Lecture Book
Electrodynamics

D. T. McGuiness, PhD

Version: 2024.1

Topics on Fundamental Science

(2025, D. T. McGuiness, PhD)

Current version is 2024.1.

This document includes the contents of Electrodynamics taught at MCI. This document is the part of Electrodynamics.

All relevant code of the document is done using *SageMath* v10.3 and Python v3.12.5.

This document was compiled with LuaTEX and all editing were done using
GNU Emacs using AUCTEX and org-mode package.

This document is based on the books on the topics of Vector Calculus and Electrodynamics which includes *Introduction to Electrodynamics* by D. Griffiths (4th Edition), *Advanced Engineering Mathematics* by E. Kreyszig (9th Edition), *A Modern Introduction to Classical Electrodynamics* by M. Maggiore, *Field and Wave Electromagnetics* by D. K. Cheng

The current maintainer of this work along with the primary lecturer

is D. T. McGuiness, PhD (dtm@mci4me.at).

Contents

Chapter 1

The Purpose of Electromagnetism

1.1 Mechanics in Four Different Views

Newtonian mechanics is usually enough 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Einstein in 1905), for objects that are extremely small (near the size of atoms) it fails for different reasons, and is superseded by quantum mechanics (developed by Bohr, Schrodinger, Heisenberg, and many others, in the 1920's, mostly).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, but even today it cannot claim to be a completely satisfactory system. In this lecture, we shall work exclusively in the domain of classical mechanics, although electrodynamics extends with ease to the other three realms.

> Interestingly, electromagnetism was one of the main catalyst for developing general relativity.

1.1.1 Four Kinds of Forces

Mechanics tells us how behaviour is when subjected to a given force. There are four (4) basic forces known to physics. Listing them in the order of decreasing strength:

- Strong
- Electromagnetic
- Weak
- Gravitational

The **strong force**, 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

Traits	Strong Force	Electromagnetism	Weak Force	Gravity
Affected Particles	quarks[1] and gluons[2]	electrically charged	quarks and leptons[3]	all particles with mass
Force Carrying Particle	gluon (g)	photon[4] (γ)	W and Z bosons (W^+, W^-, Z^0)	graviton (unobserved)
Acting Range	short ($\sim 1\,\text{fm}$)[5]	∞	short	∞
Strength	1	$^1/_{137}$[6]	1×10^{-6}	6×10^{-39}

[1] Elementary particle responsible for making **protons** and **neutron**.
[2] Elementary particle acting as exchange particle for the **strong force** between **quarks**.
[3] Elementary particle affected by the **weak force** but not by the **strong force**.
[4] Elementary particle that is a **quantum** of the **electromagnetic** field, including electromagnetic radiation such as light and radio waves.
[5] 1 femtometre ($1\,\text{fm} = 1 \times 10^{-15}\,\text{m}$). i.e., the gold nucleus radius is approx. 8.45 fm.
[6] Known as the **fine structure constant.**

Table 1.1: The four fundamental forces and their respective properties.

than electrical forces.

The **weak force**, 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 very weak (compared to all of the others) that it is only in scale of huge mass concentrations (like the earth and the sun) that we ever notice it at all.

> For reference, the electrical repulsion between two electrons is 10^{42} times as large as their gravitational attraction

Not only are electromagnetic forces overwhelmingly dominant in everyday life, they are also, the only ones that are completely understood.

There is, of course, a classical theory of gravity and a relativistic one, but no entirely satisfactory quantum mechanical theory of gravity has been constructed. At the present time there is a very successful 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 complete and successful theory, has become a pedestal for physicists: an ideal model that other theories emulate.

The laws of classical electrodynamics were discovered in bits and pieces by Franklin, Coulomb, Ampere, 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.

1.1.2 Unifying Physical Theories

In the beginning, electricity and magnetism were entirely separate subjects. The one dealt with glass rods and cat's fur, batteries, currents, electrolysis, and lightning, whereas 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, Ampere 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 justification for this hypothesis, and soon optics, which is 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, said:

The connection between light and electricity is now established . . . In every flame, in every luminous parti- cle, 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 organthe eye.

By 1900, then, three great branches of physics: **electricity**, **magnetism**, and **optics** had merged into a single unified theory.

> 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 worked on 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 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.

At each step in this hierarchy, the mathematical difficulties mount, and the gap between inspired conjecture and experimental test widens; nevertheless, it is clear the unification of forces initiated by electrodynamics has become a major theme in the progress of physics.

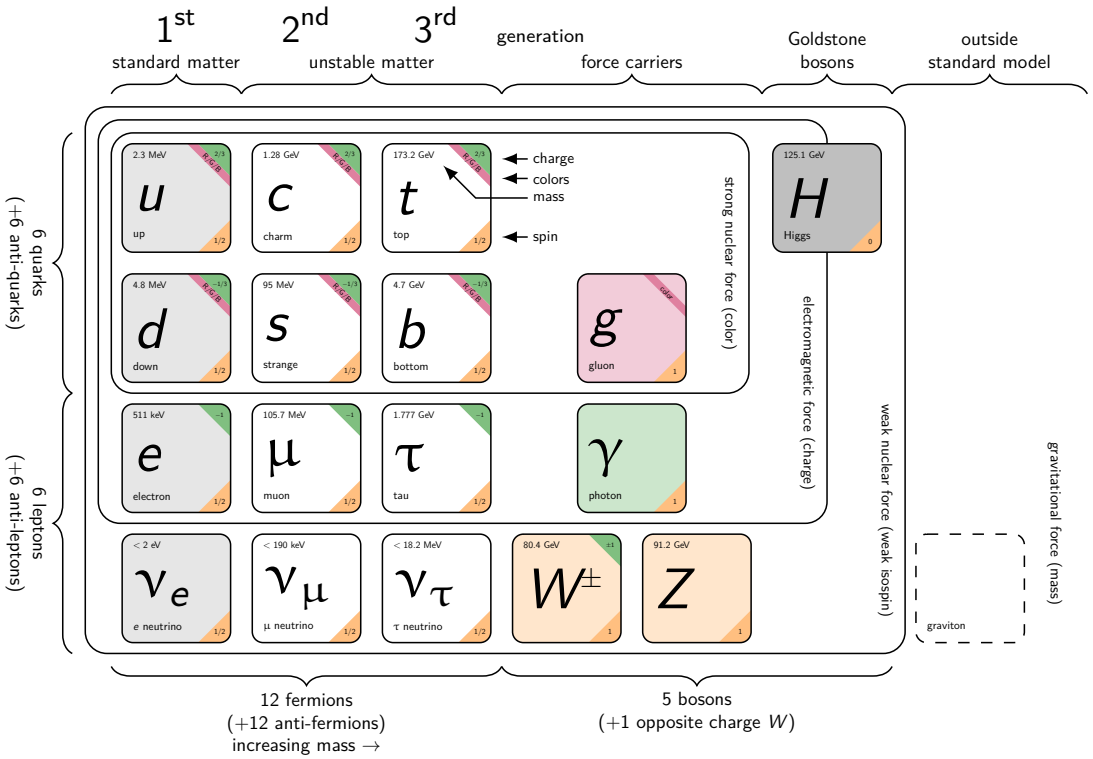


Figure 1.1: The Standard Model of particle physics is the theory describing three of the four known fundamental forces (electromagnetic, weak and strong interactions excluding gravity) in the universe and classifying all known elementary particles. It was developed in stages throughout the latter half of the 20th century, through the work of many scientists worldwide, with the current formulation being finalized in the mid-1970s upon experimental confirmation of the existence of quarks.

1.1.3 Fields of Electrodynamics

The essential problem the theory of electromagnetism hopes to solve is this:

If there exists 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. A second charge, in the presence of these fields, experiences a force. The fields, then, transmit the influence from one charge to the otherthey mediate the interaction.

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

1.1.4 Electric Charge

1. **Charge comes in two varieties:** We call **plus** and **minus**, as 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). The interesting fact is that plus and minus charges occur in exactly equal amounts, to high precision, in bulk matter, so that their effects are almost completely neutralised.

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**. Charges also cannot disappear and then reappear in some other place, the charge, if moved must follow a continuous path. This is called **local conservation of charge**.

3. **Charge is quantised:** Although nothing in classical electrodynamics requires that it be so, electric charge comes only in integer multiples of the basic unit of charge. If we call the charge on the proton $+e$, then the electron carries charge e and never 7.392e, or even 1/2e.

Chapter 2

Vector Calculus

Table of Contents **List of Examples**

2.1 Vector Algebra

2.1.1 Vector Operations

Walking 5 kilometers north and then 12 kilometers east, you will have gone a total of 17 kilometers, but you're not 13 kilometers from where you set out, which is only 7. We need a set of mathematics principles to describe quantities like this, which evidently do not add in the ordinary way.

The reason they don't, is **displacements** have *direction* as well as *magnitude*, and it is essential to take both into account when you combine them. Such objects are called **vectors**.

> Examples include: velocity, acceleration, force, momentum ...

By contrast, quantities that have magnitude but no direction are called **scalars**.

> Examples include: mass, charge, density, temperature, ..

We shall use **boldface** ($\boldsymbol{A}$, $\boldsymbol{B}$, and so on) for vectors and ordinary type for scalars. The magnitude of a vector $\boldsymbol{A}$ is written $|\boldsymbol{A}|$ 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 arrowed indicates its direction.

Minus $\boldsymbol{A}$ ($-\boldsymbol{A}$) is a vector with the same magnitude as $\boldsymbol{A}$ but of opposite direction.

> Vectors have magnitude and direction but *not location*

We define four (4) vector operations: addition and three kinds of multiplication.

(i) Addition of two vectors: Place the tail of $\boldsymbol{B}$ at the head of $\boldsymbol{A}$. The sum, $\boldsymbol{A}+\boldsymbol{B}$, is the vector from the tail of $\boldsymbol{A}$ to the head of $\boldsymbol{B}$. Addition is *commutative*:

$$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A},$$

5 kilometers east followed by 12 kilometers north gets you to the same place as 12 kilometers north followed by 5 kilometers east. Addition is also *associative*:

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

To subtract a vector, add its opposite

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

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

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

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

$$\boldsymbol{A}\cdot\boldsymbol{B}\equiv AB\cos\theta \tag{2.1}$$

where θ is the angle they form when placed tail-to-tail.

> $\boldsymbol{A}\cdot\boldsymbol{B}$ is itself a *scalar* (hence the alternative name **scalar product**)

The dot product is *commutative*,

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

and *distributive*,

$$\boldsymbol{A}\cdot(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{A}\cdot\boldsymbol{B}+\boldsymbol{A}\cdot\boldsymbol{C}, \tag{2.2}$$

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

$$\boldsymbol{A}\cdot\boldsymbol{A}=A^2 \tag{2.3}$$

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

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

$$\boldsymbol{A}\times\boldsymbol{B}\equiv AB\sin\theta\,\hat{\boldsymbol{n}} \tag{2.4}$$

where $\hat{\boldsymbol{n}}$ is a **unit vector** (vector of magnitude 1) pointing perpendicular to the plane of $\boldsymbol{A}$ and $\boldsymbol{B}$. 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 $\hat{\boldsymbol{n}}$.

> $\boldsymbol{A} \times \boldsymbol{B}$ is itself a *vector* and it is also known as **vector product**.

The cross product is *distributive*,

$$\boldsymbol{A} \times (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} \times \boldsymbol{B}) + (\boldsymbol{A} \times \boldsymbol{C}) \tag{2.5}$$

but **NOT** commutative:

$$(\boldsymbol{B} \times \boldsymbol{A}) = -(\boldsymbol{A} \times \boldsymbol{B}) \tag{2.6}$$

If two vectors are parallel, their cross product is zero. In particular,

$$\boldsymbol{A} \times \boldsymbol{A} = 0,$$

for any vector $\boldsymbol{A}$.

2.1.2 Vector Component Forms

In the previous section, we defined the four (4) vector operations in abstract form, 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 $\hat{x}$, $\hat{y}$, $\hat{z}$ be unit vectors parallel to the x, y, and z axes, respectively.

An arbitrary vector $\boldsymbol{A}$ can be expanded in terms of these **basis vectors**:

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

The symbols A_x, A_y, A_z, are the components of $\boldsymbol{A}$. In geometrical terms they are the **projections** of $\boldsymbol{A}$ along the three (3) coordinate axes (i.e., $A_x = \boldsymbol{A} \cdot \hat{x}$, $A_y = \boldsymbol{A} \cdot \hat{y}$, $A_z = \boldsymbol{A} \cdot \hat{z}$). We can now reformulate each of the four vector operations as a rule for manipulating components:

$$\begin{aligned} \boldsymbol{A} + \boldsymbol{B} &= \left(A_x \hat{x} + A_y \hat{y} + A_z \hat{z}\right) + \left(B_x \hat{x} + B_y \hat{y} + B_z \hat{z}\right) \\ &= \left(A_x + B_x\right) \hat{x} + \left(A_y + B_y\right) \hat{y} + \left(A_z + B_z\right) \hat{z} \end{aligned}$$

The operation rules are as follows:

(i) To add vectors, add like components.

$$a\boldsymbol{A} = \left(aA_x\right) \hat{x} + \left(aA_y\right) \hat{y} + \left(aA_z\right) \hat{z}$$

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

As $\hat{x}$, $\hat{y}$, $\hat{z}$ are mutually perpendicular unit vectors, the following properties are valid:

$$\hat{x} \cdot \hat{x} = \hat{y} \cdot \hat{y} = \hat{z} \cdot \hat{z} = 1 \qquad \hat{x} \cdot \hat{y} = \hat{x} \cdot \hat{z} = \hat{y} \cdot \hat{z} = 0$$

Accordingly,

$$\boldsymbol{A} \cdot \boldsymbol{B} = \left(A_x \hat{x} + A_y \hat{y} + A_z \hat{z}\right) \cdot \left(B_x \hat{x} + B_y \hat{y} + B_z \hat{z}\right)$$

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

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

so

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

Similarly,

$$\hat{x}\times\hat{x} = \hat{y}\times\hat{y} = \hat{z}\times\hat{z} = 0,$$

$$\hat{x}\times\hat{y} = -\hat{y}\times\hat{x} = \hat{z},$$

$$\hat{y}\times\hat{z} = -\hat{z}\times\hat{y} = \hat{x},$$

$$\hat{z}\times\hat{x} = -\hat{x}\times\hat{z} = \hat{y}.$$

Therefore,

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

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

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

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

2.1.3 Triple Products

As Since the cross product of two (2) vectors is itself a vector, it can be dotted or crossed with a 3rd vector to form a *triple* product.

(i) Scalar triple product: Evidently,

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

for they all correspond to the same value. Note that "alphabetical" order is preserved. The "nonalphabetical" triple products,

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

have the opposite sign. In component form,

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

Note that the dot and cross can be interchanged:

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

however, the placement of the parentheses is critical:

> $(\boldsymbol{A}\cdot\boldsymbol{B})\times\boldsymbol{C}$ is a meaningless expression. You can't make a cross product from a scalar and a vector.

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

$$\boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C})=\boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C})-\boldsymbol{C}(\boldsymbol{A}\cdot\boldsymbol{B}).$$

Notice that

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

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

$$(\boldsymbol{A}\times\boldsymbol{B})\cdot(\boldsymbol{C}\times\boldsymbol{D})=(\boldsymbol{A}\cdot\boldsymbol{C})\,(\boldsymbol{B}\cdot\boldsymbol{D})-(\boldsymbol{A}\cdot\boldsymbol{D})\,(\boldsymbol{B}\cdot\boldsymbol{C})\,,$$
$$\boldsymbol{A}\times\left[\boldsymbol{B}\times(\boldsymbol{C}\times\boldsymbol{D})\right]=\boldsymbol{B}\left[\boldsymbol{A}\cdot(\boldsymbol{C}\times\boldsymbol{D})\right]-(\boldsymbol{A}\cdot\boldsymbol{B})\,(\boldsymbol{C}\times\boldsymbol{D})\,.$$

2.1.4 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**:

$$\boldsymbol{r}\equiv x\,\hat{x}+y\,\hat{y}+z\,\hat{z}$$

Throughout this course, $\boldsymbol{r}$ will be used to measure **distance**. Its magnitude:

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

is the distance from the origin, and

$$\hat{r}=\frac{\boldsymbol{r}}{r}=\frac{(x)\;\hat{x}+(y)\;\hat{y}+(z)\;\hat{z}}{\sqrt{x^2+y^2+z^2}}$$

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

$$\mathrm{d}\boldsymbol{l}=(\,\mathrm{d}x)\;\hat{x}+(\,\mathrm{d}y)\;\hat{y}+(\,\mathrm{d}z)\;\hat{z}.$$

In electrodynamics, one frequently encounters problems involving two (2) points:

1. a **source point**, $\boldsymbol{r}'$, where an electric charge is located
2. a **field point**, $\boldsymbol{r}$, at which you are calculating the electric or magnetic field

To make this redundant calculations simpler and to save on ink, let's use the following short-hand notation:

$$\boldsymbol{\mathscr{r}}\equiv\boldsymbol{r}-\boldsymbol{r}'$$

Its magnitude is:

$$\mathscr{r}=\boldsymbol{pr}-\mathbf{r}'$$

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

$$\hat{\boldsymbol{\imath}} = \frac{\boldsymbol{\imath}}{\imath} = \frac{\boldsymbol{r} - \boldsymbol{r}'}{\boldsymbol{pr} - \boldsymbol{r}'}$$

In Cartesian coordinates,

$$\boldsymbol{\imath} = \left(x - x'\right)\,\hat{x} + \left(y - y'\right)\,\hat{y} + \left(z - z'\right)\,\hat{z}$$
$$\imath = \sqrt{\left(x - x'\right)^2 + \left(y - y'\right)^2 + \left(z - z'\right)^2}$$
$$\hat{\boldsymbol{\imath}} = \frac{\left(x - x'\right)\,\hat{x} + \left(y - y'\right)\,\hat{y} + \left(z - z'\right)\,\hat{z}}{\sqrt{\left(x - x'\right)^2 + \left(y - y'\right)^2 + \left(z - z'\right)^2}}$$

2.2 Differential Calculus

2.2.1 Ordinary Derivatives

Assume a function of one variable: $f(x)$. Therefore, what does the derivative, df/dx, do. It tells us how rapidly the function $f(x)$ varies when we change the argument x by a tiny amount, $\mathrm{d}x$:

$$\mathrm{d}f = \left(\frac{df}{dt}\right)\,\mathrm{d}x$$

If we increment x by an infinitesimal amount dx, then f changes by an amount df.

> the derivative is the proportionality factor

> Geometrically, the derivative df/dt is the *slope* of the graph of f versus x.

2.2.2 Gradient

Assume a function of three (3) variables, for example, the temperature $T(x, y, z)$ in the lecture 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 generalise the notion of "derivative" to functions like T, which depend not on *one* but on *three* variables.

A derivative tells 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:

1. If we go straight up, then the temperature will probably increase fairly rapidly,
2. 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:

$$\mathrm{d}T = \left(\frac{dT}{dx}\right)\,\mathrm{d}x + \left(\frac{dT}{dy}\right)\,\mathrm{d}y + \left(\frac{dT}{dz}\right)\,\mathrm{d}z$$

This tells us how T changes when we alter all three variables by the infinitesimal amounts dx, dy, dz. We can write the aforementioned equation as a dot product:

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

where

$$\nabla T \equiv \frac{dT}{dx}\,\hat{x} + \frac{dT}{dy}\,\hat{y} + \frac{dT}{dz}\,\hat{z}$$

is the **gradient** of T. Note that ∇T is a **vector quantity**, with three (3) components.

> This is the generalized derivative we have been looking for.

Finding Vector Components 1 **Exam**

1. Find the components of the vector $\boldsymbol{v}$ with given initial point P and terminal point Q. Find $|\boldsymbol{v}|$ and unit vector $\hat{v}$.

$P\,(3,\,2,\,0)\,,$	$Q\,(5,\,-2,\,2)\,,$	$P\,(1,\,1,\,1)\,,$	$Q\,(-4,\,-4,\,-4)$
$P\,(1,\,0,\,1.2)\,,$	$Q\,(0,\,0,\,6.2)\,,$	$P\,(2,\,-2,\,0)\,,$	$Q\,(0,\,4,\,6)$
$P\,(4,\,3,\,2)\,,$	$Q\,(-4,\,-3,\,2)\,,$	$P\,(0,\,0,\,0)\,,$	$Q\,(6,\,8,\,10)$

2. Given the components of a vector $\boldsymbol{v} = \left[v_x,\,v_y,\,v_z\right]$ and a particular initial point P, find the corresponding terminal point Q and the length of $\boldsymbol{v}$ (i.e., $|\boldsymbol{v}|$).

$\boldsymbol{v} = [3,\,-1,\,0]\,;$	$P\,(4,\,6,\,0)\,,$	$\boldsymbol{v} = [8,\,4,\,2]\,;$	$P\,(-8,\,-4,\,-2)\,,$
$\boldsymbol{v} = [0.25,\,2,\,0.75]\,;$	$P\,\{\cdot\}\,0,\,-0.5,\,0,$	$\boldsymbol{v} = [3,\,2,\,6]\,;$	$P\,(4,\,6,\,0)\,,$
$\boldsymbol{v} = [4,\,2,\,-2]\,;$	$P\,(4,\,6,\,0)\,,$	$\boldsymbol{v} = [3,\,-3,\,3]\,;$	$P\,(4,\,6,\,0)\,,$

Solution

Finding Vector Components The solution is as follows:

$$\boldsymbol{v} = (5-3)\,\hat{x} + (-2-2)\,\hat{y} + (2-0)\,\hat{z} = (2)\,\hat{x} + (-4)\,\hat{y} + (2)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(2)^2 + (-4)^2 + (2)^2} = 2\sqrt{6}.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(2)\,\hat{x} + (-4)\,\hat{y} + (2)\,\hat{z}}{2\sqrt{6}} = \left(\frac{1}{\sqrt{6}}\right)\hat{x} + \left(-\frac{2}{\sqrt{6}}\right)\hat{y} + \left(\frac{1}{\sqrt{6}}\right)\hat{z} \quad \blacksquare$$
$$\boldsymbol{v} = (-4-1)\,\hat{x} + (-4-1)\,\hat{y} + (-4-1)\,\hat{z} = (-5)\,\hat{x} + (-5)\,\hat{y} + (-5)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(-5)^2 + (-5)^2 + (-5)^2} = 5\sqrt{3}.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(-5)\,\hat{x} + (-5)\,\hat{y} + (-5)\,\hat{z}}{5\sqrt{3}} = \left(-\frac{1}{\sqrt{3}}\right)\hat{x} + \left(-\frac{1}{\sqrt{3}}\right)\hat{y} + \left(-\frac{1}{\sqrt{3}}\right)\hat{z} \quad \blacksquare$$
$$\boldsymbol{v} = (0-1)\,\hat{x} + (0-0)\,\hat{y} + (6.2-1.2)\,\hat{z} = (-1)\,\hat{x} + (0)\,\hat{y} + (5)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(-1)^2 + (0)^2 + (5)^2} = \sqrt{26}.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(-1)\,\hat{x} + (0)\,\hat{y} + (5)\,\hat{z}}{\sqrt{26}} = \left(-\frac{1}{\sqrt{26}}\right)\hat{x} + (0)\,\hat{y} + \left(-\frac{5}{\sqrt{26}}\right)\hat{z} \quad \blacksquare$$
$$\boldsymbol{v} = (0-2)\,\hat{x} + \big(4-(-2)\big)\,\hat{y} + (6-0)\,\hat{z} = (-2)\,\hat{x} + (6)\,\hat{y} + (6)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(-2)^2 + (6)^2 + (6)^2} = 2\sqrt{19}.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(-2)\,\hat{x} + (6)\,\hat{y} + (6)\,\hat{z}}{2\sqrt{19}} = \left(-\frac{1}{\sqrt{19}}\right)\hat{x} + \left(\frac{3}{\sqrt{19}}\right)\hat{y} + \left(\frac{3}{\sqrt{19}}\right)\hat{z} \quad \blacksquare$$
$$\boldsymbol{v} = (-4-4)\,\hat{x} + (-3-3)\,\hat{y} + (2-2)\,\hat{z} = (-8)\,\hat{x} + (-6)\,\hat{y} + (0)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(-8)^2 + (-6)^2 + (0)^2} = 10.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(-6)\,\hat{x} + (-8)\,\hat{y} + (0)\,\hat{z}}{10} = \left(-\frac{3}{5}\right)\hat{x} + \left(-\frac{4}{5}\right)\hat{y} + (0)\,\hat{z} \quad \blacksquare$$
$$\boldsymbol{v} = (6-0)\,\hat{x} + (8-0)\,\hat{y} + (10-0)\,\hat{z} = (6)\,\hat{x} + (8)\,\hat{y} + (10)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(6)^2 + (8)^2 + (10)^2} = 10\sqrt{2}.$$
$$\hat{v} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|} = \frac{(6)\,\hat{x} + (8)\,\hat{y} + (0)\,\hat{z}}{10\sqrt{2}} = \left(\frac{3}{5\sqrt{2}}\right)\hat{x} + \left(\frac{4}{5\sqrt{(2)}}\right)\hat{y} + \left(\frac{1}{\sqrt{2}}\right)\hat{z} \quad \blacksquare$$

Previously we have defined $\boldsymbol{v} = Q - P$. Here we have $\boldsymbol{v}$ and P. To calculate Q we only need to add individual components of the vector with the initial point P.

$$Q = \boldsymbol{v} + P = (3+4)\,\hat{x} + (-1+6)\,\hat{y} + (0+0)\,\hat{z} = (7)\,\hat{x} + (5)\,\hat{y} + (0)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(3)^2 + (-1)^2 + (0)^2} = \sqrt{10} \quad \blacksquare$$
$$Q = \boldsymbol{v} + P = \big(8+(-8)\big)\,\hat{x} + \big(4+(-4)\big)\,\hat{y} + (-2+2)\,\hat{z} = (0)\,\hat{x} + (0)\,\hat{y} + (0)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(8)^2 + (4)^2 + (2)^2} = 2\sqrt{21} \quad \blacksquare$$
$$Q = \boldsymbol{v} + P = (0.25+0)\,\hat{x} + \big(2+(-0.5)\big)\,\hat{y} + (0.75+0)\,\hat{z} = (0.25)\,\hat{x} + (1.5)\,\hat{y} + (0.75)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(0.25)^2 + (1.5)^2 + (0.75)^2} = \sqrt{74}/4 \quad \blacksquare$$
$$Q = \boldsymbol{v} + P = (3+4)\,\hat{x} + (2+6)\,\hat{y} + (6+0)\,\hat{z} = (7)\,\hat{x} + (8)\,\hat{y} + (6)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(7)^2 + (8)^2 + (6)^2} = \sqrt{149} \quad \blacksquare$$
$$Q = \boldsymbol{v} + P = (4+4)\,\hat{x} + (2+6)\,\hat{y} + (-2+0)\,\hat{z} = (8)\,\hat{x} + (8)\,\hat{y} + (-2)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(8)^2 + (8)^2 + (-2)^2} = 2\sqrt{33} \quad \blacksquare$$
$$Q = \boldsymbol{v} + P = (3+4)\,\hat{x} + (-3+6)\,\hat{y} + (3+0)\,\hat{z} = (7)\,\hat{x} + (3)\,\hat{y} + (3)\,\hat{z},$$
$$|\boldsymbol{v}| = \sqrt{(7)^2 + (3)^2 + (3)^2} = 2\sqrt{67} \quad \blacksquare$$

Example Vector Addition and Scalar Multiplication 2

- Let $\boldsymbol{a} = [2,\ 1,\ 0]$, $\boldsymbol{b} = [-4,\ 2,\ 5]$ and $\boldsymbol{c} = [0,\ 0,\ 3]$. Calculate the following vector operations:

$$\begin{array}{lll} 2\boldsymbol{a}, & -\boldsymbol{a}, & -1/2\boldsymbol{a}, \\ 5\left(\boldsymbol{a}-\boldsymbol{c}\right), & 5\boldsymbol{a}-5\boldsymbol{c}, & (3\boldsymbol{a}-5\boldsymbol{b})+2\boldsymbol{c}, \\ 3\boldsymbol{a}+(-5\boldsymbol{b}+2\boldsymbol{c}), & \boldsymbol{a}+2\boldsymbol{b}, & 2\boldsymbol{b}+\boldsymbol{a}. \end{array}$$

  reset

- Find the dot product (i.e., $\boldsymbol{a}\cdot\boldsymbol{b}$) on the lengths of $\boldsymbol{a} = [1,\ 2,\ 0]$ and $\boldsymbol{b} = [3,\ -2,\ 1]$ as well as the angle (θ) between vectors.

- Let $\boldsymbol{a} = [2,\ 1,\ 4]$, $\boldsymbol{b} = [-4,\ 0,\ 3]$ and $\boldsymbol{c} = [3,\ -2,\ 1]$. Find the following descriptions.

$$\begin{array}{ll} |\boldsymbol{a}|\,,|\boldsymbol{b}|\,,|\boldsymbol{c}|\,, & \boldsymbol{a}\cdot(\boldsymbol{b}+\boldsymbol{c})\,,\ \boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{a}\cdot\boldsymbol{c}, \\ \boldsymbol{a}\cdot\boldsymbol{b}+\boldsymbol{b}\cdot\boldsymbol{c}+\boldsymbol{c}\cdot\boldsymbol{a}, & 4\boldsymbol{a}\cdot 3\boldsymbol{c}, 12\boldsymbol{a}\cdot\boldsymbol{c}, \\ |\boldsymbol{b}+\boldsymbol{c}|\,,|\boldsymbol{b}|+|\boldsymbol{c}|\,, & \boldsymbol{a}\cdot\boldsymbol{c}, |\boldsymbol{a}||\boldsymbol{c}|\,. \end{array}$$

  reset

- Let $\boldsymbol{a} = [1,\ 1,\ 1]$, $\boldsymbol{b} = [2,\ 3,\ 1]$ and $\boldsymbol{c} = [-1,\ 1,\ 0]$. Find the angle between the following:

$$(\boldsymbol{a}-\boldsymbol{c}) \text{ and } (\boldsymbol{b}-\boldsymbol{c})\,, \qquad (\boldsymbol{a}) \text{ and } (\boldsymbol{b}-\boldsymbol{c})\,.$$

- Find the vector product $\boldsymbol{a}\times\boldsymbol{b}$ of $\boldsymbol{a} = [1,\ 1,\ 0]$ and $\boldsymbol{b} = [3,\ 0,\ 0]$.

- Let $\boldsymbol{a} = [1, 2, 0]$, $\boldsymbol{b} = [3,\ -4, 0]$, $\boldsymbol{c} = [3, 5, 2]$, $\boldsymbol{d} = [6, 2, 0]$. Calculate the cross product of:

$$\begin{array}{ll} \boldsymbol{a}\times\boldsymbol{b}, \quad \boldsymbol{b}\times\boldsymbol{a}, & \boldsymbol{a}\times\boldsymbol{c}, \quad |\boldsymbol{a}\times\boldsymbol{c}|\,, \quad \boldsymbol{a}\cdot\boldsymbol{c}, \\ (\boldsymbol{c}+\boldsymbol{d})\times\boldsymbol{d}, \quad \boldsymbol{c}\times\boldsymbol{d}, & \boldsymbol{b}\times\boldsymbol{c}+\boldsymbol{c}\times\boldsymbol{b}, \\ (\boldsymbol{a}+\boldsymbol{b})\times(\boldsymbol{b}+\boldsymbol{a})\,, & (\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c}, \quad \boldsymbol{a}\times(\boldsymbol{b}\times\boldsymbol{c}). \end{array}$$

Solution

Vector Addition and Scalar Multiplication

2.2.3 The Del Operator

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

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

The term in parentheses is called **del** operator:

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

Del is **NOT** a vector, in the usual sense. 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, $\boldsymbol{V}$ mimics the behaviour of an ordinary vector in virtually every way; almost anything that can be done with other vectors can also be done with ∇

Now, an ordinary vector $\boldsymbol{A}$ can multiply in three (3) ways:

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

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

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

It is time to examine the other two vector derivatives: divergence and curl.

2.2.4 Divergence

From the definition of ∇ we construct the divergence:

$$\begin{aligned}\nabla \cdot \boldsymbol{v} &= \left(\left(\frac{\partial}{\partial x}\right)\hat{x} + \left(\frac{\partial}{\partial y}\right)\hat{y} + \left(\frac{\partial}{\partial z}\right)\hat{z}\right) \cdot \left(\left(v_x\right)\hat{x} + \left(v_y\right)\hat{y} + \left(v_z\right)\hat{z}\right) \\ &= \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}.\end{aligned}$$

Observe that the divergence of a vector function $\boldsymbol{v}$ is itself a scalar $\nabla \cdot \boldsymbol{v}$.

2.2.5 Curl

From the definition of ∇ we construct the curl:

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

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

Example **Curl Example** 3

Find he curl ($\nabla \times$) of the following functions.

$$\boldsymbol{v} = (y)\,\hat{x} + \left(2x^2\right)\,\hat{y} + (0)\,\hat{z}, \qquad \boldsymbol{v} = \left(y^n\right)\,\hat{x} + \left(z^n\right)\,\hat{y} + \left(x^n\right)\,\hat{z},$$
$$\boldsymbol{v} = (\sin y)\,\hat{x} + (\cos z)\,\hat{y} + (-\tan x)\,\hat{z}, \qquad \boldsymbol{v} = \left(x^2 - z\right)\,\hat{x} + \left(xe^z\right)\,\hat{y} + (xy)\,\hat{z}.$$

Solution

Curl Example

The curl ($\nabla \times$) of the functions are as follows:

$$
\begin{aligned}
f(x, y, z) &= (y)\ \hat{x} + \left(2x^2\right)\ \hat{y} + (0)\ \hat{z}, \\
\nabla \times f &= (0)\ \hat{x} + (0)\ \hat{y} + (-1 + 4x)\ \hat{z}. \\
f(x, y, z) &= \left(y^n\right)\ \hat{x} + \left(z^n\right)\ \hat{y} + \left(x^n\right)\ \hat{z}, \\
\nabla \times f &= \left(-nz^{n-1}\right)\ \hat{x} + \left(-nx^{n-1}\right)\ \hat{y} + \left(-ny^{n-1}\right)\ \hat{z}. \\
f(x, y, z) &= (\sin y)\ \hat{x} + (\cos z)\ \hat{y} + (-\tan x)\ \hat{z}, \\
\nabla \times f &= (\sin z)\ \hat{x} + \left(\sec^2 x\right)\ \hat{y} + (-\cos y)\ \hat{z}. \\
f(x, y, z) &= \left(x^2 - z\right)\ \hat{x} + \left(xe^{\,z}\right)\ \hat{y} + (xy)\ \hat{z}, \\
\nabla \times f &= \left(x - e^z x\right)\ \hat{x} + (-1 - y)\ \hat{y} + \left(e^z\right)\ \hat{z}.
\end{aligned}
$$

2.2.6 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:

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

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:

$$fg \quad \text{(product of two scalar functions)},$$

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

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)

$$\mathbf{V}(fg) = f\mathbf{V}g + g\mathbf{V}f,$$

(ii) $\mathbf{V}(\mathbf{A}\cdot\mathbf{B}) = \mathbf{A}\times(\mathbf{V}\times\mathbf{B}) + \mathbf{B}\times(\mathbf{V}\times\mathbf{A}) + (\mathbf{A}\cdot\mathbf{V})\mathbf{B} + (\mathbf{B}\cdot\mathbf{V})\mathbf{A}$, two for divergences:
(iii) $\mathbf{V}\cdot(f\mathbf{A}) = f(\mathbf{V}\cdot\mathbf{A}) + \mathbf{A}\cdot(\mathbf{V}f)$,
(iv) $\mathbf{V}\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\mathbf{V}\times\mathbf{A}) - \mathbf{A}\cdot(\mathbf{V}\times\mathbf{B})$,
and two for curls:
(v)

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

(vi)

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

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,

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

It is also possible to formulate three quotient rules:

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

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

2.2.7 Second Derivatives

The gradient, the divergence, and the curl are the only first derivatives we can make with $\nabla\cdot\boldsymbol{v}$ by applying ∇ twice, we can construct five (5) types of 2^{nd} derivatives.

The gradient ∇T is a vector, so we can take the divergence and curl of it:

1. Divergence of gradient: $\nabla\cdot(\nabla T)$.
2. Curl of gradient: $\nabla\times(\nabla T)$.

The divergence $\boldsymbol{V} \cdot \boldsymbol{v}$ is a scalar, therefore all we can do is take its gradient:

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

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

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

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

Divergence of a Gradient

$$\begin{aligned}\nabla \cdot (\nabla T) &= \left(\hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}\right) \cdot \left(\frac{\partial T}{\partial x}\,\hat{x} + \frac{\partial T}{\partial y}\,\hat{y} + \frac{\partial T}{\partial z}\,\hat{z}\right) \\ &= \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}.\end{aligned}$$

This object, which we write as $\nabla^2 T$ for short, is called the **Laplacian** of T, which will be our focus later.

> The Laplacian of a scalar T is a scalar.

Occasionally, we will use the Laplacian of a vector, $\nabla^2 \boldsymbol{v}$. By this we mean a **vector** quantity whose x-component is the Laplacian of v_x, and so on.

$$\nabla^2 \boldsymbol{v} \equiv \left(\nabla^2 v_x\right)\,\hat{x} + \left(\nabla^2 v_y\right)\,\hat{y} + \left(\nabla^2 v_z\right)\,\hat{z}$$

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

Laplacian of a Vector 4 **Exam**

Calculate the Laplacian of the following functions:

$$(i)\quad T_a = x_2 + 3xy + 3z + 4, \qquad (ii)\quad T_b = \sin x \sin y \sin z,$$
$$(iii)\quad T_c = e^{-5x}\sin 4y \cos 3z, \qquad (iv)\quad \boldsymbol{v} = \left(x^2\right)\,\hat{x} + \left(3xz^2\right)\,\hat{y} + (-2xz)\,\hat{z}.$$

Solution

Laplacian of a Vector

The solution to the Laplacian of the functions are as follows:

(i) $\dfrac{\partial^2 T_a}{\partial x^2} = 2;\ \dfrac{\partial^2 T_a}{\partial y^2} = 0;\ \dfrac{\partial^2 T_a}{\partial z^2} = 0 \quad \rightarrow \quad \nabla^2 T_a = 2$ ■

(ii) $\dfrac{\partial^2 T_b}{\partial x^2} = \dfrac{\partial^2 T_b}{\partial y^2} = \dfrac{\partial^2 T_b}{\partial z^2} = -3T_b \quad \rightarrow \quad \nabla^2 T_b = -3T_b = 3\sin x \sin y \sin z$ ■

(iii) $\dfrac{\partial^2 T_c}{\partial x^2} = 25T_c;$

$\dfrac{\partial^2 T_c}{\partial y^2} = -16T_c;\quad \dfrac{\partial^2 T_c}{\partial z^2} = -9T_c \quad \rightarrow \quad \nabla^2 T_c = 0$ ■

(iii) $\dfrac{\partial^2 v_x}{\partial x^2} = 2;\ \dfrac{\partial^2 v_x}{\partial y^2} = 0;\ \dfrac{\partial^2 v_x}{\partial z^2} = 0 \quad \rightarrow \quad \nabla^2 v_x = 0,$

$\dfrac{\partial^2 v_y}{\partial x^2} = 0;\ \dfrac{\partial^2 v_y}{\partial y^2} = 0;\ \dfrac{\partial^2 v_y}{\partial z^2} = 6 \quad \rightarrow \quad \nabla^2 v_y = 6x,$

$\dfrac{\partial^2 v_z}{\partial x^2} = 0;\ \dfrac{\partial^2 v_z}{\partial y^2} = 0;\ \dfrac{\partial^2 v_z}{\partial z^2} = 0 \quad \rightarrow \quad \nabla^2 v_z = 0,$

$\nabla^2 \boldsymbol{v} = 2\,\hat{x} + 6x\,\hat{y}$ ■

Curl of a Gradient

The curl of a gradient is **always** zero:

$$\nabla \times (\nabla T)$$

This is an **important fact**, which will be used repeatedly. Without going into too much detail into the proof, it relies on the following relation:

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

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

Gradient of Divergence

This operation rarely occurs in physical applications, and it has not been given any special name of its own.

Notice that $\nabla(\nabla \cdot \boldsymbol{v})$ is not the same as the Laplacian of a vector:

$$\nabla^2 = (\nabla \cdot \nabla) \neq \nabla(\nabla \cdot \boldsymbol{v})$$

Divergence of a Curl

Like the curl of a gradient, is always zero:

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

Curl of a Curl

As you can check from the definition of ∇:

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

So curl-of-curl gives nothing new; the first term is just number Divergence of a Curl, and the second is the Laplacian.

Really, then, there are just two kinds of second derivatives:

1. the Laplacian,
2. gradient-of-divergence

> It is possible to work out 3rd derivatives, but fortunately second derivatives suffice for practically all physical applications.

2.3 Integral Calculus

2.3.1 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**, which will be the focus of this section.

a **Line Integrals** an expression of the form:

$$\int_a^b \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{l}$$

where $\boldsymbol{v}$ is a vector function, $d\boldsymbol{l}$ is the infinitesimal displacement vector, and the integral is to be carried out along a prescribed path $\mathcal{P}$ from point $\boldsymbol{a}$ to point $\boldsymbol{b}$. If the path forms a closed loop (i.e., if $\boldsymbol{b} = \boldsymbol{a}$), We put a circle on the integral sign:

$$\oint \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{l}$$

At each point on the path, we take the dot product of $\boldsymbol{v}$ (evaluated at that point) with the displacement $d\boldsymbol{l}$ to the next point on the path.

> A good example of a line integral is the work done by a force $\boldsymbol{F}$:
>
> $$W = \int \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$$

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

Fluid Flow 5 **Exam**

A fluid's velocity field is $\boldsymbol{F} = (x)\ \hat{x} + (z)\ \hat{y} + (y)\ \hat{z}$.

Find the flow along the helix $\boldsymbol{\ell}(t) = (\cos t)\ \hat{x} + (\sin t)\ \hat{y} + (t)\ \hat{z}$ with a range of $0 \leq t \leq \pi/2$.

Solution

Fluid Flow
We first evaluate $\boldsymbol{F}$ on the curve:

$$\boldsymbol{F} = (x)\ \hat{x} + (z)\ \hat{y} + (y)\ \hat{z} = (\cos t)\ \hat{x} + (t)\ \hat{y} + (\sin t)\ \hat{z} \qquad \text{Substitute x} = \cos\text{t, z} = \text{t, y} = \sin\text{t.}$$

and then find $d\boldsymbol{\ell}/dt$:

$$\frac{d\boldsymbol{\ell}}{dt} = (-\sin t)\ \hat{x} + (\cos t)\ \hat{y} + (0)\ \hat{z}.$$

Then we integrate $\boldsymbol{F} \cdot \left(d\boldsymbol{\ell}/dt\right)$ from $t = 0$ to $t = \pi/2$:

$$\begin{aligned} \boldsymbol{F} \cdot \frac{d\boldsymbol{\ell}}{dt} &= (\cos t)\,(-\sin t) + (t)\,(\cos t) + (\sin t)\,(1)\,, \\ &= -\sin t \cos t + t \cos t + \sin t. \end{aligned}$$

Which makes,

$$\text{Flow} = \int_{t=a}^{t=b} \boldsymbol{F} \cdot \frac{d\boldsymbol{\ell}}{dt}\, dt = \int_0^{\pi/2} (-\sin t \cos t + t\cos t + \sin t)\ dt,$$
$$= \left[\frac{\cos^2 t}{2} + t\sin t\right]\Bigg|_0^{\pi/2} = \left(0 + \frac{\pi}{2}\right) - \left(\frac{1}{2} + 0\right) = \frac{\pi}{2} - \frac{1}{2} \quad \blacksquare$$

Example **Circulation of a Field** 6

Find the circulation of the field $\boldsymbol{F} = (x - y)\ \hat{x} + x\,\hat{y}$ around the circle $\boldsymbol{\ell}(t) = (\cos t)\ \hat{x} + (\sin t)\ \hat{y} + (0)\ \hat{z}$ with a range of $0 \leq t \leq 2\pi$.

Solution

Circulation of a Field
On the circle, $\boldsymbol{F} = (x - y)\ \hat{x} + (x)\ \hat{y} + (0)\ \hat{z} = (\cos t - \sin t)\ \hat{x} + (\cos t)\ \hat{y} + (0)\ \hat{z}$ and

$$\frac{d\boldsymbol{\ell}}{dt} = (-\sin t)\ \hat{x} + (\cos t)\ \hat{y} + (0)\ \hat{z}.$$

Then

$$\boldsymbol{F} \cdot \frac{d\boldsymbol{\ell}}{dt} = -\sin t \cos t + \underbrace{\sin^2 t + \cos^2 t}_{1},$$

Gives.

$$\text{Circulation} = \int_0^{2\pi} \boldsymbol{F} \cdot \frac{d\boldsymbol{\ell}}{dt}\, dt = \int_0^{2\pi} (1 - \sin t \cos t)\ dt$$
$$= \left[t - \frac{\sin^2 t}{2}\Bigg|_0^{2\pi}\right] = 2\pi \quad \blacksquare$$

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

$$\int_{\mathcal{S}} \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{a}$$

where $\boldsymbol{v}$ is a vector function, and the integral is over a specified surface $\mathcal{S}$. Here $\mathrm{d}a$ is an infinitesimal patch of area, with direction perpendicular to the surface. There are, two (2) directions perpendicular to any surface, so the sign of a surface integral is intrinsically ambiguous.
If the surface is closed (forming a "ballon"), we put a circle on the integral sign

$$\oint \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{a}$$

> Tradition dictates that "outward" is positive, but for open surfaces it's arbitrary.

As an example, if $\boldsymbol{v}$ describes the flow of a fluid (mass per unit area per unit time), then $\int \boldsymbol{v} \cdot \mathrm{d}\boldsymbol{a}$ represents the total mass per unit time passing through the surface.
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.

Example **Double Integrals** 7

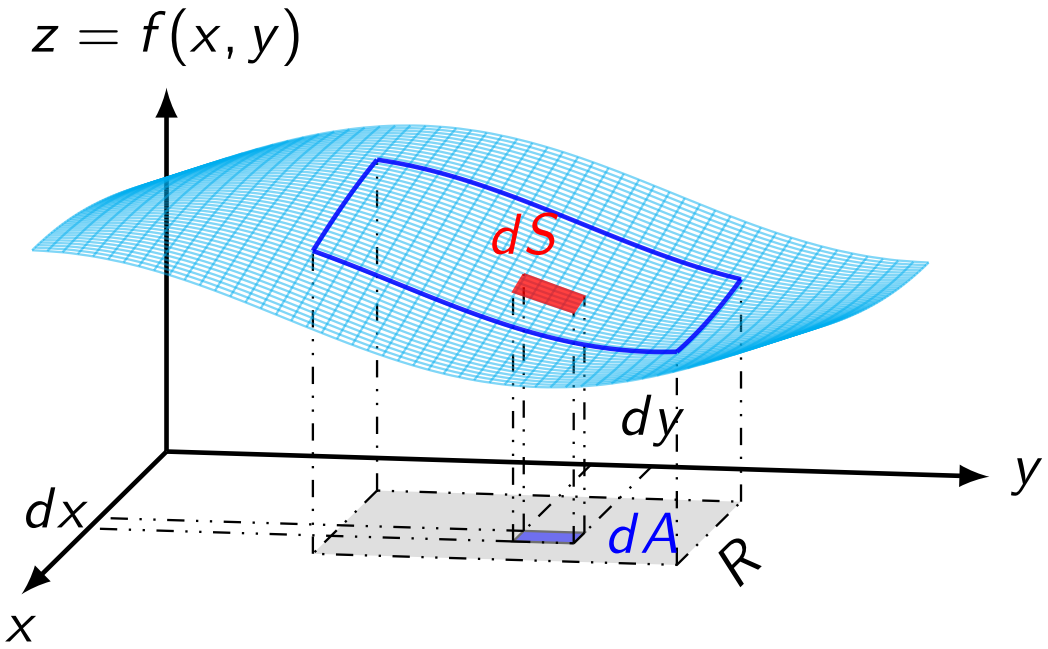


Figure 2.1

Find the following double integrals:

$$\int_0^1 \int_x^{2x} (x+y)^2 \, dy \, dx, \qquad \int_0^1 \int_y^{\sqrt{y}} (1-2xy) \, dx \, dy,$$
$$\int_0^3 \int_x^3 \cosh(x+y) \, dy \, dx, \qquad \int_0^1 \int_0^{y^3} \exp y^4 \, dx \, dy.$$

Solution

Double Integrals

The solution to integrations are as follows:

$$\begin{aligned}
\int_0^1 \int_x^{2x} (x+y)^2 \, dy \, dx &= \int_0^1 \int_x^{2x} x^2 + 2xy + y^2 \, dy \, dx, \\
&= \int_0^1 \left[yx^2 + xy^2 + \frac{y^3}{3} \right] \Bigg|_x^{2x} dx, \\
&= \int_0^1 \left(4x^3 + \frac{7x^3}{3} \right) dx, \\
&= \left[4x^3 + \frac{7x^4}{12} \right] \Bigg|_0^1 = \frac{19}{12} \quad \blacksquare \\
\int_0^1 \int_y^{\sqrt{y}} (1 - 2xy) \, dx \, dy &= \int_0^1 \left[x - x^2 y \right] \Big|_y^{\sqrt{y}} dy, \\
&= \int_0^1 \left[\left(\sqrt{y} - y^2 \right) - \left(y - y^3 \right) \right] dy = \int_0^1 \left[y^3 + \sqrt{y} - y^2 - y \right] dy, \\
&= \left[\frac{y^4}{4} + \frac{2}{3} y^{3/2} - \frac{y^3}{3} - \frac{y^2}{2} \right] \Bigg|_0^1, \\
&= \left(\frac{1}{4} + \frac{2}{3} - \frac{1}{3} - \frac{1}{2} \right) - (0) = \frac{1}{12} \quad \blacksquare \\
\int_0^3 \int_x^3 \cosh(x+y) \, dy \, dx &= \int_0^1 \left[\sinh(x+y) \right] \Big|_x^3 dx = \int_0^1 \left[\sinh(3+x) - \sinh(2x) \right] dx
\end{aligned}$$

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

$$\int_{\mathcal{V}} T \, \mathrm{d}\tau$$

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

$$\mathrm{d}\tau = \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

As an 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:

$$\int \boldsymbol{v} \, \mathrm{d}\tau = \int \left(v_x \, \hat{x} + v_y \, \hat{y} + v_z \, \hat{z} \right) \mathrm{d}\tau = \hat{x} \int v_x \, \mathrm{d}\tau + \hat{y} \int v_y \, \mathrm{d}\tau + \hat{z} \int v_z \, \mathrm{d}\tau.$$

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

2.3.2 The Fundamental Theorem of Calculus

Assume $f(x)$ is a function of one (1) variable. The **fundamental theorem of calculus** says:

Calculus Theorem

the integral of a derivative over some region is given by the value of the function at the end points (boundaries)

$$\int_a^b \left(\frac{df}{dx}\right) dx = f(x) - f(a) \qquad \text{or} \qquad \int_a^b F(x)\, dx = f(x) - f(a)$$

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.

2.3.3 The Fundamental Theorem for Gradients

Suppose we have a scalar function of three variables $T(x, y, z)$. Starting at point **a**, move a small distance $d\boldsymbol{l}_1$. The function T will change by an amount:

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

Now move an additional small displacement $d\boldsymbol{l}_2$. The incremental change in T will be:

$$dT = (\nabla T) \cdot d\boldsymbol{l}_2$$

In this manner, proceeding by infinitesimal steps, we make the journey to point $\boldsymbol{b}$. At each step we compute the gradient of T (at that point) and dot it into the displacement $dv1_{...}$ this gives us the change in T.

Gradient Theorem

The total change in T in going from $\boldsymbol{a}$ to $\boldsymbol{b}$ (along the path selected) is:

$$\int_{\boldsymbol{a}}^{\boldsymbol{b}} (\nabla T) \cdot d\boldsymbol{l} = T(\boldsymbol{b}) - T(\boldsymbol{a})$$

Similar to "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**).

Assumee you want to measure the height of GrossGlockner. You could climb the mountain from base, or take the high alpine road, or take a helicopter ride all the way up to top. Regardless or the options you take, you should get the same answer either way (that's the fundamental theorem).

Theorem 1: Incliendent,

Incidentally, as we found in Ex. 1.6, line integrals ordinarily depend on the path taken from $\boldsymbol{a}$ to $\boldsymbol{b}$. 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: $\int_{\boldsymbol{a}}^{\boldsymbol{b}} (\nabla T) \cdot d\boldsymbol{l}$ is independent of the path taken from $\boldsymbol{a}$ to $\boldsymbol{b}$.

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

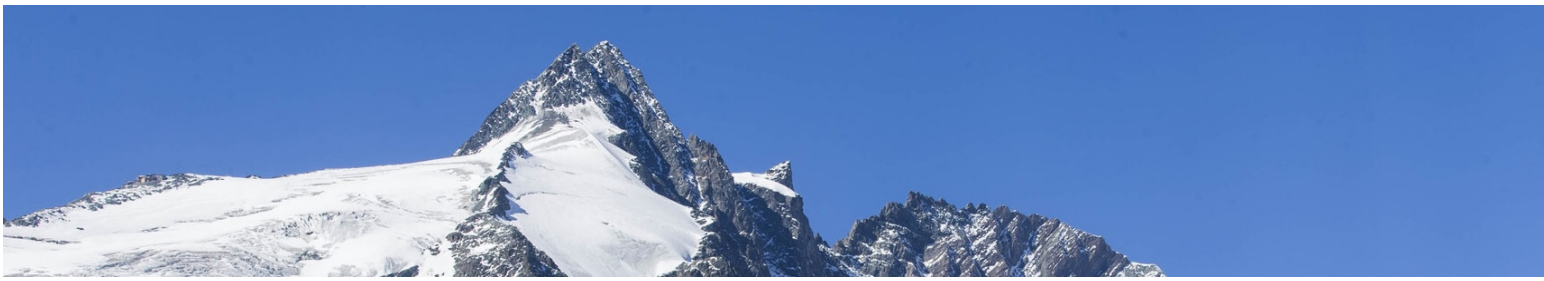


Figure 2.2: To measure the height of a mountain, it doesn't matter what way you take, as long as you know the base and the top, you will know the height.

The Fundamental Theorem for Divergences

this theorem has at least three special names: **Gauss's theorem, Green's theorem,** or simply the **divergence theorem**, The fundamental theorem for divergences states that:

Divergence Theorem

the integral of a derivative (in this case the divergence) over a region (in this case the 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).

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

The boundary term is itself an integral, more 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.

To create an analogy, if $\boldsymbol{v}$ represents the flow of an incompressible fluid, then the flux $\boldsymbol{v}$ 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 tap, pouring out liquid. If we have a bunch of tap 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 (2) ways we could determine how much is being produced:

a. we could count up all the faucets, recording how much each puts out

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:

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

Example Divergence Theorem - I 8

Evaluate both sides of the Divergence theorem for the expanding vector field $\boldsymbol{F} = (x)\,\hat{x} + (y)\,\hat{y} + (z)\,\hat{z}$ over the sphere $x^2 + y^2 + z^2 = a^2$

Solution

Divergence Theorem - I The outer unit normal to S, calculated from the gradient of $f\{\cdot\}\ x,\ y,\ z = x^2+y^2+z^2-a^2$, is:

$$\hat{n} = \frac{\nabla S}{|\nabla S|} = \frac{(2x)\ \hat{x} + (2y)\ \hat{y} + (2z)\ \hat{z}}{\sqrt{4x^2+4y^2+4z^2}} = \frac{(x)\ \hat{x} + (y)\ \hat{y} + (z)\ \hat{z}}{a}. \qquad x^2+y^2+z^2 = a^2 \text{ on S}$$

Therefore:

$$(\boldsymbol{F}\cdot\hat{n})\ da = \frac{x^2+y^2+z^2}{a}\,da = \frac{a^2}{a}\ da = a\,da.$$

This in turn gives us:

$$\iint_S (\boldsymbol{F}\cdot\hat{n})\ da = \iint_S a\,da = a\iint_S da = a\left(4\pi a^2\right) = 4\pi a^3. \qquad \text{Area of S is } 4\text{ß}a^2$$

The divergence of $\boldsymbol{F}$ is:

$$\nabla\cdot\boldsymbol{F} = \frac{\partial}{\partial x}(x) + \frac{\partial}{\partial y}(y) + \frac{\partial}{\partial z}(z) = 3,$$

So,

$$\iiint_V (\nabla\cdot\boldsymbol{v})\ d\tau = \iiint_V 3\,d\tau = 3\left(\frac{4}{3}\pi a^3\right) = 4\pi a^3 \quad \blacksquare$$

Divergence Theorem - II 9 Exam

Check the divergence theorem for the function:

$$\boldsymbol{v} = \left(r^2\cos\theta\right)\ \hat{r} + \left(r^2\cos\phi\right)\ \hat{\theta} + \left(-r^2\cos\theta\sin\phi\right)\ \hat{\phi}.$$

using as your volume one octant of the sphere of radius R.

Solution

Divergence Theorem - II

It is always useful to write the theorem we are going to work on:

$$\underbrace{\iiint_{\mathcal{V}} (\nabla\cdot\boldsymbol{v})\ dV}_{\text{Divergence integral}} = \underbrace{\iint_{\mathcal{S}} \boldsymbol{v}\cdot\boldsymbol{n}\,da}_{\text{Outward flux}}.$$

First solve the left hand side of the equation:

$$\begin{aligned}
\nabla\cdot\boldsymbol{v} &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 r^2\cos\theta\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta r^2\cos\phi\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\left(-r^2\cos\theta\sin\phi\right), \\
&= \frac{1}{r^2}4r^3\cos\theta + \frac{1}{r\sin\theta}\cos\theta r^2\cos\phi + \frac{1}{r\sin\theta}\left(-r^2\cos\theta\cos\phi\right), \\
&= \frac{r\cos\theta}{\sin\theta}\left[4\sin\theta + \cos\phi - \cos\phi\right] = 4r\cos\theta. \\
\int (\nabla\cdot\boldsymbol{v})\ d\tau &= \int (4r\cos\theta)\, r^2\sin\theta\, dr\, d\theta\, d\phi = 4\int_0^R r^3\,dr\int_0^{\pi/2}\cos\theta\sin\theta\,d\theta\int_0^{\pi/2} d\phi, \\
&= \left(R^4\right)\left(\frac{1}{2}\right)\left(\frac{\pi}{2}\right) = \frac{\pi R^4}{4} \quad \blacksquare
\end{aligned}$$

Now it is time to solve the right hand side of the question. As we are aware from the shape, an octant of the sphere has 4 sides to it: the curved surface $xyz \to \boldsymbol{a}_1$, and $xz \to \boldsymbol{a}_2$, $yz \to \boldsymbol{a}_3$ and $xy \to \boldsymbol{a}_4$. These are

$$\begin{aligned}
d\boldsymbol{a}_1 &= \hat{r}\,dl_\theta\,dl_\phi = \hat{r}\,R^2\sin\theta\,d\phi\,d\theta, & d\boldsymbol{a}_2 &= dl_r\,dl_\theta = -\hat{\phi}\,r\,dr\,d\theta, \\
d\boldsymbol{a}_3 &= \hat{\phi}\,dl_r\,dl_\theta = \hat{\phi}\,r\,dr\,d\theta, & d\boldsymbol{a}_4 &= dl_r\,dl_\phi = \hat{\theta}\,r\,dr\,d\theta. \quad \left(\theta = \pi/2\right)
\end{aligned}$$

$$\begin{aligned}
\iint_S \boldsymbol{v}\cdot d\boldsymbol{a} &= \iint_{\mathcal{S}_1} \boldsymbol{v}\cdot d\boldsymbol{a} + \iint_{\mathcal{S}_2} \boldsymbol{v}\cdot d\boldsymbol{a} + \iint_{\mathcal{S}_3} \boldsymbol{v}\cdot d\boldsymbol{a} + \iint_{\mathcal{S}_4} \boldsymbol{v}\cdot d\boldsymbol{a},\\
&= \int_0^{\pi/2}\int_0^{\pi/2} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + r^2\cos\phi\ \hat{\boldsymbol{\theta}} - r^2\cos\theta\sin\phi\ \hat{\boldsymbol{\phi}}\right]\Big|_{r=R}\cdot\left(\hat{\boldsymbol{r}}\,R^2\sin\theta\,d\phi\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + r^2\cos\phi\ \hat{\boldsymbol{\theta}} - r^2\cos\theta\sin\phi\ \hat{\boldsymbol{\phi}}\right]\Big|_{\phi=0}\cdot\left(-\hat{\boldsymbol{\phi}}\,r\,dr\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + r^2\cos\phi\ \hat{\boldsymbol{\theta}} - r^2\cos\theta\sin\phi\ \hat{\boldsymbol{\phi}}\right]\Big|_{\phi=\pi/2}\cdot\left(\hat{\boldsymbol{\phi}}\,r\,dr\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + r^2\cos\phi\ \hat{\boldsymbol{\theta}} - r^2\cos\theta\sin\phi\ \hat{\boldsymbol{\phi}}\right]\Big|_{\theta=\pi/2}\cdot\left(\hat{\boldsymbol{\theta}}\,r\,dr\,d\theta\right),
\end{aligned}$$

Time to do some integration.

$$\begin{aligned}
\iint_S \boldsymbol{v}\cdot d\boldsymbol{a} &= \int_0^{\pi/2}\int_0^{\pi/2} \left[R^2\cos\theta\,\hat{\boldsymbol{r}} + R^2\cos\phi\,\hat{\boldsymbol{\theta}} - R^2\cos\theta\sin\phi\,\hat{\boldsymbol{\phi}}\right]\cdot\left(\hat{\boldsymbol{r}}\,R^2\sin\theta\,d\phi\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + r^2(1)\ \hat{\boldsymbol{\theta}} - (0)\sin\phi\ \hat{\boldsymbol{\phi}}\right]\cdot\left(-\hat{\boldsymbol{\phi}}\,r\,dr\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[r^2\cos\theta\ \hat{\boldsymbol{r}} + (0)\phi\ \hat{\boldsymbol{\theta}} - r^2\cos\theta(1)\ \hat{\boldsymbol{\phi}}\right]\cdot\left(\hat{\boldsymbol{\phi}}\,r\,dr\,d\theta\right)\\
&+ \int_0^{\pi/2}\int_0^{R} \left[(0)\ \hat{\boldsymbol{r}} + r^2\cos\phi\ \hat{\boldsymbol{\theta}} - (0)\ \hat{\boldsymbol{\phi}}\right]\cdot\left(\hat{\boldsymbol{\theta}}\,r\,dr\,d\theta\right).
\end{aligned}$$

Final touches and cleaning up,

$$\begin{aligned}
\iint_S \boldsymbol{v}\cdot d\boldsymbol{a} &= \int_0^{\pi/2}\int_0^{\pi/2} R^4\sin\theta\cos\theta\,d\phi\,d\theta + 0 - \overbrace{\int_0^{\pi/2}\int_0^{R} r^3\cos\theta\,dr\,d\theta + \int_0^{\pi/2}\int_0^{R} r^3\cos\theta\,dr\,d\phi}^{=0},\\
&= R^4\left(\int_0^{\pi/2} d\phi\right)\left(\int_0^{\pi/2}\sin\theta\cos\theta\,d\theta\right),\\
&= R^4\left(\frac{\pi}{2}\right)\left(\frac{\pi}{2}\right),\\
&= \frac{\pi R^4}{4} \quad \blacksquare
\end{aligned}$$

2.3.4 The Fundamental Theorem for Curls

The fundamental theorem for curls, also known as **Stokes' theorem**, states:

Stokes' Theorem

the integral of a derivative over a region ($\mathcal{S}$) is equal to the value of the function at the boundary ($\mathcal{P}$).

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

Similar to the divergence theorem, the boundary term is itself an integral. Specifically, a *closed line integral*.

Remember the curl measures the *twist* of the vectors $\boldsymbol{v}$. Think of a region of high curl as a whirlpool, where if you put a wheel there, it will rotate. Now, the integral of the curl over some surface (or,

more precisely, the *flux* of the curl through the 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.

> $\oint \boldsymbol{v} \cdot d\boldsymbol{l}$ is sometimes called the **circulation** of $\boldsymbol{v}$.

There seems to be an ambiguity in Stokes' theorem: concerning the boundary line integral:

Which way are we supposed to go around (clockwise or counterclockwise)?

The answer is that it doesn't matter which way you go **as long as you are consistent**, for there is an additional sign ambiguity in the surface integral:

Which way does $d\boldsymbol{a}$ *point?*

For a closed surface (i.e., the divergence theorem), $d\boldsymbol{a}$ points in the direction of the outward normal. But for an open surface, which way would be defined as out? Consistency in Stokes' theorem is given by the right-hand rule. If your rings point in the direction of the line integral, then your thumb fixes the direction of $d\boldsymbol{a}$.

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

Preposition I $\int (\nabla \times \boldsymbol{v}) \cdot d\boldsymbol{a}$ depends only on the boundary line, not on the particular surface used.

Preposition II $\oint (\nabla \times \boldsymbol{v}) \cdot d\boldsymbol{a} = 0$ 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.

> **Exercise** 2.10: Surface Area of an Implicit Surface

Find the area of the surface cut from the bottom of the paraboloid $x^2 + y^2 - z = 0$ by the plane $z = 4$.

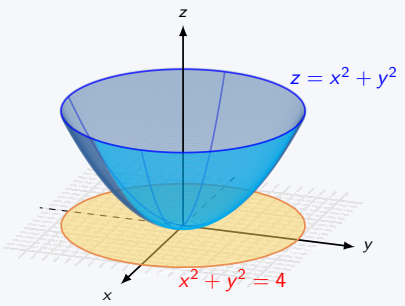


Figure 2.3: We will calculate the area of the parabolic surface in Example 10.

Solution

We sketch the surface S and the region R below it in the xy-plane (**Fig. ??**). The surface S is part of the level surface $F(x, y, z) = x^2 + y^2 - z = 0$, ans R is the disk $x^2 + y^2 \leq 4$ in the xy-plane.

To get a unit vector normal (i.e., $\hat{n}$) to the plane R, we can take $\hat{n} = \hat{z}$. At any point (x, y, z) on the surface, we have:

$$\begin{aligned} F(x, y, z) &= x^2 + y^2 - z \\ \nabla F &= (2x)\ \hat{x} + (2y)\ \hat{y} + (-1)\ \hat{z} \\ |\nabla F| &= \sqrt{(2x)^2 + (2y)^2 + (-1)^2} \\ &= \sqrt{4x^2 + 4y^2 + 1} \\ |\nabla F \cdot \hat{n}| &= |\nabla F \cdot \hat{z}| = |-1| = 1. \end{aligned}$$

In the region R, the area is defined to be $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$. Therefore:

$$\begin{aligned} \text{Surface Area} &= \iint_R \frac{|\nabla F|}{|\nabla F \cdot \hat{n}|}\,\mathrm{d}A \\ &= \iint_{x^2+y^2\leq 4} \sqrt{4x^2 + 4y^2 + 1}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{2\pi}\int_0^2 \sqrt{4r^2 + 1}\,r\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \int_0^{2\pi} \frac{1}{12}\left(4r^2 + 1\right)^{3/2}\Big|_0^2 \mathrm{d}\theta \\ &= \int_0^{2\pi} \frac{1}{12}\left(17^{3/2} - 1\right)\,\mathrm{d}\theta \\ &= \frac{\pi}{6}\left(17\sqrt{17} - 1\right) \quad \blacksquare \end{aligned}$$

Exercise 2.11: Stokes Theorem Over a Hemisphere

Evaluate Stokes's theorem for the hemisphere $S: x^2+y^2+z^2=9,\ z\geq 0$, its bounding circle $C: x^2+y^2=9, z=0$ and the field $\boldsymbol{F}=(y)\ \hat{x}+(-x)\ \hat{y}+(0)\ \hat{z}$.

Tip: Parametrisation of a circle is: $x=r\cos\theta$, $y=r\sin\theta$ and $da=\frac{3}{z}dA$)

Solution

The start by calculating the counter-clockwise circulation around C using the following parametrisation:

$$\boldsymbol{\ell}(\theta)=(3\cos\theta)\ \hat{x}+(3\sin\theta)\ \hat{y}+(0)\ \hat{z},$$
$$\text{where}\qquad 0\leq\theta\leq 2\pi.$$

Using this we can calculate the counter-clockwise circulation.

$$d\boldsymbol{\ell}=(-3\sin\theta\,d\theta)\ \hat{x}+(3\cos\theta\,d\theta)\ \hat{y}+(0)\ \hat{z},$$
$$\boldsymbol{F}=(y)\ \hat{x}+(-x)\ \hat{y}+(0)\ \hat{z}$$
$$=(3\sin\theta)\ \hat{x}+(-3\cos\theta)\ \hat{y}+(0)\ \hat{z},$$
$$\boldsymbol{F}\cdot d\boldsymbol{\ell}=-9\sin^2\theta\,d\theta-9\cos^2\theta\,d\theta=-9\,d\theta,$$
$$\oint_C \boldsymbol{F}\cdot d\boldsymbol{\ell}=\int_0^{2\pi}-9\,d\theta=-18\pi.$$

For the curl of integral we have:

$$\nabla\times\boldsymbol{F}=\begin{vmatrix}\hat{x}&\hat{y}&\hat{z}\\ \partial/\partial x&\partial/\partial y&\partial/\partial z\\ F_x&F_y&F_z\end{vmatrix}$$
$$=(0-0)\ \hat{x}+(0-0)\ \hat{y}+(-1-1)\ \hat{z}=-2\,\hat{z}$$
$$\hat{n}=\frac{\nabla S}{|\nabla S|}=\frac{(x)\ \hat{x}+(y)\ \hat{y}+(z)\ \hat{z}}{\sqrt{x^2+y^2+z^2}}$$
$$=\frac{(x)\ \hat{x}+(y)\ \hat{y}+(z)\ \hat{z}}{3}\qquad\text{Unit normal}$$

Now it is time to define the area of integration (da):

$$da=\frac{|\nabla S|}{|\nabla S\cdot\hat{z}|}\,dA$$
$$=\frac{\left|(2x)\ \hat{x}+(2y)\ \hat{y}+(2z)\ \hat{z}\right|}{2z}$$
$$=\frac{2\overbrace{\sqrt{x^2+y^2+z^2}}^{a}}{2z}$$
$$=\frac{3}{z}dA,$$
$$\nabla\times\boldsymbol{F}\cdot\boldsymbol{n}\,da=-\frac{2z}{3}\frac{3}{z}\,dA=-2\,dA$$

The cardinal direction $\hat{z}$ comes from being the direction **perpendicular** to the surface (S).

$$\iint_S\nabla\times\boldsymbol{F}\cdot\boldsymbol{n}\,da=\iint_{x^2+y^2\leq 9}-2\,dA=-18\pi$$

The circulation around the circle equals the integral of the curl over the hemisphere ■

2.4 Curvilinear Coordinates

2.4.1 Spherical Coordinate System

It is possible to label a point P in 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. **??**.

$$x=r\sin\theta\cos\phi,\qquad y=r\sin\theta\sin\phi,\qquad z=r\cos\theta.$$

Fig. **??** also shows three unit vectors, $\hat{r}$, $\hat{\theta}$, $\hat{\phi}$, pointing in the direction of increase of the corresponding coordinates.

They constitute an **orthogonal** (mutually perpendicular) basis set, similar to $\hat{x}$, $\hat{y}$, $\hat{z}$), and any vector $\boldsymbol{A}$ can be expressed in terms of them, in the usual way:

$$\boldsymbol{A}=(A_{\mathrm{r}})\ \hat{r}+(A_{\cdot})\ \hat{\theta}+(A_{\mathrm{\Phi}})\ \hat{\phi}$$

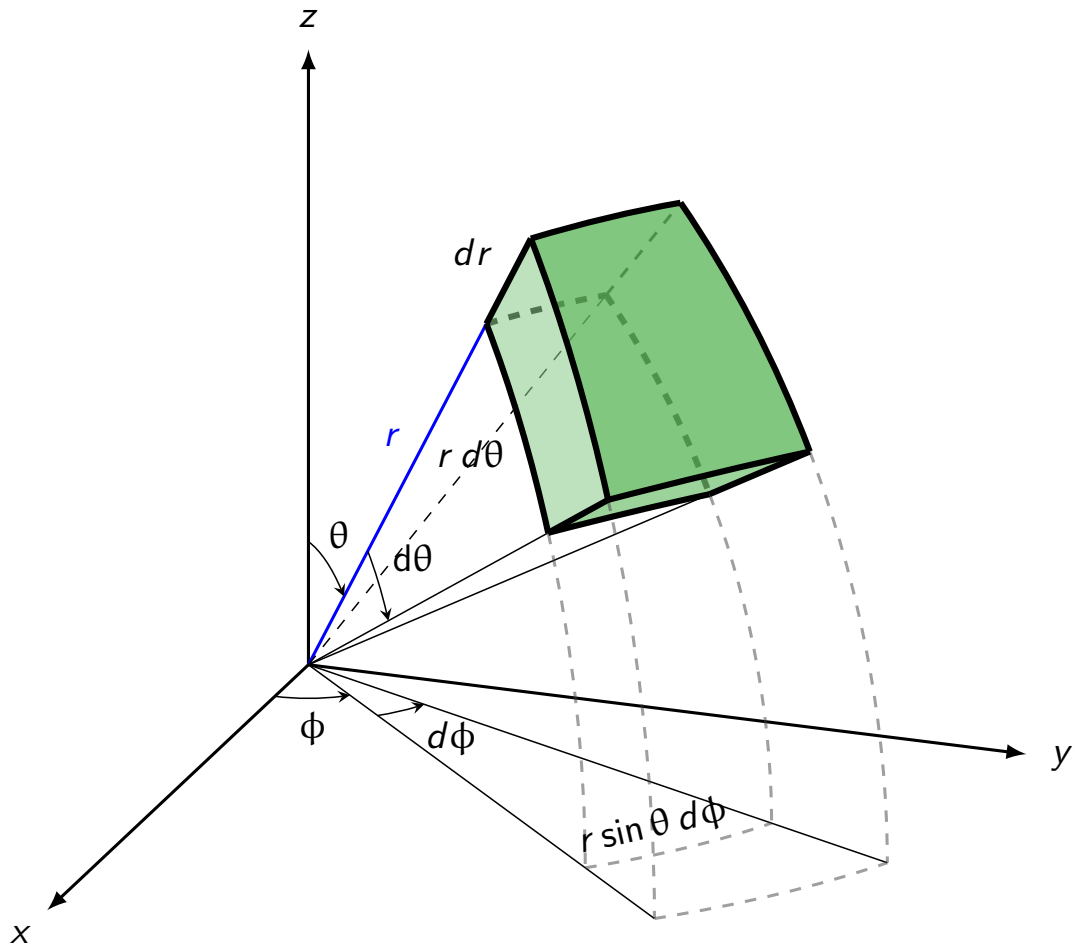


Figure 2.4: The physics convention. Spherical coordinates (r,θ , ϕ) as commonly used: (ISO 80000-2:2019): radial distance r (slant distance to origin), polar angle (theta) (angle with respect to positive polar axis), and azimuthal angle (phi) (angle of rotation from the initial meridian plane)

Here, A_r, A_θ, A_ϕ are the radial, polar, and azimuthal components of vector $\boldsymbol{A}$. In terms of the Cartesian unit vectors:

$$\hat{\boldsymbol{r}} = (\sin\theta\cos\phi)\ \hat{\boldsymbol{r}} + (\sin\theta\sin\phi)\ \hat{\boldsymbol{\theta}} + (\cos\theta)\ \hat{\boldsymbol{\phi}},$$
$$\hat{\boldsymbol{\theta}} = (\cos\theta\cos\phi)\ \hat{\boldsymbol{r}} + (\cos\theta\sin\phi)\ \hat{\boldsymbol{\theta}} + (-\sin\theta)\ \hat{\boldsymbol{\phi}},$$
$$\hat{\boldsymbol{\phi}} = (-\sin\phi)\ \hat{\boldsymbol{r}} + (\cos\phi)\ \hat{\boldsymbol{\theta}} + (0)\ \hat{\boldsymbol{\phi}}.$$

An infinitesimal displacement in the $\hat{r}$ direction is simply $\mathrm{d}r$, just as an infinitesimal element of length in the $\hat{x}$ direction is $\mathrm{d}x$:

$$\mathrm{d}l_r = \mathrm{d}r$$

On the other hand, an infinitesimal element of length in the $\hat{\theta}$ direction (Fig. 1.38b) is not just $\mathrm{d}\theta$ rather,

$$\mathrm{d}l_\theta = r\,\mathrm{d}\theta$$

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

$$\mathrm{d}l_\phi = r\sin\theta\,\mathrm{d}\phi$$

Operator	Mathematical Definition
Gradient	$\nabla T = \frac{\partial T}{\partial r}\,\hat{r} + \frac{1}{r}\frac{\partial T}{\partial \theta}\,\hat{\theta} + \frac{1}{r\sin\theta}\frac{\partial T}{\partial \phi}\,\hat{\phi}$
Divergence	$\nabla \cdot \boldsymbol{v} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 v_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta v_\theta\right) + \frac{1}{r\sin\theta}\frac{\partial v_\phi}{\partial \phi}$
Curl	$\nabla \times \boldsymbol{v} = \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}\left(\sin\theta v_\phi\right) - \frac{\partial v_\theta}{\partial \phi}\right]\hat{r} + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial v_r}{\partial \phi} - \frac{\partial}{\partial r}\left(r v_\phi\right)\right]\hat{\theta} + \frac{1}{r}\left[\frac{\partial}{\partial r}\left(r v_\theta\right) - \frac{\partial v_r}{\partial \theta}\right]\hat{\phi}$
Laplacian	$\nabla^2 T = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial T}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial T}{\partial \theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 T}{\partial \phi^2}$

Table 2.1: Defined mathematical operations in spherical coordinate system.

Thus the general infinitesimal displacement $\mathrm{d}\boldsymbol{l}$ is:

$$\mathrm{d}\boldsymbol{l} = (\,\mathrm{d}r)\;\hat{r} + (r\,\mathrm{d}\theta)\;\hat{\boldsymbol{\theta}} + (r\sin\theta)\;\hat{\boldsymbol{\phi}}$$

This plays the role $\mathrm{d}\boldsymbol{l} = (\,\mathrm{d}x)\;\hat{x} + (\,\mathrm{d}y)\;\hat{y} + (\,\mathrm{d}z)\;\hat{z}$ plays in Cartesian coordinates. The infinitesimal volume element $d\tau$, in spherical coordinates, is the product of the three (3) infinitesimal displacements:

$$\mathrm{d}\tau = \mathrm{d}l_r\,\mathrm{d}l_\theta\,\mathrm{d}l_\phi = r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

> It is not possible to give a general expression for surface elements $\mathrm{d}\boldsymbol{a}$, since these depend on the orientation of the surface. We simply have to analyze the geometry for any given case, which goes for Cartesian and curvilinear coordinates.

Integrating over the surface of a sphere, for instance, makes r constant, whereas θ and ϕ change:

$$\mathrm{d}\boldsymbol{a}_1 = \mathrm{d}l_\theta\,\mathrm{d}l_\phi\,\hat{r} = r^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,\hat{r}$$

On the other hand, if the surface lies in the xy plane, making θ is constant, while r and ϕ vary:

$$\mathrm{d}\boldsymbol{a}_2 = \mathrm{d}l_r\,\mathrm{d}l_\phi\,\hat{\boldsymbol{\theta}} = r\,\mathrm{d}r\,\mathrm{d}\phi\,\hat{\boldsymbol{\theta}}$$

Finally: r ranges from 0 to ∞, ϕ from 0 to 2π, and θ from 0 to π.

Up to know, we only talked about the geometry of spherical coordinates. Now let's translate the vector derivatives (gradient, divergence, curl, and Laplacian) into r, θ, ϕ notation.
Here, then, are the vector derivatives in spherical coordinates:

> **Exercise** 2.12: Volume of A Sphere

Find the volume of a sphere of radius R.

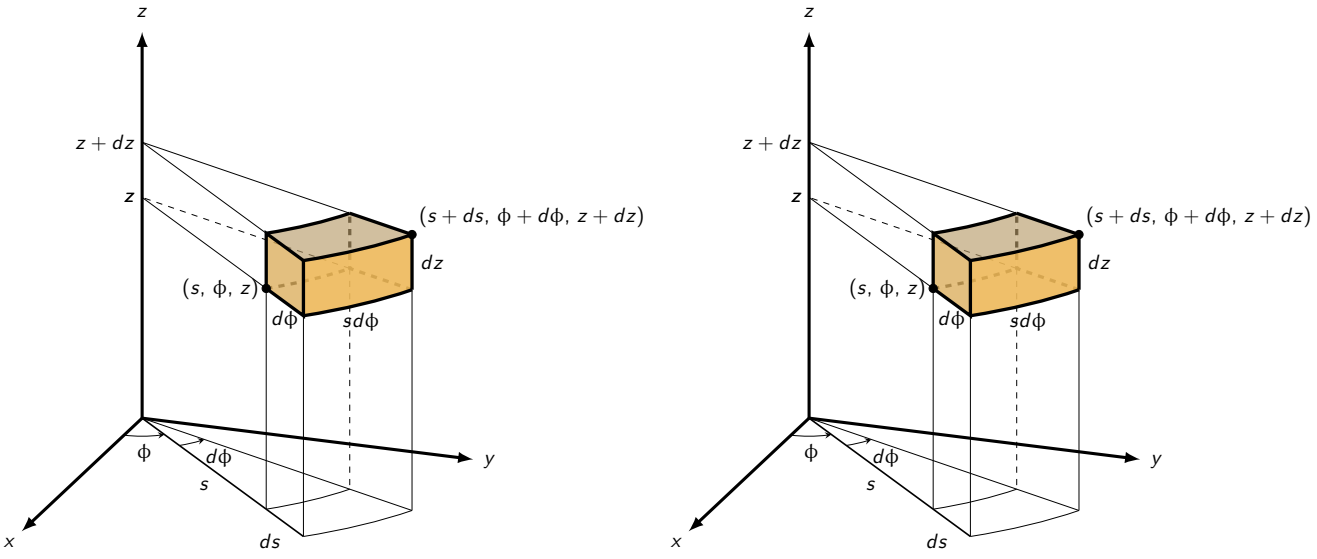
Solution

The derivation is as follows:

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

2.4.2 Cylindrical Coordinates

The cylindrical coordinates (s, ϕ, z) of a point P are defined in Fig. 1.42. Observe 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:



$$x = s\cos\phi \qquad y = s\sin\phi \qquad z = z.$$

The unit vectors are:

$$\hat{s} = \cos\phi\,\hat{x} + \sin\phi\,\hat{y}$$
$$\hat{\phi} = -\sin\phi\,\hat{x} + \cos\phi\,\hat{y}$$
$$\hat{z} = \hat{z}$$

The infinitesimal displacements are

$$dl_s = ds \qquad dl_\phi = s\,d\phi, \qquad dl_z = dz$$

which makes:

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

and the volume element is

$$d\tau' = s\,ds\,d\phi\,dz$$

The range of s is $(0, \infty)$, ϕ is from 0 to $2pi$ and z is from $-\infty$ to $+\infty$

2.5 Dirac Delta Function

2.5.1 A Mathematical Anomaly

Consider the following vector function:

$$\boldsymbol{v} = \frac{1}{r^2}\,\hat{\boldsymbol{r}}$$

At every location, $\boldsymbol{v}$ is directed radially outward (Fig. 1.44). Let's calculate its divergence:

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

This is interesting as this calculation gives us a unforeseen solution. Let's look at this closer. Suppose we integrate over a sphere of radius R, centered at the origin. The surface integral is

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

But the volume integral, $\int \nabla \cdot \boldsymbol{v}\, d\tau$, is zero if we assume the aforementioned calculation to be true.

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 $\boldsymbol{v}$ **blows up**. It is quite true that $\nabla \cdot \boldsymbol{v} = 0$ everywhere except the origin, but right at the origin is the situation is more complicated.

Observe, the surface integral is independent of R. If the divergence theorem is right, we should expect $\int \nabla \cdot \boldsymbol{v}\, d\tau = 4\pi$ for any non-zero vector and the origin.

This means the value of 4π must be coming from the point $r = 0$. Therefore, $\nabla \cdot \boldsymbol{v}$ has the unique property that it vanishes everywhere except at one point, and yet its integral is 4π.

No normal function behaves like that.

> To wrap our heads around this property think of **density**.
>
> 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 upon is called the **Dirac delta function**. It arises in numerous branches of theoretical physics and plays a central role in the theory of electrodynamics.

2.5.2 The 1D Dirac Delta Function

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

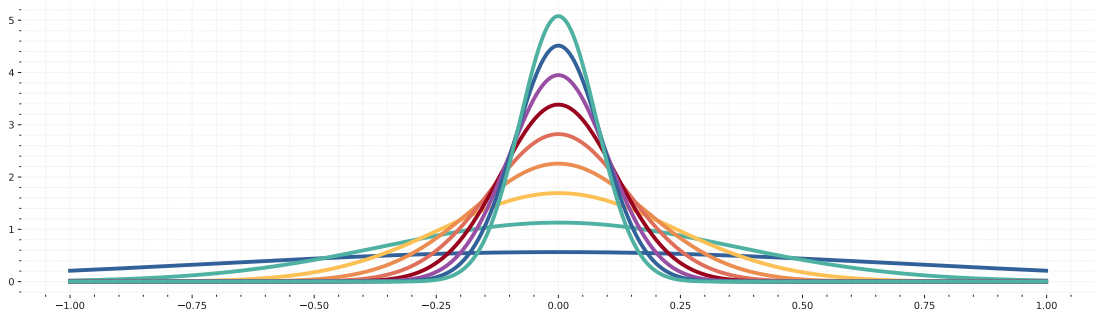


Figure 2.5: A visual representation of a 1D Dirac Delta Function. Think of it as a distribution function being squeezed to an infinitely small width.

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

and,

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

In a strict sense of definition, $\delta(x)$ is not a function at all, as its value is not finite at $x = 0$. In literature it is known as a **generalized function**.

If $f(x)$ is some "ordinary" function, then the product $f(x)\delta(x)$ is zero everywhere except at $x = 0$. It follows that:

$$f(x)\delta(x) = f(0)\delta(x).$$

> 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

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

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):

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

Equation 1.88 becomes

$$f(x)\delta(x-a) = f(a)\delta(x-a),$$

and Eq. 1.89 generalizes to

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

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]

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

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

Exercise 2.13: A Simple Dirac Integral

Evaluate the following integral:

$$\int_0^3 x^3\delta(x-2)\,\mathrm{d}x$$

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.

Exercise 2.14: 1D Dirac Delta

Evaluate the following integrals with Dirac delta functions:

$$\int_2^6 \left(3x^2 - 2x - 1\right)\,\delta(x-3)\,dx, \qquad \text{(vii)}$$

$$\int_0^5 \cos x\,\delta(x-\pi)\,dx, \qquad \text{(viii)}$$

$$\int_0^3 x^3\,\delta(x+1)\,dx, \qquad \text{(ix)}$$

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

Solution

The solution are as follows:

(a) $3\left(3^2\right) - 2(3) - 1 = 27 - 6 - 1 = 20$ ■

(b) $\cos\pi = -1$ ■

(c) 0 ■

(d) $\ln(-2+3) = \ln 1 = 0$ ■

2.5.3 The 3D Dirac Delta Function

Once we have defined the 1D Dirac, it is trivial to generalise it to 3D:

$$\delta^3(\boldsymbol{r}) = \delta(x)\,\delta(y)\,\delta(z),$$

and similar to 1D, 3D Dirac is zero everywhere except at (0, 0, 0), where it blows up. Its volume integral is 1:

$$\int_{\text{all space}} \delta^3(\boldsymbol{r})d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x)\delta(y)\delta(z)dxdydz = 1$$

And, the general form is:

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

As in the 1D case, integration with δ picks out the value of the function f at the location of the spike.

We can fix the paradox introduced in Section **??**. Remember, the divergence of $\hat{r}/r^2$ is zero everywhere except at the origin, however, its integral over any volume containing the origin is a constant.

These are precisely the defining conditions for the Dirac delta function; evidently

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

Or in a more general fashion:

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

Differentiation here is with respect to $\boldsymbol{r}$, while $\boldsymbol{r}'$ is held constant.

2.6 Vector Field Theory

2.6.1 Helmholtz Theorem

Electricity and magnetism are generally expressed as electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{B}$ and like many physical laws, these are most compactly expressed as **differential equations**.

As $\boldsymbol{E}$ and $\boldsymbol{B}$ are **vectors**, the differential equations naturally involve vector derivatives: *divergence* and *curl*. Maxwell reduced the entire theory to four (4) fundamental equations, specifying respectively the divergence and the curl of $\boldsymbol{E}$ and $\boldsymbol{B}$.

This formulation raises an interesting question:

To what extent is a vector function determined by its divergence and curl?

To study this case lets assume a vector of $\boldsymbol{F}$. if the divergence of $\boldsymbol{F}$ is a specified (scalar) function D,

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

and the curl of $\boldsymbol{F}$ is a specified (vector) function $\boldsymbol{C}$,

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

and for consistency, we assume $\boldsymbol{C}$ to have **NO** divergence,

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

> Remember, the divergence of a curl is **ALWAYS** zero.

Using this knowledge, is it possible to determine the function $\boldsymbol{F}$?

Without knowing more information, it is not really possible. There are many functions whose divergence and curl are both zero everywhere. Some examples are:

$$\boldsymbol{F} = 0,$$
$$\boldsymbol{F} = (y)\ \hat{x} + (zx)\ \hat{y} + (xy)\ \hat{z},$$
$$\boldsymbol{F} = (\sin x \cosh y)\ \hat{x} + (-\cos x \sinh y)\ \hat{y} + (.)\ \hat{z}$$

If you recall **Higher Mathematics I**, to solve a differential equation with a particular solution, you must also be supplied with appropriate **boundary conditions**.

In electrodynamics we typically require that the fields go to zero at infinity. With that extra information, the **Helmholtz theorem** guarantees the field is uniquely determined by its divergence and curl.

2.6.2 Potentials

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

$$\nabla \times \boldsymbol{F} = 0 \iff \boldsymbol{F} = -\nabla V$$

> The minus sign is purely conventional.

That's the essential burden of the following theorem:

Theorem I - Zero Curl Fields

The following conditions are equivalent.

(i) $\nabla \times \boldsymbol{F} = 0$ everywhere,

(ii) $\int_a^b \boldsymbol{F} \cdot d\boldsymbol{l}$ is independent of path, for any given end points,

(iii) $\oint \boldsymbol{F} \cdot d\boldsymbol{l} = 0$ for any closed loop,

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

The potential is **NOT** unique as any constant can be added to V, since this will not affect its gradient.

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

$$\nabla \cdot \boldsymbol{F} = 0 \iff \boldsymbol{F} = \nabla \times \boldsymbol{A}$$

That's the main conclusion of the following theorem:

Theorem II - Zero Divergence Fields

The following conditions are equivalent:

(i) . $\nabla \cdot \boldsymbol{F} = 0$ everywhere.

(ii) $f\boldsymbol{F} \cdot d\boldsymbol{a}$ is independent of surface, for any given boundary line.

(iii) $f\boldsymbol{F} \cdot d\boldsymbol{a} = 0$ for any closed surface.

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

The vector potential is **NOT** unique as the gradient of any scalar function can be added to $\boldsymbol{A}$ without affecting the curl, given the curl of a gradient is zero.

Incidentally, in all cases, a vector field $\boldsymbol{F}$ can be written as the gradient of a scalar plus the curl of a vector.

$$\boldsymbol{F} = -\nabla V + \nabla \times \boldsymbol{A}$$

Chapter 3

Electrostatics

Table of Contents

3.1 The Electric Field

The fundamental problem electrodynamics hopes to solve is this:

> We have some electric charges, q_1, q_2, q_3, ... (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 (usually as functions of time) and the trajectory of the test particle is to be calculated. General, both the source charges and the test charge are in motion.

The solution to this problem is solved with the **principle of superposition**, which states the interaction between any two (2) charges is completely unaffected by the presence of others. This means to determine the force on Q, we can first compute the force $\boldsymbol{F}_1$, due to q_1 alone then we compute the force $\boldsymbol{F}_2$, due to q_2 alone and so on.

Finally, we take the vector sum of all these individual forces:

$$\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_2 + \boldsymbol{F}_3 + \dots$$

Therefore, if we can find the force on Q due to a single source charge q, we are, in principle, done. [1]

Well, it seems easy, right? Why don't we just write down the formula for the force on Q directed towards q, and calculate it? We could, but some technicalities limit us, for not only does the force on Q depend on the separation distance $\imath$ between the charges, it also 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:

[1] Th
sup
"ob
did
sim
elec
wer
squ
cha
prin
wou
(q_1
(th
ter
Sup
logi
exp

> electromagnetic particles/waves 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 is easy to state[2], therefore we shall not answer this with a direct answer now but go at it by stages.

[2] Most complicated problem are funny enough relatively easy to state

To simplify, we assume time to be **irrelevant**, making the special case of **electrostatics** in which all the *source charges are stationary* whereas the test charge may be in motion.

3.1.1 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 $\boldsymbol{\imath}$ away? The answer is given by **Coulomb's law**[3].

[3] This is an experimental law

$$\boldsymbol{F} = \frac{1}{4\pi\varepsilon_0}\frac{qQ}{\imath^2}\hat{\boldsymbol{\imath}} \tag{3.1}$$

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

$$\varepsilon_0 8.85 \times 10^{-12} \quad \mathrm{C}^2 \cdot \mathrm{N}^{-1} \cdot \mathrm{m}^2$$

In words,

> the force is proportional to the product of the charges and inversely proportional to the square of the separation distance.

Remember, $\boldsymbol{\imath}$ is the separation vector from $\boldsymbol{r}'$ (the location of q) to $\boldsymbol{r}$ (the location of Q):

$$\boldsymbol{\imath} = \boldsymbol{r} - \boldsymbol{r}'$$

$\imath$ is its magnitude, and $\hat{\boldsymbol{\imath}}$ is its direction. The force points **along the line** from q to Q;

- it is repulsive if q and Q have the same sign,
- it is attractive if their signs are opposite

Coulomb's law and the principle of superposition form the physical foundation for electrostatics with the rest, except for some special properties of matter, is mathematical derivation of these fundamental rules.

3.1.2 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

$$\begin{aligned}\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_2 + \ldots &= \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 Q}{\imath_1^2}\hat{\boldsymbol{\imath}}_1 + \frac{q_2 Q}{\imath_1^2}\hat{\boldsymbol{\imath}}_2 + \ldots\right) \\ &= Q\frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{\imath_1^2}\hat{\boldsymbol{\imath}}_1 + \frac{q_2}{\imath_2^2}\hat{\boldsymbol{\imath}}_2 + \ldots\right)\end{aligned}$$

or in a more cleaner way:

$$F = Q\boldsymbol{E}$$

where

$$\boldsymbol{E}(\boldsymbol{r}) \equiv \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{\imath_i^2}\hat{\boldsymbol{\imath}}_i \tag{3.2}$$

$\boldsymbol{E}$ is called the **electric field** of the source charges. It is important to mention, it is a function of the position ($\boldsymbol{r}$), as the separation vectors $\hat{\boldsymbol{\imath}}_i$ depend on the location of the **field point** P.

> 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, $\boldsymbol{E}(\boldsymbol{r})$ is the force per unit charge that would be exerted on a test charge, if you were to place one at P.

Exercise 3.1: Electric Field at a Distance

Find the electric field a distance z above the midpoint between two equal charges (q), a distance d apart:

Solution

Let $\boldsymbol{E}_1$ be the field of the left charge alone, and $\boldsymbol{E}_2$ that of the right charge alone. Adding them (vectorially), the horizontal components cancel and the vertical components add up to:

$$E_z = 2\frac{1}{4\pi\varepsilon_0}\frac{q}{\imath^2}\cos\theta.$$

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

$$\boldsymbol{E} = \frac{1}{4\pi\varepsilon_0}\frac{2qz}{\left[z^2 + \left(d/2\right)^2\right]^{3/2}}\hat{z} \quad \blacksquare$$

3.1.3 Continuous Charge Distributions

Our definition of the electric field Eq. (**??**) assumes 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:

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int\frac{1}{\imath^2}\hat{\boldsymbol{\imath}}\,\mathrm{d}q$$

1. If charge is spread out along a *line*, with charge-per-unit-length λ, then $\mathrm{d}q = \lambda\,\mathrm{d}l'$ (where $\mathrm{d}l'$ is an element of length along the line)
2. If charge is smeared out over a *surface*, with charge-per-unit-area σ, then $dq = \sigma da'$ (where da' is an element of area on the surface);
3. If charge fills a *volume* with charge-per-unit-volume ρ, then $dq = \rho d\tau'$ (where $d\tau'$ is an element of volume):

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

Therefore the electric field of a line charge is

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int\frac{\lambda\left(\boldsymbol{r}'\right)}{\imath^2}\hat{\boldsymbol{\imath}}\,\mathrm{d}l'$$

for a surface charge,

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\sigma\left(\boldsymbol{r}'\right)}{\imath^2}\hat{\boldsymbol{\imath}}\,\mathrm{d}a'$$

and for a volume charge,

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho\left(\boldsymbol{r}'\right)}{\imath^2}\hat{\boldsymbol{\imath}}\,\mathrm{d}\tau' \tag{3.3}$$

Eq. (**??**) is often referred to as Coulomb's law,as it is such a short step from the original shown in Eq. (**??**), and because a volume charge is in a sense the most general and realistic case.

3.2 Divergence and Curl of Electrostatic Fields

As we have seen by now, the integrals involved in computing $\boldsymbol{E}$ can be complicate and laborious, even for reasonably simple charge distributions. Therefore, rest of electrostatics from this point onwards is devoted to assembling methods for **avoiding these integrals**. It all begins with the divergence and curl of $\boldsymbol{E}$. We shall calculate the divergence of $\boldsymbol{E}$ directly from Eq. (**??**), but first we should develop an intuitive view of $\boldsymbol{E}$.

Let's begin with the simplest possible case:

> a single point charge q, situated at the origin:

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}\hat{\boldsymbol{r}}$$

As the equation tells, the field falls off like $1/r^2$, the vectors get shorter as you go farther away from the origin and they always point radially outward. But there is a nicer way to represent this field, and that's to connect up the arrows, to form **field lines**. 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 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.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.

[4] Strictly speaking the divergence of E would not be zero, and (as we shall soon see) the curl cannot happen in empty space.

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

$$\Phi_{\mathrm{E}} = \int_{\mathcal{S}} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a}$$

is a measure of the number of field lines passing through $\mathcal{S}$. We put this in quotes as 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 $\boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a}$. is proportional to the number of lines passing through the infinitesimal area $\mathrm{d}\boldsymbol{a}$.

This means the flux through any *closed* surface is a measure of the total mass inside the surface. For the field lines that originate on a positive charge must either pass through the surface or else terminate on a negative charge inside. 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. This is the *exercise* of Gauss' law.

Now let's make it quantitative.

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

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \int \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r^2}\,\hat{\boldsymbol{r}}\right) \cdot \left(r^2 \sin\theta \,\mathrm{d}\theta d\phi\,\hat{\boldsymbol{r}}\right) = \frac{1}{\varepsilon_0} q$$

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 is 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:

$$\boldsymbol{E} = \sum_{i=1}^{n} \boldsymbol{E}_{\mathrm{i}}$$

The flux through a surface that encloses them all is:

$$\oint \boldsymbol{E} \cdot \mathrm{d}a = \sum_{i=1}^{n} \left(\oint \boldsymbol{E}_{\mathrm{i}} \cdot \mathrm{d}\boldsymbol{a}\right) = \sum_{i=1}^{n} \left(\frac{1}{\varepsilon_0} q_{\mathrm{i}}\right)$$

For any closed surface, then,

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\mathrm{enc}} \tag{3.4}$$

where Q_{enc} is the total charge enclosed within the surface. This is the quantitative statement of Gauss's law. 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:

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

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

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

So Gauss's law becomes

$$\int_{\mathcal{V}} (\nabla \cdot \boldsymbol{E}) \, \mathrm{d}\tau = \int_{\mathcal{V}} \left(\frac{\rho}{\varepsilon_0} \right) \, \mathrm{d}\tau$$

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

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\varepsilon_0} \rho \tag{3.5}$$

Eq. (**??**) carries the same message as Eq. (**??**) as it is **Gauss's law in differential form**. The differential version is idler, but the integral form has the advantage in that it accommodates point, line, and surface charges more naturally.

3.2.1 Divergence of E

Let's go back, now, and calculate the divergence of $\boldsymbol{E}$ directly from Eq. 2.8:

$$\boldsymbol{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \int_{\text{all space}} \frac{\hat{\mathbf{\imath}}}{\imath^2} \rho(\mathbf{r}') d\tau'$$

> While the interation is originally defined over the volume occupied by the charge, it is much easier to engulf all the space as $\rho = 0$ outside the volume regardless.

Noting that the $\boldsymbol{r}$-dependence is contained in the following:

$$\boldsymbol{\imath} = \boldsymbol{r} - \boldsymbol{r}',$$

we have

$$\nabla \cdot \boldsymbol{E} = \frac{1}{4\pi\varepsilon_0} \int \nabla \cdot \left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) \rho \left(\boldsymbol{r}' \right) \, \mathrm{d}\tau'$$

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

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

Using this relation we can arrive at the following:

$$\nabla \cdot \boldsymbol{E} = \frac{1}{4\pi\varepsilon_0} \int 4\pi\delta^3 \left(\boldsymbol{r} - \boldsymbol{r}' \right) \rho \left(\boldsymbol{r}' \right) \, \mathrm{d}\tau' = \frac{1}{\varepsilon_0} \rho (\boldsymbol{r})$$

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:

$$\int_{\mathcal{V}} \nabla \cdot \boldsymbol{E} \, \mathrm{d}\tau = \oint_{\mathcal{S}} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \frac{1}{\varepsilon_0} \int_{\mathcal{V}} \rho \, \mathrm{d}\tau = \frac{1}{\varepsilon_0} Q_{\text{enc}}.$$

3.2.2 Applications of Gauss's Law

Let us look at some of the use cases of Gauss's law, in integral form. When the problem allows symmetry, it allows us the quickest and easiest way of computing electric fields. Let's understand this statement with a series of examples.

Exercise 3.2: A Gaussian Sphere

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

Solution

Imagine a spherical surface at radius $r > R$. this is called a Gaussian surface. Gausss law says that:

$$\oint_{\mathcal{S}} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\text{enc}},$$

For this case it is $Q_{\text{enc}} = q$. At first glance this doesnt seem to get us very far, as 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 da, so we can drop the dot product,

$$\int_{\mathcal{S}} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \int_{\mathcal{S}} |\boldsymbol{E}| \, \mathrm{d}a$$

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

$$\int_{\mathcal{S}} |\boldsymbol{E}| \, \mathrm{d}a = |\boldsymbol{E}| \int_{\mathcal{S}} \mathrm{d}a = |E| \, 4\pi r^2$$

Therefore:

$$|\boldsymbol{E}| \, 4\pi r^2 = \frac{1}{\varepsilon_0} q \qquad \text{or} \qquad \boldsymbol{E} = \frac{1}{4\pi\varepsilon_0} \frac{q}{r^2} \hat{r} \quad \blacksquare$$

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

> Symmetry is crucial to this application of Gauss's law.

There are three (3) kinds of symmetry that work:

Spherical it is a concentric sphere,

Cylindrical it is a coaxial cylinder,

Plane A pillbox which straddles the surface.

> While Cylindrical and Plane 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.

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.

3.2.3 Curl of E

Nor it is time to calculate the curl of $\boldsymbol{E}$, as we did the divergence in Sect. 2.2.1, by studying first the simplest possible configuration:

a point charge at the origin.

In this case

$$\boldsymbol{E} = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}\,\hat{\boldsymbol{r}}.$$

To get a feel of the curve of $\boldsymbol{E}$, let us calculate the line integral of this field from some point $\boldsymbol{a}$ to some other point $\boldsymbol{a}$:

$$\int_a^b \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l}$$

As we are working in spherical coordinates, $(\mathrm{d}r)\ \hat{\boldsymbol{r}} + (r\,\mathrm{d}\theta)\ \hat{\boldsymbol{\theta}} + (r\sin\theta d\phi)\ \hat{\boldsymbol{\phi}}$, so

$$\boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l} = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}\,\mathrm{d}r$$

Therefore

$$\int_a^b = \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l} = \frac{1}{4\pi\varepsilon_0}\int_a^b \frac{q}{r^2}\,\mathrm{d}r = -\frac{1}{4\pi\varepsilon_0}\frac{q}{r}\bigg|_{r_1}^{r_2} = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r_\mathrm{a}} - \frac{q}{r_\mathrm{b}}\right)$$

where r_a is the distance from the origin to the point a and rb is the distance to b. The integral around a closed path is evidently zero (for then ra = rb):

$$\oint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{l} = 0 \tag{3.6}$$

And through the use of Stokes' Theorem one could arrive its differential form:

$$\nabla\times\boldsymbol{E} = 0 \tag{3.7}$$

Now, we proved Eq. (**??**) and Eq. (**??**) only for the field of a single point charge at the origin, but these results make no reference to what is, 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:

$$\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2 + \ldots$$

which means

$$\nabla\times\boldsymbol{E} = \nabla\times\left(\boldsymbol{E}_1 + \boldsymbol{E}_2 + \ldots\right)$$

Therefore, Eq. (**??**) and Eq. (**??**) hold for any static charge distribution whatever.

3.3 Electric Potential

3.3.1 Defining the Potential

The electric field $\boldsymbol{E}$ is a special kind of vector function, one whose curl is zero.

For example, $\boldsymbol{E} = y\,\hat{x}$, could **NOT** possibly be an electrostatic field as **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 $\boldsymbol{E}$) to a **scalar** problem. We know from the **Vector Calculus** chapter that any vector whose curl is zero is equal to the gradient of some scalar.

What we are going to do now amounts to a proof in the context of electrostatics.

As $\nabla \times \boldsymbol{E} = 0$, the line integral of $\boldsymbol{E}$ around any closed loop is zero (remember the *Stokes' theorem*). As $\oint \boldsymbol{E} \cdot \mathrm{d}l = 0$, the line integral of $\boldsymbol{E}$ from point $\boldsymbol{a}$ to point $\boldsymbol{b}$ is the same for all paths.

Because the line integral is independent of path, we can define a function

$$V(\mathbf{r}) \equiv -\int_{\mathcal{O}}^{r} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \tag{3.8}$$

Here $\mathcal{O}$ is some standard reference point on which we have agreed beforehand. V then depends only on the point $\boldsymbol{r}$. It is called the **electric potential**.

The potential difference between two points $\boldsymbol{a}$ and $\boldsymbol{b}$ is

$$\begin{aligned} V(\mathbf{b}) - V(\mathbf{a}) &= -\int_{\mathcal{O}}^{\mathrm{b}} \boldsymbol{E} \cdot d\mathbf{l} + \int_{\mathcal{O}}^{\mathrm{a}} \boldsymbol{E} \cdot d\mathbf{l} \\ &= -\int_{\mathcal{O}}^{\mathrm{b}} \boldsymbol{E} \cdot d\mathbf{l} - \int_{\mathrm{a}}^{\mathcal{O}} \boldsymbol{E} \cdot d\mathbf{l} = -\int_{\mathrm{a}}^{\mathrm{b}} \boldsymbol{E} \cdot d\mathbf{l}. \end{aligned}$$

Now, the fundamental theorem for gradients states that

$$V(\mathbf{b}) - V(\mathbf{a}) = \int_{\mathrm{a}}^{\mathrm{b}} (\nabla V) \cdot d\mathbf{l},$$

so

$$\int_{\mathrm{a}}^{\mathrm{b}} (\nabla V) \cdot d\mathbf{l} = -\int_{\mathrm{a}}^{\mathrm{b}} \boldsymbol{E} \cdot d\mathbf{l}.$$

Since, finally, this is true for any points **a** and **b**, the integrands must be equal:

$$\boldsymbol{E} = -\nabla V \tag{3.9}$$

Eq. (**??**) is the differential version of Eq. (**??**) which 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 $\nabla \times \boldsymbol{E} = 0$) in this argument. If the line integral of $\boldsymbol{E}$ depended on the path taken, then the "definition" of V, Eq. (**??**), would not be useful. It simply would not define a function, since changing the path would alter the value of $V(\boldsymbol{r})$.

> The minus sign in $\boldsymbol{E} = -\nabla V$ is there due to historical conventions.

3.3.2 Some Digressions on Potential

Naming

The naming of V as potential is unfortunate as it inevitably reminds us of potential *energy*. This is particularly problematic, as there is a connection between **potential** and **potential energy**. Incidentally, a surface over which the potential is constant is called an **equipotential**.

Advantages of the Potential

If we know V, you can easily get $\boldsymbol{E}$. Just take the gradient: $\boldsymbol{E} = -\nabla \boldsymbol{V}$. This is due to the three components of $\boldsymbol{E}$ not being independent. In fact, they are explicitly interrelated by the very condition we started with, $\nabla \times \boldsymbol{E} = 0$. In terms of components,

$$\frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x}, \qquad \frac{\partial E_z}{\partial y} = \frac{\partial E_y}{\partial z}, \qquad \frac{\partial E_x}{\partial z} = \frac{\partial E_z}{\partial x}.$$

The Reference Point

There is a necessary ambiguity in the definition of potential, as the choice of reference point $\mathcal{O}$ was arbitrary. Changing reference points amounts to adding a constant K to the potential:

$$V'(\boldsymbol{r}) = -\int_{\mathcal{O}}^{r} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\int_{\mathcal{O}'}^{\mathcal{O}} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} - \int_{\mathcal{O}}^{r} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = K + V(\boldsymbol{r})$$

where K is the line integral of $\boldsymbol{E}$ 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 $\boldsymbol{E}$ from $\mathbf{a}$ to $\mathbf{b}$, with no reference to $\mathcal{O}$.) Nor does the ambiguity affect the gradient of V:

$$\nabla V' = \nabla V,$$

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 $\boldsymbol{E}$.

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 $\mathcal{O}$., however, therefore V's a "reference level to use for $\mathcal{O}$ in electrostatics –almost-one-like equations 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 warm 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 ($\sigma/2\varepsilon_0$

$$V(z) = -\int_{\infty}^{z} \frac{1}{2\varepsilon_0}\sigma \, dz = -\frac{1}{2\varepsilon_0}\sigma(z - \infty).$$

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:

$$\boldsymbol{F} = \boldsymbol{F}_1 + \boldsymbol{F}_1 + \boldsymbol{F}_2 \ldots$$

Dividing through by Q, we see that the electric field, too, obeys the superposition principle:

$$\boldsymbol{E} = \boldsymbol{E}_1 + \boldsymbol{E}_2 + \ldots$$

Integrating from the common reference point to $\boldsymbol{r}$, it follows that the potential also satisfies such a principle:

$$V = V_1 + V_2 + \ldots$$

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 known-meters per coulomb, or joules per coulomb. A joule per coulomb is a **volt**.

3.3.3 Poisson's Equation and Laplace's Equation

We now know the electric field can be written as the gradient of a scalar potential:

$$\boldsymbol{E} = -\nabla V$$

The question arises:

How are the divergence and curl of $\boldsymbol{E}$ *related to our new definition of electric potential ?*

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon_0} \qquad \text{and} \qquad \nabla \times \boldsymbol{E} = 0$$

Well:

$$\nabla \cdot \boldsymbol{E} = \nabla \cdot -\nabla V = -\nabla^2 V$$

so, apart from that persistent minus sign, the divergence of $\boldsymbol{E}$ is the **Laplacian** of V. Gauss's law, then, says:

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0} \tag{3.10}$$

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**:

$$\nabla^2 V = 0$$

We'll explore this equation more fully in Chapter 3 when we work with on **Advanced Mathematical Methods**.

What about the curl law? This says that

$$\nabla \times \boldsymbol{E} = \nabla \times (-\nabla V) = 0$$

But that's no condition on V as curl of gradient is always zero.
It takes only one (1) differential equation (Poisson's) to determine V, because V is a scalar; for $\boldsymbol{E}$ we needed two (2), the divergence and the curl.

Chapter 4

Advanced Mathematical Methods

Table of Contents

4.1 Introduction

The primary task of electrostatics is to find the electric field of a given stationary charge distribution. In terms of mathematical expression, this is accomplished by Coulomb's law, in the following form:

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\hat{\boldsymbol{\imath}}}{\imath^2}\rho\left(\hat{\boldsymbol{\imath}}'\right)\,\mathrm{d}\tau'.$$

Unfortunately, these types integrals can be difficult to calculate. 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 slightly less complicated expression:

$$V\left(\boldsymbol{r}\right) = \frac{1}{4\pi\varepsilon_0}\int_\tau \frac{1}{\imath}\rho\left(\hat{\boldsymbol{\imath}}'\right)\,\mathrm{d}\tau'. \tag{4.1}$$

Even this integral is often too tough to handle analytically. Moreover, in problems involving conductors ρ itself may not be known in advance as charge is free to move around, the only thing we control directly is the **total charge**[1] of each conductor. For these cases, it is fruitful to recast the problem in differential form, using Poisson's equation:

$$\nabla^2 V = -\frac{1}{\varepsilon_0}\rho,$$

which, together with appropriate boundary conditions, is equivalent to Eq. (**??**). Often we are interested in finding the potential in a region where $\rho = 0$ [2] . In this case, Poisson's equation reduces to **Laplace's equation**:

$$\nabla^2 V = 0,$$

written out in Cartesian coordinates,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0.$$

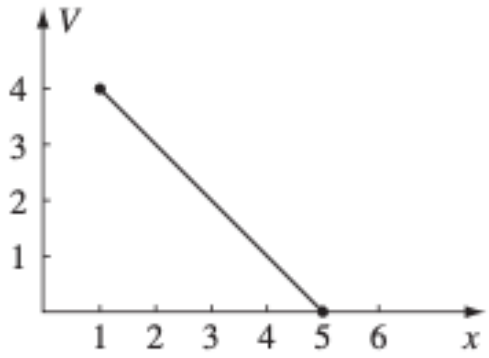


Figure 4.1: A particular solution to a 1D laplace

> The importance of this equation can't be downplayed as one might consider the field of electrostatics as the study of Laplace's equation. This equation is quite common in physics as it appears also in gravitation and magnetism, the theory of heat, and the study of bubbles.

To get an understanding of Laplace's equation and its solutions[3], let's begin with the 1D and 2D versions, which are easier to picture, and illustrate all the essential properties of the 3D case.

[3] known as harmonic functions

4.1.1 Laplace's Equation in One Dimension

Let us assume V depends on only one variable, x. Then Laplace's equation becomes:

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

The general solution is

$$V(x) = mx + b, \tag{4.2}$$

as can be seen, the equation geometrically represents a straight line. It contains two (2) undetermined constants (m, b), as is appropriate for a second-order Ordinary Differential Equation (ODE). 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).
Let's try to formulate the two (2) features of this result, while obvious in 1D, where we can write down the general solution explicitly, but the analogs in 2D and 3D dimensions are powerful and by no means obvious:

1. $V(x)$ is the average of $V(x + a)$ and $V(x - a)$, for any a:

$$V(x) = \frac{1}{2}\left[V(x + a) + V(x - a)\right]$$

Laplace's equation can be taught of as an averaging instruction as it tells to assign to the point x the average of the values to the left and to the right of x.

2. Laplace's equation tolerates no local maxima or minima [4] . Extreme values of V **must occur at the end points**. Actually, this is a consequence of the previous statement, 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.

[4] This is because the second derivative must be greater than or less than zero for a maxima or minima but it is equal to zero as per laplace equation

4.1.2 Laplaces Equation in Two Dimensions

When V depends on two (2) variables, Laplace's equation becomes:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0. \tag{4.3}$$

We can no longer say we are working with an ODE [5] . It is a Partial Differential Equation (PDE). As a consequence, some of the simple rules we may be familiar with do not apply. For instance, the general solution to this equation doesn't contain just two (2) arbitrary constants[6] . Indeed, one cannot write down a "general solution" [7] . Nevertheless, it is possible to deduce certain properties common to all solutions.

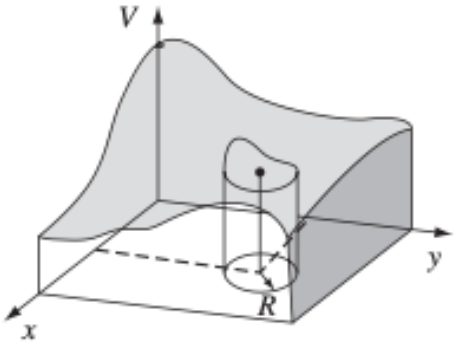


Figure 4.2

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 way 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 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 2D 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. 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 centre:

$$V(x, y) = \frac{1}{2\pi R} \oint_{\text{circle}} V \, dl.$$

This suggests the method of relaxation[8], 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.)

[8] Which in mathematics, relaxation methods are iterative methods for solving systems of equations, including nonlinear systems.

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 roll over to one side and fall offit 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.

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.

1. The value of V at point $\mathbf{r}$ is the average value of V over a spherical surface of radius R centered at $\mathbf{r}$:
$$V(\boldsymbol{r}) = \frac{1}{4\pi R^2} \oint_{\text{sphere}} V \, da.$$
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 $\boldsymbol{r}$, then by the very nature of maximum I could draw a sphere around $\boldsymbol{r}$ over which all values of V–and a fortiori the average–would be less than at $\boldsymbol{r}$.)

4.1.3 Boundary Conditions and Uniqueness Theorems

Laplace's equation does not by itself determine V:

suitable boundary conditions must be supplied.

This raises an important question. What are considered appropriate conditions to determine the answer? For the case of 1D, it is easy, for here the general solution:

$$V = mx + b$$

contains two (2) arbitrary constants, and we require two (2) boundary conditions. We could:

- specify the value of the function at each end,
- give the value of the function and its derivative at one end,
- give 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 or specifying the derivatives at both ends as 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.**

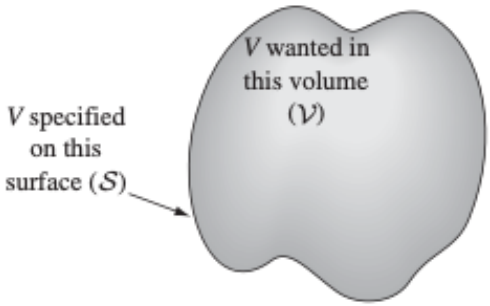
> Is the shape of a number membrane uniquely determined by the frame over which it is stretched, or, like a scanning jar lid, can it snap from one stable configuration to another?

The answer, is that V is **uniquely determined by its value at the boundary**. However, other boundary conditions can also be used. 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. Let's look at the two (2) most useful ones.

Theorem: *First Uniqueness*

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 you can see such a region and its boundary[9] . Assume there were *two* solutions to Laplace's equation:

$$\nabla^2 V_1 = 0, \qquad \text{and} \qquad \nabla^2 V_2 = 0,$$

both expression assume the specified value on the surface. Let's prove they must be equal. The main idea 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 as 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

$$V_1 = V_2 \quad \blacksquare$$

The uniqueness theorem is a license to your imagination. It doesnt matter how you come by your solution; if (a) it satisfies Laplaces 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

$$\nabla^2 V_1 = -\frac{1}{\varepsilon_0}\rho, \qquad \nabla^2 V_2 = -\frac{1}{\varepsilon_0}\rho,$$

which makes:

$$\nabla^2 V_3 = \nabla^2 V_1 - \nabla^2 V_2 = -\frac{1}{\varepsilon_0}\rho + \frac{1}{\varepsilon_0}\rho = 0.$$

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 $V_1 = V_2$.

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.

4.1.4 Conductors and the Second Uniqueness Theorem

The easiest 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 experimental work 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 Qa on the first conductor, Qb on the second, and so onIm 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, lets say there is some specified charge density 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?

Theorem: *Second Uniqueness*

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 [10] (Fig. 3.6).

[10] The region as a whole can be bounded by another conductor, or else unbounded.

Proof

Suppose there are two (2) fields satisfying the conditions of the problem. They both obey Gauss's law in differential form in the space between the conductors:

$$\nabla \cdot \boldsymbol{E}_1 = \frac{1}{\varepsilon_0}\rho, \qquad \nabla \cdot \boldsymbol{E}_2 = \frac{1}{\varepsilon_0}\rho.$$

And they both obey Gauss's law in integral form for a Gaussian surface enclosing each conductor:

$$\oint_{\substack{i\text{th conducting}\\ \text{surface}}} \boldsymbol{E}_1 \cdot \mathrm{d}\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\mathrm{i}}, \qquad \oint_{\substack{i\text{th conducting}\\ \text{surface}}} \boldsymbol{E}_2 \cdot \mathrm{d}\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\mathrm{i}}.$$

Similarly, for the outer boundary (whether this is just inside an enclosing conductor or at infinity),

$$\oint_{\substack{\text{outer}\\ \text{boundary}}} \boldsymbol{E}_1 \cdot d\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\mathrm{tot}}, \qquad \oint_{\substack{\text{outer}\\ \text{boundary}}} \boldsymbol{E}_2 \cdot d\boldsymbol{a} = \frac{1}{\varepsilon_0} Q_{\mathrm{tot}}.$$

As before, we examine the difference

$$\boldsymbol{E}_3 \equiv \boldsymbol{E}_1 - \boldsymbol{E}_2,$$

which obeys

$$\nabla \cdot \boldsymbol{E}_3 = 0$$

in the region between the conductors, and

$$\oint \boldsymbol{E}_3 \cdot d\boldsymbol{a} = 0$$

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

$$\nabla \cdot (V_3 \boldsymbol{E}_3) = V_3(\nabla \cdot \boldsymbol{E}_3) + \boldsymbol{E}_3 \cdot (\nabla V_3) = -(E_3)^2.$$

Here I have used Eq. 3.7, and $\boldsymbol{E}_3 = -v\nabla V_3$. Integrating this over $\mathcal{V}$, and applying the divergence theorem to the left side:

$$\int_{\mathcal{V}} v\nabla \cdot (V_3 \boldsymbol{E}_3) d\tau = \oint_{\mathcal{S}} V_3 \boldsymbol{E}_3 \cdot d\boldsymbol{a} = -\int_{\mathcal{V}} (E_3)^2 d\tau.$$

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,

$$\int_V (E_3)^2 d\tau = 0.$$

But this integrand is never negative; the only way the integral can vanish is if $E_3 = 0$ everywhere. Consequently, $\boldsymbol{E}_1 = \boldsymbol{E}_2$, 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 Q, situated so that the plauses are near the minuses.

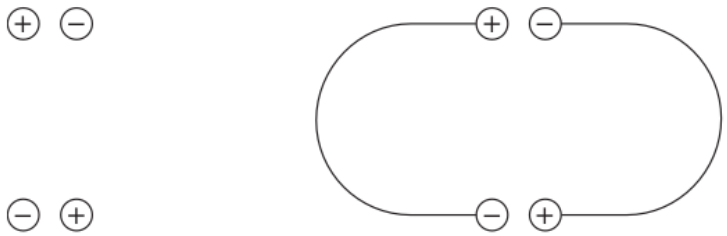


Figure 4.3

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 below will, 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 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.

4.2 Method of Images

4.2.1 A Classic 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 $(1/4\pi\varepsilon_0)q/a$, 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 different? 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 (that is, for $x^2 + y^2 + z^2 \gg d^2$).

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
$(0, 0, d)$ and $-q$ at $(0, 0, -d)$, and *no* conducting plane (Fig. 3.11). For this configuration, I can easily write down the potential:

$$V(x, y, z) = \frac{1}{4\pi\varepsilon_0}\left[\frac{q}{\sqrt{x^2 + y^2 + (z - d)^2}} - \frac{q}{\sqrt{x^2 + y^2 + (z + d)^2}}\right].$$

(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.

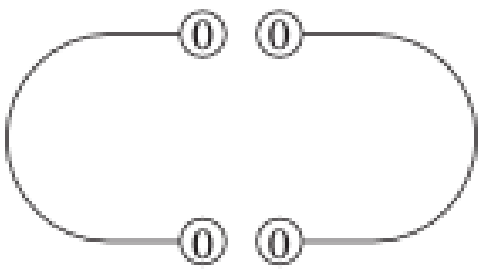


Figure 4.4

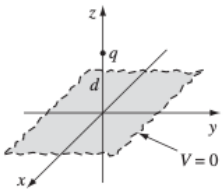


FIGURE 3.10

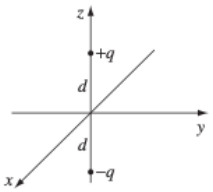


FIGURE 3.11

Figure 4.5

4.2.2 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,

$$\sigma = -\varepsilon_0 \frac{\partial V}{\partial n},$$

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

$$\sigma = -\varepsilon_0 \left.\frac{\partial V}{\partial z}\right|_{z=0}.$$

From Eq. 3.9,

$$\frac{\partial V}{\partial z} = \frac{1}{4\pi\varepsilon_0}\left[\frac{-q\,(z-d)}{[x^2+y^2+(z-d)^2]^{3/2}} + \frac{q\,(z+d)}{[x^2+y^2+(z+d)^2]^{3/2}}\right],$$

so:

$$\sigma\,(x,\,y) = \frac{-qd}{2\pi\left(x^2+y^2+d^2\right)^{3/2}}.$$

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 $da = dxdy$, but it's a little easier to use polar coordinates (r, ϕ), with $r^2 = x^2 + y^2$ and $da = rdrd\phi$. Then

$$\sigma(r) = \frac{-qd}{2\pi(r^2+d^2)^{3/2}},$$

and

The total charge induced on the plane is $-q$, as (with benefit of hindsight) you can perhaps convince yourself it *had* to be.

4.2.3 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:

$$\boldsymbol{F} = -\frac{1}{4\pi\varepsilon_0}\frac{q^2}{(2d)^2}\,\hat{z}.$$

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

$$W = -\frac{1}{4\pi\varepsilon_0}\frac{q^2}{2d}.$$

But for a single charge and conducting plane, the energy is *half* of this:

$$W = -\frac{1}{4\pi\varepsilon_0}\frac{q^2}{4d}.$$

Why half? Think of the energy stored in the fields (Eq. 2.45):

$$W = \frac{\varepsilon_0}{2}\int E^2 d\tau.$$

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 $(1/4\pi\varepsilon_0)(q^2/4z^2)\,\hat{z}$, so 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.

4.2.4 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.

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.

4.3 Separation of Variables

In separation of variables[11] we will attack Laplace's equation **directly**. The method is applicable in situations where the potential (V) or the charge density (σ) is specified on the boundaries of

some region, and we are tasked with finding 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, therefore let's try to understand this through exercises. We'll start with Cartesian coordinates and then do spherical coordinates.

Exercise 4.2: Spherical Separation of Variables

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

$$V(r,\theta) = \sum_{l=0}^{\infty} \frac{B_l}{r^{l+1}} P_l(\cos\theta).$$

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

$$\frac{B_{l'}}{R^{l'+1}} \frac{2}{2l'+1} = \int_0 \pi V_0(\theta)$$

$'](\cos\theta)\sin\theta d\theta$,
or

$$B_l = \frac{2l+1}{2} R^{l+1} \int_0 \pi V_0(\theta) P_l(\cos\theta)\sin\theta d\theta.$$

Equation 3.72, with the coefficients given by Eq. 3.73, is the solution to our problem. ■

Exercise 4.3: Charge Density

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:

$$V = \frac{1}{4\pi\varepsilon_0}\int \frac{\sigma_0}{\lambda}\,da,$$

but separation of variables is often easier. For the interior region, we have

$$V(r,\theta) = \sum_{l=0}^{\infty} A_l r^l P_l(\cos\theta) \qquad (r \le R)$$

(no B_l terms–they blow up at the origin); in the exterior region

$$V(r,\theta) = \sum_{l=0}^{\infty} \frac{B_l}{r^{l+1}} P_l(\cos\theta) \qquad (r \ge R)$$

(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):

$$\sum_{l=0}^{\infty}$$

$R^l \;\; P_l(\cos\theta) = \sum_{l=0[\infty]} \frac{B_l}{R^{l+1}} P_l(\cos\theta).$

It follows that the coefficients of like Legendre polynomials are equal:

$$B_l = A_l R^{2l+1}.$$

(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):

$$\left(\frac{\partial V_{\text{out}}}{\partial r} - \frac{\partial V_{\text{in}}}{\partial r}\right)\Big|_{r=R} = -\frac{1}{\varepsilon_0}\sigma_0(\theta).$$

Thus

$$-\sum_{l=0}^{\infty}(l+1)$$

$R^{l+2} \;\; P_l(\cos\theta) - \sum_{l=0[\infty]} l A_l R^{l-1} P_l(\cos\theta) = -\frac{1}{\varepsilon_0}\sigma_0(\theta),$

or, using Eq. 3.81,

$$\sum_{l=0}^{\infty}(2l+1)A_l R^{l-1} P_l(\cos\theta) = \frac{1}{\varepsilon_0}\sigma_0(\theta).$$

From here, the coefficients can be determined using Fourier's trick:

$$A_l = \frac{1}{2^{l-1}}\int_0^{\pi} \sigma_0(\theta) P_l(\cos\theta)\sin\theta\, d\theta.$$

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

$$\sigma_0(\theta) = k\cos\theta = kP_1(\cos\theta),$$

for some constant k, then all the A_i's are zero except for $l = 1$, and

$= \frac{k}{2\varepsilon_0}\int_0^{\pi}\left[P_1(\cos\theta)\right]^2 \sin\theta\, d\theta = \frac{k}{3\varepsilon_0}.$
The potential inside the sphere is therefore

$$V(r,\theta) = \frac{k}{3\varepsilon_0} r\cos\theta \qquad (r \le R),$$

.
whereas outside the sphere

$$V(r,\theta) = \frac{kR^3}{3\varepsilon_0}\frac{1}{r^2}\cos\theta \qquad (r \ge R).$$

In particular, if $\sigma_0(\theta)$ is the induced charge on a metal sphere in an external field $E_0\hat{z}$, so that $k = 3\varepsilon_0 E_0$ (Eq. 3.77), then the potential inside is $E_0 r\cos\theta = E_0 z$, and the field is $-E_0\hat{z}$–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.

(3.2) This is a 3D problem with the following Laplace equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0,$$

which are subject to the following boundary conditions:

(i)	$V = 0$	when $y = 0$,	**(ii)**	$V = 0$	when $y = a$,
(iii)	$V = 0$	when $z = 0$,	**(iv)**	$V = 0$	when $z = b$,
(iii)	$V \to 0$	as $x \to \infty$,	**(iv)**	$V = V_{0(x,y)}$	when $x = 0$.

As usual, we are looking for solutions that are products:

$$V(x,y,z) = X(x)\,Y(y)\,Z(z).$$

Putting this into the Laplace equation presents us:

$$\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} = 0.$$

It follows that:

$$\frac{1}{X}\frac{d^2X}{dx^2} = C_1, \quad \frac{1}{Y}\frac{d^2Y}{dy^2} = C_2 \quad \frac{1}{Z}\frac{d^2Z}{dz^2} = C_3, \quad \text{with} \quad C_1 + C_2 + C_3 = 0.$$

We will assume C_1 must be positive and C_2, C_3 are negative. Setting $C_2 = -k^2$ and $C_3 = -l^2$, we have $C_1 = k^2 + l^2$, and therefore:

$$\frac{d^2X}{dx^2} = \left(k^2 + l^2\right) X, \qquad \frac{d^2Y}{dy^2} = -k^2 Y, \qquad \frac{d^2Z}{dz^2} = -l^2 Z$$

We have yet again turned a partial differential equation into a series of ordinary differential equations. The solutions are:

$$\begin{aligned} X(x) &= A\exp\left(\sqrt{k^2 + l^2}x\right) + B\left(-\sqrt{k^2 + l^2}x\right), \\ Y(y) &= C\sin ky + D\cos ky, \\ Z(z) &= E\sin lz + F\cos lz. \end{aligned}$$

Boundary condition **(v)** implies $A = 0$, **(i)** gives $D = 0$ and **(iii)** yields $F = 0$, whereas **(ii)** and **(iv)** require $k = n\pi/a$ and $l = m\pi/b$, where n and m are positive integers. Combining the remaining constants, we are presented with:

$$V(x, y, z) = C\exp\left(-\pi\sqrt{(n/a)^2 + (m/b)^2}x\right)\sin\left(n\pi y/a\right)\sin\left(m\pi z/b\right).$$

This solution meets all the boundary condition except **(vi)**. It contains two unspecified integers (m and n) and the most general linear combination is a double sum:

$$V(x, y, z) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m}\exp\left(-\pi\sqrt{(n/a)^2 + (m/b)^2}x\right)\sin\left(n\pi y/a\right)\sin\left(m\pi z/b\right).$$

We hope to fit the remaining boundary condition:

$$V(0, y, z) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m}\sin\left(n\pi y/a\right)\sin\left(m\pi z/b\right) = V_0(y, z).$$

by choosing appropriate coefficients of $C_{n,m}$.
To determine these constants, we multiply by $\sin\left(n'\pi y/a\right)\sin\left(m'\pi z/b\right)$, where n' and m' are arbitrary positive integers, and integrate:

$$\begin{aligned} \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} C_{n,m}\int_0^b \sin\left(n\pi y/a\right)\sin\left(n'\pi y/a\right)dy\int_0^b \sin\left(m\pi z/b\right)\sin\left(m'\pi z/b\right)dz \\ = \int_0^a\int_0^b V_0(y, z)\sin\left(n'\pi z/b\right)\sin\left(m'\pi z/b\right)dydz. \end{aligned}$$

So:

$$C_{n,m} = \frac{4}{ab}\int_0^a\int_0^b V_0(y, z)\sin\left(n\pi z/b\right)\sin\left(m\pi z/b\right)dydz.$$

We are almost at the end of our problem. For example, if the end of the tube is a conductor at constant period V_0:

$$C_{n,m} = \frac{4V_0}{ab}\int_0^a \sin\left(n\pi y/a\right)dy \int_0^b \sin\left(m\pi z/b\right)dz$$
$$= \begin{cases} 0, & \text{if } n \text{ or } m \text{ is even,} \\ \dfrac{16V_0}{\pi^2 nm}, & \text{if } n \text{ and } m \text{ are odd.} \end{cases}$$

In this case:

$$V(x,y,z) = \frac{16V_0}{\pi^2}\sum_{n,m=1,3,5,\ldots}^{\infty}\frac{1}{nm}\exp\left(-\pi\sqrt{(n/a)^2+(m/b)^2}\,x\right)\sin\left(n\pi y/a\right)\sin\left(m\pi z/b\right) \quad \blacksquare$$

As the successive terms decrease rapidly, a reasonable approximation would be obtained by keeping only the first few terms.

The success of this method hinged on two (2) 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:

$$f(y) = \sum_{n=1}^{\infty} C_n f_n(y).$$

The functions $\sin(n\pi y/a)$ are complete on the interval $0 \le y \le a$. 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:

$$\int_0^a f_n(y)$$

$']](y)dy = 0 \qquad \text{for} n' \neq n.$

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.)

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:

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

I shall assume the problem has **azimuthal symmetry**, so that $V_i sindependentof\phi$,[12] in that case, Eq. 3.53 reduces to

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

As before, we look for solutions that are products:

$$V(r,\theta)=R(r)\,\Theta(\theta).$$

Chapter 5

Electric Fields in Matter

Table of Contents

5.1 Polarisation

5.1.1 Dielectrics

Here, we will study electric fields in matter. Matter, comes in many varieties, such as:

- solids, liquids, gasses
- metals, woods, glasses

and these substances do not all respond in the same way to electrostatic fields. Regardless, most everyday objects belong[1] to one of two (2) large classes: **conductors** and **insulators**. We have talked about conductors in Chapter 2. These substances contain an unlimited supply of charges that are free to move about through the material. In practice, what this means is that many of the electrons[2] are **not associated** with any particular nucleus, but room around at will.

Dielectrics, by contrast, all charges are attached to specific atoms or molecules. This means they're on a tight less, 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 (2) principal mechanisms by which electric fields can distort the charge distribution of a dielectric atom or molecule:

1. stretching,
2. rotating

5.1.2 Induced Dipoles

Let's think about what happens to a neutral atom when it is placed in an electric field $\boldsymbol{E}$? First guess might well be:

Absolutely nothing as the atom is not charged, the field has no effect on it.

But that is **incorrect**.

While the atom as a whole is electrically neutral, there is a positively charged core[3] and a negatively charged electron cloud surrounding it. These two (2) regions of charge within the atom are influenced by the field:

[3] the nucleus

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, "ionising" 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, those positive and negative charges attract one another, and that holds the atom together. The two (2) opposing forces–$\boldsymbol{E}$ pulling the electrons and nucleus apart, their mutual attraction drawing them back together–reach a balance, leaving the atom **polarized**, with plus charge shifted slightly one way, and minus the other. The atom now has a tiny dipole moment $\boldsymbol{p}$, which points in the *same direction as* $\boldsymbol{E}$. Typically, this induced dipole moment is approximately proportional to the field (as long as the latter is not too strong):

$$\boldsymbol{p} = \alpha \boldsymbol{E}.$$

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.

Exercise 5.1: Atomic Polarisability

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.

Solution

In the presence of an external field $\boldsymbol{E}$, 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

$$E_e = \frac{1}{4\pi\varepsilon_0}\frac{qd}{a^3}$$

(Prob. 2.12). At equilibrium, then,

$$E = \frac{1}{4\pi\varepsilon_0}\frac{qd}{a^3} \text{or} p = qd = (4\pi\varepsilon_0 a^3)E.$$

The atomic polarizability is therefore

$$\alpha = 4\pi\varepsilon_0 a^3 = 3\varepsilon_0 v \quad \blacksquare$$

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/W 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:

$$\boldsymbol{p} = \alpha_\perp \boldsymbol{E}_\perp + \alpha_\parallel \boldsymbol{E}_\parallel.$$

In this case, the induced dipole moment may not even be in the same *direction* as $\boldsymbol{E}$. 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 $\boldsymbol{E}$ and $\boldsymbol{p}$:

$$\begin{aligned} p_x &= \alpha_{xx} E_x + \alpha_{xy} E_y + \alpha_{xz} E_z \\ p_y &= \alpha_{yx} E_x + \alpha_{yy} E_y + \alpha_{yz} E_z \\ p_z &= \alpha_{zx} E_x + \alpha_{zy} E_y + \alpha_{zz} E_z \end{aligned}$$

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_{zz}, etc.) vanish, leaving just three nonzero polarizabilities: α_{xx}, α_{yy}, and α_{zz}.

5.1.3 Alignment of Polar Molecules

The neutral atom discussed in Sect. 4.1.2 had no dipole moment to start with–$\boldsymbol{p}$ 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 and $\boldsymbol{F}_+ = q\boldsymbol{E}$, exactly cancels the force on the negative end, $\boldsymbol{F}_- = -q\boldsymbol{E}$ (Fig. 4.5). However, there will be a *torque*:

$\boldsymbol{N} = (\boldsymbol{r}_+ \times \boldsymbol{F}_+) + (\boldsymbol{r}_- \times \boldsymbol{F}_-)$

$= \left[(\boldsymbol{d}/2) \times (q\boldsymbol{E})\right] + \left[(-\boldsymbol{d}/2) \times (-q\boldsymbol{E})\right] = q\boldsymbol{d} \times \boldsymbol{E}.$

Thus a dipole $\boldsymbol{p} = q\boldsymbol{d}$ in a uniform field $\boldsymbol{E}$ experiences a torque

$$\boldsymbol{N} = \boldsymbol{p} \times \boldsymbol{E}. \tag{5.1}$$

Notice that $\boldsymbol{N}$ is in such a direction as to line $\boldsymbol{p}$ up *parallel* to $\boldsymbol{E}$; 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 *nonuniform*, so that $\boldsymbol{F}_+$ does not exactly balance $\boldsymbol{F}_-$, there will be a *net force* on the dipole, in addition to the torque. Of course, $\boldsymbol{E}$ must change rather abruptly for there to be significant variation in the space of one molecule, so this is not ordinary 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:

$$\boldsymbol{F} = \boldsymbol{F}_+ + \boldsymbol{F}_- = q\,(\boldsymbol{E}_+ - \boldsymbol{E}_-) = q\,(\Delta \boldsymbol{E})\,,$$

where $\Delta \boldsymbol{E}$ 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_z:*test*

$$\Delta E_x \equiv (\nabla E_x) \cdot \boldsymbol{d},$$

with corresponding formulas for E_y and E_z. More compactly,

$$\Delta \boldsymbol{E} = (\boldsymbol{d} \cdot \nabla)\boldsymbol{E},$$

and therefore[2]

$$\boldsymbol{F} = (\boldsymbol{p} \cdot \nabla)\boldsymbol{E}.$$

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 $\boldsymbol{N} = (\boldsymbol{p} \times \boldsymbol{E}) + (\boldsymbol{r} \times \boldsymbol{F})$.

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 (quantitatively) the original atomic flow. What happens to a piece of dielectric matrix or when it is placed in an electric field? If the substance consists of normal, 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.)

Note that these two conditions 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

$$\boldsymbol{P} = \textit{dipole moment per unit volume},$$

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 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 *iself* 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**.

5.2 The Field of a Polarised Matter

5.2.1 Bound Charges

Suppose we have a piece of polarized materialthat is, an object containing a lot of microscopic dipoles lined up. The dipole moment per unit volume $\boldsymbol{P}$ is given. *Question:* What is the field produced by this object (not the field that may have *caused* the polarization, but the field of the polarization *is 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 $\boldsymbol{p}$ (Eq. 3.99),

$$V(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\frac{\boldsymbol{p}\cdot\hat{\boldsymbol{\alpha}}}{\hat{\boldsymbol{\alpha}}^2},$$

where $\boldsymbol{\alpha}$ 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 $\boldsymbol{p} = \boldsymbol{P}\,\mathrm{d}\tau'$ in each volume element $d\tau'$, so the total potential is

$$V(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int_V \frac{\boldsymbol{P}(\boldsymbol{r}')\cdot v\hat{\phi}}$$

$^2\,\mathrm{d}\tau'$.

That *does* it, in principle. But a little sleight-of-hand casts this integral into a much more illuminating form. Observing that

$$\nabla'\left(\frac{1}{s}\right) = \frac{v\hat{\phi}}{s^2},$$

where (unlike Prob. 1.13) the differentiation is with respect to the *source* coordinates ($\boldsymbol{r}'$), we have

$$V = \frac{1}{4\pi\varepsilon_0}\int_V \boldsymbol{P}\cdot\nabla'\left(\frac{1}{\epsilon}\right)\,\mathrm{d}\tau'.$$

Integrating by parts, using product rule number 5 (in the front cover), gives

$$V = \frac{1}{4\pi\varepsilon_0}\left[\int_V \nabla'\cdot\left(\frac{\boldsymbol{P}}{\epsilon}\right)\,\mathrm{d}\tau' - \int_V \frac{1}{\epsilon}(\nabla'\cdot\boldsymbol{P})\,\mathrm{d}\tau'\right],$$

or, invoking the divergence theorem,

$$V = \frac{1}{4\pi\varepsilon_0}\oint_S \frac{1}{\boldsymbol{P}\cdot d\boldsymbol{a}'} - \frac{1}{4\pi\varepsilon_0}\int_V \frac{1}{v\hat{\diamond}}(v\nabla'\cdot\boldsymbol{P})\,\mathrm{d}\tau'.$$

The first term looks like the potential of a surface charge

$$\boxed{\sigma_b \equiv \boldsymbol{P}\cdot v\hat{n}}$$

(where^is the normal unit vector), while the second term looks like the potential of a volume charge

$$\boxed{\rho_b \equiv -\boldsymbol{V}\cdot\boldsymbol{P}.}$$

With these definitions, Eq. 4.10 becomes

$$V(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\oint_S \frac{\rho_b}{\epsilon}da' + \frac{1}{4\pi\varepsilon_0}\int_V \frac{\rho_b}{\epsilon}\,\mathrm{d}\tau'.$$

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 $\rho_b = -v\nabla\cdot\boldsymbol{P}$ plus a surface charge density $\sigma_b = \boldsymbol{P}\cdot\hat{}$. 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).

5.2.2 Understanding 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 let 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"–more 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;

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

$$\sigma_b = \frac{q}{A_{\text{end}}} = P\cos\theta = \boldsymbol{P}\cdot\hat{}$$

The effect of the polarization, then, is to paint a bound charge $\sigma_b = \boldsymbol{P}\cdot\hat{}$ 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*.

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 $\boldsymbol{P}$ results in a pileup of negative charge. Indeed, the net bound charge $\int \rho_b d\tau$
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 $\boldsymbol{P} \cdot v\hat{n}$ per unit area, so

$$\int_V \rho_b d\tau = -\oint_S \boldsymbol{P} \cdot d\boldsymbol{a} = -\int_V (\nabla \cdot \boldsymbol{P})\, d\tau.$$

Since this is true for *any* volume, we have

$$\rho_b = -\nabla \cdot \boldsymbol{P},$$

confirming, again, the more rigorous conclusion of Sect. 4.2.1.

5.2.3 Field Inside a Dielectric

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 ($\approx$ is many times greater than the separation distance between plus and minus charges), so the dipole potential dominates overwhelmingly and the detailed "spininess" 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 fastltically 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
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 most."

In order 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 along an. Suppose I want to calculate the macroscopic field at some point $\boldsymbol{r}$ 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 $\boldsymbol{r}$, of radius, say, a thousand times the size of a molecule. The macroscopic field at $\boldsymbol{r}$, then, consists of two parts: the average field over the sphere due to all charges *outside*, plus the average due to all charges *inside*:

$$\boldsymbol{E} = \boldsymbol{E}_{\text{out}} + \boldsymbol{E}_{\text{in}}.$$

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 $\boldsymbol{E}_{\text{out}}$ is the field at $\boldsymbol{r}$ due to the dipoles exterior to the sphere. These are far enough away that we can safely use Eq. 4.9:

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

$$\boldsymbol{E}_{\text{in}} = -\frac{1}{4\pi\varepsilon_0}\frac{\boldsymbol{p}}{R^3},$$

regardless of the details of the charge distribution within the sphere. The only relevant quantity is the total dipole moment, $\boldsymbol{p} = (\frac{4}{3}\pi R^3)\boldsymbol{P}$:

$$\boldsymbol{E}_{\text{in}} = -\frac{1}{3\varepsilon_0}\boldsymbol{P}.$$

Now, by assumption, the sphere is small enough that $\boldsymbol{P}$ 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: $-(1/3\varepsilon_0)\boldsymbol{P}$ (Eq. 4.14). But this is precisely what $\boldsymbol{E}_{\text{in}}$ (Eq. 4.18) puts back in! The macroscopic field, then, is given by the potential

$$V(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\boldsymbol{P}(\boldsymbol{r}')\cdot\hat{\boldsymbol{\delta}}}{s^2}\,d\tau',$$

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.

Now we have to reread the last couple of paragraphs for the argument to sink in your object that it revolves around the curious fact the average field over *any* sphere that it revolves around the same 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

4.31 ■ 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, $\rho_b = -\boldsymbol{V}\cdot\boldsymbol{P}$ within the dielectric and $\sigma_b = \boldsymbol{P}\cdot\hat{}$ 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 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:

$$\rho = \rho_b + \rho_f,$$

and Gauss's law reads

$$\varepsilon_0 \nabla \cdot \boldsymbol{E} = \rho = \rho_b + \rho_f = -\nabla \cdot \boldsymbol{P} + \rho_f,$$

where $\boldsymbol{E}$ is now the total field, not just that portion generated by polarization.
It is convenient to combine the two divergence terms:

$$\boldsymbol{V} \cdot (\varepsilon_0 \boldsymbol{E} + \boldsymbol{P}) = \rho_f.$$

The expression in parentheses, designated by the letter $\boldsymbol{D}$,

$$\boldsymbol{D} \equiv \varepsilon_0 \boldsymbol{E} + \boldsymbol{P},$$

is known as the **electric displacement**. In terms of $\boldsymbol{D}$, Gauss's law reads

$$\left[\ \boldsymbol{V} \cdot \boldsymbol{D} = \rho_f, \ \right]$$

or, in integral form,

$$\oint \boldsymbol{D} \cdot d\boldsymbol{a} =$$

where Q_{fw} 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 $\boldsymbol{D}$ by the standard Gauss's law methods.

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 $\boldsymbol{D}$ is substituted for $\varepsilon_0 \boldsymbol{E}$. For this reason, you may be tempted to conclude that $\boldsymbol{D}$ is "just like" $\boldsymbol{E}$ (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 $\boldsymbol{D}$ instead of $\boldsymbol{E}$." This reasoning is deductive, but the conclusion is false; in particular, there is no "Coulomb's law" for $\boldsymbol{D}$:

$$\boldsymbol{D}(\boldsymbol{r}) \neq \frac{1}{4\pi} \int \frac{v\hat{\epsilon}}{v\hat{\epsilon}^2} \rho_f(\boldsymbol{r}') \, \mathrm{d}\tau'.$$

The parallel between $\boldsymbol{E}$ and $\boldsymbol{D}$ 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 $\boldsymbol{E}$ is always zero. But the curl of $\boldsymbol{D}$ is not always zero.

$$\nabla \times \boldsymbol{D} = \varepsilon_0 (\nabla \times \boldsymbol{E}) + (\nabla \times \boldsymbol{P}) = \nabla \times \boldsymbol{P},$$

and there is no reason, in general, to suppose that the curl of $\boldsymbol{P}$ 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 this is no free charge anywhere, so if you really believe that the only source of $\boldsymbol{D}$ is ρ_f, you will be forced to conclude that $\boldsymbol{D} = v0$ everywhere, and hence that $\boldsymbol{E} = (-1/\varepsilon_0)\boldsymbol{P}$ inside and $\boldsymbol{E} = v0$ outside the electret, which is obviously wrong. (I leave it for you to find the place where $v\nabla \times \boldsymbol{P} \neq v0$ in this problem.) Because $v\nabla \times \boldsymbol{D} \neq v0$, moreover, $\boldsymbol{D}$ cannot be expressed as the gradient of a scalarthere is no "potential" for $\boldsymbol{D}$.

Additive: 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 $\nabla \times \boldsymbol{P}$ 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 $\boldsymbol{D}$. Equation 4.23 tells us the discontinuity in the component perpendicular to an interface:

$$D^{\perp}_{\text{above}} - D^{\perp}_{\text{below}} = \sigma_f,$$

while Eq. 4.25 gives the discontinuity in parallel components:

$$\boldsymbol{D}^{\parallel}_{\text{above}} - \boldsymbol{D}^{\parallel}_{\text{below}} = \boldsymbol{P}^{\parallel}_{\text{above}} - \boldsymbol{P}^{\parallel}_{\text{below}}.$$

In the presence of dielectrics, these are sometimes more useful than the corresponding boundary conditions on $\boldsymbol{E}$ (Eqs. 2.31 and 2.32):

$$E^{\perp}_{\text{above}} - E^{\perp}_{\text{below}} = \frac{1}{\varepsilon_0}\sigma,$$

and

$$\boldsymbol{E}^{\parallel}_{\text{above}} - \boldsymbol{E}^{\parallel}_{\text{below}} = v0.$$

You might try applying them, for example, to Probs. 4.16 and 4.17.

4.4 ■ LINEAR DIELECTICS 4.4.1 ■ Susceptibility, Permittivity, Dielectric Constant

In Sects. 4.2 and 4.3 we did not commit ourselves as to the cause of $\boldsymbol{P}$, 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 ordinary 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 $\boldsymbol{E}$ is not too strong:

$$\boldsymbol{P} = \varepsilon_0 \chi_e \boldsymbol{E}.$$

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 $\boldsymbol{E}$ 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 $\boldsymbol{E}_0$, we

cannot compute $\boldsymbol{P}$ 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 $\boldsymbol{D}$ can be deduced directly from the free charge distribution.
In linear media we have

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} = \varepsilon_0 \boldsymbol{E} + \varepsilon_0 \chi_e \boldsymbol{E} = \varepsilon_0 (1 + \chi_e) \boldsymbol{E},$$

so $\boldsymbol{D}$ is also proportional to $\boldsymbol{E}$:

$$\boldsymbol{D} = \epsilon \boldsymbol{E},$$

where

$$\epsilon \equiv \varepsilon_0 (1 + \chi_e).$$

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 8.85×10^{-12} $C^2/N{\cdot}m^2$.) If you remove a factor of ε_0, the remaining dimensionless quantity

$$\epsilon_r \equiv 1 + \chi_e = \frac{\epsilon}{\varepsilon_0}$$

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]
You might suppose that linear dielectrics escape the defect in the parallel between $\boldsymbol{E}$ and $\boldsymbol{D}$. Since $\boldsymbol{P}$ and $\boldsymbol{D}$ are now proportional to $\boldsymbol{E}$, does it not follow that their curls, like $\boldsymbol{E}$'s, must vanish? Unfortunately, it does not, for the line integral of $\boldsymbol{P}$ around a closed path that strraddles the boundary between one type of material and another need not be zero, even though the integral of $\boldsymbol{E}$ around the same loop must be. The reason is that the proportionality factor $\varepsilon_0 \chi_e$ is different on the two sides. For instance, at the interface between a polarized dielectric and the vacuum (Fig. 4.21), $\boldsymbol{P}$ is zero on one side but not on the other. Around this
loop $\oint \boldsymbol{P} \cdot d\boldsymbol{l} \neq 0$, and hence, by Stokes' theorem, the curl of $\boldsymbol{P}$ 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

$$v\nabla \cdot \boldsymbol{D} = \rho_f \quad \text{and} \quad v\nabla \times \boldsymbol{D} = v0,$$

so $\boldsymbol{D}$ can be found from the free charge just as though the dielectric were not there:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E}_{\text{vac}},$$

where $\boldsymbol{E}_{\text{vac}}$ 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,

$$\boldsymbol{E} = \frac{1}{\epsilon}\boldsymbol{D} = \frac{1}{\epsilon_r}\boldsymbol{E}_{\text{vac}}.$$

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

$$\boldsymbol{E} = \frac{1}{4\pi\epsilon}\frac{q}{r^2}\hat{\boldsymbol{r}}$$

(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]
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

$$[] \; P_x = \varepsilon_0(\chi_{e_x}E_x + \chi_{e_y}E_y + \chi_{e_x}E_z)$$
$$[] \; P_y = \varepsilon_0(\chi_{e_y}E_x + \chi_{e_y}E_y + \chi_{e_y}E_z),$$
$$[] \; P_z = \varepsilon_0(\chi_{e_x}E_x + \chi_{e_y}E_y + \chi_{e_z}E_z)$$

just as Eq. 4.1 was superseded by Eq. 4.3 for asymmetrical molecules. The nine coefficients, χ_{e_x}, χ_{e_y},..., constitute the **susceptibility tensor**.

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]

$$\rho_b = -v\nabla\cdot\boldsymbol{P} = -v\nabla\cdot\left(\varepsilon_0\frac{\chi_e}{\epsilon}\boldsymbol{D}\right) = -\left(\frac{\chi_e}{1+\chi_e}\right)\rho_f.$$

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

$$\epsilon_{\text{above}}E^{\perp}{}_{\text{above}} - \epsilon_{\text{below}}E^{\perp}{}_{\text{below}} = \sigma_f,$$

or (in terms of the potential),

$$\epsilon_{\text{above}}\frac{\partial V_{\text{above}}}{\partial n} - \epsilon_{\text{below}}\frac{\partial V_{\text{below}}}{\partial n} = -\sigma_f,$$

whereas the potential itself is, of course, continuous (Eq. 2.34):

$$V_{\text{above}} = V_{\text{below}}.$$

Chapter 6

Magnetostatics

Table of Contents

6.1 Lorentz Force Law

6.1.1 Magnetic Fields

Let us recall the fundamental problem of electrodynamics:

> We have a collection of charges q_1, q_2, $\cdots$ (called the source charges), and we want to calculate the force they exert on some other charge, say Q (the test charge)

Based on the principle of superposition[1] , it is sufficient to find the force of a single source charge, which is the vector sum of all the individual forces.

Till now, we have confined our attention to the simplest case, electrostatics, in which the source charge is at rest[2] . 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 apartthey 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. (Its true that electrons are flowing down the linethats what a current isbut 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!

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 $\boldsymbol{E}$ in the space around it, a *moving* charge generates, in addition, a magnetic field $\boldsymbol{B}$. In fact, magnetic fields are a lot easier to detect, in practice–all you need is a Boy Scott 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.

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.

6.1.2 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 $\boldsymbol{v}$ in a magnetic field $\boldsymbol{B}$, is[3]

[3] As $\boldsymbol{F}$ and $\boldsymbol{v}$ are vectors, $\boldsymbol{B}$ is actually a

$$\boldsymbol{F}_{\text{mag}} = Q\left(\boldsymbol{v} \times \boldsymbol{B}\right) \tag{6.1}$$

Theory 6.0: Psuedovector

A Psuedovector is a quantity which behaves like a vector in many situations, but its direction does not conform when the object is rigidly transformed by rotation, translation, reflection, etc.

For example: A loop of wire (black), carrying a current I, creates a magnetic field B (blue). If the position and current of the wire are reflected across the plane indicated by the dashed line, the magnetic field it generates would not be reflected: Instead, it would be reflected and reversed. The position and current at any point in the wire are "true" vectors, but the magnetic field B is a pseudovector.

This is known as the **Lorentz force law**. In the presence of both electric ($\boldsymbol{E}$) *and* magnetic fields $\boldsymbol{B}$, the net force on Q would be

$$\boldsymbol{F} = Q\left[\boldsymbol{E} + \left(\boldsymbol{v} \times \boldsymbol{B}\right)\right].$$

Eq. (**??**) is a fundamental axiom of the theory of electromagnetism, whose justification is to be found in experiments.

Our primary job is to calculate the magnetic field $\boldsymbol{B}$. But before proceeding, 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.

Exercise 6.1: Motion of a Cyclotron

The archetypical 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:

$$QvB = m\frac{v^2}{R}, \quad or \quad p = QBR,$$

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 $\mathbf{B}$. If it starts out with some additional speed $v_{\parallel}$ *parallel* to $\mathbf{B}$, 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 $\mathbf{B}$, $v_{\perp}$.

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

$$dW_{\mathbf{mag}} = \mathbf{F}_{\mathrm{mag}} \cdot d\mathbf{l} = Q\,(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}dt = 0.$$

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 water. When a magnetic crate 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 a

6.1.3 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 $qand\boldsymbol{v}$, 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 originally 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 unfortunately convention.[5] Current is measured in coulombs-per-second, or **amperes** (A):

1 A = 1 C/s.

A line charge λ traveling down a wire at speed v (Fig. 5.9) constitutes a current

$$I = \lambda v,$$

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*:

$$\boldsymbol{I} = \lambda \boldsymbol{v}.$$

Exercise 6.2: Magnetic Field of a Straight Wire

Find the magnetic field a distance s from a long straight wire carrying a steady current I.

Solution

In the diagram it can be seen ($d\boldsymbol{l}' \times \hat{\boldsymbol{r}}$) points out of the page and has the magnitude:

$$dl' \sin\alpha = dl' \cos\theta$$

Also from geometry we can observe $l' = s\tan\theta$, therefore:

$$\frac{1}{r^2} = \frac{\cos^2\theta}{s^2}$$

Which gives:

$$\begin{aligned} B &= \frac{\mu_0 I}{4\pi}\int_{\theta_1}^{\theta_2}\left(\frac{\cos^2\theta}{s^2}\right)\left(\frac{s}{\cos^2\theta}\right)\cos\theta\,d\theta \\ &= \frac{\mu_0 I}{4\pi s}\int_{\theta_1}^{\theta_2}\cos\theta\,d\theta \\ &= \frac{\mu_0 I}{4\pi s}\left(\sin\theta_2 - \sin\theta_1\right) \end{aligned}$$

6.2 Divergence and Curl of B

Exercise 6.3: Magnetic Field at a Distance

Find the magnetic field a distance z above the center of a circular loop of radius R, which carries a steady current I

Solution

The field dB attributable to the segment dl' points as shown. As we integrate dl' around the loop, dB sweeps out a cone. The horizontal components cancel, and the vertical components combine, to give

$$B(z) = \frac{\mu_0}{4\pi} I \int \frac{dl'}{\lambda^2}\cos\theta.$$

(Notice that dl' and $v\phi$ are perpendicular, in this case; the factor of $\cos\theta$ projects out the vertical component.) Now, $\cos\theta$ and s^2 are constants, and $\int dl'$ is simply the circumference, $2\pi R$, so

$$B(z) = \frac{\mu_0 I}{4\pi}\left(\frac{\cos\theta}{s^2}\right)2\pi R = \frac{\mu_0 I}{2}\frac{R^2}{(R^2+z^2)^{3/2}}.$$

6.2.1 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 seen in *electrostatic*); let's calculate it.

According to Eq. 5.38, the integral of $\boldsymbol{B}$ around a circular path of radius s, centered at the wire, is

$$\oint \boldsymbol{B}\cdot d\boldsymbol{l} = \oint \frac{\mu_0 I}{2\pi s}\,dl = \frac{\mu_0 I}{2\pi s}\oint dl = \mu_0 I.$$

Notice that the answer is independent of s; that's because B decreases at the same rate as the circumference *increases*. In fact, it doesn't have to be a circle; *any* old 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, $\boldsymbol{B} = (\mu_0 I/2\pi s)\,\hat{\boldsymbol{\phi}}$ and $d\boldsymbol{l} = ds\,\hat{\boldsymbol{s}} + s\,d\phi\,\hat{\boldsymbol{\phi}} + dz\,\hat{\boldsymbol{z}}$, so

$$\oint \boldsymbol{B}\cdot d\boldsymbol{l} = \frac{\mu_0 I}{2\pi}\oint \frac{1}{s} s\,d\phi = \frac{\mu_0 I}{2\pi}\int_0^{2\pi} d\phi = \mu_0 I.$$

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

$$\oint \boldsymbol{B} \cdot d\boldsymbol{l} = \mu_0 I_{\text{exc}},$$

where I_{exc} stands for the total current enclosed by the integration path. If the flow of charge is represented by a volume current density $\boldsymbol{J}$, the enclosed current is

$$I_{\text{exc}} = \int \boldsymbol{J} \cdot d\boldsymbol{a},$$

with the integral taken over any surface bounded by the loop. Applying Stokes' theorem to Eq. 5.44, then,

$$\int (\nabla \times \boldsymbol{B}) \cdot d\boldsymbol{a} = \mu_0 \int \boldsymbol{J} \cdot d\boldsymbol{a},$$

and hence

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} \quad \blacksquare \tag{6.2}$$

With minimal effort, we have actually obtained the general formula for the curl of $\boldsymbol{B}$. But our derivation has a gross assumption of infinite straight line currents[4] .

Most current configurations *cannot* be constructed out of infinite straight wires, and we have no right to assume Eq. (**??**) applies to them.
So the next section is devoted to the formal derivation of the divergence and curl of $\boldsymbol{B}$, starting from the Biot-Savart law itself[5] .

6.2.2 Divergence and Curl of B

The Biot-Savart law for the general case of a volume current reads:

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{\mu_0}{4\pi} \int \frac{\boldsymbol{J}\left(\boldsymbol{r}'\right)}{\boldsymbol{\imath}^2} d\tau'. \tag{6.3}$$

This formula gives the magnetic field at a point $\boldsymbol{r} = (x, y, z)$ in terms of an integral over the current distribution $\boldsymbol{J}(x', y', z')$. To explain:

- $\boldsymbol{B}$ is a function of (x, y, z),
- $\boldsymbol{J}$ is a function of (x', y', z'),
- $\boldsymbol{\imath} = \left(x - x'\right)\hat{x} + \left(y - y'\right)\hat{y} + \left(z - z'\right)\hat{z}$ is our separation vector
- $d\tau' = dx'\,dy'\,dz'$ is the infinitesimal volume unit.

The integration is over primed coordinates with the divergence and the curl of $\boldsymbol{B}$ are with respect to the un-primed coordinates.

Applying the divergence to Eq. (**??**), we obtain the following:

$$\nabla \cdot \boldsymbol{B} = \frac{\mu_0}{4\pi} \int \nabla \cdot \left(\boldsymbol{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) \mathrm{d}\tau'.$$

Invoking product rule number 6,

Product Rule 6

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

$$\nabla \cdot \left(\boldsymbol{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \cdot (\nabla \times \boldsymbol{J}) - \boldsymbol{J} \cdot \left(\nabla \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right).$$

But $\nabla \times \boldsymbol{J} = 0$, as $\boldsymbol{J}$ doesn't depend on the unprimed variables, while: $\nabla \times rac\hat{\imath}\imath^2 = 0$ [6], so

[6] Remember the divergence problem we in Vector s chapter.

$$\nabla \cdot \boldsymbol{B} = 0. \tag{6.4}$$

The *divergence* of the magnetic field is *zero*.

Applying the curl to Eq. (**??**), we obtain:

$$\nabla \times \boldsymbol{B} = \frac{\mu_0}{4\pi} \int \nabla \times \left(\boldsymbol{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) \mathrm{d}\tau'.$$

Our strategy here is, like previously, to expand the integrand, using the product rule 8:

Product Rule 6

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

$$\nabla \times \left(\boldsymbol{J} \times \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right) = \cancel{\left(\frac{\hat{\boldsymbol{\imath}}}{\imath^2} \cdot \nabla \right) \boldsymbol{J}}^{0} - \overbrace{(\boldsymbol{J} \cdot \nabla) \frac{\hat{\boldsymbol{\imath}}}{\imath^2}}^{\text{Part A}} + \overbrace{\boldsymbol{J} \left(\nabla \cdot \frac{\hat{\boldsymbol{\imath}}}{\imath^2} \right)}^{\text{Part B}} - \cancel{\frac{\hat{\boldsymbol{\imath}}}{\imath^2} (\nabla \cdot \boldsymbol{J})}^{0}$$

Theory 6.3: Nabla Operator on the Right

You might have observed the nabla operator appearing on the right side instead of the left which we are used to. For example, we know how to expand the following:

$$\nabla \cdot \boldsymbol{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}$$

When the nabla operator is on the right the calculation change slightly.

$$\boldsymbol{B} \cdot \nabla = B_x \frac{\partial}{\partial x} + B_y \frac{\partial}{\partial y} + B_z \frac{\partial}{\partial z}$$

As can we see from these two statement, this operation is not **commutative**.

$$\nabla \cdot \boldsymbol{B} \neq \boldsymbol{B} \cdot \nabla$$

Part B integrates to zero, as we'll see in the next paragraph[7].

[7] To simplify the expression, I have dropped terms involving derivatives of J, which are irrelevant where we needed J does not depend on x, y, z.

Part A involves the divergence of a particular expression which we needed to employ the delta function.

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

Therefore[8]:

$$\nabla \times \boldsymbol{B} = \frac{\mu_0}{4\pi} \int \boldsymbol{J}\left(\boldsymbol{r}'\right) 4\pi\delta^3\left(\boldsymbol{r} - \boldsymbol{r}'\right) \, \mathrm{d}\tau' = \mu_0 \boldsymbol{J}(\boldsymbol{r}) \, .$$

which confirms that Eq. 5.46 is not restricted to straight-line currents, but holds quite generally in magnetostatics.

To complete the derivation and to an extent our argument, we must check the second term in Eq. 5.52 integrates to zero. Because the derivative acts only on $\frac{\hat{\boldsymbol{\imath}}}{\imath^2}$, we can switch from ∇ to ∇' at the cost of a minus sign[9]:

$$-\left(\boldsymbol{J} \cdot \nabla\right) \frac{\hat{\boldsymbol{\imath}}}{\imath^2} = \left(\boldsymbol{J} \cdot \nabla'\right) \frac{\hat{\boldsymbol{\imath}}}{\imath^2}$$

The x component, in particular, is[10]:

$$\left(\boldsymbol{J} \cdot \nabla'\right) \left(\frac{x - x'}{\imath^3} \right) = \nabla' \cdot \left[\frac{(x - x')}{\imath^3} \boldsymbol{J} \right] - \left(\frac{x - x'}{\imath^3} \right) \left(\nabla' \cdot \boldsymbol{J}\right)$$

Now, for *steady* currents the divergence of $\boldsymbol{J}$ is zero (Eq. 5.33), so

$$\left[-(\boldsymbol{J} \cdot \nabla) \frac{v\lambda}{\lambda^2} \right]_x = \nabla' \cdot \left[\frac{(x - x')}{\lambda^3} \boldsymbol{J} \right] ,$$

and therefore this contribution to the integral (Eq. 5.51) can be written

$$\int_{\mathcal{V}} \nabla' \cdot \left[\frac{(x - x')}{\lambda^3} \boldsymbol{J} \right] d\tau' = \oint_{\mathcal{S}} \frac{(x - x')}{\lambda^3} \boldsymbol{J} \cdot d\boldsymbol{a}' .$$

(The reason for switching from $\boldsymbol{V}$ to $\boldsymbol{V}'$ 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, $\boldsymbol{J} = 0$ 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]

6.2.3 Amperé's Law

The equation for the curl of **B**,

$$\nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} \tag{6.5}$$

is called **Amperé's law**[11]. It can be converted to integral by applying *Stokes' theorem*:

$$\int (\nabla \times \boldsymbol{B}) \cdot d\boldsymbol{a} = \oint \boldsymbol{B} \cdot d\boldsymbol{l} = \mu_0 \int \boldsymbol{J} \cdot d\boldsymbol{a}. \tag{6.6}$$

Now, $\int \boldsymbol{J} \cdot \mathrm{d}\boldsymbol{a}$ is the total current passing through the surface, which we call I_{enc}:

> the **current enclosed** by the Amperian loop.

Therefore:

$$\oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} = \mu_0 I_{\text{enc}}. \tag{6.7}$$

This is the integral version of Ampere's law where it generalises Eq. 5.44 to *arbitrary* steady currents.

Observe that Eq. 5.57 inherits the sign ambiguity of *Stokes' theorem* which raises a question:

> Which way around the loop are we supposed to go and which *direction* through the surface corresponds to a positive current?

The answer is the **right-hand rule**. If the fingers of your right hand indicates 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 Ampere's plays the part of Gauss's:

$$\begin{cases} \text{Electrostatics} \\ \text{Magnetostatics} \end{cases}$$

In particular, for currents with appropriate symmetry, Ampere's law in integral form offers an easy way of calculating the magnetic field.

Exercise 6.4: Ampere's Law

Find the magnetic field a distance s from a long straight wire carrying a steady current I.

Solution

We know the direction of $\boldsymbol{B}$ is circumferential, circling around the wire as indicated by the right-hand rule. By symmetry, the magnitude of $\boldsymbol{B}$ is constant around an Amperian loop of radius s, centered on the wire. So Ampere's law gives

$$\begin{aligned} \oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} &= B \oint \mathrm{d}l \\ &= B 2\pi s = \mu_0 I_{\text{enc}} = \mu_0 I \end{aligned}$$

or simply:

$$B = \frac{\mu_0 I}{2\pi s} \quad \blacksquare$$

As can bee seen we have a similar answer but obtained with much less effort.

Similar to *Gauss's law*, *Ampère's law* is always *true* (for steady currents), but it is not always *useful*. It only shines when the symmetry of the problem enables you to pull B outside the integral $\boldsymbol{j}\ \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l}$ can we 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:

- Infinite straight lines,
- Infinite planes,
- Infinite solenoids,
- Toroids.

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.

6.2.4 Comparison of Magnetostatics and Electrostatics

The divergence and curl of the *electrostatic* field are

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\varepsilon_0}\rho, \qquad \text{(Gauss's law)};$$
$$\nabla \times \boldsymbol{E} = 0, \qquad \text{(no name)}.$$

These are **Maxwell's equations** for electrostatics. Together with the boundary condition $\boldsymbol{E} \to 0$ 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

$$\boldsymbol{V} \cdot \boldsymbol{B} = 0, \qquad \text{(no name)};$$

.

$$\boldsymbol{V} \times \boldsymbol{B} = \mu_0 \boldsymbol{J}, \quad \text{(Ampere's law)}.$$

These are Maxwell's equations for **magnetostatics**.

Again, together with the boundary condition:

$$\lim_{r\to\infty} \boldsymbol{B} = 0$$

far from all currents. Maxwell's equations determine the magnetic field. They are equivalent to the Biot-Savart law[12] . Maxwell's equations and the force law:

$$\boldsymbol{F} = Q\,(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$$

constitute the an elegant formulation of electrostatics and magnetostatics.

The electric field diverges away from a positive charge and the magnetic field line curls around a current. Electric field lines originate on positive charges and terminate on negative ones whereas magnetic field lines do not begin or end anywhere.

> If this were to be observed, it would create a non-zero divergence.

They typically form closed loops or extend out to infinity. To put it another way,

> There are no point sources for $\boldsymbol{B}$, as there are for $\boldsymbol{E}$ as there exists no magnetic analog to electric charge.

This is the physical content of the statement $\nabla \cdot \boldsymbol{B} = 0$. Coulomb and others believed that magnetism was produced by **magnetic charges**[13]. It was Ampere who first speculated that all magnetic effects are attributable to electric charges in motion (i.e., currents). To our current understanding of electrodynamics, Ampère was right.

Nevertheless, it remains an open experimental question whether magnetic monopoles exist in nature[14]. For our purposes, though, $\boldsymbol{B}$ 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 interact a magnetic field.

[14] they are out there pretty rare, or somebody would have found one, and in fact some current elementary particle theories require them

Typically, electric forces are enormously larger than magnetic ones. That's not something intrinsic to the theory as 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.

How is it, then, that we notice that at all the magnetic field, the electric field is a *transverse* (Lorentz), it is the *curved* of a magnetic field (B-G-Savart) and it is a *curved* (Lorentz), it is the *curved* of a magnetic field (B-G-Savart) metric 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 amrege to keep the wire *neutral*, by embedding in

6.3 The Magnetic Vector Potential

Just as $\nabla \times \boldsymbol{E} = 0$ permitted us to introduce a scalar potential (V) in electrostatics,

$$\boldsymbol{E} = -\nabla V,$$

so $\nabla \cdot \boldsymbol{B} = 0$ allows us to introduce a vector potential $\boldsymbol{A}$ in magnetostatics:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}. \tag{6.8}$$

These two (2) statements are proved by *Helmholtz Theorem*. The potential formulation automatically takes care of $\nabla \cdot \boldsymbol{B} = 0$[15]. There remains Ampere's law:

[15] since the divergence of a curl is always zero

$$\nabla \times \boldsymbol{B} = \nabla \times (\nabla \times \boldsymbol{A}) = \nabla (\nabla \cdot \boldsymbol{A}) - \nabla^2 \boldsymbol{A} = \mu_0 \boldsymbol{J}. \tag{6.9}$$

Remember, 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 $\boldsymbol{E}$.

Similar to V, we can add to $\boldsymbol{A}$ any function whose *curl* vanishes[16], with no effect on $\boldsymbol{B}$. We can exploit this freedom to eliminate the divergence of $\boldsymbol{A}$:

[16] which is to say, the *gradient of any scalar.*

$$\nabla \cdot \boldsymbol{A} = 0 \tag{6.10}$$

To prove that this is always possible, suppose that our original potential, $\boldsymbol{A}_0$, is *not* divergenceless[17]. If we add to it the gradient of λ:

[17] i.e., divergences

$$\boldsymbol{A} = \boldsymbol{A}_0 + \nabla\lambda$$

the new divergence is

$$\nabla \cdot \boldsymbol{A} = \nabla \cdot \boldsymbol{A}_0 + \nabla^2\lambda.$$

We can accommodate Eq. (**??**), then, if a function λ can be found that satisfies

$$\nabla^2\lambda = -\nabla \cdot \boldsymbol{A}_0$$

But this is mathematically identical to Poisson's equation (2.24),

$$\nabla^2 V = -\frac{\rho}{\varepsilon_0},$$

with $\nabla \cdot \boldsymbol{A}_0$ in place of ρ/ε_0 as the source. And we *know* how to solve Poisson's equation. In particular, if ρ goes to zero at infinity, the solution is Eq. 2.29:

$$V = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho}{\mathscr{r}}\,d\tau',$$

and by the same token, if $\nabla \cdot \boldsymbol{A}_0$ goes to zero at infinity, then

$$\lambda = \frac{1}{4\pi}\int \frac{\nabla \cdot \boldsymbol{A}_0}{\mathscr{r}}\,d\tau'.$$

If $\nabla \cdot \boldsymbol{A}_0$ 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 $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ specifies the *curl* of $\boldsymbol{A}$, but it doesn't say anything about the *divergence*. This gives us an opportunity to pick as we see fit, and zero is ordinarily the simplest choice.

With this condition on $\boldsymbol{A}$, Ampere's law Eq. (**??**) becomes:

$$\nabla^2\boldsymbol{A} = -\mu_0\boldsymbol{J}. \tag{6.11}$$

This is nothing but Poisson's equation for 3D. Assuming $\boldsymbol{J}$ goes to zero at infinity, we can read off the solution:

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\int \frac{\boldsymbol{J}\left(\boldsymbol{r}'\right)}{\mathscr{r}}\,d\tau' \tag{6.12}$$

For line and surface currents,

$$\boldsymbol{A} = \frac{\mu_0}{4\pi}\int \frac{\boldsymbol{I}}{\imath} dl' = \frac{\mu_0 I}{4\pi}\int \frac{1}{\imath}\,\mathrm{d}l' \quad \text{and} \quad \boldsymbol{A} = \frac{\mu_0}{4\pi}\int \frac{\boldsymbol{K}}{\imath}\,\mathrm{d}a'. \tag{6.13}$$

If the current does *not* go to zero at infinity, we have to find other ways to get $\boldsymbol{A}$.

It is important to mention that that $\boldsymbol{A}$ is not as useful as V. For one thing, it's still a *vector*, and although Eq. (**??**) and Eq. (**??**) 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 similar to:

$$\boldsymbol{B} = -\nabla U,$$

but this is incompatible with Ampere's law, as the curl of a gradient is always zero.

In addition, as magnetic forces do no work, $\boldsymbol{A}$ does not admit a simple physical interpretation in terms of potential energy per unit charge. Nevertheless, the vector potential has substantial theoretical importance, which is generally applied to studying radiation.

Chapter 7

Magnetic Fields in Matter

Table of Contents

7.1 Magnetisation

7.1.1 Diamagnets, Paramagnets, and Ferromagnets

If you were to ask a person on the street what magnetism is, you will probably be told about refrigerator decorations, compass needles, and the North Pole[1] .

However, all magnetic phenomena are due to **electric charges in motion**.

> If you were to 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, we 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 polarised, or **magnetised**.

Unlike electric polarisation, which is almost always in the same direction as $\boldsymbol{E}$, some material can be magnetised opposite to $\boldsymbol{B}$ (**diamagnets**) or *parallel* to $\boldsymbol{B}$ (**paramagnet**) and some metals (nickels, iron, cobalt) **retain** their magnetisation even after the external field has been removed. This last category of material is called **ferromagnets**[2] .

For ferromagnets, the magnetisation 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.

7.1.2 Torques and Forces on Magnetic Dipoles

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 **magneticized**.

Unlike electric polarization, which is almost always in the same direction as $\boldsymbol{E}$, opposite to $\boldsymbol{B}$ (**damagnets**). A few substances *parallel* to $\boldsymbol{B}$ (**parameters**) and some opposite 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.

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:

$$\boldsymbol{N} = aF\sin\theta\,\hat{x}.$$

The magnitude of the force on each of these segments is

$$F = IbB,$$

and therefore

$$\boldsymbol{N} = IabB\sin\theta\,\hat{x} = mB\sin\theta\,\hat{x},$$

or

$$\boldsymbol{N} = \boldsymbol{m}\times\boldsymbol{B},$$

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.

Notice that Eq. 6.1 is identical in form to the electrical analog, Eq. 4.4: $\boldsymbol{N} = \boldsymbol{p}\times\boldsymbol{E}$. 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 **paramagneticism**. 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:

$$\boldsymbol{F} = I \oint (d\boldsymbol{l} \times \boldsymbol{B}) = I \left(\oint d\boldsymbol{l} \right) \times \boldsymbol{B} = 0;$$

the constant $\boldsymbol{B}$ comes outside the integral, and the net displacement $\oint \mathrm{d}\boldsymbol{l}$ 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 $\boldsymbol{B}$ has a radial component, and there is a net downward force on the loop (Fig. 6.4):

$$F = 2\pi I R B \cos\theta.$$

For an *infinitesimal* loop, with dipole moment $\boldsymbol{m}$, in a field $\boldsymbol{B}$, the force is

$$\boldsymbol{F} = \nabla(\boldsymbol{m} \cdot \boldsymbol{B})$$

(see Prob. 6.4). Once again the magnetic formula is identical to its electrical "twin", if we write the taper in the form $\boldsymbol{F} = \boldsymbol{y}$ (or $\boldsymbol{F}$). See footnote to Eq. 4.5.)

If you're starting to get a sense of *deira*, perhaps you will have more favored for short wavelengths which would have the dipoles considered of positive and negative "charges" (north and south "poles," they called them), expanded 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 $\boldsymbol{m}$ bear any resemblance to the corresponding formulas for $\boldsymbol{p}$. Sometimes it is easier to think in terms of the "Gilbert" model of a magnetic dipole (separated monopoles), instead of the physically correct "Ampere" 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 $\boldsymbol{p}$ to $\boldsymbol{m}$, $1/\varepsilon_0$ to μ_0, and $\boldsymbol{E}$ to $\boldsymbol{B}$). 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.

7.1.3 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

$$\boldsymbol{m} = -\frac{1}{2}evR\,\hat{z}.$$

Like any other magnetic dipole, this one is subject to a torque ($\boldsymbol{m}\times\boldsymbol{B}$) 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 $\boldsymbol{B}$. For whereas the centripetal acceleration v^2/R is ordinarily sustained by electrical forces alone,

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{R^2} = m_e\frac{v^2}{R},$$

in the presence of a magnetic field there is an additional force, $-e(\boldsymbol{v}\times\boldsymbol{B})$. For the sake of argument, let's say that $\boldsymbol{B}$ is perpendicular to the plane of the orbit, as shown in Fig. 6.10; then

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{R^2} + e\bar{v}B = m_e\frac{v_2}{R}.$$

Under these conditions, the new speed $\bar{v}$ is *greater* than v:

$$e\bar{v}B = \frac{m_e}{R}(v_2 - v^2) = \frac{m_e}{R}(\bar{v}+v)(\bar{v}-v),$$

or, assuming the change $\Delta v = \bar{v} - v$ is small,

$$\Delta v = \frac{eRB}{2m_e}.$$

When $\boldsymbol{B}$ is turned on, then, the electron speeds up.[3]
A change in orbital speed means a change in the dipole moment (Eq. 6.4):

$$\Delta\boldsymbol{m} = -\frac{1}{2}e(\Delta v)\,Rv\hat{z} = -\frac{e^2R^2}{4m_e}\boldsymbol{B}.$$

Notice that *the change in* $\boldsymbol{m}$ *is opposite to the direction of* $\boldsymbol{B}$. (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 $\boldsymbol{B}$.) Orientiarily, 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 *antiparalle* 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 magnetos*taticisers*, however, and the details will have to await Chapter 7 (see Pro 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 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.

7.1.4 Magnetisation

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 accounts 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

$$\boldsymbol{M} \equiv \text{magnetic dipole moment per unit volume.}$$

$\boldsymbol{M}$ is called the **magnetization**; it plays a role analogous to the polarization $\boldsymbol{P}$ 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 $\boldsymbol{M}$ 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 tro (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 *paramagnetism is attracted into the field*, whereas the *diamagnetic is repelled away*. But the actual forces are pitfalls weak–in a typical experimental arrangement the force on a comparable sample of iron would be 10^7 or 10^8 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]

7.2 Field of a Magnetised Object

7.2.1 Bound Currents

Suppose we have a piece of magnetized material; the magnetic dipole moment per unit volume, $\boldsymbol{M}$, is given. What field does this object produce? Well, the vector potential of a single dipole $\boldsymbol{m}$ is given by Eq. 5.85:

$$\boldsymbol{A}(r) = \frac{\mu_0}{4\pi}\frac{\boldsymbol{m}\times\hat{z}}{\tau^2}.$$

In the magnetized object, each volume element $d\tau'$ carries a dipole moment $\boldsymbol{M}\,\mathrm{d}\tau'$, so the total vector potential is (Fig. 6.11)

Chapter 8

Electrodynamics

Table of Contents

8.1 Introduction

In order to make a current flow, we need to push the charges. How **fast** they move in response to a given push, depends on the nature of the material. For most substances, the current density $\boldsymbol{J}$ is proportional to the *force per unit charge*, $\boldsymbol{f}$:

$$\boldsymbol{J} = \sigma \boldsymbol{f}. \tag{8.1}$$

The proportionality factor σ [1] is an empirical constant which varies from one material to another.

> It's called the **conductivity** of the medium.

Most engineering literature list the *reciprocal* of σ, called the **resistivity** defined as $\rho = 1/\sigma$ [2] . Even *instulators* conduct slightly, though the conductivity of a metal is many orders of magnitude greater. In fact, for most purposes metals can be regarded as **perfect conductors**, with $\sigma \approx \infty$, while for insulators we can pretend $\sigma = 0$.
In principle, the force that drives the charges to produce the current could be anything:

> chemically induced, gravitational, or trained tardigrades with tiny sleds.

For *our* purposes, it's usually an electromagnetic force which does the job. In this case Eq. (**??**) becomes:

$$\boldsymbol{J} = \sigma(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}).$$

Ordinarily, the velocity of the charges is sufficiently small that the second term can be ignored:

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{8.2}$$

However, in plasmas, for instance, the magnetic contribution to $\boldsymbol{f}$ can be significant which makes in non-negligible.

Eq. (**??**) is called **Ohm's law**, though the physics behind it is really contained in Eq. (**??**), of which Eq. (**??**) is just a special case.

You might be confused as we said previously $\boldsymbol{E} = 0$ inside a conductor. But remember it was defined under **Electrostatics** which was for *stationary* charges ($\boldsymbol{J} = 0$). Moreover, for *perfect* conductors:

$$\boldsymbol{E} = \frac{\boldsymbol{J}}{\sigma} = 0$$

even if current *is* flowing.

In practice, metals are such good conductors that the electric field required to drive current in them is negligible. Therefore, we routinely treat the connecting wires in electric circuits as equipotentials.

Resistors, by contrast, are made from *poorly* conducting materials.
As these examples illustrate, the total current flowing from one **electrode** to the other is proportional to the potential difference between them:

$$V = IR.$$

This, of course, is the more familiar[3] version of Ohm's law. The constant of proportionality R is called the **resistance** and it's a function of the geometry of the arrangement and the conductivity of the medium between the electrodes.

[3] Ancient Greek unit.

Resistance is measured in **ohms** (Ω): an ohm is a volt per ampere. Observe that the proportionality between V and I is a direct consequence of Eq. (**??**). If we were to double V, we would double the charge on the electrodes. This in turn doubles $\boldsymbol{E}$, which (for an ohmic material) doubles $\boldsymbol{J}$, which finally doubles I.
For *steady* currents and *uniform* conductivity,

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\sigma} \nabla \cdot \boldsymbol{J} = 0,$$

which shows the charge density as zero. Any unbalanced charge resides on the surface[4]

[4] We proved this long ago, for the case of *stationary* the fact ently, it hen the owed to move.

It follows, in particular, that Laplace's equation holds within a homogeneous ohmic material carrying a steady current.

I don't suppose there is any formula in physics more familiar than Ohm's law, and yet it's not really a true law, in the sense of Coulomb's or Ampère's. Rather, it is a rule of thumb that applies pretty well to many substances. In fact, when we stop to think about it, it's a little surprising that Ohm's law *ever* holds.

After all, a given field $\boldsymbol{E}$ produces a force $q\boldsymbol{E}$ (on a charge q), and according to Newton's second law, the charge will accelerate. But if the charges are *accelerating*, why doesn't the current *increase* with time, growing larger and larger the longer you leave the field on?

Ohm's law implies, on the contrary, that a constant field produces a constant *current*, which suggests a constant *velocity*.

Isn't that a contradiction to Newton's law?

No, for we are forgetting the frequent collisions electrons make as they pass down the wire. Think of the situation as this:

Suppose you're driving down a street with a stop sign at every intersection, so that, although you accelerate constantly in between, you are obliged to start all over again with each new block. Your average speed is then a constant, in spite of the fact that (save for the periodic abrupt stops), you are always accelerating.

If the length of a block is λ and your acceleration is a, the time it takes to go a block is

$$t = \sqrt{\frac{2\lambda}{a}},$$

and hence our average velocity is:

$$v_{\text{ave}} = \frac{1}{2}at = \sqrt{\frac{\lambda a}{2}}.$$

But thus is no good *either!*. It says that the velocity is proportional to the square root of the acceleration, and therefore that the current should be proportional to the square root of the field!

There's another twist to the story:

In practice, the charges are already moving very fast because of their thermal energy. But the thermal velocities have random directions, and average to zero. The **drift velocity** we are concerned with is a tiny extra bit. So the time between collisions is actually much shorter than we supposed. If we assume for the sake of argument that all charges travel the same distance λ between collisions, then:

$$t = \frac{\lambda}{v_{\text{thermal}}},$$

which in turn makes it to:

$$v_{\text{ave}} = \frac{1}{2}at = \frac{a\lambda}{2v_{\text{thermal}}}.$$

If there are n molecules per unit volume, and f free electrons per molecule, each with charge q and mass m, the current density is:

$$\boldsymbol{J} = nfq\boldsymbol{v}_{\text{ave}} = \frac{nfq\lambda}{2v_{\text{thermal}}}\frac{\boldsymbol{F}}{m} = \left(\frac{nf\lambda q^2}{2mv_{\text{thermal}}}\right)\boldsymbol{E}. \tag{8.3}$$

I don't claim that the term in parentheses is an accurate formula for the conductivity[5], but it does indicate the basic ingredients, and it correctly predicts that conductivity is proportional to the density of the moving charges and (ordinarily) decreases with increasing temperature.

As a result of all the collisions, the work done by the electrical force is converted into heat in the resistor. Since the work done per unit charge is V and the charge flowing per unit time is I, the power delivered is:

$$P = VI = I^2R \tag{8.4}$$

This is the **Joule heating law**. With I in amperes and R in ohms, P comes out in watts (joules per second).

8.1.1 Electromotive Force

If you think about a typical electric circuita battery hocked up to a light bulb, a perplexing question arises:

> In practice, the *current is the same all the way around the loop;* why is this the case, when the only obvious driving force is inside the battery?

Off hand, you might expect a large current in the battery and none at all in the lamp. Who's doing the pushing, in the rest of the circuit, and how does it happen that this push is exactly right to produce the same current in each segment?

What's more, given that the charges in a typical wire move (literally) at a *small's* pace, why doesn't it take half an hour for the current to reach the light bulb? How do all the charges know to start moving at the same instant?

Let's think this through.

If the current were *not* the same all the way around, then charge would be piling up somewhere.[6] for instance, during the first split second after the

> here's the crucial point–the electric field of this accumulating charge is in such a direction as to even out the flow.

Suppose, for instance, that the current *into* the bend is greater than the current *out*. Then charge piles up at the "hence," and this produces a field aiming *away* from the kink.[7] This *ellipses* the current flowing in (slowing it down) and *promotes* the current flowing out (speeding it up), until these currents are equal, at which point there is no further accumulation of charge, and equilibrium is established. It's a beautiful system, automatically self-correcting to keep the current uniform, and it does it all so quickly that, in practice, you can safely assume the current is the same all around the circuit, even in systems that oscillate at radio frequencies.

There are really two (2) forces involved in driving current around a circuit:

[7] The amount of charge involved is surprisingly small see W. G. V. Rosser, Am. J. Phys. 38, 265 (1970); nevertheless, the resulting field can be detected experimentallysee R. Jacobs, A. de Salazar, and A. Nassar, Am. J. Phys. 78, 1432 (2010).

- the *source*, $\boldsymbol{f}_s$, which is ordinarily confined to one portion of the loop (a battery, say)
- an *electrostatic* force, which serves to smooth out the flow and communicate the influence of the source to distant parts of the circuit

The would accumulate into the following expression:

$$\boldsymbol{f} = \boldsymbol{f}_s + \boldsymbol{E},$$

The physical manifestation responsible for $\boldsymbol{f}_s$ can be many different things: in a battery, it's a chemical force; in a piezoelectric crystal mechanical pressure is converted into an electrical impulse; in a thermocouple it's a temperature gradient that does the job; in a photoelectric cell it's light; and in a Van de Graaff generator the electrons are literally loaded onto a conveyor belt and swept along. Whatever the *mechanism*, its net effect is determined by the line integral of $\boldsymbol{f}$ around the circuit:

$$\mathcal{E} \equiv \oint \boldsymbol{f} \cdot d\boldsymbol{l} = \oint \boldsymbol{f}_s \cdot d\boldsymbol{l}. \tag{8.5}$$

Remember, $\oint \boldsymbol{E} \cdot d\boldsymbol{l} = 0$ for electrostatic fields, it doesn't matter whether you use $\boldsymbol{f}$ or $\boldsymbol{f}_s$.

In this context, $\mathcal{E}$ is called the **electromotive force**, or *emf*, of the circuit[8] . It is an unfortunate term, as it is not a *force* at all. It's the *integral* of a *force per unit charge*.

Within an ideal source of emf (a resistances battery[9] , for instance), the *net* force on the charges is *zero* (Eq. (**??**) with $\sigma = \infty$), so $\boldsymbol{E} = -\boldsymbol{f}_s$. The potential difference between the terminals (a and b) is therefore:

$$V = -\int_a^b \boldsymbol{E} \cdot d\boldsymbol{l} = \int_a^b \boldsymbol{f}_s \cdot d\boldsymbol{l} = \oint \boldsymbol{f}_s \cdot d\boldsymbol{l} = \mathcal{E} \tag{8.6}$$

we can extend the integral to the entire loop because $\boldsymbol{f}_r = 0$ outside the source

The function of a battery, then, is to establish and maintain a voltage difference equal to the electromotive force. The resulting electrostatic field drives current around the rest of the circuit.

notice, however, that *inside* the battery $\boldsymbol{f}_s$ drives current in the direction *opposite* to $\boldsymbol{E}$.

Because it's the line integral of $\boldsymbol{f}_r$, $\mathcal{E}$ can be interpreted as the *work done per unit charge*, by the source–indeed, in some books electromotive force is *defined* this way. However, as you'll see in the next section, there is some subtlety involved in this interpretation, so I prefer Eq. 7.9.

8.1.2 Motional EMF

In the last section, we listed several possible sources of electromotive force (emf), batteries being the most familiar. But it did not mention the commonest one of all:

the electric generator.

Generator exploit **motional emf**, which arise when you *move a line through a magnetic field*.
In the dashed region there is a uniform magnetic field $\boldsymbol{B}$, pointing **into** the page, the resistor R represents whatever it is, yet're trying to derive current through. If the entire loop is pulled to the right with speed v, the cross is seen at the reference a magnetic force whose electrical component qvB drives current around the loop, in the clockwise direction. The emf is:

$$\mathcal{E} = \oint \boldsymbol{f}_{\text{mag}} \cdot \mathrm{d}\boldsymbol{l} = vBh, \tag{8.7}$$

where h is the width of the loop[10]. Notice that the integral you perform to calculate $\mathcal{E}$ (Eq. (**??**) or Eq. (**??**)) is carried out at *one instant of time*–take a "snapshot" of the loop, if you like, and work from that. Therefore $\mathrm{d}\boldsymbol{l}$, for the segment ab, points straight up, even though the loop is moving to the right. You can't quadrel with this–it's simply the way emf is *defined*- -but it is important to be clear about it.

Horizontal segments bc and ad contribute nothing, since the force there is perpendicular to the wire

In particular, although the magnetic force is responsible for establishing the emf, it is not doing any work

> magnetic forces *never* do work.

Who, then, *is* supplying the energy that heats the resistor?

It is the person who's pulling on the loop. With the current flowing, the free charges in segment ab have a vertical velocity (cal it $\boldsymbol{u}$) in addition to the horizontal velocity $\boldsymbol{v}$ they inherit from the motion of the loop. Accordingly, the magnetic force has a component $q_{\text{t}}B$ to the left. To counteract this, the person pulling on the wire must exert a force per unit charge

$$f_{\text{pull}} = uB$$

to the *right* (Fig. 7.11). This force is transmitted to the charge by the structure of the wire.
Meanwhile, the particle is actually *moving* in the direction of the resultant velocity $\boldsymbol{w}$, and the distance it goes is $(h/\cos\theta)$. The work done per unit charge is therefore

$$\int \boldsymbol{f}_{\text{pull}} \cdot d\boldsymbol{l} = (uB)\left(\frac{h}{\cos\theta}\right)\sin\theta = vBh = \mathcal{E}$$

($\sin\theta$ coming from the dot product). As it turns out, then, the *work done per unit charge is exactly equal to the emf*, though the integrals are taken along entirely different paths (Fig. 7.12), and completely different forces are involved. To calculate the emf, you integrate around the loop *at one instant*, but to calculate the work done you follow a charge in its journey around the loop; $\boldsymbol{f}_{proj}$ contributes nothing to the emf, because it is perpendicular to the wire, whereas $\boldsymbol{f}_{mag}$ contributes nothing to work because it is perpendicular to the motion of the charge.[6]
There is a particularly nice way of expressing the emf generated in a moving loop. Let Φ be the flux of $\boldsymbol{B}$ through the loop:

$$\Phi \equiv \int \boldsymbol{B} \cdot d\boldsymbol{a}.$$

For the rectangular loop in Fig. 7.10,

$$\Phi = Bhx.$$

As the loop moves, the flux decreases:

$$\frac{d\Phi}{dt} = Bh\frac{dx}{dt} = -Bhv.$$

(The minus sign accounts for the fact that dx/dt is negative.) But this is precisely the emf (Eq. 7.11); evidently the emf generated in the loop is minus the rate of change of flux through the loop:

$$\left[\ \mathcal{E} = -\tfrac{d\Phi}{dt}.\ \right]$$

This is the **flux rule** for motional emf.

Apart from its delightful simplicity, the flux rule has the virtue of applying to *nonrectangular* loops moving in *arbitrary* directions through *nonuniform* magnetic fields; in fact, the loop need not even maintain a fixed shape.

Proof.: Figure 7.13 shows a loop of wire at time t, and also a short time dt later. Suppose we compute the flux at time t, using surface $\mathcal{S}$, and the flux at time $t+dt$, using the surface consisting of $\mathcal{S}$ plus the "ribbon" that connects the new position of the loop to the old. The *change* in flux, then, is

$$d\Phi = \Phi(t+dt) - \Phi(t) = \Phi_{ribbon} = \int_{ribbon} \boldsymbol{B}\cdot d\boldsymbol{a}.$$

Focus your attention on point P: in time dt, it moves to P'. Let $\boldsymbol{v}$ be the velocity of the *wire*, and $\boldsymbol{u}$ the velocity of a charge down the wire; $\boldsymbol{w} = \boldsymbol{v} + \boldsymbol{u}$ is the resultant velocity of a charge at P. The infinitesimal element of area on the ribbon can be written as

$$d\boldsymbol{a} = (\boldsymbol{v}\times d\boldsymbol{l})dt$$

(see inset in Fig. 7.13). Therefore

$$\frac{d\Phi}{dt} = \oint \boldsymbol{B}\cdot(\boldsymbol{v}\times d\boldsymbol{l}).$$

Since $\boldsymbol{w} = (\boldsymbol{v}+\boldsymbol{u})$ and $\boldsymbol{u}$ is parallel to $d\boldsymbol{l}$, we can just as well write this as

$$\frac{d\Phi}{dt} = \oint \boldsymbol{B}\cdot(\boldsymbol{w}\times d\boldsymbol{l}).$$

Now, the scalar triple-product can be rewritten:

$$\boldsymbol{B}\cdot(\boldsymbol{w}\times\ \mathrm{d}\boldsymbol{l}) = -(\boldsymbol{w}\times\boldsymbol{B})\cdot d\boldsymbol{l},$$

so

$$\frac{d\mathfrak{P}}{dt} = -\oint(\boldsymbol{w}\times\boldsymbol{B})\cdot d\boldsymbol{l}.$$

But $(\boldsymbol{w}\times\boldsymbol{B})$ is the magnetic force per unit charge, $\boldsymbol{f}_{mag}$, so

$$\frac{d\mathfrak{P}}{dt} = -\oint \boldsymbol{f}_{mag} \cdot \mathrm{d}\boldsymbol{l},$$

and the integral of $\boldsymbol{f}_{mag}$ is the emf:

$$\mathcal{E} = -\frac{d\mathfrak{P}}{dt}. \quad \square$$

There is a sign ambiguity in the definition of emf (Eq. 7.9): Which *way* around the loop are you supposed to integrate? There is a compensatory ambiguity in the definition of *flux* (Eq. 7.12): Which is the positive direction for $d\boldsymbol{a}$? In applying
the flux rule, sign consistency is governed (as always) by your right hand: If your fingers define the positive direction around the loop, then your thumb indicates the direction of da. Should the emf come out negative, it means the current will flow in the negative direction around the circuit.
The flux rule is a nifty short-cut for calculating motional emfs. It does not contain any new physics–just the Lorentz force law. But it can lead to error or ambiguity if you're not careful. The flux rule assumes you have a single wire loop–it can move, rotate, stretch, or distortion(continuously), but because of switches, sliding, contactes, or extended conductors allowing a variety of current paths. A standard "flux rule paradox" involves the circuit in Figure 7.14. When the switch is thrown (from a to b) the flux through the circuit doubles, but there's no motional emf (no conductor moving through a magnetic field), and the ammeter (A) records no current.
Example 7.4 (the **Faraday disk**, or **Faraday dynamo**) involves a motional emf that you can't calculate (at least, not directly) from the flux rule. The flux rule assumes the current flows along a well-defined path, whereas in this example the current spreads out over the whole disk. It's not even clear what the "flux through the circuit" would *mean* in this context.
Even more tricky is the case of **divisor currents**. Take a chunk of aluminum (say), and shake it around in a powerful magnetic field. Currents will be generated in the material, and you will feel a kind of "viscous drang"–as though you even in practice both. Findly currents are notoriously difficult to calculate," but easy on dramatic demonstrate. You may have witnessed the classic experiment in which an aluminum disk mounted as a pendulum on a horizontal axis swings down and passes between the poles of a magnet (Fig. 7.16a). When it enters the field region it suddenly slows way down. To confirm that eddy currents are responsible, one repeats the demonstration using a disk that has many slots cut in it, to prevent the flow of large-scale currents (Fig. 7.16b). This time the disk swings freely, unimpeded by the field.

8.2 Electromagnetic Induction

8.2.1 Faraday's Law

In 1831 Michael Faraday reported on a series of experiments, including three (3) that, with some simplification, can be characterized as follows:

1. He pulled a loop of wire to the right through a magnetic field (Fig. 7.21a). A current flowing in the loop.

2. He moved the *magnet to the left*, holding the loop still. Again, a current flowed in the loop.

3. With both the loop and the magnet at rest (Fig. 7.21c), he changed the *strength* of the field (the used an electromagnet, and varied the current in the coil). Once again, current flowed in the loop.

The first experiment, of course, is a straightforward case of motional emf; according to the flux rule:

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

I don't think it will surprise you to learn that exactly the same emf arises in 2. all that really matters is the *relative* motion of the magnet and the loop. Indeed, in the light of special relativity it *has* to be so. But Faraday knew nothing of relativity, and in classical electrodynamics this simple reciprocity is a remarkable coincidence.

For if the *loop* moves, it's a *magnetic* force that sets up the emf, but if the loop is *stationary*, the force *cannot* be magnetic as stationary charges experience no magnetic forces.

In that case, what *is* responsible? What sort of field exerts a force on charges at rest? Well, *electric* fields do, of course, but in this case there doesn't seem to be any electric field in sight.

Faraday had an incredible idea:

A changing magnetic field induces an electric field.

It is this induced electric field that accounts for the emf in Experiment 2. Indeed, if (as Faraday found empirically) the emf is again equal to the rate of change of the flux,

$$\mathcal{E} = \oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\frac{d\Phi}{dt},$$

then $\boldsymbol{E}$ is related to the change in $\boldsymbol{B}$ by the equation

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\int \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{a}.$$

This is **Faraday's law**, in integral form. We can convert it to differential form by applying Stokes' theorem:

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}. \tag{8.8}$$

> Note that Faraday's law reduces to the old rule $\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = 0$ (or, in differential form, $\nabla \times \boldsymbol{E} = 0$) in the static case (constant $\boldsymbol{B}$) as, of course, it should.

In Experiment 3, the magnetic field changes for entirely different reasons, but according to Faraday's law an electric field will again be induced, giving rise to an emf $-d\Phi/dt$. Indeed, one can subsume all three (3) cases (and for that matter any combination of them) into a kind of **universal flux rule**:

Whenever (and for whatever reason) the magnetic flux through a loop changes, an emf:

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

will appear in the loop.

Many people call *this* "Faraday's law." Maybe I'm overly fastiuous, but I find this confusing. There are really *two* totally different mechanisms underlying Eq. 7.17, and to identify them both as "Faraday's law" is a little like saying that because identical twins look alike we ought to call them by the same name. In Faraday's first experiment it's the Lorentz force law at work; the emf is *magnetic*. But in the other two it's an *electric* field (induced by the changing magnetic field) that does the job. Viewed in this light, it is quite astonishing that all three processes yield the same formula for the emf. In fact, it was precisely this "coincidence" that led Einstein to the special theory of relativity–he sought a deeper understanding of what is, in classical electrodynamics, a peculiar accident. But that's a story for Chapter 12. In the meantime, I shall reserve the term "Faraday's law" for electric fields induced by changing magnetic fields, and I do *not* regard Experiment 1 as an instance of Faraday's law.

Keeping track of the *signs* in Faraday's law can be a real headache. For instance, in Ex. 7.5 we would like to know which *way* around the ring the induced current flows. In principle, the right-hand rule does the job (we called Φ positive to the left, in Fig. 7.22, so the positive direction for current in the ring is counter-clockwise, as viewed from the left; since the first spike in Fig. 7.23b is *negative*, the first current pulse flows *clockwise*, and the second counterclockwise). But there's a handy rule, called **Lenz's law**, whose sole purpose is to help you get the directions right:[10]

Nature abbars a change in flux.

The induced current will flow in such a direction that the flux *it* produces tends to cancel the change. (As the front end of the magnet in Ex. 7.5 enters the ring, the flux increases, so the current in the ring must generate a field to the *right*–it therefore flows *clockwise*.) Notice that it is the *change* in flux, not the flux itself, that nature abbars (when the tail end of the magnet exists the ring, the flux *drops*, so the induced current flows *counterclockwise*, in an effort to restore it). Faraday induction is a kind of "inertial" phenomenon: A conducting loop "likes" to maintain a constant flux through it; if you try to *change* the flux, the loop responds by sending a current around in such a direction as to frustrate your efforts. (It doesn't *sticced* completely; the flux produced by the induced current is typically only a tiny fraction of the original. All Lenz's law tells you is the *direction* of the flow.)

7.2.2: **The Induced Electric Field**

Faraday's law generalizes the electrostatic rule $\nabla \times \boldsymbol{E} = 0$ to the time-dependent regime. The *divergence* of $\boldsymbol{E}$ is still given by Gauss's law ($\nabla \cdot \boldsymbol{E} = \frac{1}{\epsilon_0}\rho$). If $\boldsymbol{E}$ is a *pure* Faraday field (due exclusively to a changing $\boldsymbol{B}$, with $\rho = 0$), then

$$\nabla \cdot \boldsymbol{E} = 0, \quad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}.$$

This is mathematically identical to magnetostatics,

$$\nabla \cdot \boldsymbol{B} = 0, \quad \nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J}.$$

Conclusion: Faraday-induced electric fields are determined by $-(\partial \boldsymbol{B}/\partial t)$ in exactly the same way as magnetostatic fields are determined by $\mu_0 \boldsymbol{J}$. The analog to Biot-Savart is[13] is

$$\boldsymbol{E} = -\frac{1}{4\pi} \int \frac{(\partial \boldsymbol{B}/\partial t) \times \hat{\boldsymbol{\phi}}}{\imath^2} \, \mathrm{d}\tau = -\frac{1}{4\pi} \frac{\partial}{\partial t} \int \frac{\boldsymbol{B} \times \hat{\boldsymbol{\phi}}}{\imath^2} d\tau,$$

and if symmetry permits, we can use all the tricks associated with Ampere's law in integral form ($\oint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} = \mu_0 I_{\mathrm{enc}}$), only now it's *Faraday*'s law in integral form:

$$\oint \boldsymbol{E} \cdot d\boldsymbol{l} = -\frac{d\Phi}{dt}.$$

The rate of change of (magnetic) flux through the Amperian loop plays the role formerly assigned to $\mu_0 I_{\mathrm{enc}}$.

I must warn you, now, of a small fraud that tarinishes many applications of Faraday's law: Electromagnetic induction, of course, occurs only when the magnetic fields are *changing*, and yet we would like to use the apparatus of magnetostatics *(Ampere's law, the Biot-Savart law, and the rest) to* calculate *those magnetic fields. Technically, any result derived in this way is only approximately correct. But in practice the error is usually negligible, unless the field fluctuates extremely rapidly, or you are interested in points very far from the source. Even the case of a wire snipped by a pair of scissors (Prob. 7.18) is* static enough *for Ampere's law to apply. This regime, in which magnetostatic rules can be used to calculate t*

7.2.3 ■ Inductance

Suppose you have two loops of wire, at rest (Fig. 7.30). If you run a steady current I_1 around loop 1, it produces a magnetic field $\boldsymbol{B}_1$. Some of the field lines pass
through loop 2; let Φ_2 be the flux of $\boldsymbol{B}_1$ through 2. You might have a tough time actually *calculating* $\boldsymbol{B}_1$, but a glance at the Biot-Savart law,

$$\boldsymbol{B}_1 = \frac{\mu_0}{4\pi} I_1 \oint \frac{d\boldsymbol{l}_1 \times \hat{\boldsymbol{\imath}}}{z^2},$$

reveals one significant fact about this field: *It is proportional to the current* I_1. Therefore, so too is the flux through loop 2:

$$\Phi_2 = \int \boldsymbol{B}_1 \cdot d\boldsymbol{a}_2.$$

Thus

$$\Phi_2 = M_{21} I_1,$$

where M_{21} is the constant of proportionality; it is known as the **mutual inductance** of the two loops.

There is a cute formula for the mutual inductance, which you can derive by expressing the flux in terms of the vector potential, and invoking Stokes' theorem:

$$\Phi_2 = \int \boldsymbol{B}_1 \cdot d\boldsymbol{a}_2 = \int (\nabla \times \boldsymbol{A}_1) \cdot d\boldsymbol{a}_2 = \oint \boldsymbol{A}_1 \cdot d\boldsymbol{l}_2.$$

Now, according to Eq. 5.66,

$$\boldsymbol{A}_1 = \frac{\mu_0 I_1}{4\pi} \oint \frac{d\boldsymbol{l}_1}{\hat{z}},$$

and hence

$$\Phi_2 = \frac{\mu_0 I_1}{4\pi} \oint \left(\oint \frac{d\boldsymbol{l}_1}{\hat{z}} \right) \cdot d\boldsymbol{l}_2.$$

Evidently

$$M_{21} = \frac{\mu_0}{4\pi} \oint \oint \frac{d\boldsymbol{l}_1 \cdot d\boldsymbol{l}_2}{\hat{z}}.$$

This is the **Neumann formula**; it involves a double line integral–one integration around loop 1, the other around loop 2 (Fig. 7.31). It's not very useful for practical calculations, but it does reveal two important things about mutual inductance:

1. M_{21} is a purely geometrical quantity, having to do with the sizes, shapes, and relative positions of the two loops. 2. The integral in Eq. 7.23 is unchanged if we switch the roles of loops 1 and 2; it follows that

$$M_{21} = M_{12}.$$

(7.24) This is an astonishing conclusion: *Whatever the shapes and positions of the loops, the flux through 2 when we run a current 1 around 1 is identical to the flux through 1 when we send the same current 1 around 2.* We may as well drop the subscripts and call them both M.