **D** cannot be expressed as the gradient of a scalar—there is no "potential" for **D**.

Advice: When you are asked to compute the electric displacement, first look for symmetry. If the problem exhibits spherical, cylindrical, or plane symmetry, then you can get **D** directly from Eq. 4.23 by the usual Gauss's law methods. (Evidently in such cases $\nabla \times \mathbf{P}$ is automatically zero, but since symmetry alone dictates the answer, you're not really obliged to worry about the curl.) If the requisite symmetry is absent, you'll have to think of another approach, and, in particular, you must *not* assume that **D** is determined exclusively by the free charge.

4.3.3 ■ Boundary Conditions

The electrostatic boundary conditions of Sect. 2.3.5 can be recast in terms of **D**. Equation 4.23 tells us the discontinuity in the component perpendicular to an interface:

$$D^{\perp}_{\text{above}} - D^{\perp}_{\text{below}} = \sigma_f, \tag{4.26}$$

while Eq. 4.25 gives the discontinuity in parallel components:

$$\mathbf{D}^{\parallel}_{\text{above}} - \mathbf{D}^{\parallel}_{\text{below}} = \mathbf{P}^{\parallel}_{\text{above}} - \mathbf{P}^{\parallel}_{\text{below}}. \tag{4.27}$$

In the presence of dielectrics, these are sometimes more useful than the corresponding boundary conditions on **E** (Eqs. 2.31 and 2.32):

$$E^{\perp}_{\text{above}} - E^{\perp}_{\text{below}} = \frac{1}{\epsilon_0}\sigma, \tag{4.28}$$

and

$$\mathbf{E}^{\parallel}_{\text{above}} - \mathbf{E}^{\parallel}_{\text{below}} = \mathbf{0}. \tag{4.29}$$

You might try applying them, for example, to Probs. 4.16 and 4.17.

Problem 4.17 For the bar electret of Prob. 4.11, make three careful sketches: one of **P**, one of **E**, and one of **D**. Assume L is about $2a$. [*Hint:* **E** lines terminate on charges; **D** lines terminate on *free* charges.]

4.4 ■ LINEAR DIELECTRICS

4.4.1 ■ Susceptibility, Permittivity, Dielectric Constant

In Sects. 4.2 and 4.3 we did not commit ourselves as to the *cause* of **P**; we dealt only with the *effects* of polarization. From the qualitative discussion of Sect. 4.1, though, we know that the polarization of a dielectric ordinarily results from an electric field, which lines up the atomic or molecular dipoles. For many substances, in fact, the polarization is *proportional* to the field, provided **E** is not too strong:

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}. \tag{4.30}$$

The constant of proportionality, χ_e, is called the **electric susceptibility** of the medium (a factor of ϵ_0 has been extracted to make χ_e dimensionless). The value of χ_e depends on the microscopic structure of the substance in question (and also on external conditions such as temperature). I shall call materials that obey Eq. 4.30 **linear dielectrics.**[7]

Note that **E** in Eq. 4.30 is the *total* field; it may be due in part to free charges and in part to the polarization itself. If, for instance, we put a piece of dielectric into an external field $\mathbf{E}_0$, we cannot compute **P** directly from Eq. 4.30; the external field will polarize the material, and this polarization will produce its own field, which then contributes to the total field, and this in turn modifies the polarization, which . . . Breaking out of this infinite regress is not always easy. You'll see some examples in a moment. The simplest approach is to begin with the *displacement,* at least in those cases where **D** can be deduced directly from the free charge distribution.

In linear media we have

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = \epsilon_0\mathbf{E} + \epsilon_0\chi_e\mathbf{E} = \epsilon_0(1 + \chi_e)\mathbf{E}, \tag{4.31}$$

so **D** is *also* proportional to **E**:

$$\mathbf{D} = \epsilon\mathbf{E}, \tag{4.32}$$

where

$$\epsilon \equiv \epsilon_0(1 + \chi_e). \tag{4.33}$$

This new constant ϵ is called the **permittivity** of the material. (In vacuum, where there is no matter to polarize, the susceptibility is zero, and the permittivity is ϵ_0. That's why ϵ_0 is called the **permittivity of free space**. I dislike the term, for it suggests that the vacuum is just a special kind of linear dielectric, in which the permittivity happens to have the value $8.85 \times 10^{-12}\ \mathrm{C}^2/\mathrm{N}\cdot\mathrm{m}^2$.) If you remove a factor of ϵ_0, the remaining dimensionless quantity

$$\epsilon_r \equiv 1 + \chi_e = \frac{\epsilon}{\epsilon_0} \tag{4.34}$$

is called the **relative permittivity**, or **dielectric constant**, of the material. Dielectric constants for some common substances are listed in Table 4.2. (Notice that ϵ_r is greater than 1, for all ordinary materials.) Of course, the permittivity and the dielectric constant do not convey any information that was not already available in the susceptibility, nor is there anything essentially new in Eq. 4.32; the *physics* of linear dielectrics is all contained in Eq. 4.30.[8]

[7] In modern optical applications, especially, *non*linear materials have become increasingly important. For these there is a second term in the formula for **P** as a function of **E**—typically a *cubic* term. In general, Eq. 4.30 can be regarded as the first (nonzero) term in the Taylor expansion of **P** in powers of **E**.

[8] As long as we are engaged in this orgy of unnecessary terminology and notation, I might as well mention that formulas for **D** in terms of **E** (Eq. 4.32, in the case of linear dielectrics) are called **constitutive relations**.

Material	Dielectric Constant	Material	Dielectric Constant
Vacuum	1	Benzene	2.28
Helium	1.000065	Diamond	5.7-5.9
Neon	1.00013	Salt	5.9
Hydrogen (H_2)	1.000254	Silicon	11.7
Argon	1.000517	Methanol	33.0
Air (dry)	1.000536	Water	80.1
Nitrogen (N_2)	1.000548	Ice (-30° C)	104
Water vapor (100° C)	1.00589	$KTaNbO_3$ (0° C)	34,000

TABLE 4.2 Dielectric Constants (unless otherwise specified, values given are for 1 atm, 20° C). *Data from Handbook of Chemistry and Physics,* 91st ed. (Boca Raton: CRC Press, 2010).

Example 4.5. A metal sphere of radius a carries a charge Q (Fig. 4.20). It is surrounded, out to radius b, by linear dielectric material of permittivity ϵ. Find the potential at the center (relative to infinity).

Solution

To compute V, we need to know $\mathbf{E}$; to find $\mathbf{E}$, we might first try to locate the bound charge; we could get the bound charge from $\mathbf{P}$, but we can't calculate $\mathbf{P}$ unless we already know $\mathbf{E}$ (Eq. 4.30). We seem to be in a bind. What we *do* know is the *free* charge Q, and fortunately the arrangement is spherically symmetric, so let's begin by calculating $\mathbf{D}$, using Eq. 4.23:

$$\mathbf{D} = \frac{Q}{4\pi r^2}\hat{\mathbf{r}}, \quad \text{for all points } r > a.$$

(Inside the metal sphere, of course, $\mathbf{E} = \mathbf{P} = \mathbf{D} = \mathbf{0}$.) Once we know $\mathbf{D}$, it is a trivial matter to obtain $\mathbf{E}$, using Eq. 4.32:

$$\mathbf{E} = \begin{cases} \dfrac{Q}{4\pi\epsilon r^2}\hat{\mathbf{r}}, & \text{for } a < r < b, \\ \dfrac{Q}{4\pi\epsilon_0 r^2}\hat{\mathbf{r}}, & \text{for } r > b. \end{cases}$$

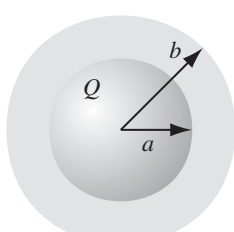


FIGURE 4.20

The potential at the center is therefore

$$V = -\int_{\infty}^{0} \mathbf{E}\cdot d\mathbf{l} = -\int_{\infty}^{b}\left(\frac{Q}{4\pi\epsilon_0 r^2}\right)dr - \int_{b}^{a}\left(\frac{Q}{4\pi\epsilon r^2}\right)dr - \int_{a}^{0}(0)\,dr$$

$$= \frac{Q}{4\pi}\left(\frac{1}{\epsilon_0 b} + \frac{1}{\epsilon a} - \frac{1}{\epsilon b}\right).$$

As it turns out, it was not necessary for us to compute the polarization or the bound charge explicitly, though this can easily be done:

$$\mathbf{P} = \epsilon_0\chi_e\mathbf{E} = \frac{\epsilon_0\chi_e Q}{4\pi\epsilon r^2}\hat{\mathbf{r}},$$

in the dielectric, and hence

$$\rho_b = -\boldsymbol{\nabla}\cdot\mathbf{P} = 0,$$

while

$$\sigma_b = \mathbf{P}\cdot\hat{\mathbf{n}} = \begin{cases} \dfrac{\epsilon_0\chi_e Q}{4\pi\epsilon b^2}, & \text{at the outer surface,} \\ \dfrac{-\epsilon_0\chi_e Q}{4\pi\epsilon a^2}, & \text{at the inner surface.} \end{cases}$$

Notice that the surface bound charge at a is *negative* ($\hat{\mathbf{n}}$ points outward *with respect to the dielectric,* which is $+\hat{\mathbf{r}}$ at b but $-\hat{\mathbf{r}}$ at a). This is natural, since the charge on the metal sphere attracts its opposite in all the dielectric molecules. It is this layer of negative charge that reduces the field, within the dielectric, from $1/4\pi\epsilon_0(Q/r^2)\hat{\mathbf{r}}$ to $1/4\pi\epsilon(Q/r^2)\hat{\mathbf{r}}$. In this respect, a dielectric is rather like an imperfect conductor: on a *conducting* shell the induced surface charge would be such as to cancel the field of Q *completely* in the region $a < r < b$; the dielectric does the best it can, but the cancellation is only partial.

You might suppose that linear dielectrics escape the defect in the parallel between **E** and **D**. Since **P** and **D** are now proportional to **E**, does it not follow that their curls, like **E**'s, must vanish? Unfortunately, it does *not,* for the line integral of **P** around a closed path that *straddles the boundary between one type of material and another* need not be zero, even though the integral of **E** around the same loop *must* be. The reason is that the proportionality factor $\epsilon_0\chi_e$ is different on the two sides. For instance, at the interface between a polarized dielectric and the vacuum (Fig. 4.21), **P** is zero on one side but not on the other. Around this

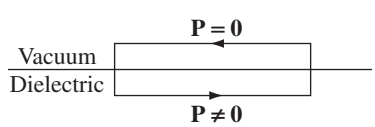


FIGURE 4.21

loop $\oint \mathbf{P} \cdot d\mathbf{l} \neq 0$, and hence, by Stokes' theorem, the curl of $\mathbf{P}$ cannot vanish everywhere within the loop (in fact, it is *infinite* at the boundary).[9]

Of course, if the space is *entirely* filled with a homogeneous[10] linear dielectric, then this objection is void; in this rather special circumstance

$$\nabla \cdot \mathbf{D} = \rho_f \quad \text{and} \quad \nabla \times \mathbf{D} = \mathbf{0},$$

so $\mathbf{D}$ can be found from the free charge just as though the dielectric were not there:

$$\mathbf{D} = \epsilon_0 \mathbf{E}_{\text{vac}},$$

where $\mathbf{E}_{\text{vac}}$ is the field the same free charge distribution would produce in the absence of any dielectric. According to Eqs. 4.32 and 4.34, therefore,

$$\mathbf{E} = \frac{1}{\epsilon}\mathbf{D} = \frac{1}{\epsilon_r}\mathbf{E}_{\text{vac}}. \tag{4.35}$$

Conclusion: When all space is filled with a homogeneous linear dielectric, the field everywhere is simply reduced by a factor of one over the dielectric constant. (Actually, it is not necessary for the dielectric to fill *all* space: in regions where the field is zero anyway, it can hardly matter whether the dielectric is present or not, since there's no polarization in any event.)

For example, if a free charge q is embedded in a large dielectric, the field it produces is

$$\mathbf{E} = \frac{1}{4\pi\epsilon}\frac{q}{r^2}\hat{\mathbf{r}} \tag{4.36}$$

(that's ϵ, not ϵ_0), and the force it exerts on nearby charges is reduced accordingly. But it's not that there is anything wrong with Coulomb's law; rather, the polarization of the medium partially "shields" the charge, by surrounding it with bound charge of the opposite sign (Fig. 4.22).[11]

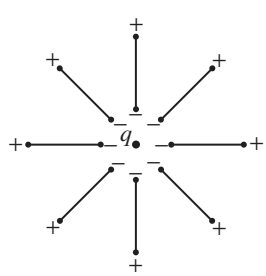


FIGURE 4.22

[9]Putting that argument in differential form, Eq. 4.30 and product rule 7 yield $\nabla \times \mathbf{P} = -\epsilon_0 \mathbf{E} \times (\nabla \chi_e)$, so the problem arises when $\nabla \chi_e$ is not parallel to $\mathbf{E}$.

[10]A **homogeneous** medium is one whose properties (in this case the susceptibility) do not vary with position.

[11]In *quantum* electrodynamics, the vacuum itself can be polarized, and this means that the effective (or "renormalized") charge of the electron, as you might measure it in the laboratory, is not its true ("bare") value, and in fact depends slightly on how far away you are!

Example 4.6. A parallel-plate capacitor (Fig. 4.23) is filled with insulating material of dielectric constant ϵ_r. What effect does this have on its capacitance?

Solution
Since the field is confined to the space between the plates, the dielectric will reduce **E**, and hence also the potential difference V, by a factor $1/\epsilon_r$. Accordingly, the capacitance $C = Q/V$ is *increased by a factor of the dielectric constant,*

$$C = \epsilon_r C_{\text{vac}}. \tag{4.37}$$

This is, in fact, a common way to beef up a capacitor.

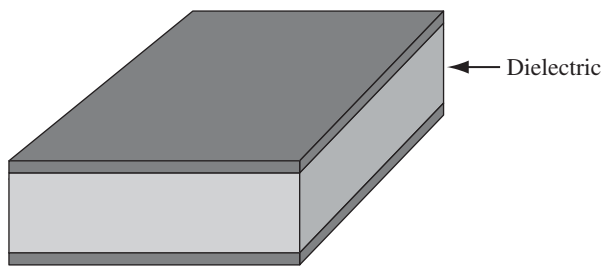


FIGURE 4.23

A *crystal* is generally easier to polarize in some directions than in others,[12] and in this case Eq. 4.30 is replaced by the general linear relation

$$\left.\begin{aligned} P_x &= \epsilon_0(\chi_{e_{xx}} E_x + \chi_{e_{xy}} E_y + \chi_{e_{xz}} E_z) \\ P_y &= \epsilon_0(\chi_{e_{yx}} E_x + \chi_{e_{yy}} E_y + \chi_{e_{yz}} E_z) \\ P_z &= \epsilon_0(\chi_{e_{zx}} E_x + \chi_{e_{zy}} E_y + \chi_{e_{zz}} E_z) \end{aligned}\right\}, \tag{4.38}$$

just as Eq. 4.1 was superseded by Eq. 4.3 for asymmetrical molecules. The nine coefficients, $\chi_{e_{xx}}$, $\chi_{e_{xy}}$, ..., constitute the **susceptibility tensor.**

Problem 4.18 The space between the plates of a parallel-plate capacitor (Fig. 4.24) is filled with two slabs of linear dielectric material. Each slab has thickness a, so the total distance between the plates is $2a$. Slab 1 has a dielectric constant of 2, and slab 2 has a dielectric constant of 1.5. The free charge density on the top plate is σ and on the bottom plate $-\sigma$.

[12] A medium is said to be **isotropic** if its properties (such as susceptibility) are the same in all directions. Thus Eq. 4.30 is the special case of Eq. 4.38 that holds for isotropic media. Physicists tend to be sloppy with their language, and unless otherwise indicated the term "linear dielectric" implies "isotropic linear dielectric," and suggests "homogeneous isotropic linear dielectric." But technically, "linear" just means that at *any given point*, and for **E** in a *given direction*, the components of **P** are proportional to E—the proportionality factor could vary with position and/or direction.

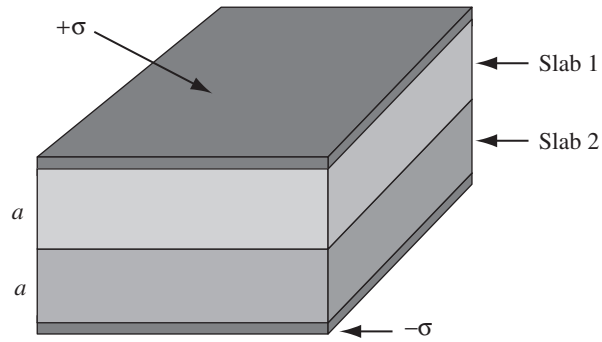


FIGURE 4.24

(a) Find the electric displacement **D** in each slab.

(b) Find the electric field **E** in each slab.

(c) Find the polarization **P** in each slab.

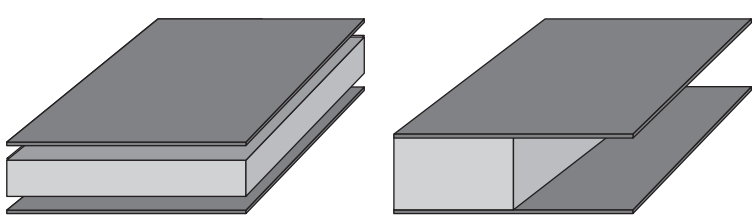
(d) Find the potential difference between the plates.

(e) Find the location and amount of all bound charge.

(f) Now that you know all the charge (free and bound), recalculate the field in each slab, and confirm your answer to (b).

Problem 4.19 Suppose you have enough linear dielectric material, of dielectric constant ϵ_r, to *half*-fill a parallel-plate capacitor (Fig. 4.25). By what fraction is the capacitance increased when you distribute the material as in Fig. 4.25(a)? How about Fig. 4.25(b)? For a given potential difference V between the plates, find **E**, **D**, and **P**, in each region, and the free and bound charge on all surfaces, for both cases.

Problem 4.20 A sphere of linear dielectric material has embedded in it a uniform free charge density ρ. Find the potential at the center of the sphere (relative to infinity), if its radius is R and the dielectric constant is ϵ_r.



(a) (b)

FIGURE 4.25

Problem 4.21 A certain coaxial cable consists of a copper wire, radius a, surrounded by a concentric copper tube of inner radius c (Fig. 4.26). The space between is partially filled (from b out to c) with material of dielectric constant ϵ_r, as shown. Find the capacitance per unit length of this cable.

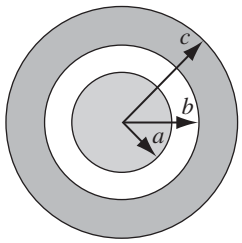


FIGURE 4.26

4.4.2 ■ Boundary Value Problems with Linear Dielectrics

In a (homogeneous isotropic) linear dielectric, the bound charge density (ρ_b) is proportional to the free charge density (ρ_f):[13]

$$\rho_b = -\nabla \cdot \mathbf{P} = -\nabla \cdot \left(\epsilon_0 \frac{\chi_e}{\epsilon}\mathbf{D}\right) = -\left(\frac{\chi_e}{1+\chi_e}\right)\rho_f. \tag{4.39}$$

In particular, unless free charge is actually embedded in the material, $\rho = 0$, and any net charge must reside at the surface. Within such a dielectric, then, the potential obeys Laplace's equation, and all the machinery of Chapter 3 carries over. It is convenient, however, to rewrite the boundary conditions in a way that makes reference only to the free charge. Equation 4.26 says

$$\epsilon_{\text{above}} E^{\perp}_{\text{above}} - \epsilon_{\text{below}} E^{\perp}_{\text{below}} = \sigma_f, \tag{4.40}$$

or (in terms of the potential),

$$\epsilon_{\text{above}} \frac{\partial V_{\text{above}}}{\partial n} - \epsilon_{\text{below}} \frac{\partial V_{\text{below}}}{\partial n} = -\sigma_f, \tag{4.41}$$

whereas the potential itself is, of course, continuous (Eq. 2.34):

$$V_{\text{above}} = V_{\text{below}}. \tag{4.42}$$

[13]This does not apply to the surface charge (σ_b), because χ_e is not independent of position (obviously) at the boundary.

Example 4.7. A sphere of homogeneous linear dielectric material is placed in an otherwise uniform electric field $\mathbf{E}_0$ (Fig. 4.27). Find the electric field inside the sphere.

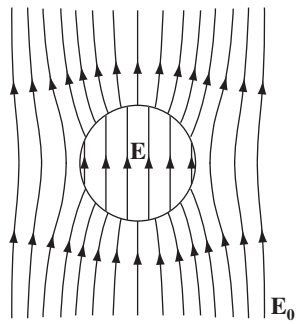


FIGURE 4.27

Solution

This is reminiscent of Ex. 3.8, in which an uncharged *conducting* sphere was introduced into a uniform field. In that case, the field of the induced charge canceled $\mathbf{E}_0$ within the sphere; in a *dielectric,* the cancellation (from the bound charge) is incomplete.

Our problem is to solve Laplace's equation, for $V_{\text{in}}(r, \theta)$ when $r \leq R$, and $V_{\text{out}}(r, \theta)$ when $r \geq R$, subject to the boundary conditions

$$\left.\begin{aligned} &\text{(i)} && V_{\text{in}} = V_{\text{out}}, && \text{at } r = R, \\ &\text{(ii)} && \epsilon \frac{\partial V_{\text{in}}}{\partial r} = \epsilon_0 \frac{\partial V_{\text{out}}}{\partial r}, && \text{at } r = R, \\ &\text{(iii)} && V_{\text{out}} \to -E_0 r \cos\theta, && \text{for } r \gg R. \end{aligned}\right\} \tag{4.43}$$

(The second of these follows from Eq. 4.41, since there is no free charge at the surface.) Inside the sphere, Eq. 3.65 says

$$V_{\text{in}}(r, \theta) = \sum_{l=0}^{\infty} A_l \, r^l \, P_l(\cos\theta); \tag{4.44}$$

outside the sphere, in view of (iii), we have

$$V_{\text{out}}(r, \theta) = -E_0 r \cos\theta + \sum_{l=0}^{\infty} \frac{B_l}{r^{l+1}} P_l(\cos\theta). \tag{4.45}$$

Boundary condition (i) requires that

$$\sum_{l=0}^{\infty} A_l \, R^l \, P_l(\cos\theta) = -E_0 R \cos\theta + \sum_{l=0}^{\infty} \frac{B_l}{R^{l+1}} P_l(\cos\theta),$$

so[14]

$$\left.\begin{aligned} A_l R^l &= \frac{B_l}{R^{l+1}}, \quad \text{for } l \neq 1, \\ A_1 R &= -E_0 R + \frac{B_1}{R^2}. \end{aligned}\right\} \tag{4.46}$$

Meanwhile, condition (ii) yields

$$\epsilon_r \sum_{l=0}^{\infty} l A_l R^{l-1} P_l(\cos\theta) = -E_0 \cos\theta - \sum_{l=0}^{\infty} \frac{(l+1)B_l}{R^{l+2}} P_l(\cos\theta),$$

so

$$\left.\begin{aligned} \epsilon_r l A_l R^{l-1} &= -\frac{(l+1)B_l}{R^{l+2}}, \text{ for } l \neq 1, \\ \epsilon_r A_1 &= -E_0 - \frac{2B_1}{R^3}. \end{aligned}\right\} \tag{4.47}$$

It follows that

$$\left.\begin{aligned} A_l = B_l = 0, &\qquad \text{for } l \neq 1, \\ A_1 = -\frac{3}{\epsilon_r + 2} E_0 &\quad B_1 = \frac{\epsilon_r - 1}{\epsilon_r + 2} R^3 E_0. \end{aligned}\right\} \tag{4.48}$$

Evidently

$$V_{\text{in}}(r, \theta) = -\frac{3E_0}{\epsilon_r + 2} r \cos\theta = -\frac{3E_0}{\epsilon_r + 2} z,$$

and hence the field inside the sphere is (surprisingly) *uniform*:

$$\mathbf{E} = \frac{3}{\epsilon_r + 2} \mathbf{E}_0. \tag{4.49}$$

Example 4.8. Suppose the entire region below the plane $z = 0$ in Fig. 4.28 is filled with uniform linear dielectric material of susceptibility χ_e. Calculate the force on a point charge q situated a distance d above the origin.

[14]Remember, $P_1(\cos\theta) = \cos\theta$, and the coefficients must be equal for each l, as you could prove by multiplying by $P_{l'}(\cos\theta)\sin\theta$, integrating from 0 to π, and invoking the orthogonality of the Legendre polynomials (Eq. 3.68).

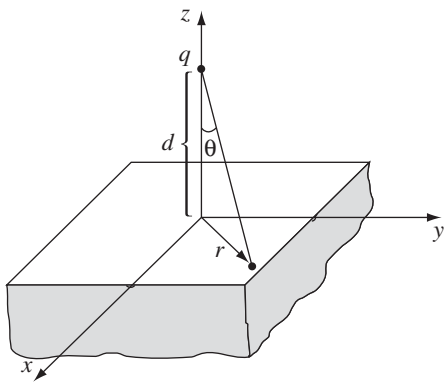


FIGURE 4.28

Solution

The surface bound charge on the xy plane is of opposite sign to q, so the force will be attractive. (In view of Eq. 4.39, there is no volume bound charge.) Let us first calculate σ_b, using Eqs. 4.11 and 4.30.[15]

$$\sigma_b = \mathbf{P} \cdot \hat{\mathbf{n}} = P_z = \epsilon_0 \chi_e E_z,$$

where E_z is the z-component of the total field just inside the dielectric, at $z = 0$. This field is due in part to q and in part to the bound charge itself. From Coulomb's law, the former contribution is

$$-\frac{1}{4\pi\epsilon_0}\frac{q}{(r^2+d^2)}\cos\theta = -\frac{1}{4\pi\epsilon_0}\frac{qd}{(r^2+d^2)^{3/2}},$$

where $r = \sqrt{x^2+y^2}$ is the distance from the origin. The z component of the field of the bound charge, meanwhile, is $-\sigma_b/2\epsilon_0$ (see footnote after Eq. 2.33). Thus

$$\sigma_b = \epsilon_0 \chi_e \left[-\frac{1}{4\pi\epsilon_0}\frac{qd}{(r^2+d^2)^{3/2}} - \frac{\sigma_b}{2\epsilon_0} \right],$$

which we can solve for σ_b:

$$\sigma_b = -\frac{1}{2\pi}\left(\frac{\chi_e}{\chi_e+2}\right)\frac{qd}{(r^2+d^2)^{3/2}}. \tag{4.50}$$

Apart from the factor $\chi_e/(\chi_e+2)$, this is exactly the same as the induced charge on an infinite *conducting* plane under similar circumstances (Eq. 3.10).[16] Evidently the *total* bound charge is

$$q_b = -\left(\frac{\chi_e}{\chi_e+2}\right)q. \tag{4.51}$$

[15]This method mimics Prob. 3.38.

[16]For some purposes a conductor can be regarded as the limiting case of a linear dielectric, with $\chi_e \to \infty$. This is often a useful check—try applying it to Exs. 4.5, 4.6, and 4.7.

We could, of course, obtain the field of σ_b by direct integration

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\int \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2}\right)\sigma_b\, da.$$

But, as in the case of the conducting plane, there is a nicer solution by the method of images. Indeed, if we replace the dielectric by a single point charge q_b at the image position $(0, 0, -d)$, we have

$$V = \frac{1}{4\pi\epsilon_0}\left[\frac{q}{\sqrt{x^2+y^2+(z-d)^2}} + \frac{q_b}{\sqrt{x^2+y^2+(z+d)^2}}\right], \tag{4.52}$$

in the region $z > 0$. Meanwhile, a charge $(q + q_b)$ at $(0, 0, d)$ yields the potential

$$V = \frac{1}{4\pi\epsilon_0}\left[\frac{q+q_b}{\sqrt{x^2+y^2+(z-d)^2}}\right], \tag{4.53}$$

for the region $z < 0$. Taken together, Eqs. 4.52 and 4.53 constitute a function that satisfies Poisson's equation with a point charge q at $(0, 0, d)$, which goes to zero at infinity, which is continuous at the boundary $z = 0$, and whose normal derivative exhibits the discontinuity appropriate to a surface charge σ_b at $z = 0$:

$$-\epsilon_0\left(\left.\frac{\partial V}{\partial z}\right|_{z=0^+} - \left.\frac{\partial V}{\partial z}\right|_{z=0^-}\right) = -\frac{1}{2\pi}\left(\frac{\chi_e}{\chi_e+2}\right)\frac{qd}{(x^2+y^2+d^2)^{3/2}}.$$

Accordingly, this is the correct potential for our problem. In particular, the force on q is:

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{qq_b}{(2d)^2}\hat{\mathbf{z}} = -\frac{1}{4\pi\epsilon_0}\left(\frac{\chi_e}{\chi_e+2}\right)\frac{q^2}{4d^2}\hat{\mathbf{z}}. \tag{4.54}$$

I do not claim to have provided a compelling *motivation* for Eqs. 4.52 and 4.53—like all image solutions, this one owes its justification to the fact that it *works:* it solves Poisson's equation, and it meets the boundary conditions. Still, discovering an image solution is not entirely a matter of guesswork. There are at least two "rules of the game": (1) You must never put an image charge into the region where you're computing the potential. (Thus Eq. 4.52 gives the potential for $z > 0$, but this image charge q_b is at $z = -d$; when we turn to the region $z < 0$ (Eq. 4.53), the image charge $(q + q_b)$ is at $z = +d$.) (2) The image charges must add up to the correct total in each region. (That's how I knew to use q_b to account for the charge in the region $z \leq 0$, and $(q + q_b)$ to cover the region $z \geq 0$.)

Problem 4.22 A very long cylinder of linear dielectric material is placed in an otherwise uniform electric field $\mathbf{E}_0$. Find the resulting field within the cylinder. (The radius is a, the susceptibility χ_e, and the axis is perpendicular to $\mathbf{E}_0$.)

Problem 4.23 Find the field inside a sphere of linear dielectric material in an otherwise uniform electric field $\mathbf{E}_0$ (Ex. 4.7) by the following method of successive approximations: First pretend the field inside is just $\mathbf{E}_0$, and use Eq. 4.30 to write down the resulting polarization $\mathbf{P}_0$. This polarization generates a field of its own, $\mathbf{E}_1$ (Ex. 4.2), which in turn modifies the polarization by an amount $\mathbf{P}_1$, which further changes the field by an amount $\mathbf{E}_2$, and so on. The resulting field is $\mathbf{E}_0 + \mathbf{E}_1 + \mathbf{E}_2 + \cdots$. Sum the series, and compare your answer with Eq. 4.49.

Problem 4.24 An uncharged conducting sphere of radius a is coated with a thick insulating shell (dielectric constant ϵ_r) out to radius b. This object is now placed in an otherwise uniform electric field $\mathbf{E}_0$. Find the electric field in the insulator.

! **Problem 4.25** Suppose the region *above* the xy plane in Ex. 4.8 is *also* filled with linear dielectric but of a different susceptibility χ_e'. Find the potential everywhere.

4.4.3 ■ Energy in Dielectric Systems

It takes work to charge up a capacitor (Eq. 2.55):

$$W = \tfrac{1}{2} C V^2.$$

If the capacitor is filled with linear dielectric, its capacitance exceeds the vacuum value by a factor of the dielectric constant,

$$C = \epsilon_r C_{\text{vac}},$$

as we found in Ex. 4.6. Evidently the work necessary to charge a dielectric-filled capacitor is increased by the same factor. The reason is pretty clear: you have to pump on more (free) charge, to achieve a given potential, because part of the field is canceled off by the bound charges.

In Chapter 2, I derived a general formula for the energy stored in any electrostatic system (Eq. 2.45):

$$W = \frac{\epsilon_0}{2} \int E^2 \, d\tau. \tag{4.55}$$

The case of the dielectric-filled capacitor suggests that this should be changed to

$$W = \frac{\epsilon_0}{2} \int \epsilon_r E^2 \, d\tau = \frac{1}{2} \int \mathbf{D} \cdot \mathbf{E} \, d\tau,$$

in the presence of linear dielectrics. To *prove* it, suppose the dielectric material is fixed in position, and we bring in the free charge, a bit at a time. As ρ_f is increased by an amount $\Delta\rho_f$, the polarization will change and with it the bound charge distribution; but we're interested only in the work done on the incremental *free* charge:

$$\Delta W = \int (\Delta\rho_f) V \, d\tau. \tag{4.56}$$

Since $\nabla \cdot \mathbf{D} = \rho_f$, $\Delta\rho_f = \nabla \cdot (\Delta\mathbf{D})$, where $\Delta\mathbf{D}$ is the resulting change in $\mathbf{D}$, so

$$\Delta W = \int [\nabla \cdot (\Delta\mathbf{D})] V \, d\tau.$$

Now

$$\nabla \cdot [(\Delta\mathbf{D})V] = [\nabla \cdot (\Delta\mathbf{D})]V + \Delta\mathbf{D} \cdot (\nabla V),$$

and hence (integrating by parts):

$$\Delta W = \int \nabla \cdot [(\Delta\mathbf{D})V] \, d\tau + \int (\Delta\mathbf{D}) \cdot \mathbf{E} \, d\tau.$$

The divergence theorem turns the first term into a surface integral, which vanishes if we integrate over all space. Therefore, the work done is equal to

$$\Delta W = \int (\Delta\mathbf{D}) \cdot \mathbf{E} \, d\tau. \tag{4.57}$$

So far, this applies to *any* material. Now, if the medium is a linear dielectric, then $\mathbf{D} = \epsilon\mathbf{E}$, so

$$\tfrac{1}{2}\Delta(\mathbf{D} \cdot \mathbf{E}) = \tfrac{1}{2}\Delta(\epsilon E^2) = \epsilon(\Delta\mathbf{E}) \cdot \mathbf{E} = (\Delta\mathbf{D}) \cdot \mathbf{E}$$

(for infinitesimal increments). Thus

$$\Delta W = \Delta\left(\frac{1}{2}\int \mathbf{D} \cdot \mathbf{E} \, d\tau\right).$$

The total work done, then, as we build the free charge up from zero to the final configuration, is

$$W = \frac{1}{2}\int \mathbf{D} \cdot \mathbf{E} \, d\tau, \tag{4.58}$$

as anticipated.[17]

It may puzzle you that Eq. 4.55, which we derived quite generally in Chapter 2, does not seem to apply in the presence of dielectrics, where it is replaced by Eq. 4.58. The point is not that one or the other of these equations is *wrong,* but rather that they address somewhat different questions. The distinction is subtle, so let's go right back to the beginning: What do we *mean* by "the energy of a system"? *Answer:* It is the work required to assemble the system. Very

[17] In case you are wondering why I did not do this more simply by the method of Sect. 2.4.3, starting with $W = \frac{1}{2}\int \rho_f V \, d\tau$, the reason is that *this* formula is untrue, in general. Study the derivation of Eq. 2.42, and you will see that it applies only to the *total* charge. For *linear* dielectrics it happens to hold for the free charge alone, but this is scarcely obvious a priori and, in fact, is most easily confirmed by working backward from Eq. 4.58.

well—but when dielectrics are involved, there are two quite different ways one might construe this process:

1. We bring in all the charges (free *and* bound), one by one, with tweezers, and glue each one down in its proper final location. If *this* is what you mean by "assemble the system," then Eq. 4.55 is your formula for the energy stored. Notice, however, that this will *not* include the work involved in stretching and twisting the dielectric molecules (if we picture the positive and negative charges as held together by tiny springs, it does not include the spring energy, $\frac{1}{2}kx^2$, associated with polarizing each molecule).[18]

2. With the unpolarized dielectric in place, we bring in the *free* charges, one by one, allowing the dielectric to respond as it sees fit. If *this* is what you mean by "assemble the system" (and ordinarily it *is,* since free charge is what we actually push around), then Eq. 4.58 is the formula you want. In this case the "spring" energy *is* included, albeit indirectly, because the force you must apply to the *free* charge depends on the disposition of the *bound* charge; as you move the free charge, you are automatically stretching those "springs."

Example 4.9. A sphere of radius R is filled with material of dielectric constant ϵ_r and uniform embedded free charge ρ_f. What is the energy of this configuration?

Solution

From Gauss's law (in the form of Eq. 4.23), the displacement is

$$\mathbf{D}(r) = \begin{cases} \dfrac{\rho_f}{3}\mathbf{r} & (r < R), \\ \dfrac{\rho_f}{3}\dfrac{R^3}{r^2}\hat{\mathbf{r}} & (r > R). \end{cases}$$

So the electric field is

$$\mathbf{E}(r) = \begin{cases} \dfrac{\rho_f}{3\epsilon_0\epsilon_r}\mathbf{r} & (r < R), \\ \dfrac{\rho_f}{3\epsilon_0}\dfrac{R^3}{r^2}\hat{\mathbf{r}} & (r > R). \end{cases}$$

The purely *electrostatic* energy (Eq. 4.55) is

$$\begin{aligned} W_1 &= \frac{\epsilon_0}{2}\left[\left(\frac{\rho_f}{3\epsilon_0\epsilon_r}\right)^2 \int_0^R r^2\, 4\pi r^2\, dr + \left(\frac{\rho_f}{3\epsilon_0}\right)^2 R^6 \int_R^\infty \frac{1}{r^4}\, 4\pi r^2\, dr\right] \\ &= \frac{2\pi}{9\epsilon_0}\rho_f^2 R^5\left(\frac{1}{5\epsilon_r^2} + 1\right). \end{aligned}$$

[18]The "spring" itself may be electrical in nature, but it is still not included in Eq. 4.55, if $\mathbf{E}$ is taken to be the *macroscopic* field.

But the *total* energy (Eq. 4.58) is

$$W_2 = \frac{1}{2}\left[\left(\frac{\rho_f}{3}\right)\left(\frac{\rho_f}{3\epsilon_0\epsilon_r}\right)\int_0^R r^2\, 4\pi r^2\, dr + \left(\frac{\rho_f R^3}{3}\right)\left(\frac{\rho_f R^3}{3\epsilon_0}\right)\int_R^\infty \frac{1}{r^4}\, 4\pi r^2\, dr\right]$$
$$= \frac{2\pi}{9\epsilon_0}\rho_f^2 R^5\left(\frac{1}{5\epsilon_r} + 1\right).$$

Notice that $W_1 < W_2$—that's because W_1 does not include the energy involved in stretching the molecules.

Let's check that W_2 is the work done on the *free* charge in assembling the system. We start with the (uncharged, unpolarized) dielectric sphere, and bring in the free charge in infinitesimal installments (dq), filling out the sphere layer by layer. When we have reached radius r', the electric field is

$$\mathbf{E}(r) = \begin{cases} \dfrac{\rho_f}{3\epsilon_0\epsilon_r}\mathbf{r} & (r < r'), \\ \dfrac{\rho_f}{3\epsilon_0\epsilon_r}\dfrac{r'^3}{r^2}\hat{\mathbf{r}} & (r' < r < R), \\ \dfrac{\rho_f}{3\epsilon_0}\dfrac{r'^3}{r^2}\hat{\mathbf{r}} & (r > R). \end{cases}$$

The work required to bring the next dq in from infinity to r' is

$$dW = -dq\left[\int_\infty^R \mathbf{E}\cdot d\mathbf{l} + \int_R^{r'} \mathbf{E}\cdot d\mathbf{l}\right]$$
$$= -dq\left[\frac{\rho_f r'^3}{3\epsilon_0}\int_\infty^R \frac{1}{r^2}\, dr + \frac{\rho_f r'^3}{3\epsilon_0\epsilon_r}\int_R^{r'} \frac{1}{r^2}\, dr\right]$$
$$= \frac{\rho_f r'^3}{3\epsilon_0}\left[\frac{1}{R} + \frac{1}{\epsilon_r}\left(\frac{1}{r'} - \frac{1}{R}\right)\right] dq.$$

This increases the radius (r'):

$$dq = \rho_f 4\pi r'^2\, dr',$$

so the *total* work done, in going from $r' = 0$ to $r' = R$, is

$$W = \frac{4\pi\rho_f^2}{3\epsilon_0}\left[\frac{1}{R}\left(1 - \frac{1}{\epsilon_r}\right)\int_0^R r'^5\, dr' + \frac{1}{\epsilon_r}\int_0^R r'^4\, dr'\right]$$
$$= \frac{2\pi}{9\epsilon_0}\rho_f^2 R^5\left(\frac{1}{5\epsilon_r} + 1\right) = W_2. \checkmark$$

Evidently the energy "stored in the springs" is

$$W_{\text{spring}} = W_2 - W_1 = \frac{2\pi}{45\epsilon_0\epsilon_r^2}\rho_f^2 R^5\,(\epsilon_r - 1)\,.$$

I would like to confirm this in an explicit model. Picture the dielectric as a collection of tiny proto-dipoles, each consisting of $+q$ and $-q$ attached to a spring of constant k and equilibrium length 0, so in the absence of any field the positive and negative ends coincide. One end of each dipole is nailed in position (like the nuclei in a solid), but the other end is free to move in response to any imposed field. Let $d\tau$ be the volume assigned to each proto-dipole (the dipole itself may occupy only a small portion of this space).

With the field turned on, the electric force on the free end is balanced by the spring force;[19] the charges separate by a distance d: $qE = kd$. In our case

$$\mathbf{E} = \frac{\rho_f}{3\epsilon_0\epsilon_r}\mathbf{r}.$$

The resulting dipole moment is $p = qd$, and the polarization is $P = p/d\tau$, so

$$k = \frac{\rho_f}{3\epsilon_0\epsilon_r d^2} Pr\, d\tau.$$

The energy of this particular spring is

$$dW_{\text{spring}} = \frac{1}{2}kd^2 = \frac{\rho_f}{6\epsilon_0\epsilon_r} Pr\, d\tau,$$

and hence the total is

$$W_{\text{spring}} = \frac{\rho_f}{6\epsilon_0\epsilon_r}\int Pr\, d\tau.$$

Now

$$\mathbf{P} = \epsilon_0\chi_e\mathbf{E} = \epsilon_0\chi_e\frac{\rho_f}{3\epsilon_0\epsilon_r}\mathbf{r} = \frac{(\epsilon_r - 1)\rho_f}{3\epsilon_r}\mathbf{r},$$

so

$$W_{\text{spring}} = \frac{\rho_f}{6\epsilon_0\epsilon_r}\frac{(\epsilon_r - 1)\rho_f}{3\epsilon_r}4\pi\int_0^R r^4\, dr = \frac{2\pi}{45\epsilon_0\epsilon_r^2}\rho_f^2 R^5\,(\epsilon_r - 1)\,,$$

and it works out perfectly.

It is sometimes alleged that Eq. 4.58 represents the energy even for *non*linear dielectrics, but this is false: To proceed beyond Eq. 4.57, one must assume linearity. In fact, for *dissipative* systems the whole notion of "stored energy" loses its meaning, because the work done depends not only on the final configuration but on *how it got there.* If the molecular "springs" are allowed to have some

[19]Note that the "spring" here is a surrogate for whatever holds the molecule together—it *includes* the electrical attraction of the other end. If it bothers you that the force is taken to be proportional to the separation, look again at Example 4.1.

friction, for instance, then W_{spring} can be made as large as you like, by assembling the charges in such a way that the spring is obliged to expand and contract many times before reaching its final state. In particular, you get nonsensical results if you try to apply Eq. 4.58 to electrets, with frozen-in polarization (see Prob. 4.27).

Problem 4.26 A spherical conductor, of radius a, carries a charge Q (Fig. 4.29). It is surrounded by linear dielectric material of susceptibility χ_e, out to radius b. Find the energy of this configuration (Eq. 4.58).

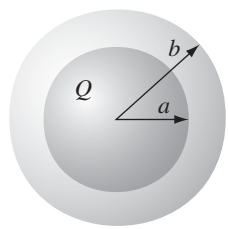


FIGURE 4.29

Problem 4.27 Calculate W, using both Eq. 4.55 and Eq. 4.58, for a sphere of radius R with frozen-in uniform polarization $\mathbf{P}$ (Ex. 4.2). Comment on the discrepancy. Which (if either) is the "true" energy of the system?

4.4.4 ■ Forces on Dielectrics

Just as a conductor is attracted into an electric field (Eq. 2.51), so too is a dielectric—and for essentially the same reason: the bound charge tends to accumulate near the free charge of the opposite sign. But the calculation of forces on dielectrics can be surprisingly tricky. Consider, for example, the case of a slab of linear dielectric material, partially inserted between the plates of a parallel-plate capacitor (Fig. 4.30). We have always pretended that the field is uniform inside a parallel-plate capacitor, and zero outside. If this were literally true, there would be no net force on the dielectric at all, since the field everywhere would be perpendicular to the plates. However, there is in reality a **fringing field** around the edges, which for most purposes can be ignored but in this case is responsible for the whole effect. (Indeed, the field *could* not terminate abruptly at the edge of the capacitor, for if it did, the line integral of $\mathbf{E}$ around the closed loop shown in Fig. 4.31 would not be zero.) It is this nonuniform fringing field that pulls the dielectric into the capacitor.

Fringing fields are notoriously difficult to calculate; luckily, we can avoid this altogether, by the following ingenious method.[20] Let W be the energy of the

[20]For a direct calculation from the fringing fields, see E. R. Dietz, *Am. J. Phys.* **72**, 1499 (2004).

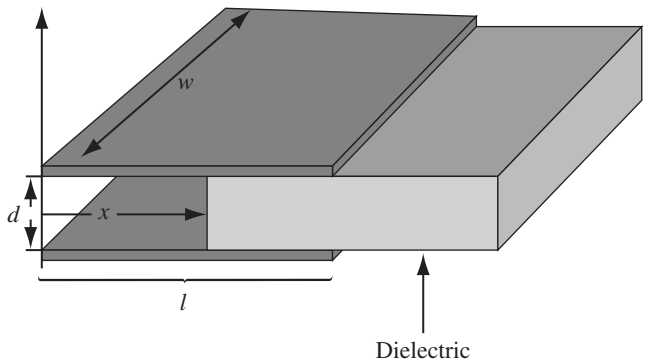


FIGURE 4.30

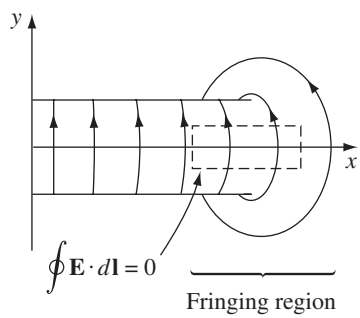


FIGURE 4.31

system—it depends, of course, on the amount of overlap. If I pull the dielectric out an infinitesimal distance dx, the energy is changed by an amount equal to the work done:

$$dW = F_{\text{me}}\, dx, \tag{4.59}$$

where F_{me} is the force I must exert, to counteract the electrical force F on the dielectric: $F_{\text{me}} = -F$. Thus the electrical force on the slab is

$$F = -\frac{dW}{dx}. \tag{4.60}$$

Now, the energy stored in the capacitor is

$$W = \tfrac{1}{2} C V^2, \tag{4.61}$$

and the capacitance in this case is

$$C = \frac{\epsilon_0 w}{d}(\epsilon_r l - \chi_e x), \tag{4.62}$$

where l is the length of the plates (Fig. 4.30). Let's assume that the total charge on the plates ($Q = CV$) is held constant, as the dielectric moves. In terms of Q,

$$W = \frac{1}{2}\frac{Q^2}{C}, \tag{4.63}$$

so

$$F = -\frac{dW}{dx} = \frac{1}{2}\frac{Q^2}{C^2}\frac{dC}{dx} = \frac{1}{2}V^2\frac{dC}{dx}. \tag{4.64}$$

But

$$\frac{dC}{dx} = -\frac{\epsilon_0 \chi_e w}{d},$$

and hence

$$F = -\frac{\epsilon_0 \chi_e w}{2d}V^2. \tag{4.65}$$

(The minus sign indicates that the force is in the negative x direction; the dielectric is pulled *into* the capacitor.)

It is a common error to use Eq. 4.61 (with V constant), rather than Eq. 4.63 (with Q constant), in computing the force. One then obtains

$$F = -\frac{1}{2}V^2\frac{dC}{dx},$$

which is off by a sign. It is, of course, *possible* to maintain the capacitor at a fixed potential, by connecting it up to a battery. But in that case the *battery also does work* as the dielectric moves; instead of Eq. 4.59, we now have

$$dW = F_{\text{me}}\,dx + V\,dQ, \tag{4.66}$$

where $V\,dQ$ is the work done by the battery. It follows that

$$F = -\frac{dW}{dx} + V\frac{dQ}{dx} = -\frac{1}{2}V^2\frac{dC}{dx} + V^2\frac{dC}{dx} = \frac{1}{2}V^2\frac{dC}{dx}, \tag{4.67}$$

the same as before (Eq. 4.64), with the *correct* sign.

Please understand: The force on the dielectric cannot possibly depend on whether you plan to hold Q constant or V constant—it is determined entirely by the distribution of charge, free and bound. It's simpler to *calculate* the force assuming constant Q, because then you don't have to worry about work done by the battery; but if you insist, it can be done correctly either way.

Notice that we were able to determine the force *without knowing anything about the fringing fields that are ultimately responsible for it!* Of course, it's built into the whole structure of electrostatics that $\boldsymbol{\nabla} \times \mathbf{E} = \mathbf{0}$, and hence that the fringing fields must be present; we're not really getting something for nothing here—just cleverly exploiting the internal consistency of the theory. The energy stored in the fringing fields themselves (which was not accounted for in this derivation) stays constant, as the slab moves; what *does* change is the energy well *inside* the capacitor, where the field is nice and uniform.

Problem 4.28 Two long coaxial cylindrical metal tubes (inner radius a, outer radius b) stand vertically in a tank of dielectric oil (susceptibility χ_e, mass density ρ). The inner one is maintained at potential V, and the outer one is grounded (Fig. 4.32). To what height (h) does the oil rise, in the space between the tubes?

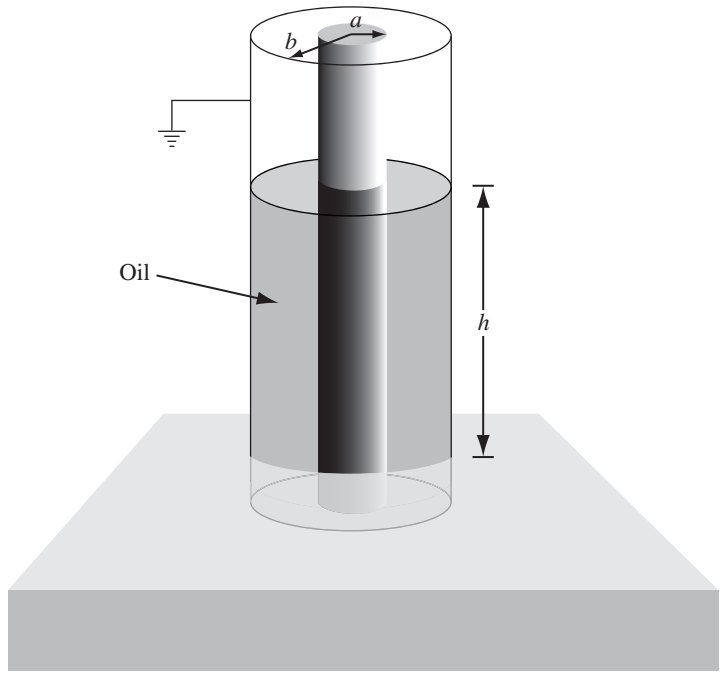


FIGURE 4.32

More Problems on Chapter 4

Problem 4.29

(a) For the configuration in Prob. 4.5, calculate the *force* on $\mathbf{p}_2$ due to $\mathbf{p}_1$, and the force on $\mathbf{p}_1$ due to $\mathbf{p}_2$. Are the answers consistent with Newton's third law?

(b) Find the total torque on $\mathbf{p}_2$ *with respect to the center of* $\mathbf{p}_1$, and compare it with the torque on $\mathbf{p}_1$ *about that same point*. [*Hint:* combine your answer to (a) with the result of Prob. 4.5.]

Problem 4.30 An electric dipole $\mathbf{p}$, pointing in the y direction, is placed midway between two large conducting plates, as shown in Fig. 4.33. Each plate makes a

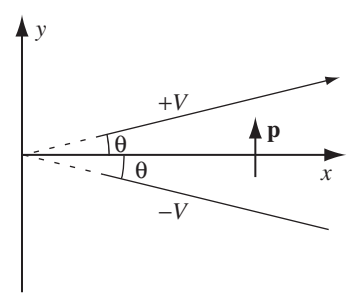


FIGURE 4.33

small angle θ with respect to the x axis, and they are maintained at potentials $\pm V$. What is the *direction* of the net force on $\mathbf{p}$? (There's nothing to *calculate*, here, but do explain your answer qualitatively.)

Problem 4.31 A point charge Q is "nailed down" on a table. Around it, at radius R, is a frictionless circular track on which a dipole $\mathbf{p}$ rides, constrained always to point tangent to the circle. Use Eq. 4.5 to show that the electric force on the dipole is

$$\mathbf{F} = \frac{Q}{4\pi\epsilon_0}\frac{\mathbf{p}}{R^3}.$$

Notice that this force is always in the "forward" direction (you can easily confirm this by drawing a diagram showing the forces on the two ends of the dipole). Why isn't this a perpetual motion machine?[21]

! **Problem 4.32** Earnshaw's theorem (Prob. 3.2) says that you cannot trap a charged particle in an electrostatic field. *Question:* Could you trap a neutral (but polarizable) atom in an electrostatic field?

(a) Show that the force on the atom is $\mathbf{F} = \frac{1}{2}\alpha\boldsymbol{\nabla}(E^2)$.

(b) The question becomes, therefore: Is it possible for E^2 to have a local maximum (in a charge-free region)? In that case the force would push the atom back to its equilibrium position. Show that the answer is *no*. [*Hint:* Use Prob. 3.4(a).][22]

Problem 4.33 A dielectric cube of side a, centered at the origin, carries a "frozen-in" polarization $\mathbf{P} = k\mathbf{r}$, where k is a constant. Find all the bound charges, and check that they add up to zero.

Problem 4.34 The space between the plates of a parallel-plate capacitor is filled with dielectric material whose dielectric constant varies linearly from 1 at the bottom plate ($x = 0$) to 2 at the top plate ($x = d$). The capacitor is connected to a battery of voltage V. Find all the bound charge, and check that the total is zero.

Problem 4.35 A point charge q is imbedded at the center of a sphere of linear dielectric material (with susceptibility χ_e and radius R). Find the electric field, the polarization, and the bound charge densities, ρ_b and σ_b. What is the total bound charge on the surface? Where is the compensating negative bound charge located?

Problem 4.36 At the interface between one linear dielectric and another, the electric field lines bend (see Fig. 4.34). Show that

$$\tan\theta_2/\tan\theta_1 = \epsilon_2/\epsilon_1, \tag{4.68}$$

assuming there is no *free* charge at the boundary. [*Comment:* Eq. 4.68 is reminiscent of Snell's law in optics. Would a convex "lens" of dielectric material tend to "focus," or "defocus," the electric field?]

[21] This charming paradox was suggested by K. Brownstein.

[22] Interestingly, it *can* be done with *oscillating* fields. See K. T. McDonald, *Am. J. Phys.* **68**, 486 (2000).

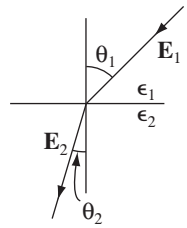


FIGURE 4.34

! **Problem 4.37** A point dipole $\mathbf{p}$ is imbedded at the center of a sphere of linear dielectric material (with radius R and dielectric constant ϵ_r). Find the electric potential inside and outside the sphere.

$$\left[\textit{Answer: } \frac{p\cos\theta}{4\pi\epsilon r^2}\left(1+2\frac{r^3}{R^3}\frac{(\epsilon_r-1)}{(\epsilon_r+2)}\right),\ (r\le R);\ \frac{p\cos\theta}{4\pi\epsilon_0 r^2}\left(\frac{3}{\epsilon_r+2}\right),\ (r\ge R)\right]$$

Problem 4.38 Prove the following uniqueness theorem: A volume $\mathcal{V}$ contains a specified free charge distribution, and various pieces of linear dielectric material, with the susceptibility of each one given. If the potential is specified on the boundaries $\mathcal{S}$ of $\mathcal{V}$ ($V = 0$ at infinity would be suitable) then the potential throughout $\mathcal{V}$ is uniquely determined. [*Hint:* Integrate $\boldsymbol{\nabla}\cdot(V_3\mathbf{D}_3)$ over $\mathcal{V}$.]

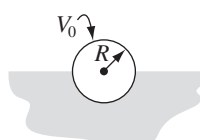


FIGURE 4.35

Problem 4.39 A conducting sphere at potential V_0 is half embedded in linear dielectric material of susceptibility χ_e, which occupies the region $z < 0$ (Fig. 4.35). *Claim:* the potential everywhere is exactly the same as it would have been in the absence of the dielectric! Check this claim, as follows:

(a) Write down the formula for the proposed potential $V(r)$, in terms of V_0, R, and r. Use it to determine the field, the polarization, the bound charge, and the free charge distribution on the sphere.

(b) Show that the resulting charge configuration would indeed produce the potential $V(r)$.

(c) Appeal to the uniqueness theorem in Prob. 4.38 to complete the argument.

(d) Could you solve the configurations in Fig. 4.36 with the same potential? If not, explain *why.*

FIGURE 4.36

Problem 4.40 According to Eq. 4.5, the force on a single dipole is $(\mathbf{p}\cdot\boldsymbol{\nabla})\mathbf{E}$, so the *net* force on a dielectric object is

$$\mathbf{F} = \int (\mathbf{P}\cdot\boldsymbol{\nabla})\mathbf{E}_{\text{ext}}\, d\tau. \tag{4.69}$$

[Here $\mathbf{E}_{\text{ext}}$ is the field of everything *except* the dielectric. You might assume that it wouldn't matter if you used the *total* field; after all, the dielectric can't exert a force on *itself.* However, because the field of the dielectric is discontinuous at the location of any bound surface charge, the derivative introduces a spurious delta function, and it is safest to stick with $\mathbf{E}_{\text{ext}}$.] Use Eq. 4.69 to determine the force on a tiny sphere, of radius R, composed of linear dielectric material of susceptibility χ_e, which is situated a distance s from a fine wire carrying a uniform line charge λ.

! **Problem 4.41** In a linear dielectric, the polarization is proportional to the field: $\mathbf{P} = \epsilon_0\chi_e\mathbf{E}$. If the material consists of atoms (or nonpolar molecules), the induced dipole moment of each one is likewise proportional to the field $\mathbf{p} = \alpha\mathbf{E}$. *Question:* What is the relation between the atomic polarizability α and the susceptibility χ_e?

Since $\mathbf{P}$ (the dipole moment per unit volume) is $\mathbf{p}$ (the dipole moment per atom) times N (the number of atoms per unit volume), $\mathbf{P} = N\mathbf{p} = N\alpha\mathbf{E}$, one's first inclination is to say that

$$\chi_e = \frac{N\alpha}{\epsilon_0}. \tag{4.70}$$

And in fact this is not far off, if the density is low. But closer inspection reveals a subtle problem, for the field $\mathbf{E}$ in Eq. 4.30 is the *total macroscopic* field in the medium, whereas the field in Eq. 4.1 is due to everything *except* the particular atom under consideration (polarizability was defined for an isolated atom subject to a specified external field); call this field $\mathbf{E}_{\text{else}}$. Imagine that the space allotted to each atom is a sphere of radius R, and show that

$$\mathbf{E} = \left(1 - \frac{N\alpha}{3\epsilon_0}\right)\mathbf{E}_{\text{else}}. \tag{4.71}$$

Use this to conclude that

$$\chi_e = \frac{N\alpha/\epsilon_0}{1 - N\alpha/3\epsilon_0},$$

or

$$\alpha = \frac{3\epsilon_0}{N}\left(\frac{\epsilon_r - 1}{\epsilon_r + 2}\right). \tag{4.72}$$

Equation 4.72 is known as the **Clausius-Mossotti** formula, or, in its application to optics, the **Lorentz-Lorenz** equation.

Problem 4.42 Check the Clausius-Mossotti relation (Eq. 4.72) for the gases listed in Table 4.1. (Dielectric constants are given in Table 4.2.) (The densities here are so small that Eqs. 4.70 and 4.72 are indistinguishable. For experimental data that confirm the Clausius-Mossotti correction term see, for instance, the first edition of Purcell's *Electricity and Magnetism,* Problem 9.28.)[23]

! **Problem 4.43** The Clausius-Mossotti equation (Prob. 4.41) tells you how to calculate the susceptibility of a *nonpolar* substance, in terms of the atomic polarizability α. The **Langevin equation** tells you how to calculate the susceptibility of a *polar* substance, in terms of the permanent molecular dipole moment p. Here's how it goes:

(a) The energy of a dipole in an external field $\mathbf{E}$ is $u = -\mathbf{p} \cdot \mathbf{E} = -pE\cos\theta$ (Eq. 4.6), where θ is the usual polar angle, if we orient the z axis along $\mathbf{E}$. Statistical mechanics says that for a material in equilibrium at absolute temperature T, the probability of a given molecule having energy u is proportional to the Boltzmann factor,

$$\exp(-u/kT).$$

The average energy of the dipoles is therefore

$$<u> = \frac{\int u e^{-(u/kT)}\, d\Omega}{\int e^{-(u/kT)}\, d\Omega},$$

where $d\Omega = \sin\theta\, d\theta\, d\phi$, and the integration is over all orientations ($\theta: 0 \to \pi$; $\phi: 0 \to 2\pi$). Use this to show that the polarization of a substance containing N molecules per unit volume is

$$P = Np[\coth(pE/kT) - (kT/pE)]. \tag{4.73}$$

That's the Langevin formula. Sketch P/Np as a function of pE/kT.

(b) Notice that for large fields/low temperatures, virtually *all* the molecules are lined up, and the material is *non*linear. Ordinarily, however, kT is much greater than pE. Show that in this régime the material *is* linear, and calculate its susceptibility, in terms of N, p, T, and k. Compute the susceptibility of water at 20°C, and compare the experimental value in Table 4.2. (The dipole moment of water is 6.1×10^{-30} C·m.) This is rather far off, because we have again neglected the distinction between $\mathbf{E}$ and $\mathbf{E}_{\text{else}}$. The agreement is better in low-density gases, for which the difference between $\mathbf{E}$ and $\mathbf{E}_{\text{else}}$ is negligible. Try it for water vapor at 100°C and 1 atm.

---

[23] E. M. Purcell, *Electricity and Magnetism* (Berkeley Physics Course, Vol. 2), (New York: McGraw-Hill, 1963).

CHAPTER

5 Magnetostatics

5.1 ■ THE LORENTZ FORCE LAW

5.1.1 ■ Magnetic Fields

Remember the basic problem of classical electrodynamics: We have a collection of charges $q_1, q_2, q_3, \ldots$ (the "source" charges), and we want to calculate the force they exert on some other charge Q (the "test" charge). (See Fig. 5.1.) According to the principle of superposition, it is sufficient to find the force of a *single* source charge—the total is then the vector sum of all the individual forces. Up to now, we have confined our attention to the simplest case, *electrostatics*, in which the source charge is *at rest* (though the test charge need not be). The time has come to consider the forces between charges *in motion*.

To give you some sense of what is in store, imagine that I set up the following demonstration: Two wires hang from the ceiling, a few centimeters apart; when I turn on a current, so that it passes up one wire and back down the other, the wires jump apart—they evidently repel one another (Fig. 5.2(a)). How do we explain this? You might suppose that the battery (or whatever drives the current) is actually charging up the wire, and that the force is simply due to the electrical repulsion of like charges. But this is incorrect. I could hold up a test charge near these wires, and there would be *no* force on it,[1] for the wires are in fact electrically neutral. (It's true that electrons are flowing down the line—that's what a current *is*—but there are just as many stationary plus charges as moving minus charges on any given segment.) Moreover, if I hook up my demonstration so as to make the current flow up *both* wires (Fig. 5.2(b)), they are found to *attract!*

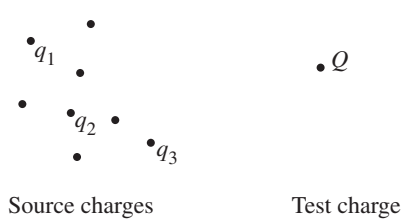
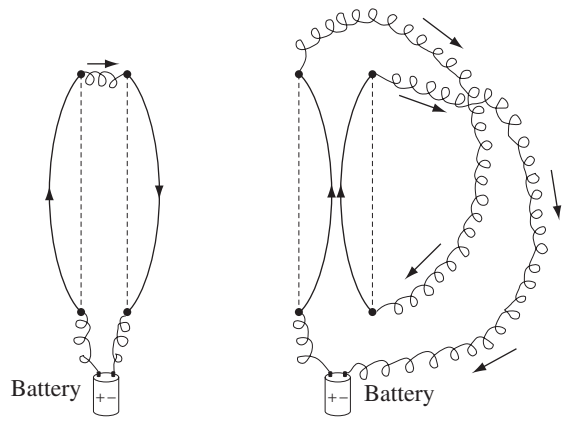


FIGURE 5.1

[1] This is not precisely true, as we shall see in Prob. 7.43.



(a) Currents in opposite directions repel.

(b) Currents in same directions attract.

FIGURE 5.2

Whatever force accounts for the attraction of parallel currents and the repulsion of antiparallel ones is *not* electrostatic in nature. It is our first encounter with a *magnetic* force. Whereas a *stationary* charge produces only an electric field **E** in the space around it, a *moving* charge generates, in addition, a magnetic field **B**. In fact, magnetic fields are a lot easier to detect, in practice—all you need is a Boy Scout compass. How these devices work is irrelevant at the moment; it is enough to know that the needle points in the direction of the local magnetic field. Ordinarily, this means *north*, in response to the earth's magnetic field, but in the laboratory, where typical fields may be hundreds of times stronger than that, the compass indicates the direction of whatever magnetic field is present.

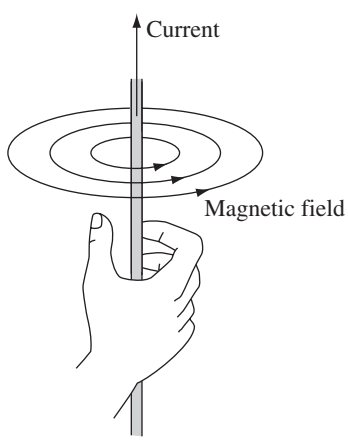


FIGURE 5.3

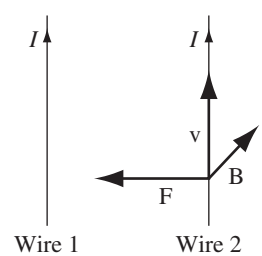


FIGURE 5.4

Now, if you hold up a tiny compass in the vicinity of a current-carrying wire, you quickly discover a very peculiar thing: The field does not point *toward* the wire, nor *away* from it, but rather it *circles around the wire*. In fact, if you grab the wire with your right hand—thumb in the direction of the current—your fingers curl around in the direction of the magnetic field (Fig. 5.3). How can such a field lead to a force of attraction on a nearby parallel current? At the second wire, the magnetic field points *into the page* (Fig. 5.4), the current is *upward*, and yet the resulting force is *to the left*! It's going to take a strange law to account for these directions.

5.1.2 ■ Magnetic Forces

In fact, this combination of directions is just right for a cross product: the magnetic force on a charge Q, moving with velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$, is[2]

$$\boxed{\mathbf{F}_{\text{mag}} = Q(\mathbf{v} \times \mathbf{B}).} \tag{5.1}$$

This is known as the **Lorentz force law**.[3] In the presence of both electric *and* magnetic fields, the net force on Q would be

$$\mathbf{F} = Q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})]. \tag{5.2}$$

I do not pretend to have *derived* Eq. 5.1, of course; it is a fundamental axiom of the theory, whose justification is to be found in experiments such as the one I described in Sect. 5.1.1.

Our main job from now on is to calculate the magnetic field $\mathbf{B}$ (and for that matter the electric field $\mathbf{E}$ as well; the rules are more complicated when the source charges are in motion). But before we proceed, it is worthwhile to take a closer look at the Lorentz force law itself; it is a peculiar law, and it leads to some truly bizarre particle trajectories.

Example 5.1. Cyclotron motion. The archtypical motion of a charged particle in a magnetic field is circular, with the magnetic force providing the centripetal acceleration. In Fig. 5.5, a uniform magnetic field points *into* the page; if the charge Q moves counterclockwise, with speed v, around a circle of radius R, the magnetic force points *inward*, and has a fixed magnitude QvB—just right to sustain uniform circular motion:

$$QvB = m\frac{v^2}{R}, \quad \text{or} \quad p = QBR, \tag{5.3}$$

[2] Since $\mathbf{F}$ and $\mathbf{v}$ are vectors, $\mathbf{B}$ is actually a *pseudo*vector.
[3] Actually, it is due to Oliver Heaviside.

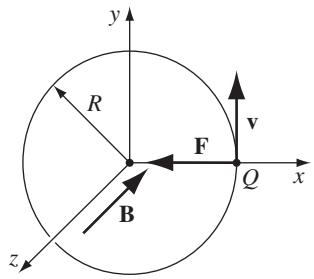


FIGURE 5.5

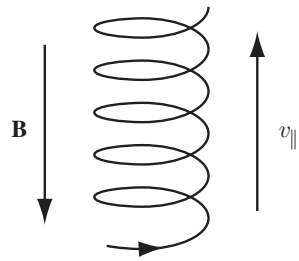


FIGURE 5.6

where m is the particle's mass and $p = mv$ is its momentum. Equation 5.3 is known as the **cyclotron formula** because it describes the motion of a particle in a cyclotron—the first of the modern particle accelerators. It also suggests a simple experimental technique for finding the momentum of a charged particle: send it through a region of known magnetic field, and measure the radius of its trajectory. This is in fact the standard means for determining the momenta of elementary particles.

I assumed that the charge moves in a plane perpendicular to **B**. If it starts out with some additional speed $v_\parallel$ *parallel* to **B**, this component of the motion is unaffected by the magnetic field, and the particle moves in a *helix* (Fig. 5.6). The radius is still given by Eq. 5.3, but the velocity in question is now the component perpendicular to **B**, $v_\perp$.

Example 5.2. Cycloid Motion. A more exotic trajectory occurs if we include a uniform electric field, at right angles to the magnetic one. Suppose, for instance, that **B** points in the x-direction, and **E** in the z-direction, as shown in Fig. 5.7. A positive charge is released from the origin; what path will it follow?

Solution

Let's think it through qualitatively, first. Initially, the particle is at rest, so the magnetic force is zero, and the electric field accelerates the charge in the z-direction. As it picks up speed, a magnetic force develops which, according to Eq. 5.1, pulls the charge around to the right. The faster it goes, the stronger F_{mag} becomes; eventually, it curves the particle back around towards the y axis. At this point the charge is moving *against* the electrical force, so it begins to slow down—the magnetic force then decreases, and the electrical force takes over, bringing the particle to rest at point a, in Fig. 5.7. There the entire process commences anew, carrying the particle over to point b, and so on.

Now let's do it quantitatively. There being no force in the x-direction, the position of the particle at any time t can be described by the vector $(0, y(t), z(t))$; the velocity is therefore

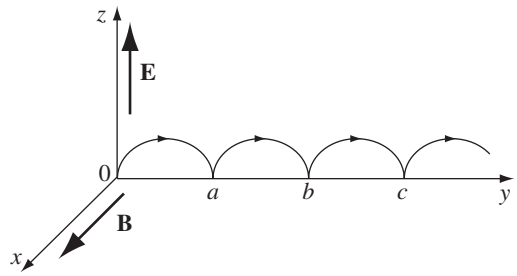


FIGURE 5.7

$$\mathbf{v} = (0, \dot{y}, \dot{z}),$$

where dots indicate time derivatives. Thus

$$\mathbf{v} \times \mathbf{B} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 0 & \dot{y} & \dot{z} \\ B & 0 & 0 \end{vmatrix} = B\dot{z}\,\hat{\mathbf{y}} - B\dot{y}\,\hat{\mathbf{z}},$$

and hence, applying Newton's second law,

$$\mathbf{F} = Q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = Q(E\,\hat{\mathbf{z}} + B\dot{z}\,\hat{\mathbf{y}} - B\dot{y}\,\hat{\mathbf{z}}) = m\mathbf{a} = m(\ddot{y}\,\hat{\mathbf{y}} + \ddot{z}\,\hat{\mathbf{z}}).$$

Or, treating the $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ components separately,

$$QB\dot{z} = m\ddot{y}, \qquad QE - QB\dot{y} = m\ddot{z}.$$

For convenience, let

$$\omega \equiv \frac{QB}{m}. \tag{5.4}$$

(This is the **cyclotron frequency**, at which the particle would revolve in the absence of any electric field.) Then the equations of motion take the form

$$\ddot{y} = \omega\dot{z}, \quad \ddot{z} = \omega\left(\frac{E}{B} - \dot{y}\right). \tag{5.5}$$

Their general solution[4] is

$$\left.\begin{aligned} y(t) &= C_1 \cos\omega t + C_2 \sin\omega t + (E/B)t + C_3, \\ z(t) &= C_2 \cos\omega t - C_1 \sin\omega t + C_4. \end{aligned}\right\} \tag{5.6}$$

[4]As coupled differential equations, they are easily solved by differentiating the first and using the second to eliminate $\ddot{z}$.

But the particle started from rest ($\dot{y}(0) = \dot{z}(0) = 0$), at the origin ($y(0) = z(0) = 0$); these four conditions determine the constants C_1, C_2, C_3, and C_4:

$$y(t) = \frac{E}{\omega B}(\omega t - \sin \omega t), \quad z(t) = \frac{E}{\omega B}(1 - \cos \omega t). \tag{5.7}$$

In this form, the answer is not terribly enlightening, but if we let

$$R \equiv \frac{E}{\omega B}, \tag{5.8}$$

and eliminate the sines and cosines by exploiting the trigonometric identity $\sin^2 \omega t + \cos^2 \omega t = 1$, we find that

$$(y - R\omega t)^2 + (z - R)^2 = R^2. \tag{5.9}$$

This is the formula for a *circle*, of radius R, whose center $(0, R\omega t, R)$ travels in the y-direction at a constant speed

$$u = \omega R = \frac{E}{B}. \tag{5.10}$$

The particle moves as though it were a spot on the rim of a wheel rolling along the y axis. The curve generated in this way is called a **cycloid**. Notice that the overall motion is *not* in the direction of **E**, as you might suppose, but perpendicular to it.

One implication of the Lorentz force law (Eq. 5.1) deserves special attention:

Magnetic forces do no work.

For if Q moves an amount $d\mathbf{l} = \mathbf{v}\,dt$, the work done is

$$dW_{\text{mag}} = \mathbf{F}_{\text{mag}} \cdot d\mathbf{l} = Q(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\,dt = 0. \tag{5.11}$$

This follows because $(\mathbf{v} \times \mathbf{B})$ is perpendicular to $\mathbf{v}$, so $(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = 0$. Magnetic forces may alter the *direction* in which a particle moves, but they cannot speed it up or slow it down. The fact that magnetic forces do no work is an elementary and direct consequence of the Lorentz force law, but there are many situations in which it *appears* so manifestly false that one's confidence is bound to waver. When a magnetic crane lifts the carcass of a junked car, for instance, *something* is obviously doing work, and it seems perverse to deny that the magnetic force is responsible. Well, perverse or not, deny it we must, and it can be a very subtle matter to figure out who *does* deserve the credit in such circumstances. We'll see a cute example in the next section, but the full story will have to await Chapter 8.

Problem 5.1 A particle of charge q enters a region of uniform magnetic field **B** (pointing *into* the page). The field deflects the particle a distance d above the original line of flight, as shown in Fig. 5.8. Is the charge positive or negative? In terms of a, d, B and q, find the momentum of the particle.

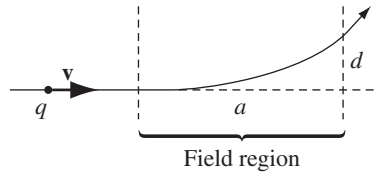


FIGURE 5.8

Problem 5.2 Find and sketch the trajectory of the particle in Ex. 5.2, if it starts at the origin with velocity

(a) $\mathbf{v}(0) = (E/B)\hat{\mathbf{y}}$,

(b) $\mathbf{v}(0) = (E/2B)\hat{\mathbf{y}}$,

(c) $\mathbf{v}(0) = (E/B)(\hat{\mathbf{y}} + \hat{\mathbf{z}})$.

Problem 5.3 In 1897, J. J. Thomson "discovered" the electron by measuring the charge-to-mass ratio of "cathode rays" (actually, streams of electrons, with charge q and mass m) as follows:

(a) First he passed the beam through uniform crossed electric and magnetic fields **E** and **B** (mutually perpendicular, and both of them perpendicular to the beam), and adjusted the electric field until he got zero deflection. What, then, was the speed of the particles (in terms of E and B)?

(b) Then he turned off the electric field, and measured the radius of curvature, R, of the beam, as deflected by the magnetic field alone. In terms of E, B, and R, what is the charge-to-mass ratio (q/m) of the particles?

5.1.3 ■ Currents

The **current** in a wire is the *charge per unit time* passing a given point. By definition, negative charges moving to the left count the same as positive ones to the right. This conveniently reflects the *physical* fact that almost all phenomena involving moving charges depend on the *product* of charge and velocity—if you reverse the signs of q *and* $\mathbf{v}$, you get the same answer, so it doesn't really matter which you have. (The Lorentz force law is a case in point; the Hall effect (Prob. 5.41) is a notorious exception.) In practice, it is ordinarily the negatively charged electrons that do the moving—in the direction *opposite* to the electric current. To avoid the petty complications this entails, I shall often pretend it's the positive charges that move, as in fact everyone assumed they did for a century or so after Benjamin Franklin established his unfortunate convention.[5] Current is measured in coulombs-per-second, or **amperes** (A):

$$1\ \text{A} = 1\ \text{C/s}. \tag{5.12}$$

[5]If we called the electron plus and the proton minus, the problem would never arise. In the context of Franklin's experiments with cat's fur and glass rods, the choice was completely arbitrary.

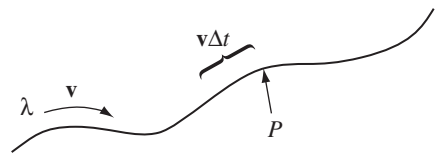


FIGURE 5.9

A line charge λ traveling down a wire at speed v (Fig. 5.9) constitutes a current

$$I = \lambda v, \tag{5.13}$$

because a segment of length $v\Delta t$, carrying charge $\lambda v \Delta t$, passes point P in a time interval Δt. Current is actually a *vector:*

$$\mathbf{I} = \lambda \mathbf{v}. \tag{5.14}$$

Because the path of the flow is dictated by the shape of the wire, one doesn't ordinarily bother to display the direction of $\mathbf{I}$ explicitly,[6] but when it comes to surface and volume currents we cannot afford to be so casual, and for the sake of notational consistency it is a good idea to acknowledge the vectorial character of currents right from the start. A neutral wire, of course, contains as many stationary positive charges as mobile negative ones. The former do not contribute to the current—the charge density λ in Eq. 5.13 refers only to the *moving* charges. In the unusual situation where *both* types move, $\mathbf{I} = \lambda_+ \mathbf{v}_+ + \lambda_- \mathbf{v}_-$.

The magnetic force on a segment of current-carrying wire is

$$\mathbf{F}_{\text{mag}} = \int (\mathbf{v} \times \mathbf{B})\, dq = \int (\mathbf{v} \times \mathbf{B}) \lambda\, dl = \int (\mathbf{I} \times \mathbf{B})\, dl. \tag{5.15}$$

Inasmuch as $\mathbf{I}$ and $d\mathbf{l}$ both point in the same direction, we can just as well write this as

$$\boxed{\mathbf{F}_{\text{mag}} = \int I\, (d\mathbf{l} \times \mathbf{B}).} \tag{5.16}$$

Typically, the current is constant (in magnitude) along the wire, and in that case I comes outside the integral:

$$\mathbf{F}_{\text{mag}} = I \int (d\mathbf{l} \times \mathbf{B}). \tag{5.17}$$

Example 5.3. A rectangular loop of wire, supporting a mass m, hangs vertically with one end in a uniform magnetic field $\mathbf{B}$, which points into the page in the shaded region of Fig. 5.10. For what current I, in the loop, would the magnetic force upward exactly balance the gravitational force downward?

[6]For the same reason, if you are describing a locomotive constrained to move along a specified track, you would probably speak of its *speed*, rather than its velocity.

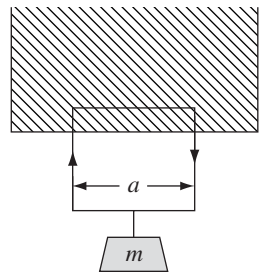


FIGURE 5.10

Solution
First of all, the current must circulate clockwise, in order for ($\mathbf{I} \times \mathbf{B}$) in the horizontal segment to point upward. The force is

$$F_{\text{mag}} = IBa,$$

where a is the width of the loop. (The magnetic forces on the two vertical segments cancel.) For F_{mag} to balance the weight (mg), we must therefore have

$$I = \frac{mg}{Ba}. \tag{5.18}$$

The weight just *hangs* there, suspended in mid-air!

What happens if we now *increase* the current? Then the upward magnetic force *exceeds* the downward force of gravity, and the loop rises, lifting the weight. *Somebody's* doing work, and it sure looks as though the magnetic force is responsible. Indeed, one is tempted to write

$$W_{\text{mag}} = F_{\text{mag}}h = IBah, \tag{5.19}$$

where h is the distance the loop rises. But we know that magnetic forces *never* do work. What's going on here?

Well, when the loop starts to rise, the charges in the wire are no longer moving horizontally—their velocity now acquires an upward component u, the speed of the loop (Fig. 5.11), in addition to the horizontal component w associated with the current ($I = \lambda w$). The magnetic force, which is always perpendicular to the velocity, no longer points straight up, but tilts back. It is perpendicular to the *net* displacement of the charge (which is in the direction of $\mathbf{v}$), and therefore *it does no work on* q. It does have a vertical *component* (qwB); indeed, the net vertical force on all the charge (λa) in the upper segment of the loop is

$$F_{\text{vert}} = \lambda awB = IBa \tag{5.20}$$

(as before); but now it also has a *horizontal* component (quB), which opposes the flow of current. Whoever is in charge of maintaining that current, therefore, must now *push* those charges along, against the backward component of the magnetic force.

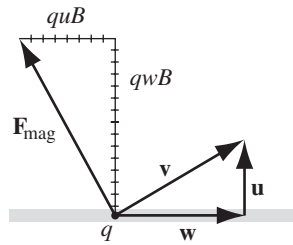


FIGURE 5.11

The total horizontal force on the top segment is

$$F_{\text{horiz}} = \lambda a u B. \tag{5.21}$$

In a time dt, the charges move a (horizontal) distance $w\,dt$, so the work done by this agency (presumably a battery or a generator) is

$$W_{\text{battery}} = \lambda a B \int u w\,dt = I B a h,$$

which is precisely what we naïvely attributed to the *magnetic* force in Eq. 5.19. Was work done in this process? Absolutely! Who did it? The battery! What, then, was the role of the magnetic force? Well, it *redirected* the horizontal force of the battery into the *vertical* motion of the loop and the weight.[7]

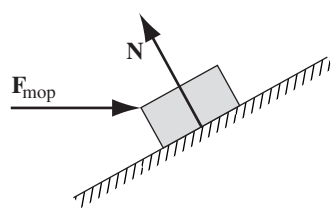


FIGURE 5.12

It may help to consider a mechanical analogy. Imagine you're sliding a trunk up a frictionless ramp, by pushing on it horizontally with a mop (Fig. 5.12). The normal force (**N**) does no work, because it is perpendicular to the displacement. But it *does* have a vertical component (which in fact is what lifts the trunk), and a (backward) horizontal component (which you have to overcome by pushing on the mop). Who is doing the work here? *You* are, obviously—and yet your *force* (which is purely horizontal) is not (at least, not directly) what lifts the box. The

[7]If you like, the *vertical* component of $\mathbf{F}_{\text{mag}}$ does work lifting the car, but the *horizontal* component does equal *negative* work opposing the current. However you look at it, the *net* work done by the magnetic force is *zero*.

normal force plays the same passive (but crucial) role as the magnetic force in Ex. 5.3: while doing no work itself, it *redirects* the efforts of the active agent (you, or the battery, as the case may be), from horizontal to vertical.

When charge flows over a *surface*, we describe it by the **surface current density**, **K**, defined as follows: Consider a "ribbon" of infinitesimal width $dl_\perp$, running parallel to the flow (Fig. 5.13). If the current in this ribbon is $d\mathbf{I}$, the surface current density is

$$\mathbf{K} \equiv \frac{d\mathbf{I}}{dl_\perp}. \tag{5.22}$$

In words, K is the *current per unit width*. In particular, if the (mobile) surface charge density is σ and its velocity is $\mathbf{v}$, then

$$\mathbf{K} = \sigma \mathbf{v}. \tag{5.23}$$

In general, **K** will vary from point to point over the surface, reflecting variations in σ and/or $\mathbf{v}$. The magnetic force on the surface current is

$$\mathbf{F}_{\text{mag}} = \int (\mathbf{v} \times \mathbf{B}) \sigma \, da = \int (\mathbf{K} \times \mathbf{B}) \, da. \tag{5.24}$$

Caveat: Just as **E** suffers a discontinuity at a surface *charge*, so **B** is discontinuous at a surface *current*. In Eq. 5.24, you must be careful to use the *average* field, just as we did in Sect. 2.5.3.

When the flow of charge is distributed throughout a three-dimensional region, we describe it by the **volume current density**, **J**, defined as follows: Consider a "tube" of infinitesimal cross section $da_\perp$, running parallel to the flow (Fig. 5.14). If the current in this tube is $d\mathbf{I}$, the volume current density is

$$\mathbf{J} \equiv \frac{d\mathbf{I}}{da_\perp}. \tag{5.25}$$

In words, J is the *current per unit area*. If the (mobile) volume charge density is ρ and the velocity is $\mathbf{v}$, then

$$\mathbf{J} = \rho \mathbf{v}. \tag{5.26}$$

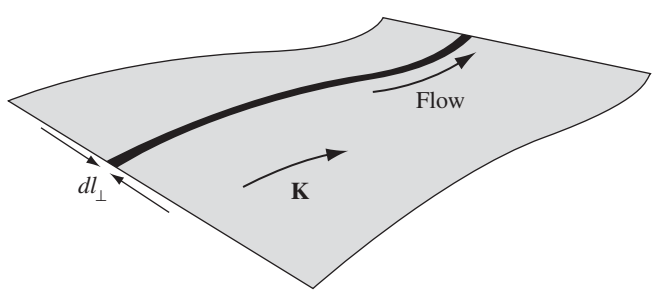


FIGURE 5.13

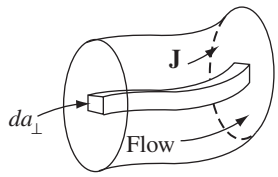


FIGURE 5.14

The magnetic force on a volume current is therefore

$$\mathbf{F}_{\text{mag}} = \int (\mathbf{v} \times \mathbf{B})\rho \, d\tau = \int (\mathbf{J} \times \mathbf{B}) \, d\tau. \tag{5.27}$$

Example 5.4.

(a) A current I is uniformly distributed over a wire of circular cross section, with radius a (Fig. 5.15). Find the volume current density J.

Solution
The area (perpendicular to the flow) is πa^2, so

$$J = \frac{I}{\pi a^2}.$$

This was trivial because the current density was uniform.

(b) Suppose the current density in the wire is proportional to the distance from the axis,

$$J = ks$$

(for some constant k). Find the total current in the wire.

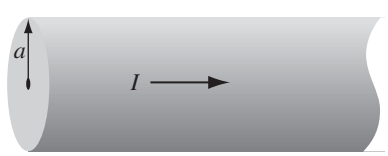


FIGURE 5.15

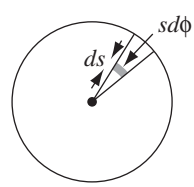


FIGURE 5.16

Solution
Because J varies with s, we must *integrate* Eq. 5.25. The current through the shaded patch (Fig. 5.16) is $J\,da_\perp$, and $da_\perp = s\,ds\,d\phi$. So

$$I = \int (ks)(s\,ds\,d\phi) = 2\pi k \int_0^a s^2\,ds = \frac{2\pi k a^3}{3}.$$

According to Eq. 5.25, the total current crossing a surface $\mathcal{S}$ can be written as

$$I = \int_{\mathcal{S}} J\, da_\perp = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{a}. \tag{5.28}$$

(The dot product serves neatly to pick out the appropriate component of $d\mathbf{a}$.) In particular, the charge per unit time leaving a volume $\mathcal{V}$ is

$$\oint_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{a} = \int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \mathbf{J})\, d\tau.$$

Because charge is conserved, whatever flows out through the surface must come at the expense of what remains inside:

$$\int_{\mathcal{V}} (\boldsymbol{\nabla} \cdot \mathbf{J})\, d\tau = -\frac{d}{dt} \int_{\mathcal{V}} \rho\, d\tau = -\int_{\mathcal{V}} \left(\frac{\partial \rho}{\partial t} \right) d\tau.$$

(The minus sign reflects the fact that an *outward* flow *decreases* the charge left in $\mathcal{V}$.) Since this applies to *any* volume, we conclude that

$$\boxed{\boldsymbol{\nabla} \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}.} \tag{5.29}$$

This is the precise mathematical statement of local charge conservation; it is called the **continuity equation**.

For future reference, let me summarize the "dictionary" we have implicitly developed for translating equations into the forms appropriate to point, line, surface, and volume currents:

$$\sum_{i=1}^{n} (\quad) q_i \mathbf{v}_i \sim \int_{\text{line}} (\quad) \mathbf{I}\, dl \sim \int_{\text{surface}} (\quad) \mathbf{K}\, da \sim \int_{\text{volume}} (\quad) \mathbf{J}\, d\tau. \tag{5.30}$$

This correspondence, which is analogous to $q \sim \lambda\, dl \sim \sigma\, da \sim \rho\, d\tau$ for the various charge distributions, generates Eqs. 5.15, 5.24, and 5.27 from the original Lorentz force law (5.1).

Problem 5.4 Suppose that the magnetic field in some region has the form

$$\mathbf{B} = kz\, \hat{\mathbf{x}}$$

(where k is a constant). Find the force on a square loop (side a), lying in the yz plane and centered at the origin, if it carries a current I, flowing counterclockwise, when you look down the x axis.

Problem 5.5 A current I flows down a wire of radius a.

(a) If it is uniformly distributed over the surface, what is the surface current density K?

(b) If it is distributed in such a way that the volume current density is inversely proportional to the distance from the axis, what is $J(s)$?

Problem 5.6

(a) A phonograph record carries a uniform density of "static electricity" σ. If it rotates at angular velocity ω, what is the surface current density K at a distance r from the center?

(b) A uniformly charged solid sphere, of radius R and total charge Q, is centered at the origin and spinning at a constant angular velocity ω about the z axis. Find the current density $\mathbf{J}$ at any point (r, θ, ϕ) within the sphere.

Problem 5.7 For a configuration of charges and currents confined within a volume $\mathcal{V}$, show that

$$\int_{\mathcal{V}} \mathbf{J}\, d\tau = d\mathbf{p}/dt, \tag{5.31}$$

where $\mathbf{p}$ is the total dipole moment. [*Hint:* evaluate $\int_{\mathcal{V}} \boldsymbol{\nabla} \cdot (x\mathbf{J})\, d\tau$.]

5.2 ■ THE BIOT-SAVART LAW

5.2.1 ■ Steady Currents

Stationary charges produce electric fields that are constant in time; hence the term **electrostatics**.[8] *Steady currents* produce magnetic fields that are constant in time; the theory of steady currents is called **magnetostatics**.

Stationary charges	$\Rightarrow$	**constant electric fields: electrostatics.**
Steady currents	$\Rightarrow$	**constant magnetic fields: magnetostatics.**

By **steady current** I mean a continuous flow that has been going on forever, without change and without charge piling up anywhere. (Some people call them "stationary currents"; to my ear, that's a contradiction in terms.) Formally, electro/magnetostatics is the régime

$$\frac{\partial \rho}{\partial t} = 0, \quad \frac{\partial \mathbf{J}}{\partial t} = \mathbf{0}, \tag{5.32}$$

at all places and all times. Of course, there's no such thing in practice as a *truly* steady current, any more than there is a *truly* stationary charge. In this sense, both electrostatics and magnetostatics describe artificial worlds that exist only in textbooks. However, they represent suitable *approximations* as long as the actual fluctuations are remote, or gradual—in fact, for most purposes magnetostatics applies very well to household currents, which alternate 120 times a second!

[8] Actually, it is not necessary that the charges be stationary, but only that the charge *density* at each point be constant. For example, the sphere in Prob. 5.6(b) produces an electrostatic field $1/4\pi\epsilon_0(Q/r^2)\hat{\mathbf{r}}$, even though it is rotating, because ρ does not depend on t.

Notice that a moving *point* charge *cannot possibly constitute a steady current*. If it's here one instant, it's gone the next. This may seem like a minor thing to *you*, but it's a major headache for *me*. I developed each topic in electrostatics by starting out with the simple case of a point charge at rest; then I generalized to an arbitrary charge distribution by invoking the superposition principle. This approach is not open to us in magnetostatics because a moving point charge does not produce a static field in the first place. We are *forced* to deal with extended current distributions right from the start, and, as a result, the arguments are bound to be more cumbersome.

When a steady current flows in a wire, its magnitude I must be the same all along the line; otherwise, charge would be piling up somewhere, and it wouldn't be a steady current. More generally, since $\partial\rho/\partial t = 0$ in magnetostatics, the continuity equation (5.29) becomes

$$\nabla \cdot \mathbf{J} = 0. \tag{5.33}$$

5.2.2 ■ The Magnetic Field of a Steady Current

The magnetic field of a steady line current is given by the **Biot-Savart law**:

$$\boxed{\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{I} \times \hat{\boldsymbol{\imath}}}{\imath^2}\, dl' = \frac{\mu_0}{4\pi} I \int \frac{d\mathbf{l}' \times \hat{\boldsymbol{\imath}}}{\imath^2}.} \tag{5.34}$$

The integration is along the current path, in the direction of the flow; $d\mathbf{l}'$ is an element of length along the wire, and $\boldsymbol{\imath}$, as always, is the vector from the source to the point $\mathbf{r}$ (Fig. 5.17). The constant μ_0 is called the **permeability of free space**:[9]

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{N/A^2}. \tag{5.35}$$

These units are such that $\mathbf{B}$ itself comes out in newtons per ampere-meter (as required by the Lorentz force law), or **teslas** (T):[10]

$$1\ \mathrm{T} = 1\ \mathrm{N/(A \cdot m)}. \tag{5.36}$$

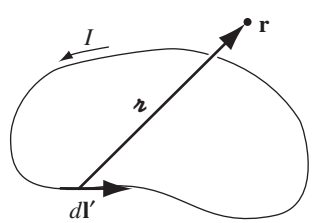


FIGURE 5.17

[9]This is an exact number, not an empirical constant. It serves (via Eq. 5.40) to define the ampere, and the ampere in turn defines the coulomb.

[10]For some reason, in this one case the cgs unit (the **gauss**) is more commonly used than the SI unit: 1 tesla = 10^4 gauss. The earth's magnetic field is about half a gauss; a fairly strong laboratory magnetic field is, say, 10,000 gauss.

As the starting point for magnetostatics, the Biot-Savart law plays a role analogous to Coulomb's law in electrostatics. Indeed, the $1/\imath^2$ dependence is common to both laws.

Example 5.5. Find the magnetic field a distance s from a long straight wire carrying a steady current I (Fig. 5.18).

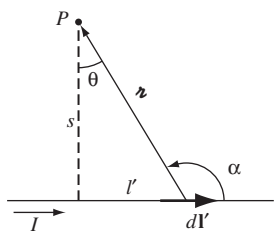


FIGURE 5.18

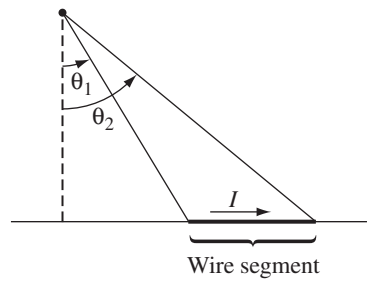


FIGURE 5.19

Solution
In the diagram, $(d\mathbf{l}' \times \hat{\boldsymbol{\imath}})$ points *out* of the page, and has the magnitude

$$dl' \sin\alpha = dl' \cos\theta.$$

Also, $l' = s\tan\theta$, so

$$dl' = \frac{s}{\cos^2\theta}\, d\theta,$$

and $s = \imath\cos\theta$, so

$$\frac{1}{\imath^2} = \frac{\cos^2\theta}{s^2}.$$

Thus

$$B = \frac{\mu_0 I}{4\pi}\int_{\theta_1}^{\theta_2}\left(\frac{\cos^2\theta}{s^2}\right)\left(\frac{s}{\cos^2\theta}\right)\cos\theta\, d\theta$$

$$= \frac{\mu_0 I}{4\pi s}\int_{\theta_1}^{\theta_2}\cos\theta\, d\theta = \frac{\mu_0 I}{4\pi s}(\sin\theta_2 - \sin\theta_1). \tag{5.37}$$

Equation 5.37 gives the field of any straight segment of wire, in terms of the initial and final angles θ_1 and θ_2 (Fig. 5.19). Of course, a finite segment by itself

could never support a steady current (where would the charge go when it got to the end?), but it might be a *piece* of some closed circuit, and Eq. 5.37 would then represent its contribution to the total field. In the case of an *infinite* wire, $\theta_1 = -\pi/2$ and $\theta_2 = \pi/2$, so we obtain

$$B = \frac{\mu_0 I}{2\pi s}. \tag{5.38}$$

Notice that the field is inversely proportional to the distance from the wire—just like the electric field of an infinite line charge. In the region *below* the wire, **B** points *into* the page, and in general, it "circles around" the wire, in accordance with the right-hand rule (Fig. 5.3):

$$\mathbf{B} = \frac{\mu_0 I}{2\pi s}\,\hat{\boldsymbol{\phi}}. \tag{5.39}$$

As an application, let's find the force of attraction between two long, parallel wires a distance d apart, carrying currents I_1 and I_2 (Fig. 5.20). The field at (2) due to (1) is

$$B = \frac{\mu_0 I_1}{2\pi d},$$

and it points into the page. The Lorentz force law (in the form appropriate to line currents, Eq. 5.17) predicts a force directed towards (1), of magnitude

$$F = I_2 \left(\frac{\mu_0 I_1}{2\pi d}\right) \int dl.$$

The *total* force, not surprisingly, is infinite, but the force per unit length is

$$f = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{d}. \tag{5.40}$$

If the currents are antiparallel (one up, one down), the force is repulsive—consistent again with the qualitative observations in Sect. 5.1.1.

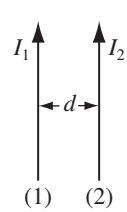


FIGURE 5.20

Example 5.6. Find the magnetic field a distance z above the center of a circular loop of radius R, which carries a steady current I (Fig. 5.21).

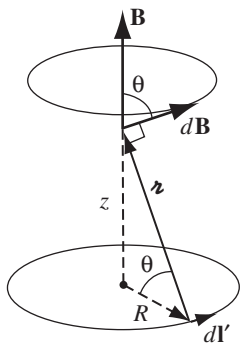


FIGURE 5.21

Solution
The field $d\mathbf{B}$ attributable to the segment $d\mathbf{l}'$ points as shown. As we integrate $d\mathbf{l}'$ around the loop, $d\mathbf{B}$ sweeps out a cone. The horizontal components cancel, and the vertical components combine, to give

$$B(z) = \frac{\mu_0}{4\pi} I \int \frac{dl'}{\imath^2} \cos\theta.$$

(Notice that $d\mathbf{l}'$ and $\boldsymbol{\imath}$ are perpendicular, in this case; the factor of $\cos\theta$ projects out the vertical component.) Now, $\cos\theta$ and $\imath^2$ are constants, and $\int dl'$ is simply the circumference, $2\pi R$, so

$$B(z) = \frac{\mu_0 I}{4\pi} \left(\frac{\cos\theta}{\imath^2} \right) 2\pi R = \frac{\mu_0 I}{2} \frac{R^2}{(R^2 + z^2)^{3/2}}. \tag{5.41}$$

For surface and volume currents, the Biot-Savart law becomes

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{K}(\mathbf{r}') \times \hat{\boldsymbol{\imath}}}{\imath^2} da' \qquad \text{and} \qquad \mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}') \times \hat{\boldsymbol{\imath}}}{\imath^2} d\tau', \tag{5.42}$$

respectively. You might be tempted to write down the corresponding formula for a moving point charge, using the "dictionary" (Eq. 5.30):

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{q\mathbf{v} \times \hat{\boldsymbol{\imath}}}{\imath^2}, \tag{5.43}$$

but this is simply *wrong*.[11] As I mentioned earlier, a point charge does not constitute a steady current, and the Biot-Savart law, which only holds for steady currents, does *not* correctly determine its field.

The superposition principle applies to magnetic fields just as it does to electric fields: if you have a *collection* of source currents, the net field is the (vector) sum of the fields due to each of them taken separately.

Problem 5.8

(a) Find the magnetic field at the center of a square loop, which carries a steady current I. Let R be the distance from center to side (Fig. 5.22).

(b) Find the field at the center of a regular n-sided polygon, carrying a steady current I. Again, let R be the distance from the center to any side.

(c) Check that your formula reduces to the field at the center of a circular loop, in the limit $n \to \infty$.

Problem 5.9 Find the magnetic field at point P for each of the steady current configurations shown in Fig. 5.23.

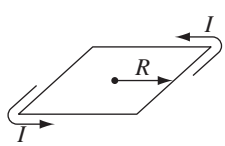


FIGURE 5.22

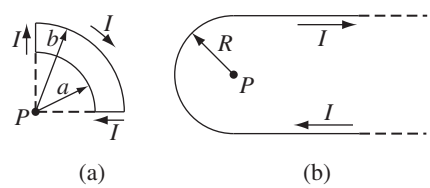


FIGURE 5.23

Problem 5.10

(a) Find the force on a square loop placed as shown in Fig. 5.24(a), near an infinite straight wire. Both the loop and the wire carry a steady current I.

(b) Find the force on the triangular loop in Fig. 5.24(b).

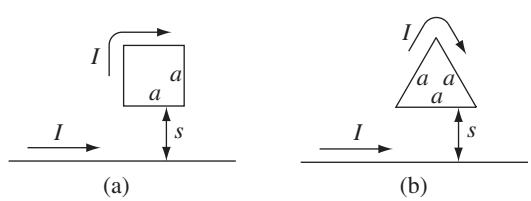


FIGURE 5.24

[11] I say this loud and clear to emphasize the point of principle; actually, Eq. 5.43 is *approximately* right for nonrelativistic charges ($v \ll c$), under conditions where retardation can be neglected (see Ex. 10.4).

Problem 5.11 Find the magnetic field at point P on the axis of a tightly wound **solenoid** (helical coil) consisting of n turns per unit length wrapped around a cylindrical tube of radius a and carrying current I (Fig. 5.25). Express your answer in terms of θ_1 and θ_2 (it's easiest that way). Consider the turns to be essentially circular, and use the result of Ex. 5.6. What is the field on the axis of an *infinite* solenoid (infinite in both directions)?

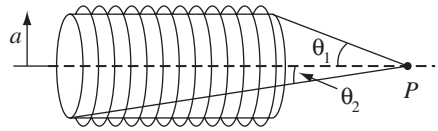


FIGURE 5.25

Problem 5.12 Use the result of Ex. 5.6 to calculate the magnetic field at the center of a uniformly charged spherical shell, of radius R and total charge Q, spinning at constant angular velocity ω.

Problem 5.13 Suppose you have two infinite straight line charges λ, a distance d apart, moving along at a constant speed v (Fig. 5.26). How great would v have to be in order for the magnetic attraction to balance the electrical repulsion? Work out the actual number. Is this a reasonable sort of speed?[12]

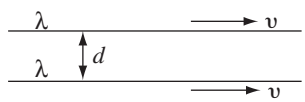


FIGURE 5.26

5.3 ■ THE DIVERGENCE AND CURL OF B

5.3.1 ■ Straight-Line Currents

The magnetic field of an infinite straight wire is shown in Fig. 5.27 (the current is coming *out* of the page). At a glance, it is clear that this field has a nonzero curl (something you'll never see in an *electro*static field); let's calculate it.

According to Eq. 5.38, the integral of **B** around a circular path of radius s, centered at the wire, is

$$\oint \mathbf{B} \cdot d\mathbf{l} = \oint \frac{\mu_0 I}{2\pi s}\, dl = \frac{\mu_0 I}{2\pi s} \oint dl = \mu_0 I.$$

Notice that the answer is independent of s; that's because B *de*creases at the same rate as the circumference *in*creases. In fact, it doesn't have to be a circle; *any* old

[12]If you've studied special relativity, you may be tempted to look for complexities in this problem that are not really there—λ and v are both measured *in the laboratory frame*, and this is *ordinary electrostatics*.

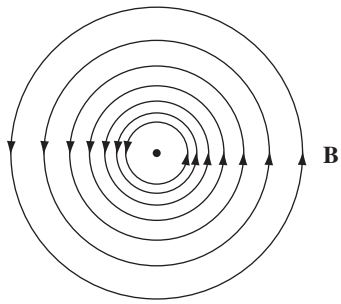


FIGURE 5.27

loop that encloses the wire would give the same answer. For if we use cylindrical coordinates (s, ϕ, z), with the current flowing along the z axis, $\mathbf{B} = (\mu_0 I/2\pi s)\hat{\boldsymbol{\phi}}$ and $d\mathbf{l} = ds\,\hat{\mathbf{s}} + s\,d\phi\,\hat{\boldsymbol{\phi}} + dz\,\hat{\mathbf{z}}$, so

$$\oint \mathbf{B} \cdot d\mathbf{l} = \frac{\mu_0 I}{2\pi} \oint \frac{1}{s}\, s\, d\phi = \frac{\mu_0 I}{2\pi} \int_0^{2\pi} d\phi = \mu_0 I.$$

This assumes the loop encircles the wire exactly once; if it went around twice, then ϕ would run from 0 to 4π, and if it didn't enclose the wire at all, then ϕ would go from ϕ_1 to ϕ_2 and back again, with $\int d\phi = 0$ (Fig. 5.28).

Now suppose we have a *bundle* of straight wires. Each wire that passes through our loop contributes $\mu_0 I$, and those outside contribute nothing (Fig. 5.29). The line integral will then be

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 I_{\text{enc}}, \tag{5.44}$$

where I_{enc} stands for the total current enclosed by the integration path. If the flow of charge is represented by a volume current density $\mathbf{J}$, the enclosed current is

$$I_{\text{enc}} = \int \mathbf{J} \cdot d\mathbf{a}, \tag{5.45}$$

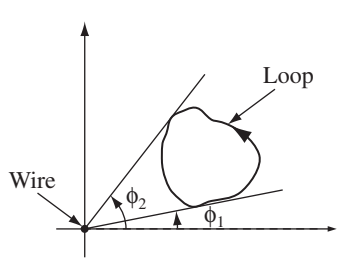


FIGURE 5.28

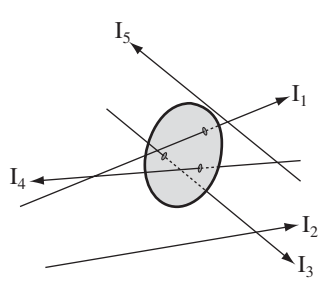


FIGURE 5.29

with the integral taken over any surface bounded by the loop. Applying Stokes' theorem to Eq. 5.44, then,

$$\int (\nabla \times \mathbf{B}) \cdot d\mathbf{a} = \mu_0 \int \mathbf{J} \cdot d\mathbf{a},$$

and hence

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \tag{5.46}$$

With minimal labor, we have actually obtained the general formula for the curl of **B**. But our derivation is seriously flawed by the restriction to infinite straight line currents (and combinations thereof). Most current configurations *cannot* be constructed out of infinite straight wires, and we have no right to assume that Eq. 5.46 applies to them. So the next section is devoted to the formal derivation of the divergence and curl of **B**, starting from the Biot-Savart law itself.

5.3.2 ■ The Divergence and Curl of B

The Biot-Savart law for the general case of a volume current reads

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(\mathbf{r}') \times \hat{\boldsymbol{\imath}}}{\imath^2} \, d\tau'. \tag{5.47}$$

This formula gives the magnetic field at a point $\mathbf{r} = (x, y, z)$ in terms of an integral over the current distribution $\mathbf{J}(x', y', z')$ (Fig. 5.30). It is best to be absolutely explicit at this stage:

$$\mathbf{B} \text{ is a function of } (x, y, z),$$

$$\mathbf{J} \text{ is a function of } (x', y', z'),$$

$$\boldsymbol{\imath} = (x - x')\,\hat{\mathbf{x}} + (y - y')\,\hat{\mathbf{y}} + (z - z')\,\hat{\mathbf{z}},$$

$$d\tau' = dx'\, dy'\, dz'.$$

The integration is over the *primed* coordinates; the divergence and the curl of **B** are with respect to the *unprimed* coordinates.

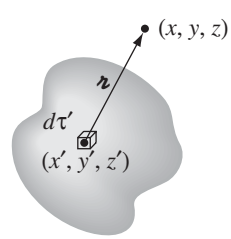


FIGURE 5.30

Applying the divergence to Eq. 5.47, we obtain:

$$\boldsymbol{\nabla} \cdot \mathbf{B} = \frac{\mu_0}{4\pi} \int \boldsymbol{\nabla} \cdot \left(\mathbf{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) d\tau'. \tag{5.48}$$

Invoking product rule number 6,

$$\boldsymbol{\nabla} \cdot \left(\mathbf{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \cdot (\boldsymbol{\nabla} \times \mathbf{J}) - \mathbf{J} \cdot \left(\boldsymbol{\nabla} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right). \tag{5.49}$$

But $\boldsymbol{\nabla} \times \mathbf{J} = 0$, because $\mathbf{J}$ doesn't depend on the unprimed variables, while $\boldsymbol{\nabla} \times (\hat{\boldsymbol{\imath}}/\imath^2) = \mathbf{0}$ (Prob. 1.63), so

$$\boxed{\boldsymbol{\nabla} \cdot \mathbf{B} = 0.} \tag{5.50}$$

Evidently the *divergence* of the magnetic field is *zero*.

Applying the curl to Eq. 5.47, we obtain:

$$\boldsymbol{\nabla} \times \mathbf{B} = \frac{\mu_0}{4\pi} \int \boldsymbol{\nabla} \times \left(\mathbf{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) d\tau'. \tag{5.51}$$

Again, our strategy is to expand the integrand, using the appropriate product rule—in this case number 8:

$$\boldsymbol{\nabla} \times \left(\mathbf{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = \mathbf{J} \left(\boldsymbol{\nabla} \cdot \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) - (\mathbf{J} \cdot \boldsymbol{\nabla}) \frac{\hat{\boldsymbol{\imath}}}{\imath^2}. \tag{5.52}$$

(I have dropped terms involving derivatives of $\mathbf{J}$, because $\mathbf{J}$ does not depend on x, y, z.) The second term integrates to zero, as we'll see in the next paragraph. The first term involves the divergence we were at pains to calculate in Chapter 1 (Eq. 1.100):

$$\boldsymbol{\nabla} \cdot \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = 4\pi \delta^3(\boldsymbol{\imath}). \tag{5.53}$$

Thus

$$\boldsymbol{\nabla} \times \mathbf{B} = \frac{\mu_0}{4\pi} \int \mathbf{J}(\mathbf{r}') 4\pi \delta^3(\mathbf{r} - \mathbf{r}')\, d\tau' = \mu_0 \mathbf{J}(\mathbf{r}),$$

which confirms that Eq. 5.46 is not restricted to straight-line currents, but holds quite generally in magnetostatics.

To complete the argument, however, we must check that the second term in Eq. 5.52 integrates to zero. Because the derivative acts only on $\hat{\boldsymbol{\imath}}/\imath^2$, we can switch from $\boldsymbol{\nabla}$ to $\boldsymbol{\nabla}'$ at the cost of a minus sign:[13]

$$-(\mathbf{J} \cdot \boldsymbol{\nabla}) \frac{\hat{\boldsymbol{\imath}}}{\imath^2} = (\mathbf{J} \cdot \boldsymbol{\nabla}') \frac{\hat{\boldsymbol{\imath}}}{\imath^2}. \tag{5.54}$$

[13]The point here is that $\boldsymbol{\imath}$ depends only on the *difference* between the coordinates; note that $(\partial/\partial x)\, f(x - x') = -(\partial/\partial x') f(x - x')$.

The x component, in particular, is

$$(\mathbf{J}\cdot\boldsymbol{\nabla}')\left(\frac{x-x'}{\imath^3}\right)=\boldsymbol{\nabla}'\cdot\left[\frac{(x-x')}{\imath^3}\mathbf{J}\right]-\left(\frac{x-x'}{\imath^3}\right)(\boldsymbol{\nabla}'\cdot\mathbf{J})$$

(using product rule 5). Now, for *steady* currents the divergence of $\mathbf{J}$ is zero (Eq. 5.33), so

$$\left[-(\mathbf{J}\cdot\boldsymbol{\nabla})\frac{\boldsymbol{\imath}}{\imath^2}\right]_x=\boldsymbol{\nabla}'\cdot\left[\frac{(x-x')}{\imath^3}\mathbf{J}\right],$$

and therefore this contribution to the integral (Eq. 5.51) can be written

$$\int_{\mathcal{V}}\boldsymbol{\nabla}'\cdot\left[\frac{(x-x')}{\imath^3}\mathbf{J}\right]d\tau'=\oint_{\mathcal{S}}\frac{(x-x')}{\imath^3}\mathbf{J}\cdot d\mathbf{a}'. \tag{5.55}$$

(The reason for switching from $\boldsymbol{\nabla}$ to $\boldsymbol{\nabla}'$ was to permit this integration by parts.) But what region are we integrating over? Well, it's the volume that appears in the Biot-Savart law (Eq. 5.47)—large enough, that is, to include all the current. You can make it *bigger* than that, if you like; $\mathbf{J}=0$ out there anyway, so it will add nothing to the integral. The essential point is that *on the boundary* the current is *zero* (all current is safely *inside*) and hence the surface integral (Eq. 5.55) vanishes.[14]

5.3.3 ■ Ampère's Law

The equation for the curl of **B**,

$$\boxed{\boldsymbol{\nabla}\times\mathbf{B}=\mu_0\mathbf{J},} \tag{5.56}$$

is called **Ampère's law** (in differential form). It can be converted to integral form by the usual device of applying one of the fundamental theorems—in this case Stokes' theorem:

$$\int(\boldsymbol{\nabla}\times\mathbf{B})\cdot d\mathbf{a}=\oint\mathbf{B}\cdot d\mathbf{l}=\mu_0\int\mathbf{J}\cdot d\mathbf{a}.$$

Now, $\int\mathbf{J}\cdot d\mathbf{a}$ is the total current passing through the surface (Fig. 5.31), which we call I_{enc} (the **current enclosed** by the **Amperian loop**). Thus

$$\boxed{\oint\mathbf{B}\cdot d\mathbf{l}=\mu_0 I_{\text{enc}}.} \tag{5.57}$$

[14] If **J** itself extends to infinity (as in the case of an infinite straight wire), the surface integral is still typically zero, though the analysis calls for greater care.

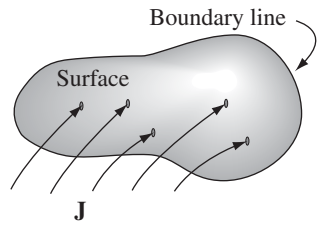


FIGURE 5.31

This is the integral version of Ampère's law; it generalizes Eq. 5.44 to *arbitrary* steady currents. Notice that Eq. 5.57 inherits the sign ambiguity of Stokes' theorem (Sect. 1.3.5): Which *way* around the loop am I supposed to go? And which *direction* through the surface corresponds to a "positive" current? The resolution, as always, is the right-hand rule: If the fingers of your right hand indicate the direction of integration around the boundary, then your thumb defines the direction of a positive current.

Just as the Biot-Savart law plays a role in magnetostatics that Coulomb's law assumed in electrostatics, so Ampère's plays the part of Gauss's:

$$\begin{cases} \text{Electrostatics}: & \text{Coulomb} & \rightarrow & \text{Gauss}, \\ \text{Magnetostatics}: & \text{Biot}-\text{Savart} & \rightarrow & \text{Ampère}. \end{cases}$$

In particular, for currents with appropriate symmetry, Ampère's law in integral form offers a lovely and extraordinarily efficient way of calculating the magnetic field.

Example 5.7. Find the magnetic field a distance s from a long straight wire (Fig. 5.32), carrying a steady current I (the same problem we solved in Ex. 5.5, using the Biot-Savart law).

Solution

We know the direction of $\mathbf{B}$ is "circumferential," circling around the wire as indicated by the right-hand rule. By symmetry, the magnitude of $\mathbf{B}$ is constant around an Amperian loop of radius s, centered on the wire. So Ampère's law gives

$$\oint \mathbf{B} \cdot d\mathbf{l} = B \oint dl = B 2\pi s = \mu_0 I_{\text{enc}} = \mu_0 I,$$

or

$$B = \frac{\mu_0 I}{2\pi s}.$$

This is the same answer we got before (Eq. 5.38), but it was obtained this time with far less effort.

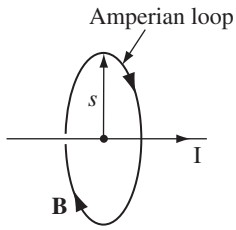


FIGURE 5.32

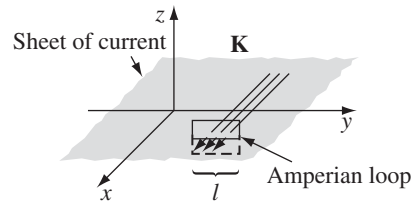


FIGURE 5.33

Example 5.8. Find the magnetic field of an infinite uniform surface current $\mathbf{K} = K\,\hat{\mathbf{x}}$, flowing over the xy plane (Fig. 5.33).

Solution

First of all, what is the *direction* of **B**? Could it have any x component? *No*: A glance at the Biot-Savart law (Eq. 5.42) reveals that **B** is *perpendicular* to **K**. Could it have a z component? *No again*. You could confirm this by noting that any vertical contribution from a filament at $+y$ is canceled by the corresponding filament at $-y$. But there is a nicer argument: Suppose the field pointed *away* from the plane. By reversing the direction of the current, I could make it point *toward* the plane (in the Biot-Savart law, changing the sign of the current switches the sign of the field). But the z component of **B** cannot possibly depend on the *direction* of the current in the xy plane. (Think about it!) So **B** can only have a y component, and a quick check with your right hand should convince you that it points to the *left* above the plane and to the *right* below it.

With this in mind, we draw a rectangular Amperian loop as shown in Fig. 5.33, parallel to the yz plane and extending an equal distance above and below the surface. Applying Ampère's law,

$$\oint \mathbf{B} \cdot d\mathbf{l} = 2Bl = \mu_0 I_{\text{enc}} = \mu_0 K l,$$

(one Bl comes from the top segment and the other from the bottom), so $B = (\mu_0/2)K$, or, more precisely,

$$\mathbf{B} = \begin{cases} +(\mu_0/2)K\,\hat{\mathbf{y}} & \text{for } z < 0, \\ -(\mu_0/2)K\,\hat{\mathbf{y}} & \text{for } z > 0. \end{cases} \tag{5.58}$$

Notice that the field is independent of the distance from the plane, just like the electric field of a uniform surface charge (Ex. 2.5).

Example 5.9. Find the magnetic field of a very long solenoid, consisting of n closely wound turns per unit length on a cylinder of radius R, each carrying a steady current I (Fig. 5.34). [The point of making the windings so close is that one can then pretend each turn is circular. If this troubles you (after all, there is a net current I in the direction of the solenoid's axis, no matter *how* tight the

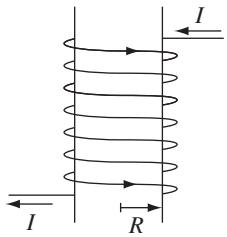


FIGURE 5.34

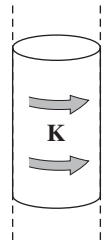


FIGURE 5.35

winding), picture instead a sheet of aluminum foil wrapped around the cylinder, carrying the equivalent uniform surface current $K = nI$ (Fig. 5.35). Or make a double winding, going up to one end and then—always in the same sense—going back down again, thereby eliminating the net longitudinal current. But, in truth, this is all unnecessary fastidiousness, for the field inside a solenoid is huge (relatively speaking), and the field of the longitudinal current is at most a tiny refinement.]

Solution

First of all, what is the *direction* of **B**? Could it have a radial component? *No*. For suppose B_s were *positive*; if we reversed the direction of the current, B_s would then be *negative*. But switching I is physically equivalent to turning the solenoid upside down, and that certainly should not alter the radial field. How about a "circumferential" component? *No*. For B_ϕ would be constant around an Amperian loop concentric with the solenoid (Fig. 5.36), and hence

$$\oint \mathbf{B} \cdot d\mathbf{l} = B_\phi(2\pi s) = \mu_0 I_{\text{enc}} = 0,$$

since the loop encloses no current.

So the magnetic field of an infinite, closely wound solenoid runs *parallel to the axis*. From the right-hand rule, we expect that it points upward inside the solenoid and downward outside. Moreover, it certainly approaches zero as you go very far

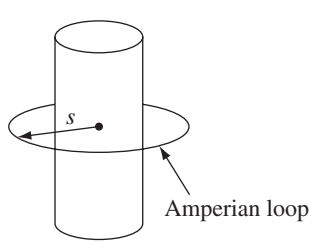


FIGURE 5.36

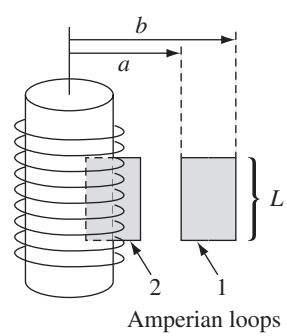


FIGURE 5.37

away. With this in mind, let's apply Ampère's law to the two rectangular loops in Fig. 5.37. Loop 1 lies entirely outside the solenoid, with its sides at distances a and b from the axis:

$$\oint \mathbf{B} \cdot d\mathbf{l} = [B(a) - B(b)]L = \mu_0 I_{\text{enc}} = 0,$$

so

$$B(a) = B(b).$$

Evidently the *field outside does not depend on the distance from the axis*. But we agreed that it goes to *zero* for large s. It must therefore be zero *everywhere*! (This astonishing result can also be derived from the Biot-Savart law, of course, but it's much more difficult. See Prob. 5.46.)

As for loop 2, which is half inside and half outside, Ampère's law gives

$$\oint \mathbf{B} \cdot d\mathbf{l} = BL = \mu_0 I_{\text{enc}} = \mu_0 n I L,$$

where B is the field inside the solenoid. (The right side of the loop contributes nothing, since $B = 0$ out there.) *Conclusion:*

$$\mathbf{B} = \begin{cases} \mu_0 n I\, \hat{\mathbf{z}}, & \text{inside the solenoid,} \\ \mathbf{0}, & \text{outside the solenoid.} \end{cases} \tag{5.59}$$

Notice that the field inside is *uniform*—it doesn't depend on the distance from the axis. In this sense the solenoid is to magnetostatics what the parallel-plate capacitor is to electrostatics: a simple device for producing strong uniform fields.

Like Gauss's law, Ampère's law is always *true* (for steady currents), but it is not always *useful*. Only when the symmetry of the problem enables you to pull B outside the integral $\oint \mathbf{B} \cdot d\mathbf{l}$ can you calculate the magnetic field from Ampère's law. When it *does* work, it's by far the fastest method; when it doesn't, you have to fall back on the Biot-Savart law. The current configurations that can be handled by Ampère's law are

1. Infinite straight lines (prototype: Ex. 5.7).
2. Infinite planes (prototype: Ex. 5.8).
3. Infinite solenoids (prototype: Ex. 5.9).
4. Toroids (prototype: Ex. 5.10).

The last of these is a surprising and elegant application of Ampère's law. As in Exs. 5.8 and 5.9, the hard part is figuring out the *direction* of the field (which we will now have done, once and for all, for each of the four geometries); the actual application of Ampère's law takes only one line.

Example 5.10. A toroidal coil consists of a circular ring, or "donut," around which a long wire is wrapped (Fig. 5.38). The winding is uniform and tight enough so that each turn can be considered a plane closed loop. The cross-sectional shape of the coil is immaterial. I made it rectangular in Fig. 5.38 for the sake of simplicity, but it could just as well be circular or even some weird asymmetrical form, as in Fig. 5.39, as long as the shape remains the same all the way around the ring. In that case, it follows that the *magnetic field of the toroid is circumferential at all points, both inside and outside the coil*.

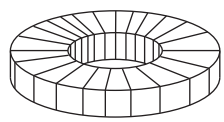


FIGURE 5.38

Proof. According to the Biot-Savart law, the field at $\mathbf{r}$ due to the current element at $\mathbf{r}'$ is

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{\mathbf{I}\times\boldsymbol{\mathcal{r}}}{\mathcal{r}^3}\,dl'.$$

We may as well put $\mathbf{r}$ in the xz plane (Fig. 5.39), so its Cartesian components are $(x, 0, z)$, while the source coordinates are

$$\mathbf{r}' = (s'\cos\phi', s'\sin\phi', z').$$

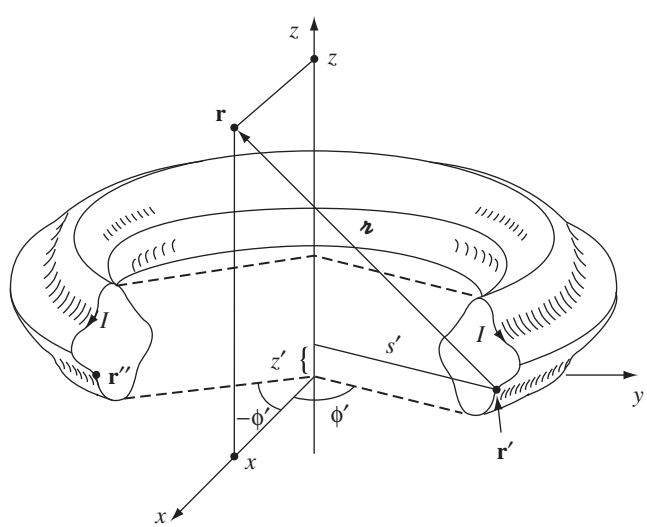


FIGURE 5.39

Then

$$\boldsymbol{\imath} = (x - s'\cos\phi', -s'\sin\phi', z - z').$$

Since the current has no ϕ component, $\mathbf{I} = I_s\,\hat{\mathbf{s}} + I_z\,\hat{\mathbf{z}}$, or (in Cartesian coordinates)

$$\mathbf{I} = (I_s\cos\phi', I_s\sin\phi', I_z).$$

Accordingly,

$$\mathbf{I}\times\boldsymbol{\imath} = \begin{bmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ I_s\cos\phi' & I_s\sin\phi' & I_z \\ (x - s'\cos\phi') & (-s'\sin\phi') & (z - z') \end{bmatrix}$$

$$= \left[\sin\phi'\left(I_s(z - z') + s'I_z\right)\right]\hat{\mathbf{x}} + \left[I_z(x - s'\cos\phi') - I_s\cos\phi'(z - z')\right]\hat{\mathbf{y}}$$

$$+ \left[-I_s x\sin\phi'\right]\hat{\mathbf{z}}.$$

But there is a symmetrically situated current element at $\mathbf{r}''$, with the same s', the same $\imath$, the same dl', the same I_s, and the same I_z, *but negative* ϕ' (Fig. 5.39). Because $\sin\phi'$ changes sign, the $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ contributions from $\mathbf{r}'$ and $\mathbf{r}''$ cancel, leaving only a $\hat{\mathbf{y}}$ term. Thus the field at $\mathbf{r}$ is in the $\hat{\mathbf{y}}$ direction, and in general the field points in the $\hat{\boldsymbol{\phi}}$ direction. □

Now that we know the field is circumferential, determining its magnitude is ridiculously easy. Just apply Ampère's law to a circle of radius s about the axis of the toroid:

$$B2\pi s = \mu_0 I_{\text{enc}},$$

and hence

$$\mathbf{B}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0 N I}{2\pi s}\hat{\boldsymbol{\phi}}, & \text{for points inside the coil,} \\ \mathbf{0}, & \text{for points outside the coil,} \end{cases} \tag{5.60}$$

where N is the total number of turns.

Problem 5.14 A steady current I flows down a long cylindrical wire of radius a (Fig. 5.40). Find the magnetic field, both inside and outside the wire, if

(a) The current is uniformly distributed over the outside surface of the wire.

(b) The current is distributed in such a way that J is proportional to s, the distance from the axis.

FIGURE 5.40

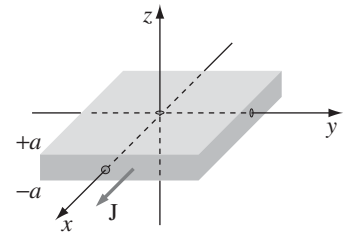


FIGURE 5.41

Problem 5.15 A thick slab extending from $z = -a$ to $z = +a$ (and infinite in the x and y directions) carries a uniform volume current $\mathbf{J} = J\,\hat{\mathbf{x}}$ (Fig. 5.41). Find the magnetic field, as a function of z, both inside and outside the slab.

Problem 5.16 Two long coaxial solenoids each carry current I, but in opposite directions, as shown in Fig. 5.42. The inner solenoid (radius a) has n_1 turns per unit length, and the outer one (radius b) has n_2. Find **B** in each of the three regions: (i) inside the inner solenoid, (ii) between them, and (iii) outside both.

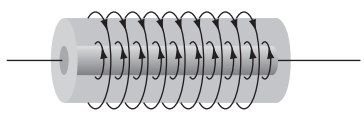


FIGURE 5.42

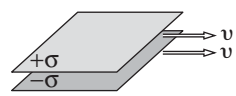


FIGURE 5.43

Problem 5.17 A large parallel-plate capacitor with uniform surface charge σ on the upper plate and $-\sigma$ on the lower is moving with a constant speed v, as shown in Fig. 5.43.

(a) Find the magnetic field between the plates and also above and below them.

(b) Find the magnetic force per unit area on the upper plate, including its direction.

(c) At what speed v would the magnetic force balance the electrical force?[15]

! **Problem 5.18** Show that the magnetic field of an infinite solenoid runs parallel to the axis, *regardless of the cross-sectional shape of the coil*, as long as that shape is constant along the length of the solenoid. What is the magnitude of the field, inside and outside of such a coil? Show that the toroid field (Eq. 5.60) reduces to the solenoid field, when the radius of the donut is so large that a segment can be considered essentially straight.

Problem 5.19 In calculating the current enclosed by an Amperian loop, one must, in general, evaluate an integral of the form

$$I_{\text{enc}} = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{a}.$$

[15] See footnote to Prob. 5.13.

The trouble is, there are infinitely many surfaces that share the same boundary line. Which one are we supposed to use?

5.3.4 ■ Comparison of Magnetostatics and Electrostatics

The divergence and curl of the *electrostatic* field are

$$\begin{cases} \nabla \cdot \mathbf{E} = \dfrac{1}{\epsilon_0}\rho, & \text{(Gauss's law);} \\ \nabla \times \mathbf{E} = \mathbf{0}, & \text{(no name).} \end{cases}$$

These are **Maxwell's equations** for electrostatics. Together with the boundary condition $\mathbf{E} \to \mathbf{0}$ far from all charges,[16] Maxwell's equations determine the field, if the source charge density ρ is given; they contain essentially the same information as Coulomb's law plus the principle of superposition. The divergence and curl of the *magnetostatic* field are

$$\begin{cases} \nabla \cdot \mathbf{B} = 0, & \text{(no name);} \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J}, & \text{(Ampère's law).} \end{cases}$$

These are Maxwell's equations for magnetostatics. Again, together with the boundary condition $\mathbf{B} \to \mathbf{0}$ far from all currents, Maxwell's equations determine the magnetic field; they are equivalent to the Biot-Savart law (plus superposition). Maxwell's equations and the force law

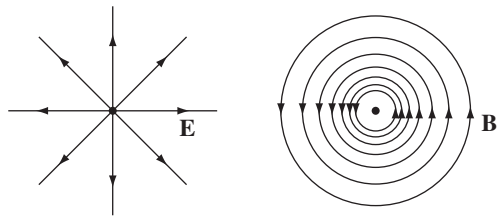
$$\mathbf{F} = Q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

constitute the most elegant formulation of electrostatics and magnetostatics.

The electric field *diverges away from* a (positive) charge; the magnetic field line *curls around* a current (Fig. 5.44). Electric field lines originate on positive charges and terminate on negative ones; magnetic field lines do not begin or end anywhere—to do so would require a nonzero divergence. They typically form closed loops or extend out to infinity.[17] To put it another way, *there are no point sources for* **B**, as there are for **E**; there exists no magnetic analog to electric charge. This is the physical content of the statement $\nabla \cdot \mathbf{B} = 0$. Coulomb and others believed that magnetism was produced by **magnetic charges** (**magnetic monopoles**, as we would now call them), and in some older books you will still find references to a magnetic version of Coulomb's law, giving the force of attraction or repulsion between them. It was Ampère who first speculated that all magnetic effects are attributable to *electric* charges *in motion* (currents). As far

[16] In those artificial problems where the charge (or current) extends to infinity—infinite planes, for example—symmetry considerations can sometimes take the place of boundary conditions.

[17] A third possibility turns out to be surprisingly common: they can form chaotic tangles. See M. Lieberherr, *Am. J. Phys.* **78**, 1117 (2010).



(a) Electrostatic field of a point charge (b) Magnetostatic field of a long wire

FIGURE 5.44

as we know, Ampère was right; nevertheless, it remains an open experimental question whether magnetic monopoles exist in nature (they are obviously pretty *rare*, or somebody would have found one[18]), and in fact some recent elementary particle theories *require* them. For our purposes, though, **B** is divergenceless, and there are no magnetic monopoles. It takes a *moving* electric charge to *produce* a magnetic field, and it takes another moving electric charge to "feel" a magnetic field.

Typically, electric forces are enormously larger than magnetic ones. That's not something intrinsic to the theory; it has to do with the sizes of the fundamental constants ϵ_0 and μ_0. In general, it is only when both the source charges and the test charge are moving at velocities comparable to the speed of light that the magnetic force approaches the electric force in strength. (Problems 5.13 and 5.17 illustrate this rule.) How is it, then, that we notice magnetic effects at all? The answer is that both in the production of a magnetic field (Biot-Savart) and in its detection (Lorentz), it is the *current* that matters, and we can compensate for a smallish velocity by pouring huge amounts of charge down the wire. Ordinarily, this charge would simultaneously generate so large an *electric* force as to swamp the magnetic one. But if we arrange to keep the wire *neutral*, by embedding in it an equal quantity of opposite charge at rest, the electric field cancels out, leaving the magnetic field to stand alone. It sounds very elaborate, but of course this is precisely what happens in an ordinary current carrying wire.

Problem 5.20

(a) Find the density ρ of mobile charges in a piece of copper, assuming each atom contributes one free electron. [Look up the necessary physical constants.]

(b) Calculate the average electron velocity in a copper wire 1 mm in diameter, carrying a current of 1 A. [*Note*: This is literally a *snail's* pace. How, then, can you carry on a long distance telephone conversation?]

[18] An apparent detection (B. Cabrera, *Phys. Rev. Lett.* **48**, 1378 (1982)) has never been reproduced—and not for want of trying. For a delightful brief history of ideas about magnetism, see Chapter 1 in D. C. Mattis, *The Theory of Magnetism* (New York: Harper & Row, 1965).

(c) What is the force of attraction between two such wires, 1 cm apart?

(d) If you could somehow remove the stationary positive charges, what would the electrical repulsion force be? How many times greater than the magnetic force is it?

Problem 5.21 Is Ampère's law consistent with the general rule (Eq. 1.46) that divergence-of-curl is always zero? Show that Ampère's law *cannot* be valid, in general, outside magnetostatics. Is there any such "defect" in the other three Maxwell equations?

Problem 5.22 Suppose there *did* exist magnetic monopoles. How would you modify Maxwell's equations and the force law to accommodate them? If you think there are several plausible options, list them, and suggest how you might decide experimentally which one is right.

5.4 ■ MAGNETIC VECTOR POTENTIAL

5.4.1 ■ The Vector Potential

Just as $\nabla \times \mathbf{E} = \mathbf{0}$ permitted us to introduce a scalar potential (V) in electrostatics,

$$\mathbf{E} = -\nabla V,$$

so $\nabla \cdot \mathbf{B} = 0$ invites the introduction of a *vector* potential $\mathbf{A}$ in magnetostatics:

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}.} \tag{5.61}$$

The former is authorized by Theorem 1 (of Sect. 1.6.2), the latter by Theorem 2 (The proof of Theorem 2 is developed in Prob. 5.31). The potential formulation automatically takes care of $\nabla \cdot \mathbf{B} = 0$ (since the divergence of a curl is always zero); there remains Ampère's law:

$$\nabla \times \mathbf{B} = \nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J}. \tag{5.62}$$

Now, the electric potential had a built-in ambiguity: you can add to V any function whose gradient is zero (which is to say, any *constant*), without altering the *physical* quantity $\mathbf{E}$. Likewise, you can add to $\mathbf{A}$ any function whose *curl* vanishes (which is to say, the *gradient of any scalar*), with no effect on $\mathbf{B}$. We can exploit this freedom to eliminate the divergence of $\mathbf{A}$:

$$\boxed{\nabla \cdot \mathbf{A} = 0.} \tag{5.63}$$

To prove that this is always possible, suppose that our original potential, $\mathbf{A}_0$, is *not* divergenceless. If we add to it the gradient of λ ($\mathbf{A} = \mathbf{A}_0 + \nabla\lambda$), the new divergence is

$$\nabla \cdot \mathbf{A} = \nabla \cdot \mathbf{A}_0 + \nabla^2 \lambda.$$

We can accommodate Eq. 5.63, then, if a function λ can be found that satisfies

$$\nabla^2\lambda = -\nabla \cdot \mathbf{A}_0.$$

But this is mathematically identical to Poisson's equation (2.24),

$$\nabla^2 V = -\frac{\rho}{\epsilon_0},$$

with $\nabla \cdot \mathbf{A}_0$ in place of ρ/ϵ_0 as the "source." And we *know* how to solve Poisson's equation—that's what electrostatics is all about ("given the charge distribution, find the potential"). In particular, if ρ goes to zero at infinity, the solution is Eq. 2.29:

$$V = \frac{1}{4\pi\epsilon_0}\int \frac{\rho}{\imath}\, d\tau',$$

and by the same token, if $\nabla \cdot \mathbf{A}_0$ goes to zero at infinity, then

$$\lambda = \frac{1}{4\pi}\int \frac{\nabla \cdot \mathbf{A}_0}{\imath}\, d\tau'.$$

If $\nabla \cdot \mathbf{A}_0$ does *not* go to zero at infinity, we'll have to use other means to discover the appropriate λ, just as we get the electric potential by other means when the charge distribution extends to infinity. But the essential point remains: *It is always possible to make the vector potential divergenceless*. To put it the other way around: the definition $\mathbf{B} = \nabla \times \mathbf{A}$ specifies the *curl* of $\mathbf{A}$, but it doesn't say anything about the *divergence*—we are at liberty to pick that as we see fit, and zero is ordinarily the simplest choice.

With this condition on $\mathbf{A}$, Ampère's law (Eq. 5.62) becomes

$$\boxed{\nabla^2\mathbf{A} = -\mu_0\mathbf{J}.} \tag{5.64}$$

This *again* is nothing but Poisson's equation—or rather, it is *three* Poisson's equations, one for each Cartesian[19] component. Assuming $\mathbf{J}$ goes to zero at infinity, we can read off the solution:

$$\boxed{\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}')}{\imath}\, d\tau'.} \tag{5.65}$$

[19] In Cartesian coordinates, $\nabla^2\mathbf{A} = (\nabla^2 A_x)\hat{\mathbf{x}} + (\nabla^2 A_y)\hat{\mathbf{y}} + (\nabla^2 A_z)\hat{\mathbf{z}}$, so Eq. 5.64 reduces to $\nabla^2 A_x = -\mu_0 J_x$, $\nabla^2 A_y = -\mu_0 J_y$, and $\nabla^2 A_z = -\mu_0 J_z$. In curvilinear coordinates the unit vectors *themselves* are functions of position, and must be differentiated, so it is *not* the case, for example, that $\nabla^2 A_r = -\mu_0 J_r$. Remember that even if you plan to *evaluate* integrals such as 5.65 using curvilinear coordinates, you must first express $\mathbf{J}$ in terms of its *Cartesian* components (see Sect. 1.4.1).

For line and surface currents,

$$\mathbf{A} = \frac{\mu_0}{4\pi}\int \frac{\mathbf{I}}{\imath}\,dl' = \frac{\mu_0 I}{4\pi}\int \frac{1}{\imath}\,d\mathbf{l}'; \qquad \mathbf{A} = \frac{\mu_0}{4\pi}\int \frac{\mathbf{K}}{\imath}\,da'. \tag{5.66}$$

(If the current does *not* go to zero at infinity, we have to find other ways to get **A**; some of these are explored in Ex. 5.12 and in the problems at the end of the section.)

It must be said that **A** is not as *useful* as V. For one thing, it's still a *vector*, and although Eqs. 5.65 and 5.66 are somewhat easier to work with than the Biot-Savart law, you still have to fuss with components. It would be nice if we could get away with a *scalar* potential

$$\mathbf{B} = -\nabla U, \tag{5.67}$$

but this is incompatible with Ampère's law, since the curl of a gradient is always zero. (A **magnetostatic scalar potential** *can* be used, if you stick scrupulously to simply-connected, current-free regions, but as a theoretical tool, it is of limited interest. See Prob. 5.29.) Moreover, since magnetic forces do no work, **A** does not admit a simple physical interpretation in terms of potential energy per unit charge. (In some contexts it can be interpreted as *momentum* per unit charge.[20]) Nevertheless, the vector potential has substantial theoretical importance, as we shall see in Chapter 10.

Example 5.11. A spherical shell of radius R, carrying a uniform surface charge σ, is set spinning at angular velocity $\boldsymbol{\omega}$. Find the vector potential it produces at point **r** (Fig. 5.45).

Solution

It might seem natural to set the polar axis along $\boldsymbol{\omega}$, but in fact the integration is easier if we let **r** lie on the z axis, so that $\boldsymbol{\omega}$ is tilted at an angle ψ. We may as well orient the x axis so that $\boldsymbol{\omega}$ lies in the xz plane, as shown in Fig. 5.46. According to Eq. 5.66,

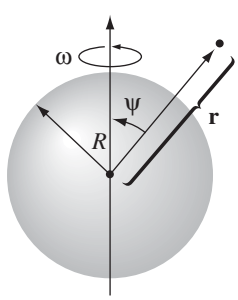


FIGURE 5.45

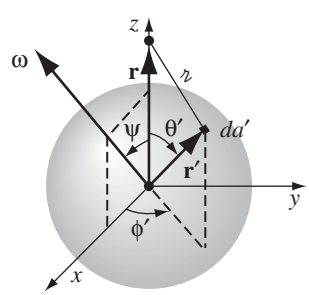


FIGURE 5.46

[20]M. D. Semon and J. R. Taylor, *Am. J. Phys.* **64**, 1361 (1996).

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{K}(\mathbf{r}')}{\imath}\, da',$$

where $\mathbf{K} = \sigma\mathbf{v}$, $\imath = \sqrt{R^2 + r^2 - 2Rr\cos\theta'}$, and $da' = R^2 \sin\theta'\, d\theta'\, d\phi'$. Now the velocity of a point $\mathbf{r}'$ in a rotating rigid body is given by $\boldsymbol{\omega} \times \mathbf{r}'$; in this case,

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}' = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \omega\sin\psi & 0 & \omega\cos\psi \\ R\sin\theta'\cos\phi' & R\sin\theta'\sin\phi' & R\cos\theta' \end{vmatrix}$$

$$= R\omega\left[-\left(\cos\psi\sin\theta'\sin\phi'\right)\hat{\mathbf{x}} + (\cos\psi\sin\theta'\cos\phi' - \sin\psi\cos\theta')\,\hat{\mathbf{y}}\right.$$

$$\left. + (\sin\psi\sin\theta'\sin\phi')\,\hat{\mathbf{z}}\right].$$

Notice that each of these terms, save one, involves either $\sin\phi'$ or $\cos\phi'$. Since

$$\int_0^{2\pi} \sin\phi'\, d\phi' = \int_0^{2\pi} \cos\phi'\, d\phi' = 0,$$

such terms contribute nothing. There remains

$$\mathbf{A}(\mathbf{r}) = -\frac{\mu_0 R^3 \sigma\omega\sin\psi}{2}\left(\int_0^{\pi} \frac{\cos\theta'\sin\theta'}{\sqrt{R^2 + r^2 - 2Rr\cos\theta'}}\, d\theta'\right)\hat{\mathbf{y}}.$$

Letting $u \equiv \cos\theta'$, the integral becomes

$$\int_{-1}^{+1} \frac{u}{\sqrt{R^2 + r^2 - 2Rru}}\, du = -\frac{(R^2 + r^2 + Rru)}{3R^2r^2}\sqrt{R^2 + r^2 - 2Rru}\,\Bigg|_{-1}^{+1}$$

$$= -\frac{1}{3R^2r^2}\left[(R^2 + r^2 + Rr)|R - r|\right.$$

$$\left. -(R^2 + r^2 - Rr)(R + r)\right].$$

If the point $\mathbf{r}$ lies *inside* the sphere, then $R > r$, and this expression reduces to $(2r/3R^2)$; if $\mathbf{r}$ lies *outside* the sphere, so that $R < r$, it reduces to $(2R/3r^2)$. Noting that $(\boldsymbol{\omega} \times \mathbf{r}) = -\omega r \sin\psi\,\hat{\mathbf{y}}$, we have, finally,

$$\mathbf{A}(\mathbf{r}) = \begin{cases} \dfrac{\mu_0 R\sigma}{3}(\boldsymbol{\omega} \times \mathbf{r}), & \text{for points } \textit{inside} \text{ the sphere,} \\ \dfrac{\mu_0 R^4\sigma}{3r^3}(\boldsymbol{\omega} \times \mathbf{r}), & \text{for points } \textit{outside} \text{ the sphere.} \end{cases} \tag{5.68}$$

Having evaluated the integral, I revert to the "natural" coordinates of Fig. 5.45, in which $\boldsymbol{\omega}$ coincides with the z axis and the point $\mathbf{r}$ is at (r, θ, ϕ):

$$\mathbf{A}(r, \theta, \phi) = \begin{cases} \dfrac{\mu_0 R\omega\sigma}{3} r\sin\theta\,\hat{\boldsymbol{\phi}}, & (r \le R), \\ \dfrac{\mu_0 R^4\omega\sigma}{3}\dfrac{\sin\theta}{r^2}\,\hat{\boldsymbol{\phi}}, & (r \ge R). \end{cases} \tag{5.69}$$

Curiously, the field inside this spherical shell is *uniform*:

$$\mathbf{B} = \nabla \times \mathbf{A} = \frac{2\mu_0 R\omega\sigma}{3}(\cos\theta\,\hat{\mathbf{r}} - \sin\theta\,\hat{\boldsymbol{\theta}}) = \frac{2}{3}\mu_0\sigma R\omega\,\hat{\mathbf{z}} = \frac{2}{3}\mu_0\sigma R\boldsymbol{\omega}. \quad (5.70)$$

Example 5.12. Find the vector potential of an infinite solenoid with n turns per unit length, radius R, and current I.

Solution

This time we cannot use Eq. 5.66, since the current itself extends to infinity. But here's a cute method that does the job. Notice that

$$\oint \mathbf{A}\cdot d\mathbf{l} = \int (\nabla\times\mathbf{A})\cdot d\mathbf{a} = \int \mathbf{B}\cdot d\mathbf{a} = \Phi, \quad (5.71)$$

where Φ is the flux of $\mathbf{B}$ through the loop in question. This is reminiscent of Ampère's law in integral form (Eq. 5.57),

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 I_{\text{enc}}.$$

In fact, it's the same equation, with $\mathbf{B}\to\mathbf{A}$ and $\mu_0 I_{\text{enc}}\to\Phi$. If symmetry permits, we can determine $\mathbf{A}$ from Φ in the same way we got $\mathbf{B}$ from I_{enc}, in Sect. 5.3.3. The present problem (with a uniform longitudinal magnetic field $\mu_0 nI$ inside the solenoid and no field outside) is analogous to the Ampère's law problem of a fat wire carrying a uniformly distributed current. The vector potential is "circumferential" (it mimics the magnetic field in the analog); using a circular "Amperian loop" at radius s *inside* the solenoid, we have

$$\oint \mathbf{A}\cdot d\mathbf{l} = A(2\pi s) = \int \mathbf{B}\cdot d\mathbf{a} = \mu_0 nI(\pi s^2),$$

so

$$\mathbf{A} = \frac{\mu_0 nI}{2} s\,\hat{\boldsymbol{\phi}}, \quad \text{for } s \le R. \quad (5.72)$$

For an Amperian loop *outside* the solenoid, the flux is

$$\int \mathbf{B}\cdot d\mathbf{a} = \mu_0 nI(\pi R^2),$$

since the field only extends out to R. Thus

$$\mathbf{A} = \frac{\mu_0 nI}{2}\frac{R^2}{s}\,\hat{\boldsymbol{\phi}}, \quad \text{for } s \ge R. \quad (5.73)$$

If you have any doubts about this answer, *check* it: Does $\nabla\times\mathbf{A} = \mathbf{B}$? Does $\nabla\cdot\mathbf{A} = 0$? If so, we're done.

Typically, the direction of **A** mimics the direction of the current. For instance, both were azimuthal in Exs. 5.11 and 5.12. Indeed, if all the current flows in *one* direction, then Eq. 5.65 suggests that **A** *must* point that way too. Thus the potential of a finite segment of straight wire (Prob. 5.23) is in the direction of the current. Of course, if the current extends to infinity you can't use Eq. 5.65 in the first place (see Probs. 5.26 and 5.27). Moreover, you can always add an arbitrary constant vector to **A**—this is analogous to changing the reference point for V, and it won't affect the divergence or curl of **A**, which is all that matters (in Eq. 5.65 we have chosen the constant so that **A** goes to zero at infinity). In principle you could even use a vector potential that is not divergenceless, in which case all bets are off. Despite these caveats, the essential point remains: *Ordinarily* the direction of **A** will match the direction of the current.

Problem 5.23 Find the magnetic vector potential of a finite segment of straight wire carrying a current I. [Put the wire on the z axis, from z_1 to z_2, and use Eq. 5.66.] Check that your answer is consistent with Eq. 5.37.

Problem 5.24 What current density would produce the vector potential, $\mathbf{A} = k\,\hat{\boldsymbol{\phi}}$ (where k is a constant), in cylindrical coordinates?

Problem 5.25 If **B** is *uniform*, show that $\mathbf{A}(\mathbf{r}) = -\frac{1}{2}(\mathbf{r} \times \mathbf{B})$ works. That is, check that $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ and $\boldsymbol{\nabla} \times \mathbf{A} = \mathbf{B}$. Is this result unique, or are there other functions with the same divergence and curl?

Problem 5.26

(a) By whatever means you can think of (short of looking it up), find the vector potential a distance s from an infinite straight wire carrying a current I. Check that $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ and $\boldsymbol{\nabla} \times \mathbf{A} = \mathbf{B}$.

(b) Find the magnetic potential *inside* the wire, if it has radius R and the current is uniformly distributed.

Problem 5.27 Find the vector potential above and below the plane surface current in Ex. 5.8.

Problem 5.28

(a) Check that Eq. 5.65 is consistent with Eq. 5.63, by applying the *divergence*.

(b) Check that Eq. 5.65 is consistent with Eq. 5.47, by applying the *curl*.

(c) Check that Eq. 5.65 is consistent with Eq. 5.64, by applying the *Laplacian*.

Problem 5.29 Suppose you want to define a magnetic scalar potential U (Eq. 5.67) in the vicinity of a current-carrying wire. First of all, you must stay away from the wire itself (there $\boldsymbol{\nabla} \times \mathbf{B} \neq \mathbf{0}$); but that's not enough. Show, by applying Ampère's law to a path that starts at **a** and circles the wire, returning to **b** (Fig. 5.47), that the scalar potential cannot be single-valued (that is, $U(\mathbf{a}) \neq U(\mathbf{b})$, even if they represent the same physical point). As an example, find the scalar potential for an infinite

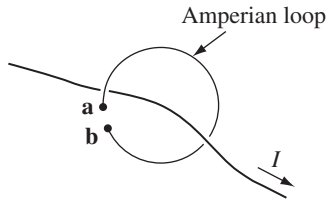


FIGURE 5.47

straight wire. (To avoid a multivalued potential, you must restrict yourself to simply-connected regions that remain on one side or the other of every wire, never allowing you to go all the way around.)

Problem 5.30 Use the results of Ex. 5.11 to find the magnetic field inside a solid sphere, of uniform charge density ρ and radius R, that is rotating at a constant angular velocity $\boldsymbol{\omega}$.

Problem 5.31

(a) Complete the proof of Theorem 2, Sect. 1.6.2. That is, show that any divergenceless vector field $\mathbf{F}$ can be written as the curl of a vector potential $\mathbf{A}$. What you have to do is find A_x, A_y, and A_z such that (i) $\partial A_z/\partial y - \partial A_y/\partial z = F_x$; (ii) $\partial A_x/\partial z - \partial A_z/\partial x = F_y$; and (iii) $\partial A_y/\partial x - \partial A_x/\partial y = F_z$. Here's one way to do it: Pick $A_x = 0$, and solve (ii) and (iii) for A_y and A_z. Note that the "constants of integration" are themselves functions of y and z—they're constant only with respect to x. Now plug these expressions into (i), and use the fact that $\boldsymbol{\nabla} \cdot \mathbf{F} = 0$ to obtain

$$A_y = \int_0^x F_z(x', y, z)\, dx'; \quad A_z = \int_0^y F_x(0, y', z)\, dy' - \int_0^x F_y(x', y, z)\, dx'.$$

(b) By direct differentiation, check that the $\mathbf{A}$ you obtained in part (a) satisfies $\boldsymbol{\nabla} \times \mathbf{A} = \mathbf{F}$. Is $\mathbf{A}$ divergenceless? [This was a very asymmetrical construction, and it would be surprising if it *were*—although we know that there *exists* a vector whose curl is $\mathbf{F}$ *and* whose divergence is zero.]

(c) As an example, let $\mathbf{F} = y\,\hat{\mathbf{x}} + z\,\hat{\mathbf{y}} + x\,\hat{\mathbf{z}}$. Calculate $\mathbf{A}$, and confirm that $\boldsymbol{\nabla} \times \mathbf{A} = \mathbf{F}$. (For further discussion, see Prob. 5.53.)

5.4.2 ■ Boundary Conditions

In Chapter 2, I drew a triangular diagram to summarize the relations among the three fundamental quantities of electrostatics: the charge density ρ, the electric field $\mathbf{E}$, and the potential V. A similar figure can be constructed for magnetostatics (Fig. 5.48), relating the current density $\mathbf{J}$, the field $\mathbf{B}$, and the potential $\mathbf{A}$. There is one "missing link" in the diagram: the equation for $\mathbf{A}$ in terms of $\mathbf{B}$. It's unlikely you would ever need such a formula, but in case you are interested, see Probs. 5.52 and 5.53.

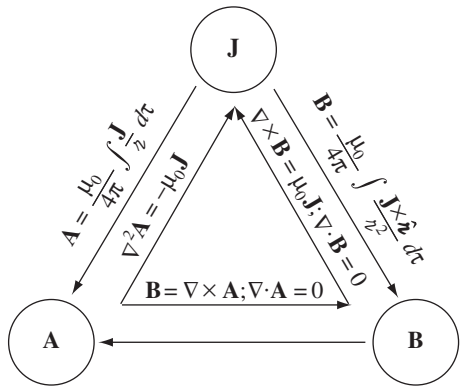


FIGURE 5.48

Just as the electric field suffers a discontinuity at a surface *charge*, so the magnetic field is discontinuous at a surface *current*. Only this time it is the *tangential* component that changes. For if we apply Eq. 5.50, in integral form,

$$\oint \mathbf{B}\cdot d\mathbf{a} = 0,$$

to a wafer-thin pillbox straddling the surface (Fig. 5.49), we get

$$B^{\perp}_{\text{above}} = B^{\perp}_{\text{below}}. \tag{5.74}$$

As for the tangential components, an Amperian loop running perpendicular to the current (Fig. 5.50) yields

$$\oint \mathbf{B}\cdot d\mathbf{l} = \left(B^{\parallel}_{\text{above}} - B^{\parallel}_{\text{below}}\right) l = \mu_0 I_{\text{enc}} = \mu_0 K l,$$

or

$$B^{\parallel}_{\text{above}} - B^{\parallel}_{\text{below}} = \mu_0 K. \tag{5.75}$$

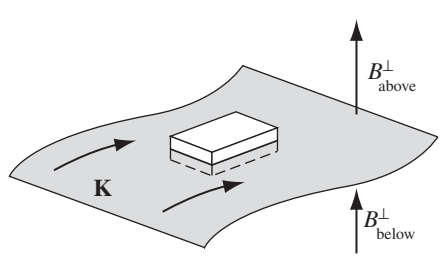


FIGURE 5.49

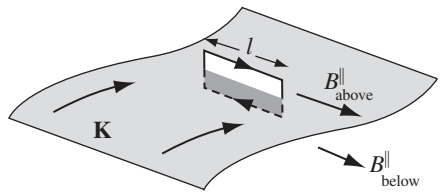


FIGURE 5.50

Thus the component of **B** that is parallel to the surface but perpendicular to the current is discontinuous in the amount $\mu_0 K$. A similar Amperian loop running *parallel* to the current reveals that the *parallel* component is *continuous*. These results can be summarized in a single formula:

$$\mathbf{B}_{\text{above}} - \mathbf{B}_{\text{below}} = \mu_0(\mathbf{K} \times \hat{\mathbf{n}}), \tag{5.76}$$

where $\hat{\mathbf{n}}$ is a unit vector perpendicular to the surface, pointing "upward."

Like the scalar potential in electrostatics, the vector potential is continuous across any boundary:

$$\mathbf{A}_{\text{above}} = \mathbf{A}_{\text{below}}, \tag{5.77}$$

for $\nabla \cdot \mathbf{A} = 0$ guarantees[21] that the *normal* component is continuous; and $\nabla \times \mathbf{A} = \mathbf{B}$, in the form

$$\oint \mathbf{A} \cdot d\mathbf{l} = \int \mathbf{B} \cdot d\mathbf{a} = \Phi,$$

means that the tangential components are continuous (the flux through an Amperian loop of vanishing thickness is zero). But the *derivative* of **A** inherits the discontinuity of **B**:

$$\frac{\partial \mathbf{A}_{\text{above}}}{\partial n} - \frac{\partial \mathbf{A}_{\text{below}}}{\partial n} = -\mu_0 \mathbf{K}. \tag{5.78}$$

Problem 5.32

(a) Check Eq. 5.76 for the configuration in Ex. 5.9.

(b) Check Eqs. 5.77 and 5.78 for the configuration in Ex. 5.11.

Problem 5.33 Prove Eq. 5.78, using Eqs. 5.63, 5.76, and 5.77. [*Suggestion:* I'd set up Cartesian coordinates at the surface, with z perpendicular to the surface and x parallel to the current.]

---

[21] Note that Eqs. 5.77 and 5.78 presuppose that **A** is divergenceless.

5.4.3 ■ Multipole Expansion of the Vector Potential

If you want an approximate formula for the vector potential of a localized current distribution, valid at distant points, a multipole expansion is in order. Remember: the idea of a multipole expansion is to write the potential in the form of a power series in $1/r$, where r is the distance to the point in question (Fig. 5.51); if r is sufficiently large, the series will be dominated by the lowest nonvanishing contribution, and the higher terms can be ignored. As we found in Sect. 3.4.1 (Eq. 3.94),

$$\frac{1}{\mathscr{r}} = \frac{1}{\sqrt{r^2 + (r')^2 - 2rr'\cos\alpha}} = \frac{1}{r}\sum_{n=0}^{\infty}\left(\frac{r'}{r}\right)^n P_n(\cos\alpha), \tag{5.79}$$

where α is the angle between $\mathbf{r}$ and $\mathbf{r}'$. Accordingly, the vector potential of a current loop can be written

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\oint \frac{1}{\mathscr{r}}\,d\mathbf{l}' = \frac{\mu_0 I}{4\pi}\sum_{n=0}^{\infty}\frac{1}{r^{n+1}}\oint (r')^n P_n(\cos\alpha)\,d\mathbf{l}', \tag{5.80}$$

or, more explicitly:

$$\begin{aligned}\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\Bigg[&\frac{1}{r}\oint d\mathbf{l}' + \frac{1}{r^2}\oint r'\cos\alpha\,d\mathbf{l}' \\ &+\frac{1}{r^3}\oint (r')^2\left(\frac{3}{2}\cos^2\alpha - \frac{1}{2}\right)d\mathbf{l}' + \cdots\Bigg].\end{aligned} \tag{5.81}$$

As in the multipole expansion of V, we call the first term (which goes like $1/r$) the **monopole** term, the second (which goes like $1/r^2$) **dipole**, the third **quadrupole**, and so on.

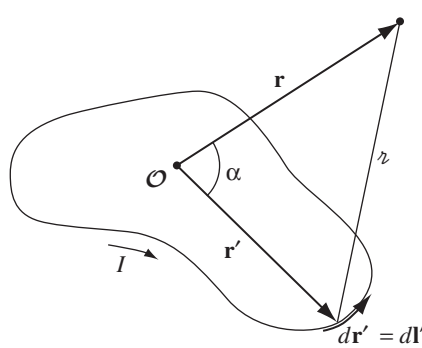


FIGURE 5.51

Now, the *magnetic monopole term is always zero*, for the integral is just the total vector displacement around a closed loop:

$$\oint d\mathbf{l}' = \mathbf{0}. \tag{5.82}$$

This reflects the fact that there are no magnetic monopoles in nature (an assumption contained in Maxwell's equation $\nabla \cdot \mathbf{B} = 0$, on which the entire theory of vector potential is predicated).

In the absence of any monopole contribution, the dominant term is the dipole (except in the rare case where it, too, vanishes):

$$\mathbf{A}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0 I}{4\pi r^2} \oint r' \cos\alpha \, d\mathbf{l}' = \frac{\mu_0 I}{4\pi r^2} \oint (\hat{\mathbf{r}} \cdot \mathbf{r}') \, d\mathbf{l}'. \tag{5.83}$$

This integral can be rewritten in a more illuminating way if we invoke Eq. 1.108, with $\mathbf{c} = \hat{\mathbf{r}}$:

$$\oint (\hat{\mathbf{r}} \cdot \mathbf{r}') \, d\mathbf{l}' = -\hat{\mathbf{r}} \times \int d\mathbf{a}'. \tag{5.84}$$

Then

$$\boxed{\mathbf{A}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \hat{\mathbf{r}}}{r^2},} \tag{5.85}$$

where $\mathbf{m}$ is the **magnetic dipole moment**:

$$\boxed{\mathbf{m} \equiv I \int d\mathbf{a} = I\mathbf{a}.} \tag{5.86}$$

Here $\mathbf{a}$ is the "vector area" of the loop (Prob. 1.62); if the loop is *flat*, $\mathbf{a}$ is the ordinary area enclosed, with the direction assigned by the usual right-hand rule (fingers in the direction of the current).

Example 5.13. Find the magnetic dipole moment of the "bookend-shaped" loop shown in Fig. 5.52. All sides have length w, and it carries a current I.

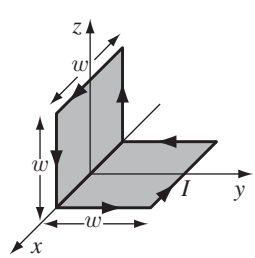


FIGURE 5.52

Solution

This wire could be considered the superposition of two plane square loops (Fig. 5.53). The "extra" sides (AB) cancel when the two are put together, since the currents flow in opposite directions. The net magnetic dipole moment is

$$\mathbf{m} = Iw^2\,\hat{\mathbf{y}} + Iw^2\,\hat{\mathbf{z}};$$

its magnitude is $\sqrt{2}Iw^2$, and it points along the 45° line $z = y$.

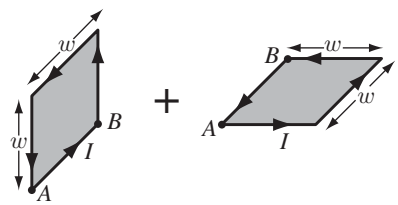


FIGURE 5.53

It is clear from Eq. 5.86 that the magnetic dipole moment is independent of the choice of origin. You may remember that the *electric* dipole moment is independent of the origin only when the total charge vanishes (Sect. 3.4.3). Since the *magnetic* monopole moment is *always* zero, it is not really surprising that the magnetic dipole moment is always independent of origin.

Although the dipole term *dominates* the multipole expansion (unless $\mathbf{m} = 0$) and thus offers a good approximation to the true potential, it is not ordinarily the *exact* potential; there will be quadrupole, octopole, and higher contributions. You might ask, is it possible to devise a current distribution whose potential is "pure" dipole—for which Eq. 5.85 is *exact*? Well, yes and no: like the electrical analog, it can be done, but the model is a bit contrived. To begin with, you must take an *infinitesimally small* loop at the origin, but then, in order to keep the dipole moment finite, you have to crank the current up to infinity, with the product $m = Ia$ held fixed. In practice, the dipole potential is a suitable approximation whenever the distance r greatly exceeds the size of the loop.

The magnetic *field* of a (perfect) dipole is easiest to calculate if we put $\mathbf{m}$ at the origin and let it point in the z-direction (Fig. 5.54). According to Eq. 5.85, the potential at point (r, θ, ϕ) is

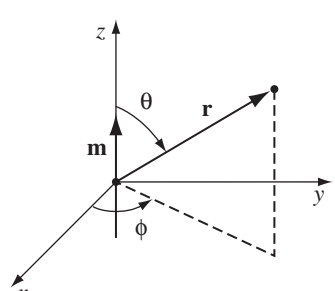
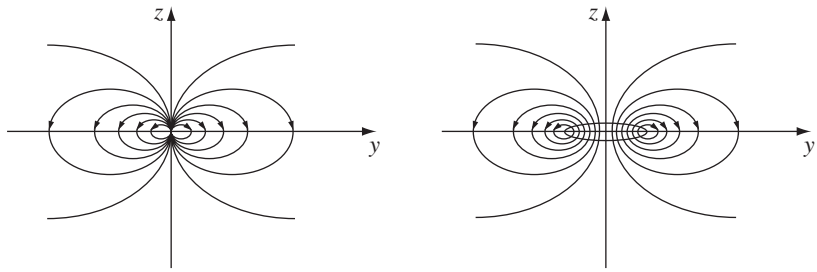


FIGURE 5.54



(a) Field of a "pure" dipole (b) Field of a "physical" dipole

FIGURE 5.55

$$\mathbf{A}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{m\sin\theta}{r^2}\,\hat{\boldsymbol{\phi}}, \tag{5.87}$$

and hence

$$\mathbf{B}_{\text{dip}}(\mathbf{r}) = \boldsymbol{\nabla}\times\mathbf{A} = \frac{\mu_0 m}{4\pi r^3}(2\cos\theta\,\hat{\mathbf{r}} + \sin\theta\,\hat{\boldsymbol{\theta}}). \tag{5.88}$$

Surprisingly, this is *identical* in structure to the field of an *electric* dipole (Eq. 3.103)! (Up close, however, the field of a *physical* magnetic dipole—a small current loop—looks quite different from the field of a physical electric dipole—plus and minus charges a short distance apart. Compare Fig. 5.55 with Fig. 3.37.)

• **Problem 5.34** Show that the magnetic field of a dipole can be written in coordinate-free form:

$$\boxed{\mathbf{B}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{1}{r^3}\big[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}\big].} \tag{5.89}$$

Problem 5.35 A circular loop of wire, with radius R, lies in the xy plane (centered at the origin) and carries a current I running counterclockwise as viewed from the positive z axis.

(a) What is its magnetic dipole moment?

(b) What is the (approximate) magnetic field at points far from the origin?

(c) Show that, for points on the z axis, your answer is consistent with the *exact* field (Ex. 5.6), when $z \gg R$.

Problem 5.36 Find the exact magnetic field a distance z above the center of a square loop of side w, carrying a current I. Verify that it reduces to the field of a dipole, with the appropriate dipole moment, when $z \gg w$.

Problem 5.37

(a) A phonograph record of radius R, carrying a uniform surface charge σ, is rotating at constant angular velocity ω. Find its magnetic dipole moment.

(b) Find the magnetic dipole moment of the spinning spherical shell in Ex. 5.11. Show that for points $r > R$ the potential is that of a perfect dipole.

Problem 5.38 I worked out the multipole expansion for the vector potential of a *line* current because that's the most common type, and in some respects the easiest to handle. For a *volume* current $\mathbf{J}$:

(a) Write down the multipole expansion, analogous to Eq. 5.80.

(b) Write down the monopole potential, and prove that it vanishes.

(c) Using Eqs. 1.107 and 5.86, show that the dipole moment can be written

$$\mathbf{m} = \frac{1}{2}\int (\mathbf{r} \times \mathbf{J})\, d\tau. \tag{5.90}$$

More Problems on Chapter 5

Problem 5.39 Analyze the motion of a particle (charge q, mass m) in the magnetic field of a long straight wire carrying a steady current I.

(a) Is its kinetic energy conserved?

(b) Find the force on the particle, in cylindrical coordinates, with I along the z axis.

(c) Obtain the equations of motion.

(d) Suppose $\dot{z}$ is constant. Describe the motion.

Problem 5.40 It may have occurred to you that since parallel currents attract, the current within a single wire should contract into a tiny concentrated stream along the axis. Yet in practice the current typically distributes itself quite uniformly over the wire. How do you account for this? If the positive charges (density ρ_+) are "nailed down," and the negative charges (density ρ_-) move at speed v (and none of these depends on the distance from the axis), show that $\rho_- = -\rho_+\gamma^2$, where $\gamma \equiv 1/\sqrt{1-(v/c)^2}$ and $c^2 = 1/\mu_0\epsilon_0$. If the wire as a whole is neutral, where is the compensating charge located?[22] [Notice that for typical velocities (see Prob. 5.20), the two charge densities are essentially unchanged by the current (since $\gamma \approx 1$). In **plasmas**, however, where the positive charges are *also* free to move, this so-called **pinch effect** can be very significant.]

Problem 5.41 A current I flows to the right through a rectangular bar of conducting material, in the presence of a uniform magnetic field $\mathbf{B}$ pointing out of the page (Fig. 5.56).

(a) If the moving charges are *positive*, in which direction are they deflected by the magnetic field? This deflection results in an accumulation of charge on the

[22] For further discussion, see D. C. Gabuzda, *Am. J. Phys.* **61**, 360 (1993).

upper and lower surfaces of the bar, which in turn produces an electric force to counteract the magnetic one. Equilibrium occurs when the two exactly cancel. (This phenomenon is known as the **Hall effect**.)

(b) Find the resulting potential difference (the **Hall voltage**) between the top and bottom of the bar, in terms of B, v (the speed of the charges), and the relevant dimensions of the bar.[23]

(c) How would your analysis change if the moving charges were *negative*? [The Hall effect is the classic way of determining the sign of the mobile charge carriers in a material.]

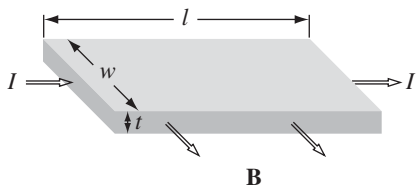


FIGURE 5.56

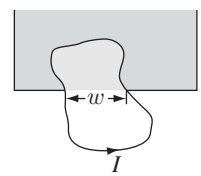


FIGURE 5.57

Problem 5.42 A plane wire loop of irregular shape is situated so that part of it is in a uniform magnetic field **B** (in Fig. 5.57 the field occupies the shaded region, and points perpendicular to the plane of the loop). The loop carries a current I. Show that the net magnetic force on the loop is $F = IBw$, where w is the chord subtended. Generalize this result to the case where the magnetic field region itself has an irregular shape. What is the direction of the force?

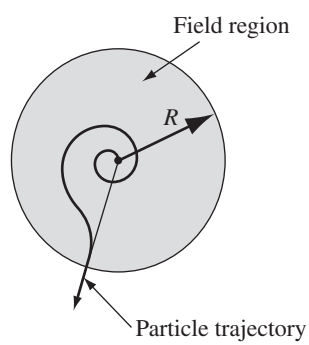


FIGURE 5.58

Problem 5.43 A circularly symmetrical magnetic field (**B** depends only on the distance from the axis), pointing perpendicular to the page, occupies the shaded region in Fig. 5.58. If the total flux ($\int \mathbf{B} \cdot d\mathbf{a}$) is zero, show that a charged particle that starts out at the center will emerge from the field region on a *radial* path (provided

[23]The potential *within* the bar makes an interesting boundary-value problem. See M. J. Moelter, J. Evans, G. Elliot, and M. Jackson, *Am. J. Phys.* **66**, 668 (1998).

it escapes at all). On the reverse trajectory, a particle fired at the center from outside will hit its target (if it has sufficient energy), though it may follow a weird route getting there. [*Hint:* Calculate the total angular momentum acquired by the particle, using the Lorentz force law.]

Problem 5.44 Calculate the magnetic force of attraction between the northern and southern hemispheres of a spinning charged spherical shell (Ex. 5.11). [*Answer:* $(\pi/4)\mu_0\sigma^2\omega^2R^4$.]

! **Problem 5.45** Consider the motion of a particle with mass m and electric charge q_e in the field of a (hypothetical) stationary magnetic *monopole* q_m at the origin:

$$\mathbf{B} = \frac{\mu_0}{4\pi}\frac{q_m}{r^2}\,\hat{\mathbf{r}}.$$

(a) Find the acceleration of q_e, expressing your answer in terms of q, q_m, m, $\mathbf{r}$ (the position of the particle), and $\mathbf{v}$ (its velocity).

(b) Show that the speed $v = |\mathbf{v}|$ is a constant of the motion.

(c) Show that the vector quantity

$$\mathbf{Q} \equiv m(\mathbf{r}\times\mathbf{v}) - \frac{\mu_0 q_e q_m}{4\pi}\,\hat{\mathbf{r}}$$

is a constant of the motion. [*Hint:* differentiate it with respect to time, and prove—using the equation of motion from (a)—that the derivative is zero.]

(d) Choosing spherical coordinates (r, θ, ϕ), with the polar (z) axis along $\mathbf{Q}$,

(i) calculate $\mathbf{Q}\cdot\hat{\boldsymbol{\phi}}$, and show that θ is a constant of the motion (so q_e moves on the surface of a cone—something Poincaré first discovered in 1896)[24];

(ii) calculate $\mathbf{Q}\cdot\hat{\mathbf{r}}$, and show that the magnitude of $\mathbf{Q}$ is

$$Q = \frac{\mu_0}{4\pi}\left|\frac{q_e q_m}{\cos\theta}\right|;$$

(iii) calculate $\mathbf{Q}\cdot\hat{\boldsymbol{\theta}}$, show that

$$\frac{d\phi}{dt} = \frac{k}{r^2},$$

and determine the constant k.

(e) By expressing v^2 in spherical coordinates, obtain the equation for the trajectory, in the form

$$\frac{dr}{d\phi} = f(r)$$

(that is: determine the function $f(r)$).

(f) Solve this equation for $r(\phi)$.

[24]In point of fact, the charge follows a *geodesic* on the cone. The original paper is H. Poincaré, *Comptes rendus de l'Academie des Sciences* **123**, 530 (1896); for a more modern treatment, see B. Rossi and S. Olbert, *Introduction to the Physics of Space* (New York: McGraw-Hill, 1970).

! **Problem 5.46** Use the Biot-Savart law (most conveniently in the form of Eq. 5.42 appropriate to surface currents) to find the field inside and outside an infinitely long solenoid of radius R, with n turns per unit length, carrying a steady current I.

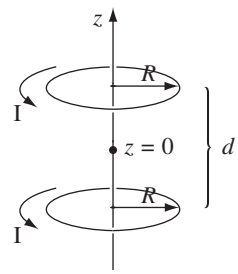


FIGURE 5.59

Problem 5.47 The magnetic field on the axis of a circular current loop (Eq. 5.41) is far from uniform (it falls off sharply with increasing z). You can produce a more nearly uniform field by using *two* such loops a distance d apart (Fig. 5.59).

(a) Find the field (B) as a function of z, and show that $\partial B/\partial z$ is zero at the point midway between them ($z = 0$).

(b) If you pick d just right, the *second* derivative of B will *also* vanish at the midpoint. This arrangement is known as a **Helmholtz coil**; it's a convenient way of producing relatively uniform fields in the laboratory. Determine d such that $\partial^2 B/\partial z^2 = 0$ at the midpoint, and find the resulting magnetic field at the center. [*Answer:* $8\mu_0 I/5\sqrt{5}R$]

Problem 5.48 Use Eq. 5.41 to obtain the magnetic field on the axis of the rotating disk in Prob. 5.37(a). Show that the dipole field (Eq. 5.88), with the dipole moment you found in Prob. 5.37, is a good approximation if $z \gg R$.

Problem 5.49 Suppose you wanted to find the field of a circular loop (Ex. 5.6) at a point $\mathbf{r}$ that is *not* directly above the center (Fig. 5.60). You might as well choose your axes so that $\mathbf{r}$ lies in the yz plane at $(0, y, z)$. The source point is $(R\cos\phi', R\sin\phi', 0)$, and ϕ' runs from 0 to 2π. Set up the integrals[25] from which you could calculate B_x, B_y, and B_z, and evaluate B_x explicitly.

Problem 5.50 Magnetostatics treats the "source current" (the one that sets up the field) and the "recipient current" (the one that experiences the force) so asymmetrically that it is by no means obvious that the magnetic force between two current loops is consistent with Newton's third law. Show, starting with the Biot-Savart law (Eq. 5.34) and the Lorentz force law (Eq. 5.16), that the force on loop 2 due to loop 1 (Fig. 5.61) can be written as

$$\mathbf{F}_2 = -\frac{\mu_0}{4\pi} I_1 I_2 \oint \oint \frac{\hat{\boldsymbol{\imath}}}{\imath^2}\, d\mathbf{l}_1 \cdot d\mathbf{l}_2. \tag{5.91}$$

[25]These are **elliptic integrals**—see R. H. Good, *Eur. J. Phys.* **22**, 119 (2001).

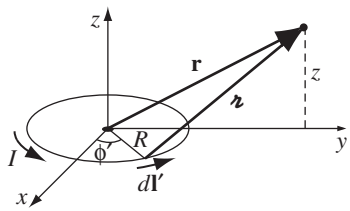


FIGURE 5.60

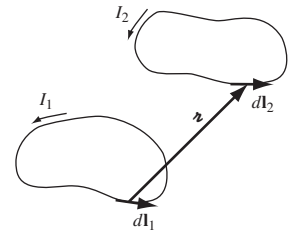


FIGURE 5.61

In this form, it is clear that $\mathbf{F}_2 = -\mathbf{F}_1$, since $\hat{\boldsymbol{\mathcal{r}}}$ changes direction when the roles of 1 and 2 are interchanged. (If you seem to be getting an "extra" term, it will help to note that $d\mathbf{l}_2 \cdot \hat{\boldsymbol{\mathcal{r}}} = d\mathcal{r}$.)

Problem 5.51 Consider a *plane* loop of wire that carries a steady current I; we want to calculate the magnetic field at a point in the plane. We might as well take that point to be the origin (it could be inside or outside the loop). The shape of the wire is given, in polar coordinates, by a specified function $r(\theta)$ (Fig. 5.62).

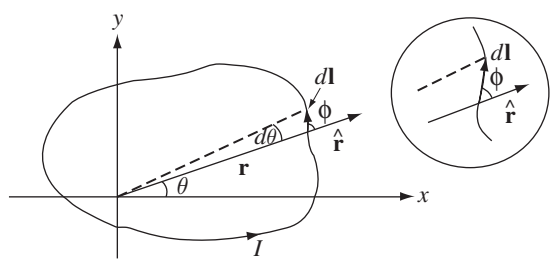


FIGURE 5.62

(a) Show that the magnitude of the field is[26]

$$B = \frac{\mu_0 I}{4\pi} \oint \frac{d\theta}{r}. \tag{5.92}$$

[*Hint:* Start with the Biot-Savart law; note that $\boldsymbol{\mathcal{r}} = -\mathbf{r}$, and $d\mathbf{l} \times \hat{\mathbf{r}}$ points perpendicular to the plane; show that $|d\mathbf{l} \times \hat{\mathbf{r}}| = dl\ \sin\phi = r\, d\theta$.]

(b) Test this formula by calculating the field at the center of a circular loop.

(c) The "lituus spiral" is defined by

$$r(\theta) = \frac{a}{\sqrt{\theta}}, \quad (0 < \theta \leq 2\pi)$$

(for some constant a). Sketch this figure, and complete the loop with a straight segment along the x axis. What is the magnetic field at the origin?

[26] J. A. Miranda, *Am. J. Phys.* **68**, 254 (2000).

(d) For a conic section with focus at the origin,

$$r(\theta) = \frac{p}{1 + e\cos\theta},$$

where p is the semilatus rectum (the y intercept) and e is the eccentricity ($e = 0$ for a circle, $0 < e < 1$ for an ellipse, $e = 1$ for a parabola). Show that the field is

$$B = \frac{\mu_0 I}{2p}$$

regardless of the eccentricity.[27]

Problem 5.52

(a) One way to fill in the "missing link" in Fig. 5.48 is to exploit the analogy between the defining equations for $\mathbf{A}$ (viz. $\nabla \cdot \mathbf{A} = 0$, $\nabla \times \mathbf{A} = \mathbf{B}$) and Maxwell's equations for $\mathbf{B}$ (viz. $\nabla \cdot \mathbf{B} = 0$, $\nabla \times \mathbf{B} = \mu_0\mathbf{J}$). Evidently $\mathbf{A}$ depends on $\mathbf{B}$ in exactly the same way that $\mathbf{B}$ depends on $\mu_0\mathbf{J}$ (to wit: the Biot-Savart law). Use this observation to write down the formula for $\mathbf{A}$ in terms of $\mathbf{B}$.

(b) The electrical analog to your result in (a) is

$$V(\mathbf{r}) = -\frac{1}{4\pi}\int \frac{\mathbf{E}(\mathbf{r}')\cdot\hat{\boldsymbol{\imath}}}{\imath^2}\,d\tau'.$$

Derive it, by exploiting the appropriate analogy.

! **Problem 5.53** Another way to fill in the "missing link" in Fig. 5.48 is to look for a magnetostatic analog to Eq. 2.21. The obvious candidate would be

$$\mathbf{A}(\mathbf{r}) = \int_{\mathcal{O}}^{\mathbf{r}} (\mathbf{B} \times d\mathbf{l}).$$

(a) Test this formula for the simplest possible case—uniform $\mathbf{B}$ (use the origin as your reference point). Is the result consistent with Prob. 5.25? You could cure this problem by throwing in a factor of $\frac{1}{2}$, but the flaw in this equation runs deeper.

(b) Show that $\int(\mathbf{B} \times d\mathbf{l})$ is *not* independent of path, by calculating $\oint(\mathbf{B} \times d\mathbf{l})$ around the rectangular loop shown in Fig. 5.63.

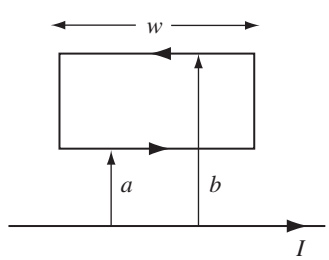


FIGURE 5.63

[27] C. Christodoulides, *Am. J. Phys.* **77**, 1195 (2009).

As far as I know,[28] the best one can do along these lines is the pair of equations

(i) $V(\mathbf{r}) = -\mathbf{r} \cdot \int_0^1 \mathbf{E}(\lambda \mathbf{r})\, d\lambda$,

(ii) $\mathbf{A}(\mathbf{r}) = -\mathbf{r} \times \int_0^1 \lambda \mathbf{B}(\lambda \mathbf{r})\, d\lambda$.

[Equation (i) amounts to selecting a *radial* path for the integral in Eq. 2.21; equation (ii) constitutes a more "symmetrical" solution to Prob. 5.31.]

(c) Use (ii) to find the vector potential for *uniform* $\mathbf{B}$.

(d) Use (ii) to find the vector potential of an infinite straight wire carrying a steady current I. Does (ii) automatically satisfy $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$? [*Answer:* $(\mu_0 I/2\pi s)(z\,\hat{\mathbf{s}} - s\,\hat{\mathbf{z}})$]

Problem 5.54

(a) Construct the scalar potential $U(\mathbf{r})$ for a "pure" magnetic dipole $\mathbf{m}$.

(b) Construct a scalar potential for the spinning spherical shell (Ex. 5.11). [*Hint:* for $r > R$ this is a pure dipole field, as you can see by comparing Eqs. 5.69 and 5.87.]

(c) Try doing the same for the interior of a *solid* spinning sphere. [*Hint:* If you solved Prob. 5.30, you already know the *field*; set it equal to $-\boldsymbol{\nabla} U$, and solve for U. What's the trouble?]

Problem 5.55 Just as $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$ allows us to express $\mathbf{B}$ as the curl of a vector potential ($\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$), so $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$ permits us to write $\mathbf{A}$ itself as the curl of a "higher" potential: $\mathbf{A} = \boldsymbol{\nabla} \times \mathbf{W}$. (And this hierarchy can be extended ad infinitum.)

(a) Find the general formula for $\mathbf{W}$ (as an integral over $\mathbf{B}$), which holds when $\mathbf{B} \to \mathbf{0}$ at ∞.

(b) Determine $\mathbf{W}$ for the case of a *uniform* magnetic field $\mathbf{B}$. [*Hint:* see Prob. 5.25.]

(c) Find $\mathbf{W}$ inside and outside an infinite solenoid. [*Hint:* see Ex. 5.12.]

Problem 5.56 Prove the following uniqueness theorem: If the current density $\mathbf{J}$ is specified throughout a volume $\mathcal{V}$, and *either* the potential $\mathbf{A}$ *or* the magnetic field $\mathbf{B}$ is specified on the surface $\mathcal{S}$ bounding $\mathcal{V}$, then the magnetic field itself is uniquely determined throughout $\mathcal{V}$. [*Hint:* First use the divergence theorem to show that

$$\int \{(\boldsymbol{\nabla} \times \mathbf{U}) \cdot (\boldsymbol{\nabla} \times \mathbf{V}) - \mathbf{U} \cdot [\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{V})]\}\, d\tau = \oint [\mathbf{U} \times (\boldsymbol{\nabla} \times \mathbf{V})] \cdot d\mathbf{a},$$

for arbitrary vector functions $\mathbf{U}$ and $\mathbf{V}$.]

Problem 5.57 A magnetic dipole $\mathbf{m} = -m_0\,\hat{\mathbf{z}}$ is situated at the origin, in an otherwise uniform magnetic field $\mathbf{B} = B_0\,\hat{\mathbf{z}}$. Show that there exists a spherical surface, centered at the origin, through which no magnetic field lines pass. Find the radius of this sphere, and sketch the field lines, inside and out.

[28] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds*, Section 4.5 (New York: Macmillan, 1968).

Problem 5.58 A thin uniform donut, carrying charge Q and mass M, rotates about its axis as shown in Fig. 5.64.

(a) Find the ratio of its magnetic dipole moment to its angular momentum. This is called the **gyromagnetic ratio** (or **magnetomechanical ratio**).

(b) What is the gyromagnetic ratio for a uniform spinning sphere? [This requires no new calculation; simply decompose the sphere into infinitesimal rings, and apply the result of part (a).]

(c) According to quantum mechanics, the angular momentum of a spinning electron is $\frac{1}{2}\hbar$, where $\hbar$ is Planck's constant. What, then, is the electron's magnetic dipole moment, in $\mathrm{A}\cdot m^2$? [This semiclassical value is actually off by a factor of almost exactly 2. Dirac's relativistic electron theory got the 2 right, and Feynman, Schwinger, and Tomonaga later calculated tiny further corrections. The determination of the electron's magnetic dipole moment remains the finest achievement of quantum electrodynamics, and exhibits perhaps the most stunningly precise agreement between theory and experiment in all of physics. Incidentally, the quantity $(e\hbar/2m)$, where e is the charge of the electron and m is its mass, is called the **Bohr magneton**.]

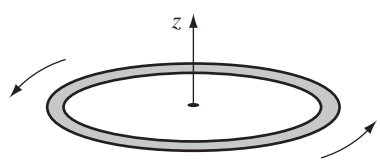


FIGURE 5.64

• **Problem 5.59**

(a) Prove that the average magnetic field, over a sphere of radius R, due to steady currents inside the sphere, is

$$\mathbf{B}_{\text{ave}} = \frac{\mu_0}{4\pi}\frac{2\mathbf{m}}{R^3}, \tag{5.93}$$

where $\mathbf{m}$ is the total dipole moment of the sphere. Contrast the electrostatic result, Eq. 3.105. [This is tough, so I'll give you a start:

$$\mathbf{B}_{\text{ave}} = \frac{1}{\frac{4}{3}\pi R^3}\int \mathbf{B}\, d\tau.$$

Write $\mathbf{B}$ as $(\boldsymbol{\nabla}\times\mathbf{A})$, and apply Prob. 1.61(b). Now put in Eq. 5.65, and do the surface integral first, showing that

$$\int \frac{1}{\mathcal{r}}\, d\mathbf{a} = \frac{4}{3}\pi\mathbf{r}'$$

(see Fig. 5.65). Use Eq. 5.90, if you like.]

(b) Show that the average magnetic field due to steady currents *outside* the sphere is the same as the field they produce at the center.

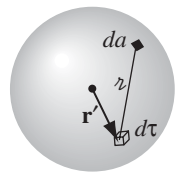


FIGURE 5.65

Problem 5.60 A uniformly charged solid sphere of radius R carries a total charge Q, and is set spinning with angular velocity ω about the z axis.

(a) What is the magnetic dipole moment of the sphere?

(b) Find the average magnetic field within the sphere (see Prob. 5.59).

(c) Find the approximate vector potential at a point (r, θ) where $r \gg R$.

(d) Find the *exact* potential at a point (r, θ) outside the sphere, and check that it is consistent with (c). [*Hint:* refer to Ex. 5.11.]

(e) Find the magnetic field at a point (r, θ) *inside* the sphere (Prob. 5.30), and check that it is consistent with (b).

Problem 5.61 Using Eq. 5.88, calculate the average magnetic field of a dipole over a sphere of radius R centered at the origin. Do the angular integrals first. Compare your answer with the general theorem in Prob. 5.59. Explain the discrepancy, and indicate how Eq. 5.89 can be corrected to resolve the ambiguity at $r = 0$. (If you get stuck, refer to Prob. 3.48.)

Evidently the *true* field of a magnetic dipole is[29]

$$\mathbf{B}_{\text{dip}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{1}{r^3}\left[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}\right] + \frac{2\mu_0}{3}\mathbf{m}\delta^3(\mathbf{r}). \tag{5.94}$$

Compare the electrostatic analog, Eq. 3.106.

Problem 5.62 A thin glass rod of radius R and length L carries a uniform surface charge σ. It is set spinning about its axis, at an angular velocity ω. Find the magnetic field at a distance $s \gg R$ from the axis, in the xy plane (Fig. 5.66). [*Hint:* treat it as a stack of magnetic dipoles.] [*Answer:* $\mu_0\omega\sigma L R^3/4[s^2 + (L/2)^2]^{3/2}$]

[29]The delta-function term is responsible for the **hyperfine splitting** in atomic spectra—see, for example, D. J. Griffiths, *Am. J. Phys.* **50**, 698 (1982).

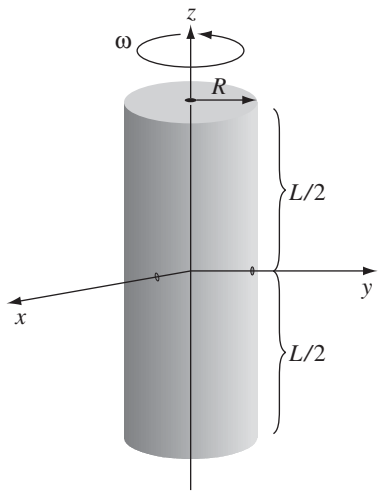


FIGURE 5.66

CHAPTER

6 Magnetic Fields in Matter

6.1 ■ MAGNETIZATION

6.1.1 ■ Diamagnets, Paramagnets, Ferromagnets

If you ask the average person what "magnetism" is, you will probably be told about refrigerator decorations, compass needles, and the North Pole—none of which has any obvious connection with moving charges or current-carrying wires. Yet all magnetic phenomena are due to electric charges in motion, and in fact, if you could examine a piece of magnetic material on an atomic scale you *would* find tiny currents: electrons orbiting around nuclei and spinning about their axes. For macroscopic purposes, these current loops are so small that we may treat them as magnetic dipoles. Ordinarily, they cancel each other out because of the random orientation of the atoms. But when a magnetic field is applied, a net alignment of these magnetic dipoles occurs, and the medium becomes magnetically polarized, or **magnetized**.

Unlike electric polarization, which is almost always in the same direction as **E**, some materials acquire a magnetization *parallel* to **B** (**paramagnets**) and some *opposite* to **B** (**diamagnets**). A few substances (called **ferromagnets**, in deference to the most common example, iron) retain their magnetization even after the external field has been removed—for these, the magnetization is not determined by the *present* field but by the whole magnetic "history" of the object. Permanent magnets made of iron are the most familiar examples of magnetism, but from a theoretical point of view they are the most complicated; I'll save ferromagnetism for the end of the chapter, and begin with qualitative models of paramagnetism and diamagnetism.

6.1.2 ■ Torques and Forces on Magnetic Dipoles

A magnetic dipole experiences a torque in a magnetic field, just as an electric dipole does in an electric field. Let's calculate the torque on a rectangular current loop in a uniform field **B**. (Since any current loop could be built up from infinitesimal rectangles, with all the "internal" sides canceling, as indicated in Fig. 6.1, there is no real loss of generality here; but if you prefer to start from scratch with an arbitrary shape, see Prob. 6.2.) Center the loop at the origin, and tilt it an angle θ from the z axis towards the y axis (Fig. 6.2). Let **B** point in the z direction. The forces on the two sloping sides cancel (they tend to *stretch* the loop, but they don't

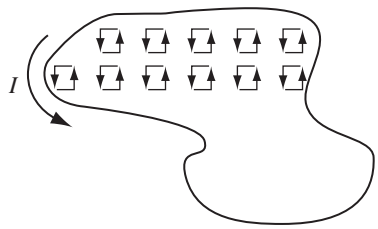


FIGURE 6.1

rotate it). The forces on the "horizontal" sides are likewise equal and opposite (so the net *force* on the loop is zero), but they do generate a torque:

$$\mathbf{N} = aF \sin\theta\, \hat{\mathbf{x}}.$$

The magnitude of the force on each of these segments is

$$F = IbB,$$

and therefore

$$\mathbf{N} = IabB \sin\theta\, \hat{\mathbf{x}} = mB \sin\theta\, \hat{\mathbf{x}},$$

or

$$\boxed{\mathbf{N} = \mathbf{m} \times \mathbf{B},} \tag{6.1}$$

where $m = Iab$ is the magnetic dipole moment of the loop. Equation 6.1 gives the torque on any localized current distribution, in the presence of a *uniform* field; in a *nonuniform* field it is the exact torque (about the center) for a *perfect* dipole of infinitesimal size.

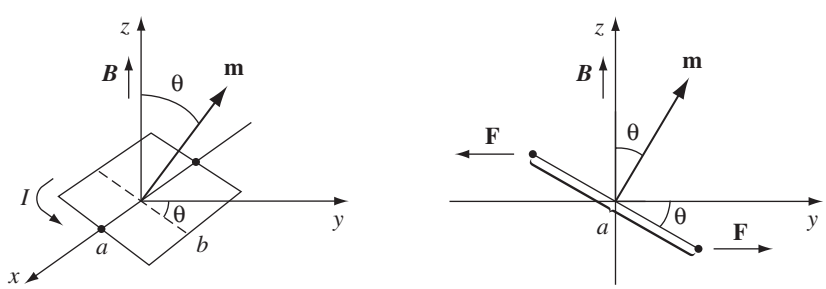


FIGURE 6.2

Notice that Eq. 6.1 is identical in form to the electrical analog, Eq. 4.4: $\mathbf{N} = \mathbf{p} \times \mathbf{E}$. In particular, the torque is again in such a direction as to line the dipole up *parallel* to the field. It is this torque that accounts for **paramagnetism**. Since every electron constitutes a magnetic dipole (picture it, if you wish, as a tiny spinning sphere of charge), you might expect paramagnetism to be a universal phenomenon. Actually, quantum mechanics (specifically, the Pauli exclusion principle) tends to lock the electrons within a given atom together in pairs with opposing spins,[1] and this effectively neutralizes the torque on the combination. As a result, paramagnetism most often occurs in atoms or molecules with an odd number of electrons, where the "extra" unpaired member is subject to the magnetic torque. Even here, the alignment is far from complete, since random thermal collisions tend to destroy the order.

In a uniform field, the net *force* on a current loop is zero:

$$\mathbf{F} = I \oint (d\mathbf{l} \times \mathbf{B}) = I \left(\oint d\mathbf{l} \right) \times \mathbf{B} = \mathbf{0};$$

the constant $\mathbf{B}$ comes outside the integral, and the net displacement $\oint d\mathbf{l}$ around a closed loop vanishes. In a *nonuniform* field this is no longer the case. For example, suppose a circular wire ring of radius R, carrying a current I, is suspended above a short solenoid in the "fringing" region (Fig. 6.3). Here $\mathbf{B}$ has a radial component, and there is a net downward force on the loop (Fig. 6.4):

$$F = 2\pi I R B \cos\theta. \tag{6.2}$$

For an *infinitesimal* loop, with dipole moment $\mathbf{m}$, in a field $\mathbf{B}$, the force is

$$\boxed{\mathbf{F} = \nabla(\mathbf{m} \cdot \mathbf{B})} \tag{6.3}$$

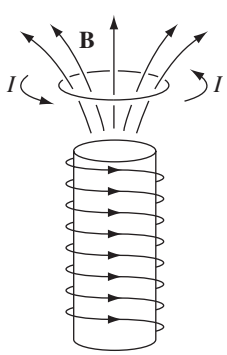


FIGURE 6.3

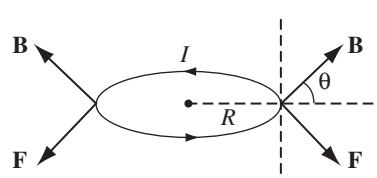
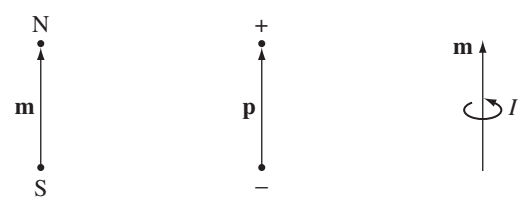


FIGURE 6.4

[1] This is not always true for the outermost electrons in unfilled shells.

(see Prob. 6.4). Once again the magnetic formula is identical to its electrical "twin," if we write the latter in the form $\mathbf{F} = \nabla(\mathbf{p} \cdot \mathbf{E})$. (See footnote to Eq. 4.5.)

If you're starting to get a sense of *déjà vu*, perhaps you will have more respect for those early physicists who thought magnetic dipoles consisted of positive and negative magnetic "charges" (north and south "poles," they called them), separated by a small distance, just like electric dipoles (Fig. 6.5(a)). They wrote down a "Coulomb's law" for the attraction and repulsion of these poles, and developed the whole of magnetostatics in exact analogy to electrostatics. It's not a bad model, for many purposes—it gives the correct field of a dipole (at least, away from the origin), the right torque on a dipole (at least, on a *stationary* dipole), and the proper force on a dipole (at least, in the absence of external currents). But it's bad physics, because *there's no such thing* as a single magnetic north pole or south pole. If you break a bar magnet in half, you don't get a north pole in one hand and a south pole in the other; you get two complete magnets. Magnetism is *not* due to magnetic monopoles, but rather to *moving electric charges*; magnetic dipoles are tiny current loops (Fig. 6.5(c)), and it's an extraordinary thing, really, that the formulas involving $\mathbf{m}$ bear any resemblance to the corresponding formulas for $\mathbf{p}$. Sometimes it is easier to think in terms of the "Gilbert" model of a magnetic dipole (separated monopoles), instead of the physically correct "Ampère" model (current loop). Indeed, this picture occasionally offers a quick and clever solution to an otherwise cumbersome problem (you just copy the corresponding result from electrostatics, changing $\mathbf{p}$ to $\mathbf{m}$, $1/\epsilon_0$ to μ_0, and $\mathbf{E}$ to $\mathbf{B}$). But whenever the *close-up* features of the dipole come into play, the two models can yield strikingly different answers. My advice is to use the Gilbert model, if you like, to get an intuitive "feel" for a problem, but never rely on it for quantitative results.



(a) Magnetic dipole (Gilbert model) (b) Electric dipole (c) Magnetic dipole (Ampère model)

FIGURE 6.5

Problem 6.1 Calculate the torque exerted on the square loop shown in Fig. 6.6, due to the circular loop (assume r is much larger than a or b). If the square loop is free to rotate, what will its equilibrium orientation be?

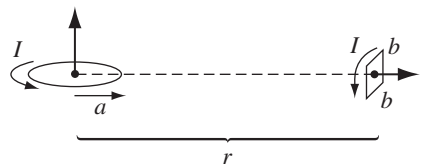


FIGURE 6.6

Problem 6.2 Starting from the Lorentz force law, in the form of Eq. 5.16, show that the torque on *any* steady current distribution (not just a square loop) in a uniform field $\mathbf{B}$ is $\mathbf{m} \times \mathbf{B}$.

Problem 6.3 Find the force of attraction between two magnetic dipoles, $\mathbf{m}_1$ and $\mathbf{m}_2$, oriented as shown in Fig. 6.7, a distance r apart, (a) using Eq. 6.2, and (b) using Eq. 6.3.

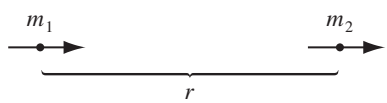


FIGURE 6.7

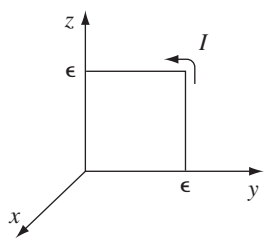


FIGURE 6.8

Problem 6.4 Derive Eq. 6.3. [Here's one way to do it: Assume the dipole is an infinitesimal square, of side ϵ (if it's not, chop it up into squares, and apply the argument to each one). Choose axes as shown in Fig. 6.8, and calculate $\mathbf{F} = I \int (d\mathbf{l} \times \mathbf{B})$ along each of the four sides. Expand $\mathbf{B}$ in a Taylor series—on the right side, for instance,

$$\mathbf{B} = \mathbf{B}(0, \epsilon, z) \cong \mathbf{B}(0, 0, z) + \epsilon \left.\frac{\partial \mathbf{B}}{\partial y}\right|_{(0,0,z)}.$$

For a more sophisticated method, see Prob. 6.22.]

Problem 6.5 A uniform current density $\mathbf{J} = J_0\,\hat{\mathbf{z}}$ fills a slab straddling the yz plane, from $x = -a$ to $x = +a$. A magnetic dipole $\mathbf{m} = m_0\,\hat{\mathbf{x}}$ is situated at the origin.

(a) Find the force on the dipole, using Eq. 6.3.

(b) Do the same for a dipole pointing in the y direction: $\mathbf{m} = m_0\hat{\mathbf{y}}$.

(c) In the *electrostatic* case, the expressions $\mathbf{F} = \nabla(\mathbf{p} \cdot \mathbf{E})$ and $\mathbf{F} = (\mathbf{p} \cdot \nabla)\mathbf{E}$ are equivalent (prove it), but this is *not* the case for the magnetic analogs (explain why). As an example, calculate $(\mathbf{m} \cdot \nabla)\mathbf{B}$ for the configurations in (a) and (b).

6.1.3 ■ Effect of a Magnetic Field on Atomic Orbits

Electrons not only *spin*; they also *revolve* around the nucleus—for simplicity, let's assume the orbit is a circle of radius R (Fig. 6.9). Although technically this orbital motion does not constitute a steady current, in practice the period $T = 2\pi R/v$ is so short that unless you blink awfully fast, it's going to *look* like a steady current:

$$I = \frac{-e}{T} = -\frac{ev}{2\pi R}.$$

(The minus sign accounts for the negative charge of the electron.) Accordingly, the orbital dipole moment ($I\pi R^2$) is

$$\mathbf{m} = -\frac{1}{2}evR\,\hat{\mathbf{z}}. \tag{6.4}$$

Like any other magnetic dipole, this one is subject to a torque ($\mathbf{m} \times \mathbf{B}$) when you turn on a magnetic field. But it's a lot harder to tilt the entire orbit than it is the spin, so the orbital contribution to paramagnetism is small. There is, however, a more significant effect on the orbital motion: The electron *speeds up* or *slows down*, depending on the orientation of $\mathbf{B}$. For whereas the centripetal acceleration v^2/R is ordinarily sustained by electrical forces alone,[2]

$$\frac{1}{4\pi\epsilon_0}\frac{e^2}{R^2} = m_e\frac{v^2}{R}, \tag{6.5}$$

in the presence of a magnetic field there is an additional force, $-e(\mathbf{v} \times \mathbf{B})$. For the sake of argument, let's say that $\mathbf{B}$ is perpendicular to the plane of the orbit, as shown in Fig. 6.10; then

$$\frac{1}{4\pi\epsilon_0}\frac{e^2}{R^2} + e\bar{v}B = m_e\frac{\bar{v}^2}{R}. \tag{6.6}$$

Under these conditions, the new speed $\bar{v}$ is *greater* than v:

$$e\bar{v}B = \frac{m_e}{R}(\bar{v}^2 - v^2) = \frac{m_e}{R}(\bar{v} + v)(\bar{v} - v),$$

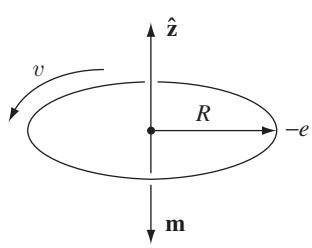


FIGURE 6.9

[2]To avoid confusion with the magnetic dipole moment m, I'll write the electron mass with subscript: m_e.

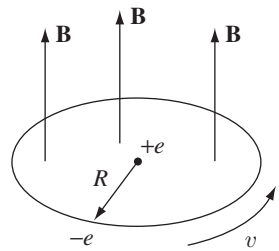


FIGURE 6.10

or, assuming the change $\Delta v = \bar{v} - v$ is small,

$$\Delta v = \frac{eRB}{2m_e}. \tag{6.7}$$

When **B** is turned on, then, the electron speeds up.[3]

A change in orbital speed means a change in the dipole moment (Eq. 6.4):

$$\Delta \mathbf{m} = -\frac{1}{2}e(\Delta v)R\,\hat{\mathbf{z}} = -\frac{e^2R^2}{4m_e}\mathbf{B}. \tag{6.8}$$

Notice that *the change in* **m** *is opposite to the direction of* **B**. (An electron circling the other way would have a dipole moment pointing upward, but such an orbit would be slowed down by the field, so the *change* is still opposite to **B**.) Ordinarily, the electron orbits are randomly oriented, and the orbital dipole moments cancel out. But in the presence of a magnetic field, each atom picks up a little "extra" dipole moment, and these increments are all *antiparallel* to the field. This is the mechanism responsible for **diamagnetism**. It is a universal phenomenon, affecting all atoms. However, it is typically much weaker than paramagnetism, and is therefore observed mainly in atoms with *even* numbers of electrons, where paramagnetism is usually absent.

In deriving Eq. 6.8, I assumed that the orbit remains circular, with its original radius R. I cannot offer a justification for this at the present stage. If the atom is stationary while the field is turned on, then my assumption can be proved—this is not magneto*statics*, however, and the details will have to await Chapter 7 (see Prob. 7.52). If the atom is moved into the field, the situation is enormously more complicated. But never mind—I'm only trying to give you a qualitative account of diamagnetism. Assume, if you prefer, that the velocity remains the same while the *radius* changes—the formula (Eq. 6.8) is altered (by a factor of 2), but the qualitative conclusion is unaffected. The truth is that this classical model is fundamentally flawed (diamagnetism is really a *quantum* phenomenon), so there's

[3]I said (Eq. 5.11) that magnetic fields do no work, and are incapable of speeding a particle up. I stand by that. However, as we shall see in Chapter 7, a *changing* magnetic field induces an *electric* field, and it is the latter that accelerates the electrons in this instance.

not much point in refining the details.[4] What *is* important is the empirical fact that in diamagnetic materials the induced dipole moments point opposite to the magnetic field.

6.1.4 ■ Magnetization

In the presence of a magnetic field, matter becomes *magnetized*; that is, upon microscopic examination, it will be found to contain many tiny dipoles, with a net alignment along some direction. We have discussed two mechanisms that account for this magnetic polarization: (1) paramagnetism (the dipoles associated with the spins of unpaired electrons experience a torque tending to line them up parallel to the field) and (2) diamagnetism (the orbital speed of the electrons is altered in such a way as to change the orbital dipole moment in a direction opposite to the field). Whatever the cause, we describe the state of magnetic polarization by the vector quantity

$$\mathbf{M} \equiv \textit{magnetic dipole moment per unit volume.} \tag{6.9}$$

M is called the **magnetization**; it plays a role analogous to the polarization **P** in electrostatics. In the following section, we will not worry about how the magnetization *got* there—it could be paramagnetism, diamagnetism, or even ferromagnetism—we shall take **M** as *given*, and calculate the field this magnetization itself produces.

Incidentally, it may have surprised you to learn that materials other than the famous ferromagnetic trio (iron, nickel, and cobalt) are affected by a magnetic field *at all*. You cannot, of course, pick up a piece of wood or aluminum with a magnet. The reason is that diamagnetism and paramagnetism are extremely weak: It takes a delicate experiment and a powerful magnet to detect them at all. If you were to suspend a piece of paramagnetic material above a solenoid, as in Fig. 6.3, the induced magnetization would be upward, and hence the force downward. By contrast, the magnetization of a diamagnetic object would be downward and the force upward. In general, when a sample is placed in a region of nonuniform field, the *paramagnet is attracted into the field*, whereas the *diamagnet is repelled away*. But the actual forces are pitifully weak—in a typical experimental arrangement the force on a comparable sample of iron would be 10^4 or 10^5 times as great. That's why it was reasonable for us to calculate the field inside a piece of copper wire, say, in Chapter 5, without worrying about the effects of magnetization.[5]

[4] S. L. O'Dell and R. K. P. Zia, *Am. J. Phys.* **54**, 32, (1986); R. Peierls, *Surprises in Theoretical Physics*, Section 4.3 (Princeton, N.J.: Princeton University Press, 1979); R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. 2, Sec. 34–36 (New York: Addison-Wesley, 1966).

[5] In 1997 Andre Geim managed to levitate a live frog (diamagnetic) for 30 minutes; he was awarded the 2000 Ig Nobel prize for this achievement, and later (2010) the Nobel prize for research on graphene. See M. V. Berry and A. K. Geim, *Eur. J. Phys.* **18**, 307 (1997) and Geim, *Physics Today*, September 1998, p. 36.

Problem 6.6 Of the following materials, which would you expect to be paramagnetic and which diamagnetic: aluminum, copper, copper chloride ($CuCl_2$), carbon, lead, nitrogen (N_2), salt (NaCl), sodium, sulfur, water? (Actually, copper is slightly *dia*magnetic; otherwise they're all what you'd expect.)

6.2 ■ THE FIELD OF A MAGNETIZED OBJECT

6.2.1 ■ Bound Currents

Suppose we have a piece of magnetized material; the magnetic dipole moment per unit volume, **M**, is given. What field does this object produce? Well, the vector potential of a single dipole **m** is given by Eq. 5.85:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\hat{\boldsymbol{\imath}}}{\imath^2}. \tag{6.10}$$

In the magnetized object, each volume element $d\tau'$ carries a dipole moment $\mathbf{M}\,d\tau'$, so the total vector potential is (Fig. 6.11)

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{M}(\mathbf{r}')\times\hat{\boldsymbol{\imath}}}{\imath^2}\,d\tau'. \tag{6.11}$$

That *does* it, in principle. But, as in the electrical case (Sect. 4.2.1), the integral can be cast in a more illuminating form by exploiting the identity

$$\boldsymbol{\nabla}'\frac{1}{\imath} = \frac{\hat{\boldsymbol{\imath}}}{\imath^2}.$$

With this,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \left[\mathbf{M}(\mathbf{r}')\times\left(\boldsymbol{\nabla}'\frac{1}{\imath}\right)\right] d\tau'.$$

Integrating by parts, using product rule 7, gives

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\left\{\int \frac{1}{\imath}[\boldsymbol{\nabla}'\times\mathbf{M}(\mathbf{r}')]\,d\tau' - \int \boldsymbol{\nabla}'\times\left[\frac{\mathbf{M}(\mathbf{r}')}{\imath}\right] d\tau'\right\}.$$

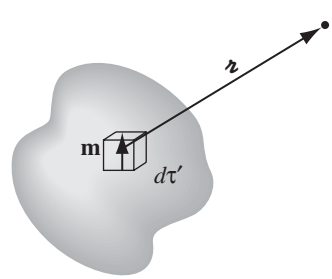


FIGURE 6.11

Problem 1.61(b) invites us to express the latter as a surface integral,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{1}{\imath}[\boldsymbol{\nabla}' \times \mathbf{M}(\mathbf{r}')]\,d\tau' + \frac{\mu_0}{4\pi}\oint \frac{1}{\imath}[\mathbf{M}(\mathbf{r}') \times d\mathbf{a}']. \tag{6.12}$$

The first term looks just like the potential of a volume current,

$$\boxed{\mathbf{J}_b = \boldsymbol{\nabla} \times \mathbf{M},} \tag{6.13}$$

while the second looks like the potential of a surface current,

$$\boxed{\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}},} \tag{6.14}$$

where $\hat{\mathbf{n}}$ is the normal unit vector. With these definitions,

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_{\mathcal{V}} \frac{\mathbf{J}_b(\mathbf{r}')}{\imath}\,d\tau' + \frac{\mu_0}{4\pi}\oint_{\mathcal{S}} \frac{\mathbf{K}_b(\mathbf{r}')}{\imath}\,da'. \tag{6.15}$$

What this means is that the potential (and hence also the field) of a magnetized object is the same as would be produced by a volume current $\mathbf{J}_b = \boldsymbol{\nabla} \times \mathbf{M}$ throughout the material, plus a surface current $\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}}$, on the boundary. Instead of integrating the contributions of all the infinitesimal dipoles, using Eq. 6.11, we first determine the **bound currents**, and then find the field *they* produce, in the same way we would calculate the field of any other volume and surface currents. Notice the striking parallel with the electrical case: there the field of a polarized object was the same as that of a bound volume charge $\rho_b = -\boldsymbol{\nabla} \cdot \mathbf{P}$ plus a bound surface charge $\sigma_b = \mathbf{P} \cdot \hat{\mathbf{n}}$.

Example 6.1. Find the magnetic field of a uniformly magnetized sphere.

Solution

Choosing the z axis along the direction of $\mathbf{M}$ (Fig. 6.12), we have

$$\mathbf{J}_b = \boldsymbol{\nabla} \times \mathbf{M} = \mathbf{0}, \quad \mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}} = M \sin\theta\, \hat{\boldsymbol{\phi}}.$$

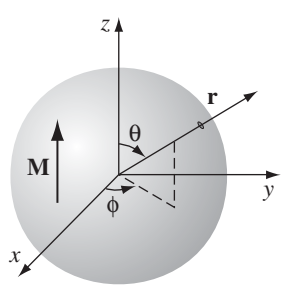


FIGURE 6.12

Now, a rotating spherical shell, of uniform surface charge σ, corresponds to a surface current density

$$\mathbf{K} = \sigma \mathbf{v} = \sigma \omega R \sin\theta \, \hat{\boldsymbol{\phi}}.$$

It follows, therefore, that the field of a uniformly magnetized sphere is identical to the field of a spinning spherical shell, with the identification $\sigma R\boldsymbol{\omega} \to \mathbf{M}$. Referring back to Ex. 5.11, I conclude that

$$\mathbf{B} = \frac{2}{3}\mu_0 \mathbf{M}, \tag{6.16}$$

inside the sphere, while the field outside is the same as that of a perfect dipole,

$$\mathbf{m} = \frac{4}{3}\pi R^3 \mathbf{M}.$$

Notice that the internal field is *uniform*, like the electric field inside a uniformly polarized sphere (Eq. 4.14), although the actual *formulas* for the two cases are curiously different ($\frac{2}{3}$ in place of $-\frac{1}{3}$).[6] The external fields are also analogous: pure dipole in both instances.

Problem 6.7 An infinitely long circular cylinder carries a uniform magnetization $\mathbf{M}$ parallel to its axis. Find the magnetic field (due to $\mathbf{M}$) inside and outside the cylinder.

Problem 6.8 A long circular cylinder of radius R carries a magnetization $\mathbf{M} = ks^2 \hat{\boldsymbol{\phi}}$, where k is a constant, s is the distance from the axis, and $\hat{\boldsymbol{\phi}}$ is the usual azimuthal unit vector (Fig. 6.13). Find the magnetic field due to $\mathbf{M}$, for points inside and outside the cylinder.

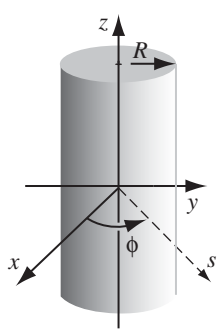


FIGURE 6.13

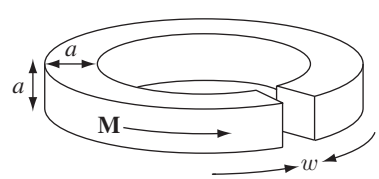


FIGURE 6.14

[6]It is no accident that the same factors appear in the "contact" term for the fields of electric and magnetic dipoles (Eqs. 3.106 and 5.94). In fact, one good way to model a perfect dipole is to take the limit ($R \to 0$) of a polarized/magnetized sphere.

Problem 6.9 A short circular cylinder of radius a and length L carries a "frozen-in" uniform magnetization $\mathbf{M}$ parallel to its axis. Find the bound current, and sketch the magnetic field of the cylinder. (Make three sketches: one for $L \gg a$, one for $L \ll a$, and one for $L \approx a$.) Compare this **bar magnet** with the bar electret of Prob. 4.11.

Problem 6.10 An iron rod of length L and square cross section (side a) is given a uniform longitudinal magnetization $\mathbf{M}$, and then bent around into a circle with a narrow gap (width w), as shown in Fig. 6.14. Find the magnetic field at the center of the gap, assuming $w \ll a \ll L$. [*Hint:* treat it as the superposition of a *complete* torus plus a square loop with reversed current.]

6.2.2 ■ Physical Interpretation of Bound Currents

In the last section, we found that the field of a magnetized object is identical to the field that would be produced by a certain distribution of "bound" currents, $\mathbf{J}_b$ and $\mathbf{K}_b$. I want to show you how these bound currents arise physically. This will be a heuristic argument—the rigorous derivation has already been given. Figure 6.15 depicts a thin slab of uniformly magnetized material, with the dipoles represented by tiny current loops. Notice that all the "internal" currents cancel: every time there is one going to the right, a contiguous one is going to the left. However, at the edge there is *no adjacent loop to do the canceling*. The whole thing, then, is equivalent to a single ribbon of current I flowing around the boundary (Fig. 6.16).

What *is* this current, in terms of $\mathbf{M}$? Say that each of the tiny loops has area a and thickness t (Fig. 6.17). In terms of the magnetization M, its dipole moment

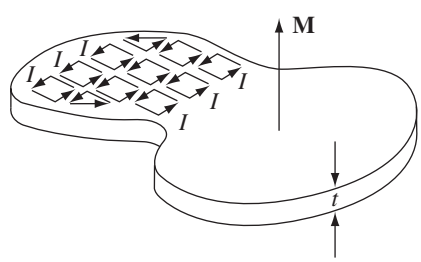


FIGURE 6.15

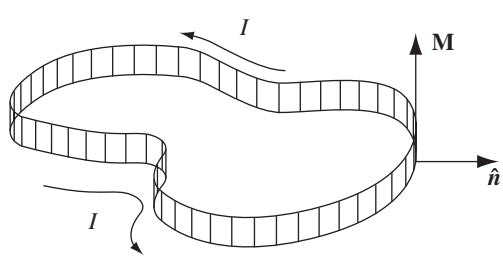


FIGURE 6.16

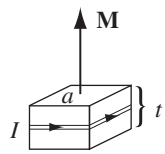


FIGURE 6.17

is $m = Mat$. In terms of the circulating current I, however, $m = Ia$. Therefore $I = Mt$, so the surface current is $K_b = I/t = M$. Using the outward-drawn unit vector $\hat{\mathbf{n}}$ (Fig. 6.16), the direction of $\mathbf{K}_b$ is conveniently indicated by the cross product:

$$\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}}.$$

(This expression also records the fact that there is *no* current on the top or bottom surface of the slab; here $\mathbf{M}$ is parallel to $\hat{\mathbf{n}}$, so the cross product vanishes.)

This bound surface current is exactly what we obtained in Sect. 6.2.1. It is a peculiar *kind* of current, in the sense that no single charge makes the whole trip—on the contrary, each charge moves only in a tiny little loop within a single atom. Nevertheless, the net effect is a macroscopic current flowing over the surface of the magnetized object. We call it a "bound" current to remind ourselves that every charge is attached to a particular atom, but it's a perfectly genuine current, and it produces a magnetic field in the same way any other current does.

When the magnetization is *non*uniform, the internal currents no longer cancel. Figure 6.18(a) shows two adjacent chunks of magnetized material, with a larger arrow on the one to the right suggesting greater magnetization at that point. On the surface where they join, there is a net current in the x direction, given by

$$I_x = [M_z(y + dy) - M_z(y)]\, dz = \frac{\partial M_z}{\partial y}\, dy\, dz.$$

The corresponding volume current density is therefore

$$(J_b)_x = \frac{\partial M_z}{\partial y}.$$

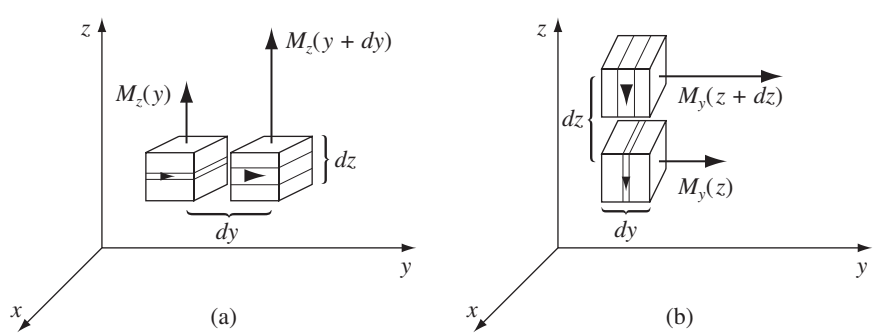


FIGURE 6.18

By the same token, a nonuniform magnetization in the y direction would contribute an amount $-\partial M_y/\partial z$ (Fig. 6.18(b)), so

$$(J_b)_x = \frac{\partial M_z}{\partial y} - \frac{\partial M_y}{\partial z}.$$

In general, then,

$$\mathbf{J}_b = \nabla \times \mathbf{M},$$

consistent, again, with the result of Sect. 6.2.1.

Incidentally, like any other steady current, $\mathbf{J}_b$ should obey the conservation law 5.33:

$$\nabla \cdot \mathbf{J}_b = 0.$$

Does it? *Yes*, for the divergence of a curl is *always* zero.

6.2.3 ■ The Magnetic Field Inside Matter

Like the electric field, the actual *microscopic* magnetic field inside matter fluctuates wildly from point to point and instant to instant. When we speak of "the" magnetic field in matter, we mean the *macroscopic* field: the average over regions large enough to contain many atoms. (The magnetization $\mathbf{M}$ is "smoothed out" in the same sense.) It is this macroscopic field that one obtains when the methods of Sect. 6.2.1 are applied to points inside magnetized material, as you can prove for yourself in the following problem.

Problem 6.11 In Sect, 6.2.1, we began with the potential of a *perfect* dipole (Eq. 6.10), whereas in *fact* we are dealing with *physical* dipoles. Show, by the method of Sect. 4.2.3, that we nonetheless get the correct macroscopic field.

6.3 ■ THE AUXILIARY FIELD H

6.3.1 ■ Ampère's Law in Magnetized Materials

In Sect. 6.2, we found that the effect of magnetization is to establish bound currents $\mathbf{J}_b = \nabla \times \mathbf{M}$ within the material and $\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}}$ on the surface. The field due to magnetization of the medium is just the field produced by these bound currents. We are now ready to put everything together: the field attributable to bound currents, plus the field due to everything else—which I shall call the **free current**. The free current might flow through wires imbedded in the magnetized substance or, if the latter is a conductor, through the material itself. In any event, the total current can be written as

$$\mathbf{J} = \mathbf{J}_b + \mathbf{J}_f. \tag{6.17}$$

There is no new physics in Eq. 6.17; it is simply a *convenience* to separate the current into these two parts, because they got there by quite different means: the

free current is there because somebody hooked up a wire to a battery—it involves actual transport of charge; the bound current is there because of magnetization—it results from the conspiracy of many aligned atomic dipoles.

In view of Eqs. 6.13 and 6.17, Ampère's law can be written

$$\frac{1}{\mu_0}(\nabla \times \mathbf{B}) = \mathbf{J} = \mathbf{J}_f + \mathbf{J}_b = \mathbf{J}_f + (\nabla \times \mathbf{M}),$$

or, collecting together the two curls:

$$\nabla \times \left(\frac{1}{\mu_0}\mathbf{B} - \mathbf{M}\right) = \mathbf{J}_f.$$

The quantity in parentheses is designated by the letter **H**:

$$\boxed{\mathbf{H} \equiv \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}.} \tag{6.18}$$

In terms of **H**, then, Ampère's law reads

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}_f,} \tag{6.19}$$

or, in integral form,

$$\oint \mathbf{H} \cdot d\mathbf{l} = I_{f_{\text{enc}}}, \tag{6.20}$$

where $I_{f_{\text{enc}}}$ is the total *free* current passing through the Amperian loop.

H plays a role in magnetostatics analogous to **D** in electrostatics: Just as **D** allowed us to write *Gauss's* law in terms of the free *charge* alone, **H** permits us to express *Ampère's* law in terms of the free *current* alone—and free current is what we control directly. Bound current, like bound charge, comes along for the ride—the material gets magnetized, and this results in bound currents; we cannot turn them on or off independently, as we can free currents. In applying Eq. 6.20, all we need to worry about is the *free* current, which we know about because we *put* it there. In particular, when symmetry permits, we can calculate **H** immediately from Eq. 6.20 by the usual Ampère's law methods. (For example, Probs. 6.7 and 6.8 can be done in one line by noting that $\mathbf{H} = \mathbf{0}$.)

Example 6.2. A long copper rod of radius R carries a uniformly distributed (free) current I (Fig. 6.19). Find **H** inside and outside the rod.

Solution

Copper is weakly diamagnetic, so the dipoles will line up opposite to the field. This results in a bound current running antiparallel to I, within the wire, and parallel to I along the surface (Fig. 6.20). Just how great these bound currents will

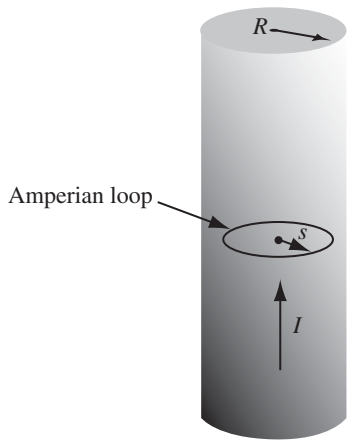


FIGURE 6.19

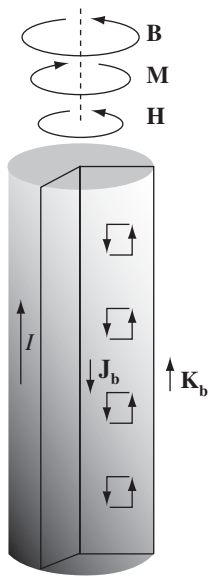


FIGURE 6.20

be we are not yet in a position to say—but in order to calculate **H**, it is sufficient to realize that all the currents are longitudinal, so **B**, **M**, and therefore also **H**, are circumferential. Applying Eq. 6.20 to an Amperian loop of radius $s < R$,

$$H(2\pi s) = I_{f_{\text{enc}}} = I\frac{\pi s^2}{\pi R^2},$$

so, inside the wire,

$$\mathbf{H} = \frac{I}{2\pi R^2} s\,\hat{\boldsymbol{\phi}} \qquad (s \leq R). \tag{6.21}$$

Outside the wire

$$\mathbf{H} = \frac{I}{2\pi s}\,\hat{\boldsymbol{\phi}} \qquad (s \geq R). \tag{6.22}$$

In the latter region (as always, in empty space) $\mathbf{M} = \mathbf{0}$, so

$$\mathbf{B} = \mu_0\mathbf{H} = \frac{\mu_0 I}{2\pi s}\,\hat{\boldsymbol{\phi}} \qquad (s \geq R),$$

the same as for a *non*magnetized wire (Ex. 5.7). *Inside* the wire **B** cannot be determined at this stage, since we have no way of knowing **M** (though in practice the magnetization in copper is so slight that for most purposes we can ignore it altogether).

As it turns out, **H** is a more useful quantity than **D**. In the laboratory, you will frequently hear people talking about **H** (more often even than **B**), but you will never hear anyone speak of **D** (only **E**). The reason is this: To build an

electromagnet you run a certain (free) current through a coil. The *current* is the thing you read on the dial, and this determines **H** (or at any rate, the line integral of **H**); **B** depends on the specific materials you used and even, if iron is present, on the history of your magnet. On the other hand, if you want to set up an *electric* field, you do *not* plaster a known free charge on the plates of a parallel plate capacitor; rather, you connect them to a battery of known *voltage*. It's the *potential difference* you read on your dial, and that determines **E** (or rather, the line integral of **E**); **D** depends on the details of the dielectric you're using. If it were easy to measure charge, and hard to measure potential, then you'd find experimentalists talking about **D** instead of **E**. So the relative familiarity of **H**, as contrasted with **D**, derives from purely practical considerations; theoretically, they're on an equal footing.

Many authors call **H**, not **B**, the "magnetic field." Then they have to invent a new word for **B**: the "flux density," or magnetic "induction" (an absurd choice, since that term already has at least two other meanings in electrodynamics). Anyway, **B** is indisputably the fundamental quantity, so I shall continue to call it the "magnetic field," as everyone does in the spoken language. **H** has no sensible name: just call it "**H**."[7]

Problem 6.12 An infinitely long cylinder, of radius R, carries a "frozen-in" magnetization, parallel to the axis,

$$\mathbf{M} = ks\,\hat{\mathbf{z}},$$

where k is a constant and s is the distance from the axis; there is no free current anywhere. Find the magnetic field inside and outside the cylinder by two different methods:

(a) As in Sect. 6.2, locate all the bound currents, and calculate the field they produce.

(b) Use Ampère's law (in the form of Eq. 6.20) to find **H**, and then get **B** from Eq. 6.18. (Notice that the second method is much faster, and avoids any explicit reference to the bound currents.)

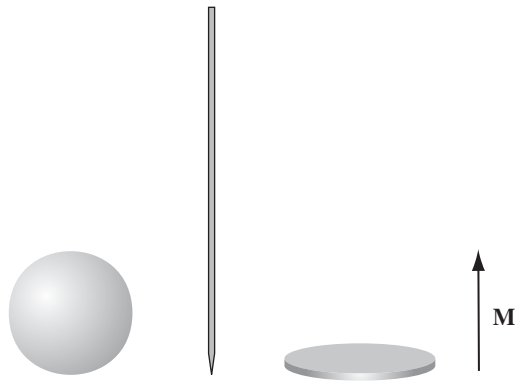
Problem 6.13 Suppose the field inside a large piece of magnetic material is $\mathbf{B}_0$, so that $\mathbf{H}_0 = (1/\mu_0)\mathbf{B}_0 - \mathbf{M}$, where **M** is a "frozen-in" magnetization.

(a) Now a small spherical cavity is hollowed out of the material (Fig. 6.21). Find the field at the center of the cavity, in terms of $\mathbf{B}_0$ and **M**. Also find **H** at the center of the cavity, in terms of $\mathbf{H}_0$ and **M**.

(b) Do the same for a long needle-shaped cavity running parallel to **M**.

(c) Do the same for a thin wafer-shaped cavity perpendicular to **M**.

[7]For those who disagree, I quote A. Sommerfeld's *Electrodynamics* (New York: Academic Press, 1952), p. 45: "The unhappy term 'magnetic field' for **H** should be avoided as far as possible. It seems to us that this term has led into error none less than Maxwell himself . . ."



(a) Sphere (b) Needle (c) Wafer

FIGURE 6.21

Assume the cavities are small enough so $\mathbf{M}$, $\mathbf{B}_0$, and $\mathbf{H}_0$ are essentially constant. Compare Prob. 4.16. [*Hint:* Carving out a cavity is the same as superimposing an object of the same shape but opposite magnetization.]

6.3.2 ■ A Deceptive Parallel

Equation 6.19 looks just like Ampère's original law (Eq. 5.56), except that the *total* current is replaced by the *free* current, and $\mathbf{B}$ is replaced by $\mu_0\mathbf{H}$. As in the case of $\mathbf{D}$, however, I must warn you against reading too much into this correspondence. It does *not* say that $\mu_0\mathbf{H}$ is "just like $\mathbf{B}$, only its source is $\mathbf{J}_f$ instead of $\mathbf{J}$." For the curl alone does not determine a vector field—you must *also* know the divergence. And whereas $\nabla \cdot \mathbf{B} = 0$, the divergence of $\mathbf{H}$ is *not*, in general, zero. In fact, from Eq. 6.18

$$\nabla \cdot \mathbf{H} = -\nabla \cdot \mathbf{M}. \tag{6.23}$$

Only when the divergence of $\mathbf{M}$ vanishes is the parallel between $\mathbf{B}$ and $\mu_0\mathbf{H}$ faithful.

If you think I'm being pedantic, consider the example of the bar magnet—a short cylinder of iron that carries a permanent uniform magnetization $\mathbf{M}$ parallel to its axis. (See Probs. 6.9 and 6.14.) In this case there is no free current anywhere, and a naïve application of Eq. 6.20 might lead you to suppose that $\mathbf{H} = \mathbf{0}$, and hence that $\mathbf{B} = \mu_0\mathbf{M}$ inside the magnet and $\mathbf{B} = \mathbf{0}$ outside, which is nonsense. It is quite true that the *curl* of $\mathbf{H}$ vanishes everywhere, but the divergence does not. (Can you see where $\nabla \cdot \mathbf{M} \neq 0$?) *Advice*: When you are asked to find $\mathbf{B}$ or $\mathbf{H}$ in a problem involving magnetic materials, first look for symmetry. If the problem exhibits cylindrical, plane, solenoidal, or toroidal symmetry, then you can get $\mathbf{H}$ directly from Eq. 6.20 by the usual Ampère's law methods. (Evidently, in such cases $\nabla \cdot \mathbf{M}$ is automatically zero, since the free current alone determines the answer.) If the requisite symmetry is absent, you'll have to think of another

approach, and in particular you must *not* assume that **H** is zero just because there is no free current in sight.

6.3.3 ■ Boundary Conditions

The magnetostatic boundary conditions of Sect. 5.4.2 can be rewritten in terms of **H** and the *free* current. From Eq. 6.23 it follows that

$$H^{\perp}_{\text{above}} - H^{\perp}_{\text{below}} = -(M^{\perp}_{\text{above}} - M^{\perp}_{\text{below}}), \tag{6.24}$$

while Eq. 6.19 says

$$\mathbf{H}^{\parallel}_{\text{above}} - \mathbf{H}^{\parallel}_{\text{below}} = \mathbf{K}_f \times \hat{\mathbf{n}}. \tag{6.25}$$

In the presence of materials, these are sometimes more useful than the corresponding boundary conditions on **B** (Eqs. 5.74 and 5.76):

$$B^{\perp}_{\text{above}} - B^{\perp}_{\text{below}} = 0, \tag{6.26}$$

and

$$\mathbf{B}^{\parallel}_{\text{above}} - \mathbf{B}^{\parallel}_{\text{below}} = \mu_0(\mathbf{K} \times \hat{\mathbf{n}}). \tag{6.27}$$

You might want to check them, for Ex. 6.2 or Prob. 6.14.

Problem 6.14 For the bar magnet of Prob. 6.9, make careful sketches of **M**, **B**, and **H**, assuming L is about $2a$. Compare Prob. 4.17.

Problem 6.15 If $\mathbf{J}_f = \mathbf{0}$ everywhere, the curl of **H** vanishes (Eq. 6.19), and we can express **H** as the gradient of a scalar potential W:

$$\mathbf{H} = -\nabla W.$$

According to Eq. 6.23, then,

$$\nabla^2 W = (\nabla \cdot \mathbf{M}),$$

so W obeys Poisson's equation, with $\nabla \cdot \mathbf{M}$ as the "source." This opens up all the machinery of Chapter 3. As an example, find the field inside a uniformly magnetized sphere (Ex. 6.1) by separation of variables. [*Hint:* $\nabla \cdot \mathbf{M} = 0$ everywhere except at the surface ($r = R$), so W satisfies Laplace's equation in the regions $r < R$ and $r > R$; use Eq. 3.65, and from Eq. 6.24 figure out the appropriate boundary condition on W.]

6.4 ■ LINEAR AND NONLINEAR MEDIA

6.4.1 ■ Magnetic Susceptibility and Permeability

In paramagnetic and diamagnetic materials, the magnetization is sustained by the field; when **B** is removed, **M** disappears. In fact, for most substances the magnetization is *proportional* to the field, provided the field is not too strong. For

notational consistency with the electrical case (Eq. 4.30), I *should* express the proportionality thus:

$$\mathbf{M} = \frac{1}{\mu_0}\chi_m \mathbf{B} \quad \text{(incorrect!)}. \tag{6.28}$$

But custom dictates that it be written in terms of **H**, instead of **B**:

$$\boxed{\mathbf{M} = \chi_m \mathbf{H}.} \tag{6.29}$$

The constant of proportionality χ_m is called the **magnetic susceptibility**; it is a dimensionless quantity that varies from one substance to another—positive for paramagnets and negative for diamagnets. Typical values are around 10^{-5} (see Table 6.1).

Materials that obey Eq. 6.29 are called **linear media**. In view of Eq. 6.18,

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(1 + \chi_m)\mathbf{H}, \tag{6.30}$$

for linear media. Thus **B** is *also* proportional to **H**:[8]

$$\mathbf{B} = \mu \mathbf{H}, \tag{6.31}$$

where

$$\mu \equiv \mu_0(1 + \chi_m). \tag{6.32}$$

μ is called the **permeability** of the material.[9] In a vacuum, where there is no matter to magnetize, the susceptibility χ_m vanishes, and the permeability is μ_0. That's why μ_0 is called the **permeability of free space**.

Material	Susceptibility	Material	Susceptibility
Diamagnetic:		*Paramagnetic:*	
Bismuth	-1.7×10^{-4}	Oxygen (O_2)	1.7×10^{-6}
Gold	-3.4×10^{-5}	Sodium	8.5×10^{-6}
Silver	-2.4×10^{-5}	Aluminum	2.2×10^{-5}
Copper	-9.7×10^{-6}	Tungsten	7.0×10^{-5}
Water	-9.0×10^{-6}	Platinum	2.7×10^{-4}
Carbon Dioxide	-1.1×10^{-8}	Liquid Oxygen ($-200°$ C)	3.9×10^{-3}
Hydrogen (H_2)	-2.1×10^{-9}	Gadolinium	4.8×10^{-1}

TABLE 6.1 Magnetic Susceptibilities (unless otherwise specified, values are for 1 atm, 20° C). *Data from Handbook of Chemistry and Physics,* 91st ed. (Boca Raton: CRC Press, Inc., 2010) and other references.

[8]Physically, therefore, Eq. 6.28 would say exactly the same as Eq. 6.29, only the constant χ_m would have a different value. Equation 6.29 is a little more convenient, because experimentalists find it handier to work with **H** than **B**.

[9]If you factor out μ_0, what's left is called the **relative permeability**: $\mu_r \equiv 1 + \chi_m = \mu/\mu_0$. By the way, formulas for **H** in terms of **B** (Eq. 6.31, in the case of linear media) are called **constitutive relations**, just like those for **D** in terms of **E**.

Example 6.3. An infinite solenoid (n turns per unit length, current I) is filled with linear material of susceptibility χ_m. Find the magnetic field inside the solenoid.

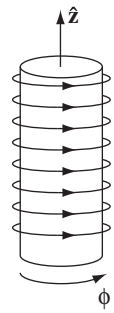


FIGURE 6.22

Solution

Since **B** is due in part to bound currents (which we don't yet know), we cannot compute it directly. However, this is one of those symmetrical cases in which we can get **H** from the free current alone, using Ampère's law in the form of Eq. 6.20:

$$\mathbf{H} = nI\,\hat{\mathbf{z}}$$

(Fig. 6.22). According to Eq. 6.31, then,

$$\mathbf{B} = \mu_0(1 + \chi_m)nI\,\hat{\mathbf{z}}.$$

If the medium is paramagnetic, the field is slightly enhanced; if it's diamagnetic, the field is somewhat reduced. This reflects the fact that the bound surface current

$$\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}} = \chi_m(\mathbf{H} \times \hat{\mathbf{n}}) = \chi_m nI\,\hat{\boldsymbol{\phi}}$$

is in the same direction as I, in the former case ($\chi_m > 0$), and opposite in the latter ($\chi_m < 0$).

You might suppose that linear media escape the defect in the parallel between **B** and **H**: since **M** and **H** are now proportional to **B**, does it not follow that their divergence, like **B**'s, must always vanish? Unfortunately, it does *not*;[10] at the *boundary* between two materials of different permeability, the divergence of **M** can actually be infinite. For instance, at the end of a cylinder of linear paramagnetic material, **M** is zero on one side but not on the other. For the "Gaussian pillbox" shown in Fig. 6.23, $\oint \mathbf{M} \cdot d\mathbf{a} \neq 0$, and hence, by the divergence theorem, $\boldsymbol{\nabla} \cdot \mathbf{M}$ cannot vanish everywhere within it.

[10]Formally, $\boldsymbol{\nabla} \cdot \mathbf{H} = \boldsymbol{\nabla} \cdot \left(\frac{1}{\mu}\mathbf{B}\right) = \frac{1}{\mu}\boldsymbol{\nabla} \cdot \mathbf{B} + \mathbf{B} \cdot \boldsymbol{\nabla}\left(\frac{1}{\mu}\right) = \mathbf{B} \cdot \boldsymbol{\nabla}\left(\frac{1}{\mu}\right)$, so **H** is *not* divergenceless (in general) at points where μ is changing.

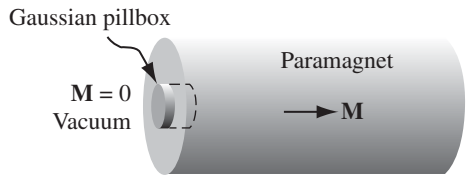


FIGURE 6.23

Incidentally, the volume bound current density in a homogeneous linear material is proportional to the *free* current density:

$$\mathbf{J}_b = \boldsymbol{\nabla} \times \mathbf{M} = \boldsymbol{\nabla} \times (\chi_m \mathbf{H}) = \chi_m \mathbf{J}_f. \tag{6.33}$$

In particular, unless free current actually flows *through* the material, all bound current will be at the surface.

Problem 6.16 A coaxial cable consists of two very long cylindrical tubes, separated by linear insulating material of magnetic susceptibility χ_m. A current I flows down the inner conductor and returns along the outer one; in each case, the current distributes itself uniformly over the surface (Fig. 6.24). Find the magnetic field in the region between the tubes. As a check, calculate the magnetization and the bound currents, and confirm that (together, of course, with the free currents) they generate the correct field.

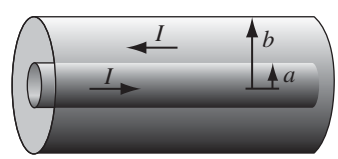


FIGURE 6.24

Problem 6.17 A current I flows down a long straight wire of radius a. If the wire is made of linear material (copper, say, or aluminum) with susceptibility χ_m, and the current is distributed uniformly, what is the magnetic field a distance s from the axis? Find all the bound currents. What is the *net* bound current flowing down the wire?

! **Problem 6.18** A sphere of linear magnetic material is placed in an otherwise uniform magnetic field $\mathbf{B}_0$. Find the new field inside the sphere. [*Hint:* See Prob. 6.15 or Prob. 4.23.]

Problem 6.19 On the basis of the naïve model presented in Sect. 6.1.3, estimate the magnetic susceptibility of a diamagnetic metal such as copper. Compare your answer with the empirical value in Table 6.1, and comment on any discrepancy.

6.4.2 ■ Ferromagnetism

In a linear medium, the alignment of atomic dipoles is maintained by a magnetic field imposed from the outside. Ferromagnets—which are emphatically *not* linear[11]—require no external fields to sustain the magnetization; the alignment is "frozen in." Like paramagnetism, ferromagnetism involves the magnetic dipoles associated with the spins of unpaired electrons. The new feature, which makes ferromagnetism so different from paramagnetism, is the interaction between nearby dipoles: In a ferromagnet, *each dipole "likes" to point in the same direction as its neighbors*. The *reason* for this preference is essentially quantum mechanical, and I shall not endeavor to explain it here; it is enough to know that the correlation is so strong as to align virtually 100% of the unpaired electron spins. If you could somehow magnify a piece of iron and "see" the individual dipoles as tiny arrows, it would look something like Fig. 6.25, with all the spins pointing the same way.

But if that is true, why isn't every wrench and nail a powerful magnet? The answer is that the alignment occurs in relatively small patches, called **domains**. Each domain contains billions of dipoles, all lined up (these domains are actually *visible* under a microscope, using suitable etching techniques—see Fig. 6.26), but the domains *themselves* are randomly oriented. The household wrench contains an enormous number of domains, and their magnetic fields cancel, so the wrench as a whole is not magnetized. (Actually, the orientation of domains is not *completely* random; within a given crystal, there may be some preferential alignment along the crystal axes. But there will be just as many domains pointing one way as the other, so there is still no large-scale magnetization. Moreover, the crystals themselves are randomly oriented within any sizable chunk of metal.)

How, then, would you produce a **permanent magnet**, such as they sell in toy stores? If you put a piece of iron into a strong magnetic field, the torque $\mathbf{N} = \mathbf{m} \times \mathbf{B}$ tends to align the dipoles parallel to the field. Since they like to stay parallel to their neighbors, most of the dipoles will resist this torque. However,

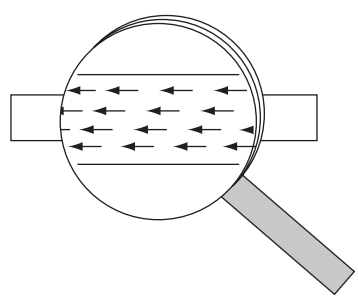
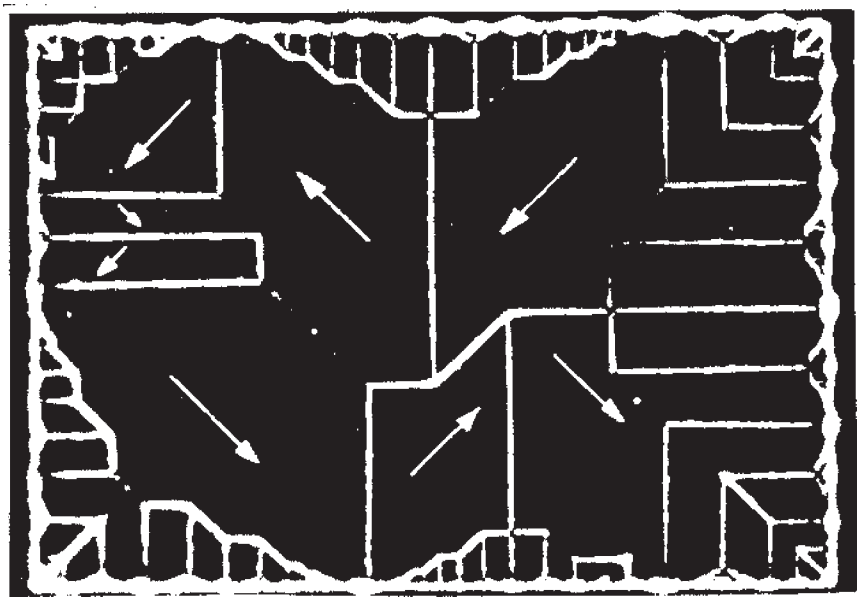


FIGURE 6.25

[11] In this sense, it is misleading to speak of the susceptibility or permeability of a ferromagnet. The terms *are* used for such materials, but they refer to the proportionality factor between a *differential* increase in **H** and the resulting *differential* change in **M** (or **B**); moreover, they are not constants, but functions of **H**.



Ferromagnetic domains. (Photo courtesy of R. W. DeBlois)

FIGURE 6.26

at the *boundary* between two domains, there are *competing* neighbors, and the torque will throw its weight on the side of the domain most nearly parallel to the field; this domain will win some converts, at the expense of the less favorably oriented one. The net effect of the magnetic field, then, is to *move the domain boundaries*. Domains parallel to the field grow, and the others shrink. If the field is strong enough, one domain takes over entirely, and the iron is said to be **saturated**.

It turns out that this process (the shifting of domain boundaries in response to an external field) is not entirely reversible: When the field is switched off, there will be *some* return to randomly oriented domains, but it is far from complete—there remains a preponderance of domains in the original direction. You now have a permanent magnet.

A simple way to accomplish this, in practice, is to wrap a coil of wire around the object to be magnetized (Fig. 6.27). Run a current I through the coil; this provides the external magnetic field (pointing to the left in the diagram). As you increase the current, the field increases, the domain boundaries move, and the magnetization grows. Eventually, you reach the saturation point, with all the dipoles aligned, and a further increase in current has no effect on $\mathbf{M}$ (Fig. 6.28, point b).

Now suppose you *reduce* the current. Instead of retracing the path back to $M = 0$, there is only a *partial* return to randomly oriented domains; M decreases, but even with the current off there is some residual magnetization (point c). The wrench is now a permanent magnet. If you want to eliminate the remaining magnetization, you'll have to run a current backwards through the coil (a negative I). Now the external field points to the right, and as you increase I (negatively),

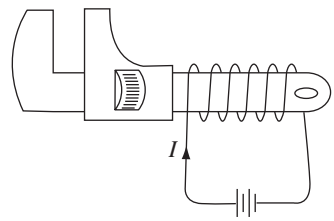


FIGURE 6.27

M drops down to zero (point d). If you turn I still higher, you soon reach saturation in the other direction—all the dipoles now pointing to the *right* (e). At this stage, switching off the current will leave the wrench with a permanent magnetization to the right (point f). To complete the story, turn I on again in the positive sense: M returns to zero (point g), and eventually to the forward saturation point (b).

The path we have traced out is called a **hysteresis loop**. Notice that the magnetization of the wrench depends not only on the applied field (that is, on I), but also on its previous magnetic "history."[12] For instance, at three different times in our experiment the current was zero (a, c, and f), yet the magnetization was different for each of them. Actually, it is customary to draw hysteresis loops as plots of B against H, rather than M against I. (If our coil is approximated by a long solenoid, with n turns per unit length, then $H = nI$, so H and I are proportional. Meanwhile, $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$, but in practice M is huge compared to H, so to all intents and purposes $\mathbf{B}$ is proportional to $\mathbf{M}$.)

To make the units consistent (teslas), I have plotted ($\mu_0 H$) horizontally (Fig. 6.29); notice, however, that the vertical scale is 10^4 times greater than the horizontal one. Roughly speaking, $\mu_0\mathbf{H}$ is the field our coil *would* have produced in the absence of any iron; $\mathbf{B}$ is what we *actually* got, and compared to $\mu_0\mathbf{H}$, it is gigantic. A little current goes a long way, when you have ferromagnetic materials

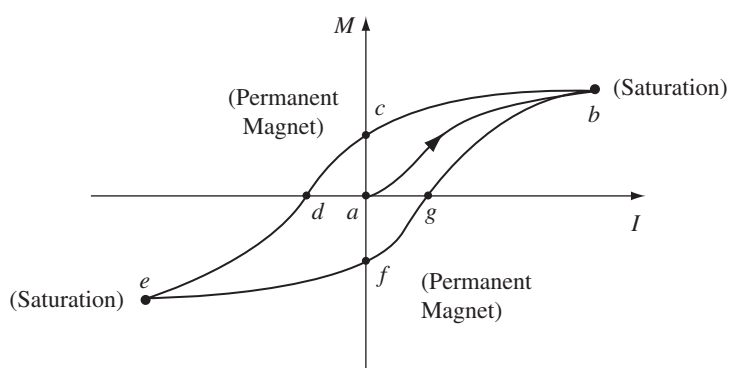


FIGURE 6.28

[12]Etymologically, the word *hysteresis* has nothing to do with the word *history*—nor with the word *hysteria*. It derives from a Greek verb meaning "lag behind."

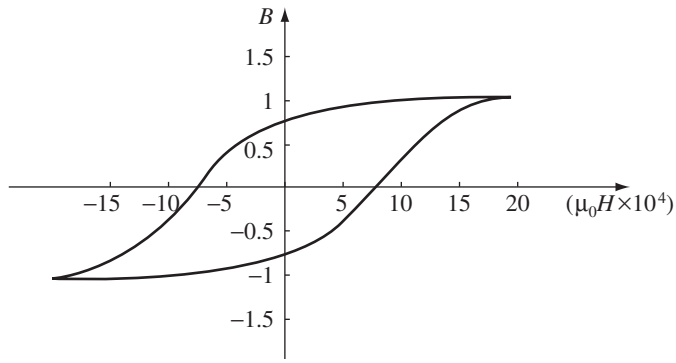


FIGURE 6.29

around. That's why anyone who wants to make a powerful electromagnet will wrap the coil around an iron core. It doesn't take much of an external field to move the domain boundaries, and when you do that, you have all the dipoles in the iron working with you.

One final point about ferromagnetism: It all follows, remember, from the fact that the dipoles within a given domain line up parallel to one another. Random thermal motions compete with this ordering, but as long as the temperature doesn't get too high, they cannot budge the dipoles out of line. It's not surprising, though, that *very* high temperatures do destroy the alignment. What *is* surprising is that this occurs at a precise temperature (770° C, for iron). Below this temperature (called the **Curie point**), iron is ferromagnetic; above, it is paramagnetic. The Curie point is rather like the boiling point or the freezing point in that there is no *gradual* transition from ferro- to para-magnetic behavior, any more than there is between water and ice. These abrupt changes in the properties of a substance, occurring at sharply defined temperatures, are known in statistical mechanics as **phase transitions**.

Problem 6.20 How would you go about *de*magnetizing a permanent magnet (such as the wrench we have been discussing, at point c in the hysteresis loop)? That is, how could you restore it to its original state, with $M = 0$ at $I = 0$?

Problem 6.21

(a) Show that the energy of a magnetic dipole in a magnetic field **B** is

$$\boxed{U = -\mathbf{m} \cdot \mathbf{B}.} \tag{6.34}$$

[Assume that the *magnitude* of the dipole moment is fixed, and all you have to do is move it into place and rotate it into its final orientation. The energy required to keep the current flowing is a different problem, which we will confront in Chapter 7.] Compare Eq. 4.6.

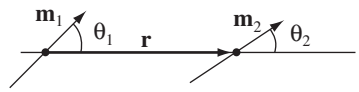


FIGURE 6.30

(b) Show that the interaction energy of two magnetic dipoles separated by a displacement **r** is given by

$$U = \frac{\mu_0}{4\pi}\frac{1}{r^3}[\mathbf{m}_1 \cdot \mathbf{m}_2 - 3(\mathbf{m}_1 \cdot \hat{\mathbf{r}})(\mathbf{m}_2 \cdot \hat{\mathbf{r}})]. \tag{6.35}$$

Compare Eq. 4.7.

(c) Express your answer to (b) in terms of the angles θ_1 and θ_2 in Fig. 6.30, and use the result to find the stable configuration two dipoles would adopt if held a fixed distance apart, but left free to rotate.

(d) Suppose you had a large collection of compass needles, mounted on pins at regular intervals along a straight line. How would they point (assuming the earth's magnetic field can be neglected)? [A rectangular array of compass needles aligns itself spontaneously, and this is sometimes used as a demonstration of "ferromagnetic" behavior on a large scale. It's a bit of a fraud, however, since the mechanism here is purely classical, and much weaker than the quantum mechanical **exchange forces** that are actually responsible for ferromagnetism.[13]]

More Problems on Chapter 6

! **Problem 6.22** In Prob. 6.4, you calculated the force on a dipole by "brute force." Here's a more elegant approach. First write $\mathbf{B}(\mathbf{r})$ as a Taylor expansion about the center of the loop:

$$\mathbf{B}(\mathbf{r}) \cong \mathbf{B}(\mathbf{r}_0) + [(\mathbf{r} - \mathbf{r}_0) \cdot \boldsymbol{\nabla}_0]\mathbf{B}(\mathbf{r}_0),$$

where $\mathbf{r}_0$ is the position of the dipole and $\boldsymbol{\nabla}_0$ denotes differentiation with respect to $\mathbf{r}_0$. Put this into the Lorentz force law (Eq. 5.16) to obtain

$$\mathbf{F} = I \oint d\mathbf{l} \times [(\mathbf{r} \cdot \boldsymbol{\nabla}_0)\mathbf{B}(\mathbf{r}_0)].$$

Or, numbering the Cartesian coordinates from 1 to 3:

$$F_i = I \sum_{j,k,l=1}^{3} \epsilon_{ijk} \left\{ \oint r_l \, dl_j \right\} \left[\nabla_{0_l} B_k(\mathbf{r}_0) \right],$$

where ϵ_{ijk} is the **Levi-Civita symbol** ($+1$ if $ijk = 123$, 231, or 312; -1 if $ijk =$ 132, 213, or 321; 0 otherwise), in terms of which the cross-product can be written $(\mathbf{A} \times \mathbf{B})_i = \sum_{j,k=1}^{3} \epsilon_{ijk} A_j B_k$. Use Eq. 1.108 to evaluate the integral. Note that

$$\sum_{j=1}^{3} \epsilon_{ijk}\epsilon_{ljm} = \delta_{il}\delta_{km} - \delta_{im}\delta_{kl},$$

where δ_{ij} is the Kronecker delta (Prob. 3.52).

[13] For an intriguing exception, see B. Parks, *Am. J. Phys.* **74**, 351 (2006), Section II.

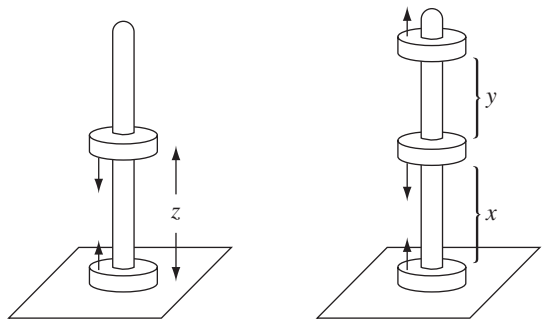


FIGURE 6.31

Problem 6.23 A familiar toy consists of donut-shaped permanent magnets (magnetization parallel to the axis), which slide frictionlessly on a vertical rod (Fig. 6.31). Treat the magnets as dipoles, with mass m_d and dipole moment $\mathbf{m}$.

(a) If you put two back-to-back magnets on the rod, the upper one will "float"—the magnetic force upward balancing the gravitational force downward. At what height (z) does it float?

(b) If you now add a *third* magnet (parallel to the bottom one), what is the *ratio* of the two heights? (Determine the actual number, to three significant digits.) [*Answer:* (a) $[3\mu_0 m^2/2\pi m_d g]^{1/4}$; (b) 0.8501]

Problem 6.24 Imagine two *charged* magnetic dipoles (charge q, dipole moment $\mathbf{m}$), constrained to move on the z axis (same as Problem 6.23(a), but without gravity). Electrically they repel, but magnetically (if both $\mathbf{m}$'s point in the z direction) they attract.

(a) Find the equilibrium separation distance.

(b) What is the equilibrium separation for two *electrons* in this orientation. [*Answer:* 4.72×10^{-13} m.]

(c) Does there exist, then, a stable bound state of two electrons?

Problem 6.25 Notice the following parallel:

$$\begin{cases} \nabla \cdot \mathbf{D} = 0, & \nabla \times \mathbf{E} = \mathbf{0}, & \epsilon_0 \mathbf{E} = \mathbf{D} - \mathbf{P}, & \text{(no free charge);} \\ \nabla \cdot \mathbf{B} = 0, & \nabla \times \mathbf{H} = \mathbf{0}, & \mu_0 \mathbf{H} = \mathbf{B} - \mu_0 \mathbf{M}, & \text{(no free current).} \end{cases}$$

Thus, the transcription $\mathbf{D} \to \mathbf{B}$, $\mathbf{E} \to \mathbf{H}$, $\mathbf{P} \to \mu_0 \mathbf{M}$, $\epsilon_0 \to \mu_0$ turns an electrostatic problem into an analogous magnetostatic one. Use this, together with your knowledge of the electrostatic results, to rederive

(a) the magnetic field inside a uniformly magnetized sphere (Eq. 6.16);

(b) the magnetic field inside a sphere of linear magnetic material in an otherwise uniform magnetic field (Prob. 6.18);

(c) the average magnetic field over a sphere, due to steady currents within the sphere (Eq. 5.93).

Problem 6.26 Compare Eqs. 2.15, 4.9, and 6.11. Notice that if ρ, **P**, and **M** are *uniform*, the *same integral* is involved in all three:

$$\int \frac{\hat{\boldsymbol{\imath}}}{\imath^2}\, d\tau'.$$

Therefore, if you happen to know the electric field of a uniformly *charged* object, you can immediately write down the scalar potential of a uniformly *polarized* object, and the vector potential of a uniformly *magnetized* object, of the same shape. Use this observation to obtain V inside and outside a uniformly polarized sphere (Ex. 4.2), and **A** inside and outside a uniformly magnetized sphere (Ex. 6.1).

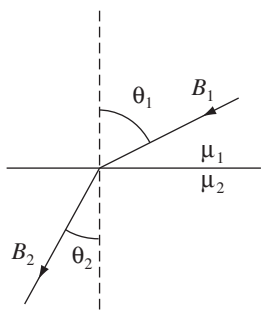


FIGURE 6.32

Problem 6.27 At the interface between one linear magnetic material and another, the magnetic field lines bend (Fig. 6.32). Show that $\tan\theta_2/\tan\theta_1 = \mu_2/\mu_1$, assuming there is no free current at the boundary. Compare Eq. 4.68.

! **Problem 6.28** A magnetic dipole **m** is imbedded at the center of a sphere (radius R) of linear magnetic material (permeability μ). Show that the magnetic field inside the sphere ($0 < r \leq R$) is

$$\frac{\mu}{4\pi}\left\{\frac{1}{r^3}[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}] + \frac{2(\mu_0 - \mu)\mathbf{m}}{(2\mu_0 + \mu)R^3}\right\}.$$

What is the field *outside* the sphere?

Problem 6.29 You are asked to referee a grant application, which proposes to determine whether the magnetization of iron is due to "Ampère" dipoles (current loops) or "Gilbert" dipoles (separated magnetic monopoles). The experiment will involve a cylinder of iron (radius R and length $L = 10R$), uniformly magnetized along the direction of its axis. If the dipoles are Ampère-type, the magnetization is equivalent to a surface bound current $\mathbf{K}_b = M\hat{\boldsymbol{\phi}}$; if they are Gilbert-type, the magnetization is equivalent to surface monopole densities $\sigma_b = \pm M$ at the two ends. Unfortunately, these two configurations produce identical magnetic fields, at exterior points. However, the *interior* fields are radically different—in the first case **B** is in the *same*

general direction as **M**, whereas in the second it is roughly *opposite* to **M**. The applicant proposes to measure this internal field by carving out a small cavity and finding the torque on a tiny compass needle placed inside.

Assuming that the obvious technical difficulties can be overcome, and that the question itself is worthy of study, would you advise funding this experiment? If so, what shape cavity would you recommend? If not, what is wrong with the proposal? [*Hint:* Refer to Probs. 4.11, 4.16, 6.9, and 6.13.]