Notes on

# *Electrodynamics*

Daniel F. Styer

*Notes on Electrodynamics*

Daniel F. Styer
Schiffer Professor of Physics, Oberlin College
Dan.Styer@oberlin.edu
http://www.oberlin.edu/physics/dstyer



*Cover photo:* Light scattering from early morning mist over the Vermilion River in Wolf Run Nature Preserve, Ohio.

In these notes "Griffiths" means the book David J. Griffiths, *Introduction to Electrodynamics*, fourth edition (Pearson, Boston, 2013).



You may freely download this book in pdf format from

http://www.oberlin.edu/physics/dstyer/Electrodynamics.

It is formatted to print nicely on either A4 or U.S. Letter paper. The author receives no monetary gain from your download: it is reward enough for him that you want to explore this strange and beautiful universe, our home.

# Contents


**1 Welcome** **5**

1.1 Electrostatics and Magnetostatics . . . . . . . . . . . . . . . . . . . . . . 5

1.2 Electrodynamics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

1.3 Electromagnetic energy . . . . . . . . . . . . . . . . . . . . . . . . . 8

**2 Vector Calculus** **10**

2.1 What is a vector? . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.2 Geometrical definition of divergence and curl . . . . . . . . . . . . . . 13

2.3 Pictorializing divergence and curl . . . . . . . . . . . . . . . . . . . . 19

2.4 Vector identities . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

2.5 The divergence of the curl is zero . . . . . . . . . . . . . . . . . . . . 22

**3 Conservation of Charge and the Maxwell Equations** **23**

3.1 Changing electric field makes magnetic field . . . . . . . . . . . . . . . 23

3.2 Do magnetic monopoles make magnetic field? . . . . . . . . . . . . . . 24

**4 Energy and Momentum** **26**

4.1 Conservation of energy . . . . . . . . . . . . . . . . . . . . . . . . . 26

4.2 The Maxwell stress tensor . . . . . . . . . . . . . . . . . . . . . . . . 27

**5 Electromagnetic Waves** **38**

5.1 One-dimensional waves . . . . . . . . . . . . . . . . . . . . . . . . . 38

5.2 From Maxwell equations to wave equations . . . . . . . . . . . . . . . 46

5.3 Waves in media . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

**6 Potentials and Gauges** **53**

6.1 Math . . . . . 53

6.2 Potentials . . . . . 54

6.3 Gauge freedom . . . . . 59

**7 Resultant Potentials and Fields** **63**

7.1 Single moving point charge . . . . . 65

7.2 Fields from a good swift kick . . . . . 70

7.3 Radiation simulations . . . . . 75

7.4 Radiation from an electric dipole . . . . . 75

**8 Relativistic Electrodynamics** **83**

8.1 Space and time in relativity . . . . . 83

8.2 Energy and momentum in relativity . . . . . 86

8.3 Current density in relativity . . . . . 86

8.4 Four-tensors . . . . . 87

8.5 Magnetism as a relativistic phenomenon . . . . . 89

8.6 Electromagnetic fields in relativity . . . . . 91

8.7 Maxwell equations in relativistic notation . . . . . 92

8.8 The stress-energy four-tensor . . . . . 94

8.9 What to do with relativistic electrodynamics . . . . . 97

**9 Atmospheric Optics** **98**

**A Pictorializing divergence and curl** **104**

# Chapter 1

# Welcome

You have been familiar with electrostatics since, as an infant, you rubbed a balloon on your hair and then squealed in delight when it stuck to a wall.

## 1.1 Electrostatics and Magnetostatics

| Electrostatics | Magnetostatics |
| --- | --- |
| $\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$<br>Coulomb's Law | $\vec{B} = \frac{\mu_0}{4\pi}\frac{q\vec{v}\times\hat{r}}{r^2}$<br>Biot–Savart Law |
| $\vec{\nabla}\cdot\vec{E} = \rho/\epsilon_0$<br>$\oint \vec{E}\cdot\hat{n}\,dA = Q_{\text{inside}}/\epsilon_0$<br>Gauss's Law | $\vec{\nabla}\cdot\vec{B} = 0$<br>$\oint \vec{B}\cdot\hat{n}\,dA = 0$ |
| $\vec{\nabla}\times\vec{E} = 0$<br>$\oint \vec{E}\cdot d\vec{\ell} = 0$ | $\vec{\nabla}\times\vec{B} = \mu_0\vec{J}$<br>$\oint \vec{B}\cdot d\vec{\ell} = \mu_0 I_{\text{linked}}$<br>Ampère's Law |
| $\vec{E}$ field lines<br>begin and end on charges<br>(or infinity)<br>they never loop | $\vec{B}$ field lines<br>loop around current<br>(right-hand-rule)<br>they begin or end only at infinity |
| stationary charge distributions | steady currents |

## 1.2 Electrodynamics

Electrodynamics

| | |
|---|---|
| NO! $\vec{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}$<br>Coulomb's Law | NO! $\vec{B} = \frac{\mu_0}{4\pi}\frac{q\vec{v}\times\hat{r}}{r^2}$<br>Biot–Savart Law |
| $\vec{\nabla}\cdot\vec{E} = \rho/\epsilon_0$<br>$\oint \vec{E}\cdot\hat{n}\,dA = Q_{\text{inside}}/\epsilon_0$<br>Gauss's Law | $\vec{\nabla}\cdot\vec{B} = 0$<br>$\oint \vec{B}\cdot\hat{n}\,dA = 0$ |
| $\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$<br>$\oint \vec{E}\cdot d\vec{\ell} = -\frac{d\Phi_B}{dt}$<br>Faraday's Law | $\vec{\nabla}\times\vec{B} = \mu_0\vec{J} + \mu_0\epsilon_0\frac{\partial\vec{E}}{\partial t}$<br>$\oint \vec{B}\cdot d\vec{\ell} = \mu_0 I_{\text{linked}} + \mu_0\epsilon_0\frac{d\Phi_E}{dt}$<br>Ampère's Law with Maxwell's term |
| $\vec{E}$ field lines<br>begin and end on charges<br>(or infinity)<br>or they loop around $\partial\vec{B}/\partial t$<br>(anti-right-hand rule) | $\vec{B}$ field lines<br>loop around current and $\partial\vec{E}/\partial t$<br>(right-hand-rule)<br>they begin or end only at infinity |

**Faraday** discovered his law $\oint \vec{E} \cdot d\vec{\ell} = -\frac{d\Phi_B}{dt}$ experimentally

Michael Faraday, *Experimental Researches in Electricity*, 24 November 1831:

Two hundred and three feet of copper wire in one length were coiled round a large block of wood; other two hundred and three feet of similar wire were interposed as a spiral between the turns of the first coil, and metallic contact everywhere prevented by twine. One of these helices was connected with a galvanometer, and the other with a battery of one hundred pairs of plates four inches square, with double coppers, and well charged. When the contact was made, there was a sudden and very slight effect at the galvanometer, and there was also a similar slight effect when the contact with the battery was broken. But whilst the voltaic current was continuing to pass through the one helix, no galvanometrical appearances nor any effect like induction upon the other helix could be perceived, although the active power of the battery was proved to be great, by its heating the whole of its own helix, and by the brilliancy of the discharge when made through charcoal.

---

**Maxwell** discovered his term $\oint \vec{B} \cdot d\vec{\ell} = \mu_0\epsilon_0 \frac{d\Phi_E}{dt}$ theoretically

In SI units, we think of $\frac{\mu_0}{4\pi}$ as "small", and of $\frac{1}{4\pi\epsilon_0}$ as "big", so $\mu_0\epsilon_0$ is "real small":

$$\begin{aligned}
\mu_0 &= 1.257 \times 10^{-6} \text{ H/m} \\
\epsilon_0 &= 8.854 \times 10^{-12} \text{ F/m} \\
\mu_0\epsilon_0 &= 11.13 \times 10^{-18} \text{ HF/m}^2
\end{aligned}$$

What is a henry times a farad?

$$\Phi_B = Li \qquad \Delta V = \frac{1}{C}Q \quad \text{so} \quad LC = \frac{\Phi_B}{i}\frac{Q}{\Delta V}.$$

Then, using square brackets to mean "units of" (for example the units of length are meters, $[\ell] = \text{m}$):

$$\text{HF} = [L][C] = \frac{[\Phi_B]}{[i]}\frac{[Q]}{[V]} = \frac{[B][\ell]^2}{[Q]/[t]}\frac{[Q]}{[E][\ell]}$$

But $\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$ so $[E] = [v][B]$ and the chain of reasoning continues as

$$\text{HF} = \frac{[B][\ell]^2[t]}{[Q]}\frac{[Q]}{[v][B][\ell]} = \frac{[\ell]^2[t]}{[v][\ell]} = \frac{[\ell][t]}{[v]} = \frac{[\ell][t]}{[\ell]/[t]} = [t]^2.$$

(The "electrical like" units cancel out, just as in $RC = \tau$ — an ohm times a farad is a second.) Finally

$$\mu_0\epsilon_0 = 1.113 \times 10^{-17} \text{ s}^2/\text{m}^2.$$

This effect is so small that it was unlikely ever to be discovered by fiddling with wire and twine and charcoal.

⟦Maxwell published his results in a series of papers between 1865 and 1875, the first direct experimental evidence did not come until 54 years later: R. van Cauwenberghe, “Vrification expérimentale de léquivalence électromagnétique entre les courants de déplacement de Maxwell et les courants de conduction” [“Experimental verification of electromagnetic equivalence between Maxwell displacement currents and conduction currents”] *Journal de Physique et le Radium* **10** (1929) 303–312. For later work see D.F. Bartlett and T.R. Corle, “Measuring Maxwell’s displacement current inside a capacitor” *Physical Review Letters* **55** (1985) 59–62.⟧

## 1.3 Electromagnetic energy

**Electrostatic potential energy:**

$$U_E = \frac{1}{4\pi\epsilon_0}\sum_{\text{pairs}} \frac{q_i q_j}{\mathscr{r}_{i,j}}. \tag{1.1}$$

- The sum is over *pairs* not over *particles*. Suppose the particles are 1, 2, and 3. Then there’s the potential energy of interaction of 1 and 2, the potential energy of interaction of 1 and 3, and the potential energy of interaction of 2 and 3. These sum to the total potential energy. There is no such thing as “the potential energy of particle 2”, so of course you can’t sum over the potential energy of 1, of 2, and of 3, because these things don’t exist.
- This potential energy can be positive or negative.
- The derivation starts with the three particles infinitely far away (zero potential energy by convention). Calculate the work done by the electric force while bringing in particle 1. (This is of course zero.). Now calculate the work done by the electric force while bringing in particle 2. (This is the negative of the potential energy of interaction of 1 and 2.) Finally calculate the work done by the electric force while bringing in particle 3. (This is the negative of the potential energy of interaction of 1 and 3 and of 2 and 3.) It is half-way to a miracle that the final result depends not on the sequence by which particles brought in, not on the route taken by each particle, not on speed at which each particle was brought in along that route, but only on the final positions.

**Magnetostatic potential energy:**

$$U_B = \frac{\mu_0}{4\pi} \sum_{\text{pairs}} \frac{q_i q_j \vec{v}_i \cdot \vec{v}_j}{\mathscr{r}_{i,j}}. \tag{1.2}$$

- The magnetic field does no work. How can there be "magnetostatic potential energy"?
- The magnetostatic potential energy is not "negative the work done by the magnetic field while bringing in particle 2". (This quantity is zero.) Instead it is the "negative of the work done by the electric field caused by changing magnetic field while bringing in particle 2". Magnetostatic potential energy exists because of Faraday's Law, a dynamical result!
- It is three-quarters-way to a miracle that the final result depends not on the route with which each particle comes in, nor on the speed taken by each particle as it comes in along its route, but only upon the final positions and velocities.

**Electrodynamic potential energy:**

$$U_{ED} = \int \left( \frac{\epsilon_0}{2} E^2(\vec{r}) + \frac{1}{2\mu_0} B^2(\vec{r}) \right) d^3r \tag{1.3}$$

- This is always positive, so obviously there is a change in sea-level involved in deriving this result.
- Equations (1.1) and (1.2) are wrong in electrodynamics.

# Chapter 2

# Vector Calculus

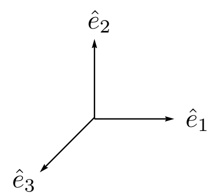
## 2.1 What is a vector?

A list of three numbers is a triplet or a 3-tuple; it is not necessarily a three dimensional vector. For example, the thermodynamic list $(p, V, T)$ is not a vector. Again, the list

$$(\text{population of Spain}, \text{cost of a cup of coffee in Caracas}, \text{height of Kilimanjaro})$$

is not a vector.

Here is an orthonormal basis:




(Basis vectors don't *have* to be orthogonal, they don't even have to be unit vectors, but that's convenient and traditional.)

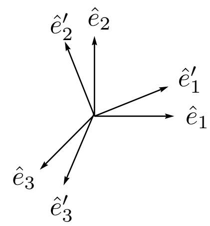
$$\begin{aligned}
\vec{r} &= (r_1, r_2, r_3) && \text{No!} \\
\vec{r} &= r_1\,\hat{e}_1 + r_2\,\hat{e}_2 + r_3\,\hat{e}_3 && \text{Yes!} \\
\vec{r} &\doteq (r_1, r_2, r_3) && \text{Yes!}
\end{aligned}$$

The last equation is read "The vector $\vec{r}$ is represented by the three numbers $(r_1, r_2, r_3)$ in the basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$. or "The name of vector $\vec{r}$ in the basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ is $(r_1, r_2, r_3)$."

The name of an object depends upon both the object and the language used: The name of a horse in the English language is "horse"; the name of a horse in the

German language is “pferd”; the name of a horse in the Swahili language is “farasi”. The horse doesn’t change if you use one language rather than another, only its name changes. Similarly the vector doesn’t change if you represent it in one basis rather than another, only its name (or “representation”) changes.

Here are two orthonormal bases:




Vector $\vec{r}$ can be written in terms of this basis as

$$\begin{aligned}\vec{r} &= r_1\,\hat{e}_1 + r_2\,\hat{e}_2 + r_3\,\hat{e}_3 \\ &= r_{1'}\,\hat{e}'_1 + r_{2'}\,\hat{e}'_2 + r_{3'}\,\hat{e}'_3 \qquad \text{(best notation)} \\ &= r'_1\,\hat{e}'_1 + r'_2\,\hat{e}'_2 + r'_3\,\hat{e}'_3 \qquad \text{(most practical notation)}\end{aligned}$$

How do we translate from one “language” (basis) to another? What is our “dictionary”?

$$\begin{aligned} r'_1 &= \vec{r}\cdot\hat{e}'_1 \\ &= (r_1\,\hat{e}_1 + r_2\,\hat{e}_2 + r_3\,\hat{e}_3)\cdot\hat{e}'_1 \\ &= r_1\,\hat{e}_1\cdot\hat{e}'_1 + r_2\,\hat{e}_2\cdot\hat{e}'_1 + r_3\,\hat{e}_3\cdot\hat{e}'_1\end{aligned}$$

or for all three components

$$\begin{pmatrix} r'_1 \\ r'_2 \\ r'_3 \end{pmatrix} = \begin{pmatrix} \hat{e}_1\cdot\hat{e}'_1 & \hat{e}_2\cdot\hat{e}'_1 & \hat{e}_3\cdot\hat{e}'_1 \\ \hat{e}_1\cdot\hat{e}'_2 & \hat{e}_2\cdot\hat{e}'_2 & \hat{e}_3\cdot\hat{e}'_2 \\ \hat{e}_1\cdot\hat{e}'_3 & \hat{e}_2\cdot\hat{e}'_3 & \hat{e}_3\cdot\hat{e}'_3 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}.$$

In other words

$$r'_i = \sum_{j=1}^{3} R_{ij} r_j \qquad \text{where} \qquad R_{ij} = \hat{e}_j\cdot\hat{e}'_i.$$

The definition of *vector* is “an entity with components that transform from one basis to another in this way.” (This is the definition produced by Gregorio Ricci-Curbastro and Tullio Levi-Civita in the year 1900. Other definitions of “vector” exist, including the 1888 “vector space” definition of Giuseppe Peano, which was generalized by David Hilbert and Erhard Schmidt in 1908.)

**How can we use this defintion?**

We can have scalar functions of vectors, $f(\vec{r})$, and vector functions of vectors, $\vec{F}(\vec{r})$.

You know how to take and interpret the gradient $\vec{\nabla} f(\vec{r})$, with components ("name")

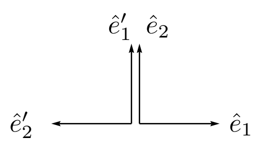
$$\vec{\nabla} f(\vec{r}) \doteq \left( \frac{\partial f}{\partial r_1}, \frac{\partial f}{\partial r_2}, \frac{\partial f}{\partial r_3} \right).$$

It points in the direction of fastest increase of $f(\vec{r})$, and its magnitude is the slope in that direction.

There is an obvious way of taking the derivative of a vector function, too. It is

$$\left( \frac{\partial F_1}{\partial r_1}, \frac{\partial F_2}{\partial r_2}, \frac{\partial F_3}{\partial r_3} \right).$$

Here are two orthonormal bases related by a 90° rotation in the $\hat{e}_1$-$\hat{e}_2$ plane (I don't show the third dimension because it doesn't change ($\hat{e}'_3 = \hat{e}_3$) and because it's hard to draw.)




If vector $\vec{r}$ is named $(r_1, r_2, r_3) = (x, y, z)$ in the first basis, it is named $(r'_1, r'_2, r'_3) = (y, -x, z)$ in the second basis.

If vector $\vec{F}$ is named $(F_1, F_2, F_3) = (F_x, F_y, F_z)$ in the first basis, it is named $(F'_1, F'_2, F'_3) = (F_y, -F_x, F_z)$ in the second basis.

In the first basis, the combination of symbols is

$$\left( \frac{\partial F_1}{\partial r_1}, \frac{\partial F_2}{\partial r_2}, \frac{\partial F_3}{\partial r_3} \right) = \left( \frac{\partial F_x}{\partial x}, \frac{\partial F_y}{\partial y}, \frac{\partial F_z}{\partial z} \right) \equiv (a, b, c).$$

In the second basis, the combination of symbols is

$$\left( \frac{\partial F'_1}{\partial r'_1}, \frac{\partial F'_2}{\partial r'_2}, \frac{\partial F'_3}{\partial r'_3} \right) = \left( \frac{\partial F_y}{\partial y}, \frac{\partial(-F_x)}{\partial(-x)}, \frac{\partial F_z}{\partial z} \right) = (b, a, c).$$

The components of a vector would transform from $(a, b, c)$ to $(b, -a, c)$,

This combination of symbols is not a vector, it is an impostor! Like the triplet $(p, V, T)$, they are three numbers that do not constitute a vector. Now that we know how *not* to take a vector derivative, we investigate how we *should* do it.

*Challenge:* Show that the transformation matrix for the change of basis above is

$$R_{ij} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

## 2.2 Geometrical definition of divergence and curl

### Derivative of a single-variable function

The derivative of function $f(x)$ at point $x_0$ is given by

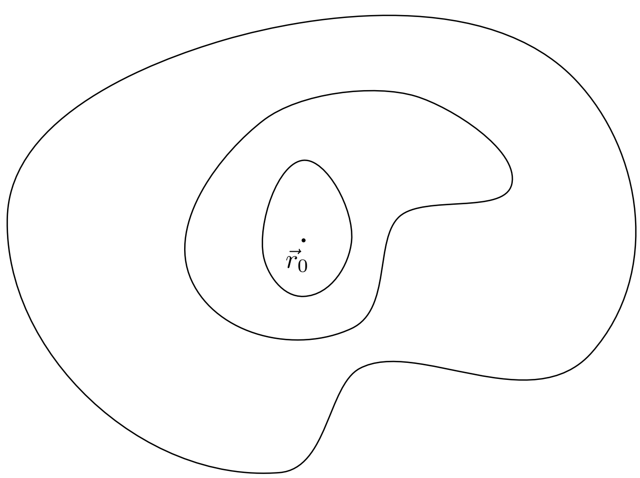
$$\lim_{L\to 0}\frac{f(x_0+L/2)-f(x_0-L/2)}{L}$$

that is, in involves the function values at the edges of an interval, divided by the magnitude of that interval.

If you're a mathematician, you'll want to prove that the limit exists, and that it is the same whether $L$ approaches zero from above (through positive numbers) or from below (through negative numbers). Physicists usually just skip over such questions, interesting though they may be. Instead we just note that the result has the correct dimensions for a slope, and that it leads to the indeterminate form 0/0.

### Divergence

So how should we define the derivative of a vector function $\vec{F}(\vec{r})$ at point $\vec{r}_0$? Here's one way. Consider a sequence of volumes $\mathcal{V}$ that enclose point $\vec{r}_0$, but that grow smaller and smaller.



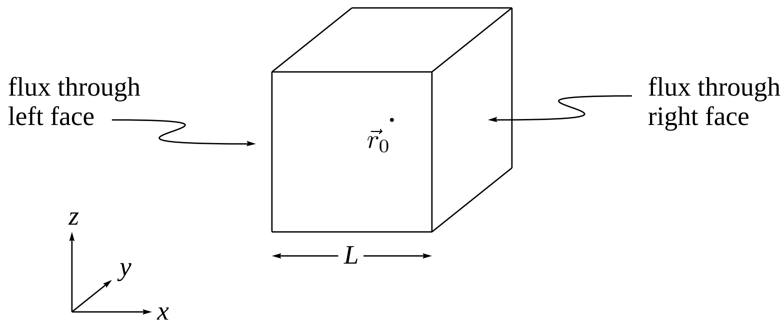

*A sequence of volumes (shown in cross-section) homing in on point $\vec{r}_0$.*

If the volume $\mathcal{V}$ is enclosed by surface $\mathcal{S}$, then

$$\lim\frac{\displaystyle\int_{\mathcal{S}\text{ of }\mathcal{V}}\vec{F}(\vec{r})\cdot\hat{n}\,dA}{\text{volume of }\mathcal{V}} \tag{2.1}$$

fits the requirements for some sort of derivative: it involves the function values at the edge of the volume divided by the magnitude of that volume, it has the correct dimensions, and it leads to the indeterminate form 0/0. Mathematicians will want to prove that the limit exists, and that it gives the same result regardless of what sequence of volumes (cubes, spheres, hemispheres, cats, etc.) is used to close in on $\vec{r}_0$. But we'll skip over such general questions and ask:

What is the result if the sequence of volumes consists of cubes centered on $\vec{r}_0$?




It's clear from the definition of flux that for a small cube

$$\begin{aligned} \text{flux through right face} &\approx F_x(\text{evaluated at center of right face})L^2 \\ &= F_x(x_0 + L/2, y_0, z_0)L^2 \end{aligned}$$

and that this approximation grows better and better as $L$ grows smaller and smaller. Similarly

$$\begin{aligned} \text{flux through left face} &\approx -F_x(\text{evaluated at center of left face})L^2 \\ &= -F_x(x_0 - L/2, y_0, z_0)L^2. \end{aligned}$$

Thus

$$\begin{aligned} \text{flux through right plus left faces} &\approx [F_x(x_0 + L/2, y_0, z_0) - F_x(x_0 - L/2, y_0, z_0)]L^2 \\ &= \left[\frac{F_x(x_0 + L/2, y_0, z_0) - F_x(x_0 - L/2, y_0, z_0)}{L}\right] L^3 \\ &\to \left[\frac{\partial F_x}{\partial x}(x_0, y_0, z_0)\right] L^3 \end{aligned}$$

where the symbol $\to$ means "in the limit as $L \to 0$".

Parallel reasoning shows that the flux through the back plus front faces is

$$\left[\frac{\partial F_y}{\partial y}(\vec{r}_0)\right] L^3$$

while the flux through the top plus bottom faces is

$$\left[\frac{\partial F_z}{\partial z}(\vec{r}_0)\right] L^3.$$

Finally, the limit presented in definition (2.1) results in

$$\frac{\partial F_x}{\partial x}(\vec{r}_0) + \frac{\partial F_y}{\partial y}(\vec{r}_0) + \frac{\partial F_z}{\partial z}(\vec{r}_0). \tag{2.2}$$

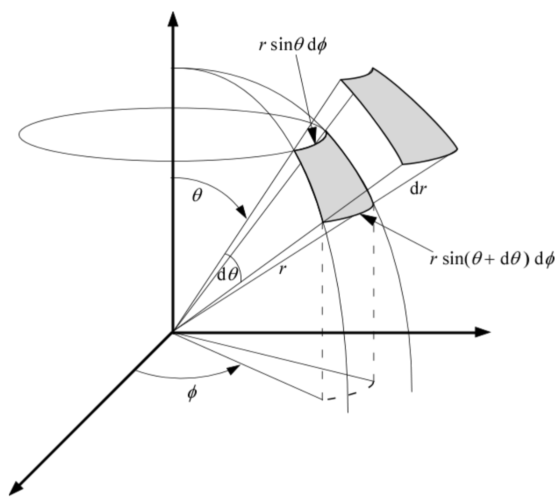
Because this derivative is the "flux per volume at a point" we call it the "divergence at a point".

Some people like to begin with equation (2.2) and call this the definition of divergence. Then they have a difficult time proving the *divergence theorem* (or Gauss's theorem), namely that if volume $\mathcal{V}$ is enclosed by surface $\mathcal{S}$, then

$$\int_{\mathcal{V}} \vec{\nabla} \cdot \vec{F}(\vec{r})\, d^3r = \int_{\mathcal{S}\text{ of }\mathcal{V}} \vec{F}(\vec{r}) \cdot \hat{n}\, dA.$$

I prefer to begin with the geometrical definition (2.1), and derive expression (2.2) for the divergence in Cartesian coordinates. In this approach, the divergence theorem just pops right out of the definition.

You could do the "flux through a shrinking volume" argument for shapes other than cubes. If you do it for the shape below




you will find the expression for divergence in spherical coordinates. If you do it for other shapes you will find the expression for divergence in cylindrical coordinates, or prolate spheroidal coordinates, or confocal paraboloidal coordinates, or any other kind of coordinates. The idea of "flux per volume at a point" is the same for all coordinate systems – the shape of the shrinking volume is different for different coordinate systems.

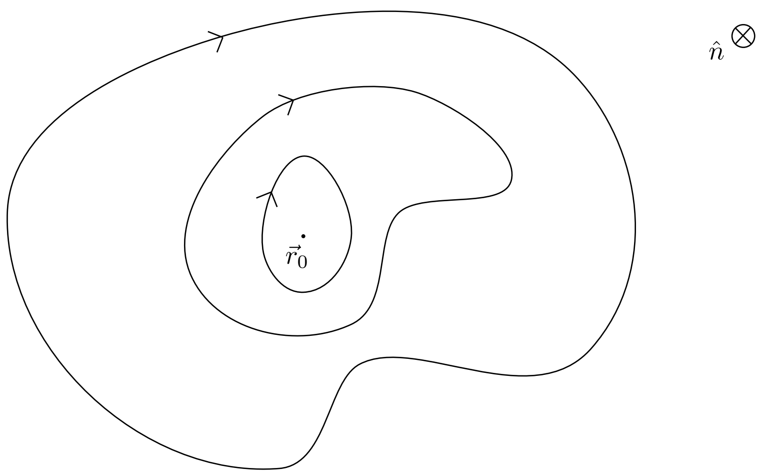
**Summary:** You know that the total mass $M$ of an object can be found by integrating the mass density $\rho(\vec{r})$ over the volume of the object:

$$M = \int_{\mathcal{V}} \rho(\vec{r})\, d^3r.$$

The divergence plays the role of "flux density" rather than mass density.

## Curl

Expression (2.1) is not the only possible derivative of a vector function. Consider a sequence of loops $\mathcal{L}$ (all within the plane perpendicular to a unit vector $\hat{n}$) that enclose point $\vec{r}_0$, but that grow smaller and smaller.



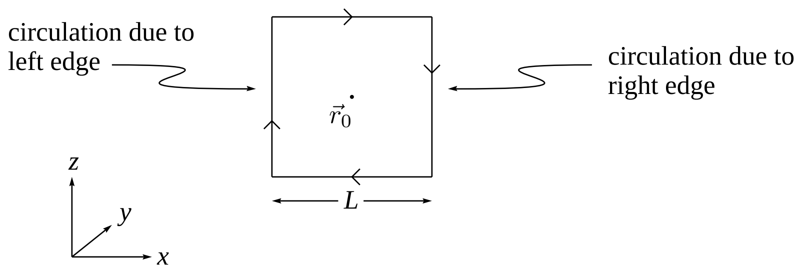

*A sequence of loops (all within the plane perpendicular to $\hat{n}$) homing in on point $\vec{r}_0$. (Direction of loop given through right-hand rule.)*

The line integral of $\vec{F}(\vec{r})$ along loop $\mathcal{L}$ is called the "circulation of $\vec{F}(\vec{r})$ along $\mathcal{L}$." If the loop $\mathcal{L}$ embraces a surface $\mathcal{S}$, then

$$\lim \frac{\displaystyle\int_{\mathcal{L}\text{ of }\mathcal{S}} \vec{F}(\vec{r}) \cdot d\vec{\ell}}{\text{area of }\mathcal{S}} \tag{2.3}$$

also fits the requirements for some sort of derivative: it involves the function values at the edge of the surface divided by the magnitude of that surface, it has the correct dimensions, and it leads to the indeterminate form 0/0. Mathematicians will want to prove that the limit exists, and that it gives the same result regardless of what sequence of shapes (squares, circles, squirrels, etc.) is used to close in on $\vec{r}_0$. But we'll skip over such general questions and ask:

What is the result if the sequence of loops consists of squares in the $x$-$z$ plane, centered on $\vec{r}_0$?




It's clear from the definition of circulation that for a small square

$$\begin{aligned}\text{circulation due to right edge} &\approx -F_z(\text{evaluated at center of right edge})L \\ &= -F_z(x_0 + L/2, y_0, z_0)L\end{aligned}$$

and that this approximation grows better and better as $L$ grows smaller and smaller. Similarly

$$\begin{aligned}\text{circulation due to left edge} &\approx F_z(\text{evaluated at center of left edge})L \\ &= F_z(x_0 - L/2, y_0, z_0)L.\end{aligned}$$

Thus

$$\begin{aligned}\text{circulation due to right plus left edges} &\approx -[F_z(x_0 + L/2, y_0, z_0) - F_z(x_0 - L/2, y_0, z_0)]L \\ &= -\left[\frac{F_z(x_0 + L/2, y_0, z_0) - F_z(x_0 - L/2, y_0, z_0)}{L}\right] L^2 \\ &\to -\left[\frac{\partial F_z}{\partial x}(x_0, y_0, z_0)\right] L^2\end{aligned}$$

where the symbol $\to$ means "in the limit as $L \to 0$". Parallel reasoning shows that the circulation due to the top plus bottom faces is

$$\left[\frac{\partial F_x}{\partial z}(\vec{r}_0)\right] L^2.$$

Finally, the limit presented in definition (2.3) results in

$$\frac{\partial F_x}{\partial z}(\vec{r}_0) - \frac{\partial F_z}{\partial x}(\vec{r}_0). \tag{2.4}$$

You can carry out this "limit of circulation per area" for an infinite number of planes passing through point $\vec{r}_0$. Find the plane with the largest "limit of circulation per area". The curl of $\vec{F}(\vec{r})$ at $\vec{r}_0$ is a vector with direction perpendicular to this plane and magnitude equal to that maximum.

This process finds the curl, but it requires finding the limit for an infinite number of planes! It is equivalent, and far more practical, to carry out the process for three planes only. The process resulting in equation (2.4) finds the "circulation in the plane perpendicular to $y$ per area at a point" which also equals the "$y$-component of curl at a point".

Parallel considerations for planes perpendicular to $x$ and to $z$ result in the traditional expression for the curl in Cartesian coordinates, namely

$$\vec{\nabla} \times \vec{F}(\vec{r}_0) \doteq \left[ \frac{\partial F_z}{\partial y}(\vec{r}_0) - \frac{\partial F_y}{\partial z}(\vec{r}_0), \frac{\partial F_x}{\partial z}(\vec{r}_0) - \frac{\partial F_z}{\partial x}(\vec{r}_0), \frac{\partial F_y}{\partial x}(\vec{r}_0) - \frac{\partial F_x}{\partial y}(\vec{r}_0) \right]. \quad (2.5)$$

Some people like to begin with equation (2.5) and call it the definition of curl. Then they have a difficult time proving the *circulation theorem* (or Stokes's theorem): If the surface $\mathcal{S}$ is bounded by loop $\mathcal{L}$, then

$$\int_{\mathcal{S}} (\vec{\nabla} \times \vec{F}(\vec{r})) \cdot \hat{n}\, dA = \int_{\mathcal{L} \text{ of } \mathcal{S}} \vec{F}(\vec{r}) \cdot d\vec{\ell}. \quad (2.6)$$

I prefer to begin with the geometrical definition

$$(\vec{\nabla} \times \vec{F}(\vec{r}_0)) \cdot \hat{n} = \lim \frac{\displaystyle\int_{\mathcal{L} \text{ of } \mathcal{S}} \vec{F}(\vec{r}) \cdot d\vec{\ell}}{\text{area of } \mathcal{S}}, \quad (2.7)$$

and derive expression (2.5) for the Cartesian coordinates of the curl. In this approach, the circulation theorem just pops right out of the definition.

**Acknowledgment:** In my multivariate calculus course, I learned the "Cartesian coordinate" definitions of divergence and curl, and these definitions left a bad taste in my mouth. Why were divergence and curl – particularly curl – defined through such bazaar combinations of derivatives? Math is supposed to be coordinate-independent: Why were Cartesian coordinates so special? A one-variable derivative has geometrical significance as a slope – what was the geometrical significance of divergence and curl? A related question – why do divergence and curl have such strange names?

I learned the much-more-satisfactory geometric approach to vector derivatives – the one outlined in this document – from my physics professor Mark Heald. When I asked him how he had learned it, he told me it was the approach used by his teacher, Carl Howe, when he took undergraduate electricity and magnetism at Oberlin College.

## 2.3 Pictorializing divergence and curl

The picture below [from Edward M. Purcell, *Electricity and Magnetism* (McGraw-Hill, 1965) page 72] helped develop my sense of the geometric meaning of divergence and curl — perhaps it will help you, too. Spend some time puzzling over these figures before turning to the answers at the back of this book.

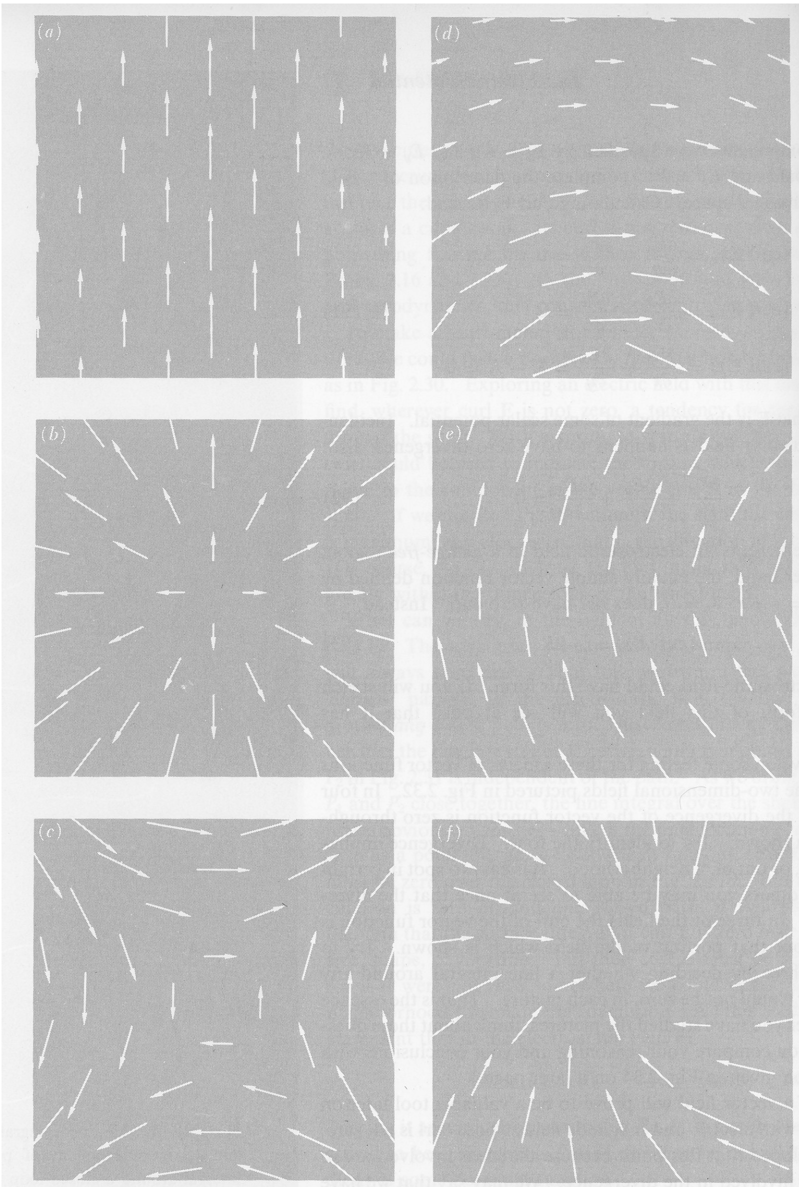



*Fig. 2.32* Four of these vector fields have zero divergence in the region shown. Three have zero curl. Can you spot them?

## 2.4 Vector identities

I am convinced that every vector calculus identity is not just a jumble of symbols: it is a statement about geometry.

For example, the Laplacian is often defined, in Cartesian coordinates, as

$$\vec{\nabla}^2 f(x,y,z) = \vec{\nabla}\cdot\vec{\nabla} f(x,y,z) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

What makes this particular combination of partial derivatives so special? Why do we so often encounter this combination and so rarely encounter, say, the combination

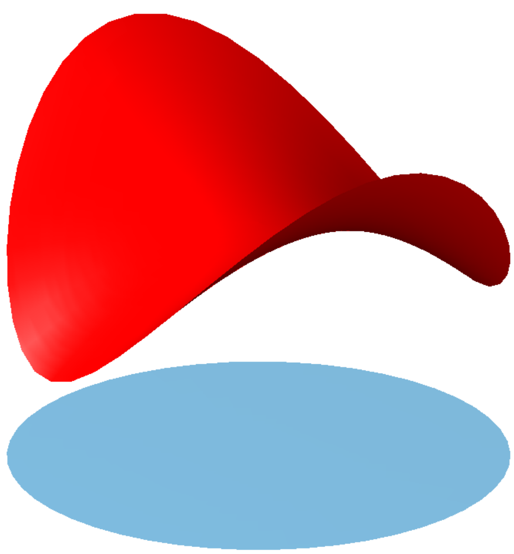
$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} \quad ?$$

I wrote a paper about this question ["The geometrical significance of the Laplacian" *American Journal of Physics*, **83** (12) 992–997 (December 2015)] in which I show that an equivalent definition of the Laplacian in dimension $d$ is

$$\vec{\nabla}^2 f(\vec{r}_0) = \lim_{R\to 0}\left\{\frac{2d}{R^2}\left[\langle f\rangle_{\text{shell}} - f(\vec{r}_0)\right]\right\}.$$

James Clerk Maxwell used a property like this to motivate the name "concentration" for $-\vec{\nabla}^2$.

If $\vec{\nabla}^2 f(\vec{r}) = 0$ in a region, then within that region there are no maxima, no minima. An example is this two dimensional function



Such a function could be used as a roof to shed rain. And if it were turned upside-down it would still shed rain!

I haven't yet figured out the geometric significance of some vector identities, for example

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \vec{\nabla} \cdot (\vec{\nabla}\vec{A}).$$

Others are subtle. For example...

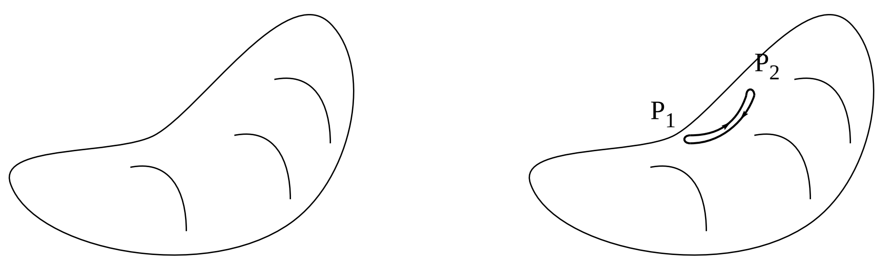
## 2.5 The divergence of the curl is zero

[Approach from Edward M. Purcell, *Electricity and Magnetism* (McGraw-Hill, 1965) problem 2.15.]

If $\vec{A}(\vec{r})$ is a vector field with continuous derivatives, then

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}(\vec{r})) = 0.$$

How to prove? You could plug-and-chug in Cartesian coordinates. But it's easier and more insightful to do it this way.




Consider (figure on the left) the volume $\mathcal{V}$ enclosed by surface $\mathcal{S}$. Apply the divergence theorem to function $\vec{F}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r})$, giving

$$\int_{\mathcal{V}} \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}(\vec{r}))\, d^3r = \int_{\mathcal{S} \text{ of } \mathcal{V}} (\vec{\nabla} \times \vec{A}(\vec{r})) \cdot \hat{n}\, dA.$$

Now slice out and remove from the surface a tiny sliver (figure on the right). Technically we've altered $\mathcal{S}$, but this tiny alteration will not affect the value of the surface integral. The edge $\mathcal{E}$ of the altered $\mathcal{S}$ is the edge of the sliver. Apply the circulation theorem to $\vec{F}(\vec{r}) = \vec{A}(\vec{r})$, giving

$$\int_{\mathcal{S}} (\vec{\nabla} \times \vec{A}(\vec{r})) \cdot \hat{n}\, dA = \int_{\mathcal{E} \text{ of } \mathcal{S}} \vec{A}(\vec{r}) \cdot d\vec{\ell}.$$

But

$$\int_{\mathcal{E} \text{ of } \mathcal{S}} \vec{A} \cdot d\vec{\ell} = \int_{P_1 \text{ to } P_2} \vec{A} \cdot d\vec{\ell} + \int_{P_2 \text{ to } P_1} \vec{A} \cdot d\vec{\ell} = \int_{P_1 \text{ to } P_2} \vec{A} \cdot d\vec{\ell} - \int_{P_1 \text{ to } P_2} \vec{A} \cdot d\vec{\ell} = 0.$$

So for any volume $\mathcal{V}$,

$$\int_{\mathcal{V}} \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}(\vec{r}))\, d^3r = 0.$$

Because this holds for *any* volume,

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}(\vec{r})) = 0.$$

(This is an example of the theorem that "the boundary of a boundary is zero," as emphasized by C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation*, box 15.1, pages 365–371.)

# Chapter 3

# Conservation of Charge and the Maxwell Equations

Local conservation of charge (charge goes into or out of region through boundary; doesn't disappear inside, simultaneously reappear outside):

$$\begin{aligned}\oint \vec{J}(\vec{r},t)\cdot\hat{n}\,dA &= -\frac{d}{dt}Q_{\text{inside}}(t)\\ \int \vec{\nabla}\cdot\vec{J}(\vec{r},t)\,d^3r &= -\frac{d}{dt}\int \rho(\vec{r},t)\,d^3r\\ \vec{\nabla}\cdot\vec{J}(\vec{r},t) &= -\frac{\partial\rho(\vec{r},t)}{\partial t}\end{aligned} \tag{3.1}$$

the "continuity equation"... equivalent to local conservation of charge.

## 3.1 Changing electric field makes magnetic field

The "pre-Maxwell equations"

$$\begin{aligned}&\vec{\nabla}\cdot\vec{E} = \rho/\epsilon_0 && \vec{\nabla}\cdot\vec{B} = 0\\ &\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} && \vec{\nabla}\times\vec{B} = \mu_0\vec{J}\end{aligned}$$

Divergence of Faraday's law

$$\vec{\nabla}\cdot(\vec{\nabla}\times\vec{E}) = -\frac{\partial}{\partial t}(\vec{\nabla}\cdot\vec{B}) = 0$$

Okay!

Divergence of Ampere's law

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{B}) = \mu_0 \vec{\nabla} \cdot \vec{J} = -\mu_0 \frac{\partial \rho}{\partial t} \neq 0$$

Not okay!

Maxwell asks: How to fix?

$$\vec{\nabla} \cdot \vec{J} = -\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial t}(\epsilon_0 \vec{\nabla} \cdot \vec{E}) = \vec{\nabla} \cdot \left( -\epsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$

So

$$\vec{\nabla} \cdot \vec{J} \neq 0$$

but

$$\vec{\nabla} \cdot \left( \vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right) = 0$$

Replace $\vec{J}$ in Ampere's law with

$$\vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t}.$$

Now

$$\vec{\nabla} \times \vec{B} = \mu_0 \left( \vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right).$$

The Maxwell term!

So now the Maxwell equations are

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \rho/\epsilon_0 & \qquad \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} & \qquad \vec{\nabla} \times \vec{B} &= \mu_0 \left( \vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \end{aligned}$$

## 3.2 Do magnetic monopoles make magnetic field?

Can we add magnetic monopoles? First shot:

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \rho_e/\epsilon_0 & \qquad \vec{\nabla} \cdot \vec{B} &= \mu_0 \rho_m \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} & \qquad \vec{\nabla} \times \vec{B} &= \mu_0 \left( \vec{J}_e + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \end{aligned}$$

with

$$\vec{\nabla} \cdot \vec{J}_m = -\frac{\partial \rho_m}{\partial t}.$$

Can't stop here!

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{E}) = -\frac{\partial}{\partial t}(\vec{\nabla} \cdot \vec{B}) = -\frac{\partial}{\partial t}(\mu_0 \rho_m) = \mu_0 \vec{\nabla} \cdot \vec{J}_m$$

Fix in the same way:

$$\vec{\nabla} \cdot \vec{J}_m = -\frac{\partial \rho_m}{\partial t} = -\frac{\partial}{\partial t}((1/\mu_0)\vec{\nabla} \cdot \vec{B}) = \vec{\nabla} \cdot \left( -(1/\mu_0)\frac{\partial \vec{B}}{\partial t} \right)$$

So

$$\vec{\nabla} \cdot \left( \mu_0 \vec{J}_m + \frac{\partial \vec{B}}{\partial t} \right) = 0$$

Replace $\partial \vec{B}/\partial t$ in Faraday's law with

$$\mu_0 \vec{J}_m + \frac{\partial \vec{B}}{\partial t}.$$

Now the revised Maxwell equations are

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \rho_e/\epsilon_0 & \qquad \vec{\nabla} \cdot \vec{B} &= \mu_0 \rho_m \\ \vec{\nabla} \times \vec{E} &= -\mu_0 \vec{J}_m - \frac{\partial \vec{B}}{\partial t} & \qquad \vec{\nabla} \times \vec{B} &= \mu_0 \vec{J}_e + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned}$$

I love the symmetry of these equations: the sources of electric field are electric charge, magnetic current, and changing magnetic field; the sources of magnetic field are magnetic charge, electric current, and changing electric field. However, nature seems not to have taken advantage of this possibility.

Whenever people access strange new places (bore holes under the sea, the surface of the moon) they look for magnetic monopoles. One was found in the strangest place of all: California. The laboratory of Blas Cabrera at Stanford University found evidence for a magnetic monopole on the night of 14 February 1982 (the "Valentine's Day Monopole"). Since then, nothing. It was probably a fluke, but who knows...

# Chapter 4

# Energy and Momentum

## 4.1 Conservation of energy

In section 8.1.2 Griffiths considers a collection of moving charged particles some region $\mathcal{V}$, with surface $\mathcal{S}$. No particles moving through the surface, and there are no interactions other than electromagnetic interactions. Let KE represent the kinetic energy of the moving charged particles. At equation (8.9), Griffths proves Poynting's theorem, that

$$\frac{d(\mathrm{KE})}{dt} + \frac{d}{dt}\int_{\mathcal{V}} \left(\frac{\epsilon_0}{2}\vec{E} + \frac{1}{2\mu_0}\vec{B}\right) d^3r + \frac{1}{\mu_0}\oint_{\mathcal{S}} (\vec{E} \times \vec{B}) \cdot \hat{n}\, dA = 0.$$

The integral over volume had previously been interpreted as the electromagnetic potential energy localized within this region (Griffiths equations (2.45) and (7.35)). But that was before we had the Maxwell term. Is this interpretation still legitimate? Yes, because the Maxwell term generates magnetic field, and the magnetic field does no work.

Comparing the above to the charge conservation equations

$$\frac{d}{dt} Q_{\text{inside}} + \oint_{\mathcal{S}} \vec{J} \cdot \hat{n}\, dA = 0$$

and

$$\frac{d}{dt}\int_{\mathcal{V}} \rho\, d^3r + \oint_{\mathcal{S}} \vec{J} \cdot \hat{n}\, dA = 0$$

makes it clear that Poynting's theorem talks about the conservation of energy. The term $(1/\mu_0)(\vec{E} \times \vec{B})$ is the current density, not of charge, but of energy. Sure enough it has the units of watts/m$^2$.

$$\vec{S} = \frac{1}{\mu_0}(\vec{E} \times \vec{B}) = \text{Poynting vector} = \text{current density of EM energy}$$

Comments:

- Energy travels from a battery into a resistor, not through the wires, but through the empty space between wires. [Griffiths example problem 8.1.]
- Circuit simulation program CircuitSurveyor, by Noah Morris. [Noah A. Morris and Daniel F. Styer, "Visualizing Poynting vector energy flow in electric circuits" *American Journal of Physics* **80** (6) June 2012, pages 552–554.]
- If fields carry energy, it's a good bet they carry momentum as well. Griffiths equation (8.29) shows that the momentum density is
$$\vec{g} = \frac{1}{c^2}\vec{S}.$$
- Feynman disk and angular momentum conservation. [Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics*, volume II (Addison-Wesley, Reading, MA, 1964) pages 17-6 and 27-11.]
- Thomson's dipole. [Griffiths problem 8.19.]

## 4.2 The Maxwell stress tensor

Fields contain energy density that moves around, and that energy flow is described through the Poynting vector $\vec{S}$. Fields also contain momentum density that moves around. How is that momentum flow described?

We begin with the familiar: the flow of charge density, which is described through current density — a vector. After this review, we generalize to the flow of momentum density, which (we will find) is described through the Maxwell stress tensor.

## Charge transport

Suppose charge flows uniformly in space and time, like a broad steady wind.


A plug of length $v\Delta t$, cross-sectional area $A_\perp$, passes through the imaginary plane perpendicular to the "wind". This plug contains charge

$$Q = \rho|\vec{v}|\Delta t A_\perp,$$

so the current density, namely

$$\frac{\text{charge/time}}{\text{cross-sectional area}}$$

has direction parallel to $\vec{v}$ and magnitude

$$|\vec{J}| = \frac{Q/\Delta t}{A_\perp} = \frac{\rho|\vec{v}|\Delta t A_\perp/\Delta t}{A_\perp} = \rho|\vec{v}|,$$

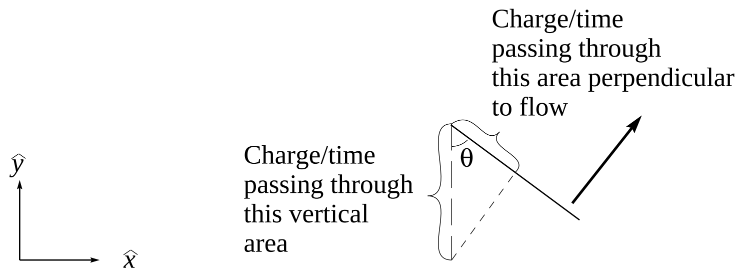
whence

$$\vec{J} = \rho\vec{v}.$$

Suppose you had a current detector shaped like a tennis racquet, which measured the current (charge per time) passing through the webbing of the racquet. If the webbing were oriented parallel to the charge flow, then the detector would read zero current. You you changed its orientation to the plane facing the flow, then the detector reading would be a maximum. This is how you could measure the current density: twist your racquet detector this way and that until its reading is a maximum, then the current density vector is oriented perpendicular to the webbing, and has magnitude of the current reading divided by the area of the webbing.

While this technique works, it's inefficient. You have to take many readings, homing in on a maximum. You have to twist your body like a contortionist in order to check out various orientations. Isn't there an easier way?

There is. In fact, you can find the current density vector by taking just three readings, one with the racquet webbing held perpendicular to $\hat{x}$, one with the racquet webbing held perpendicular to $\hat{y}$, and one with the racquet webbing held perpendicular to $\hat{z}$. The first such experiment is illustrated below:




Suppose the marked-off part of the vertical plane has area $A_x$. Then the marked-off part of the perpendicular-to-flow plane has area $A_x \cos\theta$. The two areas are equal in charge/time passing through, namely

$$|\vec{J}|A_x \cos\theta = J_x A_x = \vec{J}\cdot\hat{x}\,A_x.$$

(In general, the current flowing through area $A$ perpendicular to $\hat{n}$ is $\vec{J}\cdot\hat{n}\,A$.) Thus, if you hold the racquet detector with webbing perpendicular to $\hat{x}$, find the current passing through, and divide that current by the area, you will find the $x$-component of current density, which we call $J_x$. Similarly you can measure the $y$ and $z$ components, and with just these three measurements you will find the current density vector

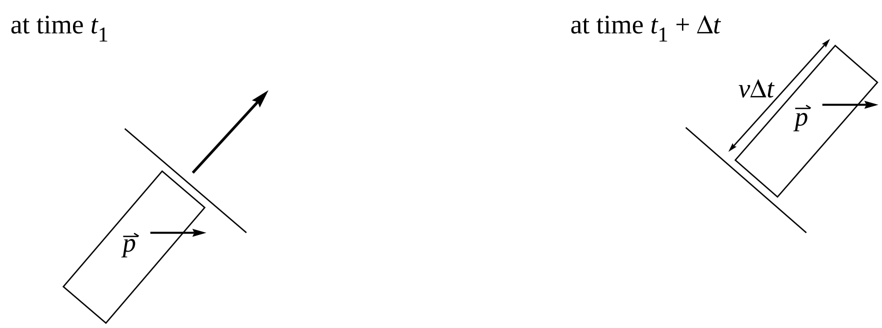
$$\vec{J} \doteq (J_x, J_y, J_z).$$

Notice that we never experimentally measure a vector: all our measurements are of components of a vector projected on various unit vectors. By putting together a series of such measurements (either through three components projected onto the basis vectors, or through homing in on the maximum), we can uncover the vector.

I hope that nothing I've said in this section is new to you. I just wanted to remind you of how current flow is defined and what it means to be a vector.

## Momentum transport

Suppose momentum flows uniformly in space and time, like a broad steady wind. If this momentum were carried by matter, like a literal wind or rain, then the motion of the air or raindrops ($\vec{v}$) would be parallel to the momentum being transported ($\vec{p}$). But the case of electromagnetic fields is less familiar, so the diagram below admits that possibility that the momentum might not be parallel to the direction the field is moving.




The reasoning proceeds exactly as it did for charge transport, except that charge density $\rho$ is replaced by momentum density $\vec{g}$. A plug of length $v\Delta t$, cross-sectional area $A_\perp$, passes through the imaginary plane perpendicular to the "wind". This plug contains momentum

$$\vec{p} = \vec{g}|\vec{v}|\Delta t A_\perp,$$

so the current density of $x$-momentum, namely

$$\frac{x\text{-momentum/time}}{\text{cross-sectional area}}$$

has direction parallel to $\vec{v}$ and magnitude

$$\frac{p_x/\Delta t}{A_\perp} = \frac{g_x|\vec{v}|\Delta t A_\perp/\Delta t}{A_\perp} = g_x|\vec{v}|,$$

whence the current density of $x$-momentum is

$$g_x\vec{v}.$$

Of course, you can do the same calculation for $y$-momentum density and $z$-momentum density. In total, the momentum current density is

$$\vec{g}\vec{v}.$$

What is this thing we get by multiplying two vectors in this way? It's not a scalar, which would result from a dot product, and it's not a vector, which would result from a cross product. This is called an “outer product” or a “tensor product” and the result is called a tensor. James Clerk Maxwell recognized the importance of this particular combination, except that he defined it with a negative sign. The “Maxwell stress tensor” is defined through

$$\overleftrightarrow{T} = -\vec{g}\vec{v}.$$

The physical meaning of the Maxwell stress tensor is exactly as described above. You could image a momentum detector shaped like a tennis racquet, and everything we said above about charge current density measurements (charge current density is a vector) would apply to momentum current density measurements (momentum current density is a tensor). In particular, if you look at components in some given basis you’ll find

$$\overleftrightarrow{T} \doteq \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix} = -\begin{pmatrix} g_x \\ g_y \\ g_z \end{pmatrix} \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} = -\begin{pmatrix} g_x v_x & g_x v_y & g_x v_z \\ g_y v_x & g_y v_y & g_y v_z \\ g_z v_x & g_z v_y & g_z v_z \end{pmatrix}.$$

Thus, for example, $-T_{xz}$ represents the current of $x$-momentum passing through a plane perpendicular to $z$.

As with vectors, we never experimentally measure tensors: all our measurements are of components of tensors. And as with vectors, this doesn’t mean that tensors are useless or unnatural. It just means that they’re less familiar.

*Exercise:* Suppose the material is indeed, say, wind, so that the momentum transported is always in the direction of the motion of the material flow. Is the tensor then represented by a diagonal matrix? Symmetric? Antisymmetric? What would the tensor components look like if the basis were oriented so that the $x$ direction points along the direction of the wind?

*Answer:* In this case the tensor would be represented by a symmetric matrix. If the basis were oriented as described, all matrix elements would vanish but for the $T_{xx}$ element.

## Force

We've considered the flow of momentum, carried in a field, across a plane. But there's another possibility. Remember that momentum is conserved only in the absence of external force. We could start with nothing, and end up with a plug of momentum, by exerting a force.


In this figure, think of the lower left portion as occupied by some charges and currents. After some time has passed, there is a plug of fields on the upper right. The force exerted by these charges and currents on the field, in order to produce this momentum, is

$$\vec{F} = \frac{\vec{p}}{\Delta t} = \frac{\vec{g}|\vec{v}|\Delta t A_\perp}{\Delta t} = \vec{g}\vec{v} \cdot \hat{n}\, A_\perp = -\overset{\leftrightarrow}{T} \cdot \hat{n}\, A_\perp .$$

By Newton's third law, the force on the charges and currents by the field is thus

$$\vec{F} = \overset{\leftrightarrow}{T} \cdot \hat{n}\, A_\perp .$$

If the surface is not uniform, then the total force on the charges and currents is found by integrating over the surface

$$\vec{F}_{\text{total}} = \int_{\mathcal{S}} \overset{\leftrightarrow}{T} \cdot \hat{n}\, dA$$

Cute trick! Instead of integrating the force density over the volume of the charges and currents, you only need to surface integrate the Maxwell stress tensor over the surface.

One of the traditional difficult problems of physics is "the problem of the damn rotating cylinder, one-third full of dielectric [fluid], in a magnetic field". [Leon Lederman, "An open letter" *Physics Today* **47** (3) March 1994, 9–10]. I am not going to assign this problem to you. But when, eventually, it is assigned to you, solve the problem using the Maxwell stress tensor. (Also, check your answer by investigating what happens when the dielectric constant $\epsilon$ equals $\epsilon_0$. That's the mistake I made.)

## Connection to Griffiths

The motivation of the Maxwell stress tensor given here differs dramatically from the one given by Griffiths in section 8.2.2. I think that this way carries a lot more insight, which is why I use it. But Griffiths's approach has two advantages. First, my way holds only for static situations. (I said "steady wind" in the first sentence.) For a non-static situation, the total force on the charges and currents is (Griffiths equation 8.20)

$$\vec{F}_{\text{total}} = \int_{\mathcal{S}} \overset{\leftrightarrow}{T} \cdot \hat{n}\, dA - \epsilon_0 \mu_0 \frac{d}{dt} \int_{\mathcal{V}} \vec{S}\, d^3r,$$

where $\vec{S}$ is the Poynting vector.

Griffiths's second advantage is that in my approach, everything depends on "the velocity of the fields". A nice, compact concept, but how are you supposed to calculate it?! Griffiths gives a formula (equation 8.17) for the stress tensor in terms of fields alone:

$$\overset{\leftrightarrow}{T} = \epsilon_0 \left( \vec{E}\vec{E} - \tfrac{1}{2} E^2 \overset{\leftrightarrow}{I} \right) + \frac{1}{\mu_0} \left( \vec{B}\vec{B} - \tfrac{1}{2} B^2 \overset{\leftrightarrow}{I} \right),$$

where $\overset{\leftrightarrow}{I}$ represents the identity tensor, with components $\delta_{ij}$. This formula makes it clear that the stress tensor for electromagnetic field, like the stress tensor for wind or rain or fluid flow, is symmetric.

## Field stress

If there's a volume of charges and currents and fields bounded by surface $\mathcal{S}$, the net force on this volume is

$$\vec{F}_{\text{total}} = \int_{\mathcal{S}} \overset{\leftrightarrow}{T} \cdot \hat{n}\, dA = \frac{d}{dt}(\vec{p}_{\text{matter}} + \vec{p}_{\text{fields}}).$$

If there is no matter within the volume, only fields, then the net force must vanish.

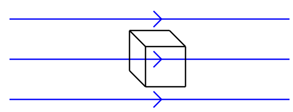
Think about a region of uniform electric field (and any field can be made uniform by thinking of a small enough region). Orient the coordinate axes so that the $x$ axis is parallel to $\vec{E}$. In this basis $\vec{E} \doteq (E, 0, 0)$ so

$$\overset{\leftrightarrow}{T} = \epsilon_0 \left[ \vec{E}\vec{E} - \tfrac{1}{2} E^2 \overset{\leftrightarrow}{I} \right] \doteq \epsilon_0 \left[ \begin{pmatrix} E^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \tfrac{1}{2} \begin{pmatrix} E^2 & 0 & 0 \\ 0 & E^2 & 0 \\ 0 & 0 & E^2 \end{pmatrix} \right]$$

or finally

$$\overset{\leftrightarrow}{T} \doteq \tfrac{1}{2} \epsilon_0 E^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{4.1}$$

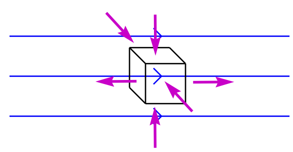
Now think of a little cube, length $L$ on a side, within the electric field.



The force on the right face of this cube (where $\hat{n} = \hat{x}$) is

$$\overleftrightarrow{T} \cdot \hat{x}\, L^2 \doteq \tfrac{1}{2}\epsilon_0 E^2 L^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \tfrac{1}{2}\epsilon_0 E^2 L^2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \doteq \tfrac{1}{2}\epsilon_0 E^2 L^2 \hat{x}.$$

Similar calculations can be performed on the other cube faces. You need to keep your signs straight. The resulting forces are shown below.



As promised, the net force cancels to zero. The pair of forces acting parallel to the electric field are tugging the cube apart; the pairs of forces acting perpendicular to the electric field are squishing their two sides together. My teacher Mark Heald liked to say that "electric field lies behave like furry rubber bands": the tugs parallel to the field lines act like a taut rubber band, but adjacent field lines don't get too close to each other because the squishing, like fur on the bands, prevents them from getting too close. They want to be as short as possible (rubbery) but don't want to be near each other (furry).

The same calculation can of course be performed for magnetic fields, with the same result except that $\epsilon_0$ is replaced with $1/\mu_0$.

Why do the field lines for electric dipoles and magnetic dipoles look so much alike? Not because the sources are similar... they're not (one source is a pair of charges and the other a loop of current). Not because Coulomb's law and the Biot-Savart law are similar... they're not (one has cross products and the other doesn't). It's because the stress tensor for $\vec{E}$ and for $\vec{B}$ are similar.

## Generalizations

The flow of a scalar (like charge density $\rho$) is described through a vector (like current density $\vec{J}$).

The flow of a vector (like momentum density $\vec{g}$) is described through a tensor (like the negative of the Maxwell stress tensor $-\overset{\leftrightarrow}{T}$).

What mathematical tool would one use to describe the flow of a tensor?

I ask this question not to make your brain hurt, but to open your mind to more and richer possibilities. The tensor that we've discussed, namely the Maxwell stress tensor, is an example of a "rank-2 tensor". In three dimensions, a rank-2 tensor can be described using 9 projections, called components, which are conveniently presented in a $3 \times 3$ matrix. The flow of a rank-2 tensor is described through a "rank-3 tensor". In three dimensions, a rank-3 tensor can be described using 27 components, and there's no real convenient way to present one on flat paper. (Sometimes I use a stack of three index cards, on each of which I write a $3 \times 3$ matrix. But even this is not really effective.) From this point of view, a vector is a rank-1 tensor and a scalar is a rank-0 tensor. In general, a rank-$r$ tensor in $d$ dimensions is specified through $d^r$ components.

People get upset by the word tensor, but they shouldn't:

The motion of a point is described by a vector
The motion of a vector is described by a (rank-2) tensor
The motion of a (rank-2) tensor is described by a rank-3 tensor

and so forth. It's nothing more than that.

## Appendix: Definition of tensor through components

Back on page 11 I defined a vector as "an entity with components $x_i$ that transform from one basis to another through

$$x'_i = \sum_{k=1}^{3} R_{ik} x_k. \text{ ''}$$

Can I generalize this definition to a tensor?

I can. We've seen how to write a tensor through an outer product as, for example,

$$\overleftrightarrow{T} = -\vec{g}\vec{v}.$$

In terms of components, this means

$$T_{ij} = -g_i v_j.$$

Because the components of $\vec{g}$ and of $\vec{v}$ transform as

$$g'_i = \sum_{k=1}^{3} R_{ik} g_k \quad \text{and} \quad v'_j = \sum_{\ell=1}^{3} R_{j\ell} v_\ell,$$

the components of a rank-2 tensor like $\overleftrightarrow{T}$ transform as

$$T'_{ij} = \sum_{k=1}^{3} \sum_{\ell=1}^{3} R_{ik} R_{j\ell} T_{k\ell}.$$

Generalizing, the components $x_{ijk}$ of a rank-3 tensor transform as

$$x'_{ijk} = \sum_{\ell=1}^{3} \sum_{m=1}^{3} \sum_{n=1}^{3} R_{i\ell} R_{jm} R_{kn} x_{\ell mn}.$$

and so forth.

## Appendix: What is stress?

How did the entity we're talking about get the name "stress" tensor?

The word "stress", in physics, means internal force per area. Grab a pencil by the eraser end and hold it horizontally. In your mind, think of a "cut plane" separating the pencil into the half closer to the eraser and the half closer to the tip. (This plane is not artificial at all: you know that if you select a plane through an eraser it will hit mostly empty space.) The tip half is being pulled down by gravity. Why doesn't it accelerate downward? Because there's an internal upward force exerted through the cut plane *by* the eraser half *on* the tip half. (This is the sum of all the forces *by* atoms in the eraser half *on* atoms in the tip half.) With a finger on your other hand, press the pencil tip downward, while keeping it horizontal. Now this internal upward force has increased. The internal force per area across a cut plane is called the "stress".

If the stress is always perpendicular to the cut plane, as it is in a static fluid, then this stress is called "pressure". Pressure is a form of stress that can be represented through a scalar. But internal forces in a solid, or in a moving fluid, are not always pressure stresses: They will have a direction that depends on the orientation of the cut plane, and they must be represented through a tensor.

If you're building a bridge, or building, or railroad locomotive, it's essential to keep track of the stresses in your structure, and to know at what stress your material will fracture. This was a big task for physics in the nineteenth century, and when James Clerk Maxwell attended classes at the Universities of Edinburgh and of Cambridge (1847–1856) he must have taken courses on this topic.

When Maxwell researched electromagnetism, it was only natural that he would apply what he knew about mechanical stresses to those researches. He considered all electromagnetic phenomena, not just light, to reflect forces and displacements within the "luminiferous ether". When Maxwell needed to keep track of internal forces within the ether, he used what we now call the "Maxwell stress tensor".

I have read part of Maxwell's 1873 *Treatise on Electricity and Magnetism.* It is remarkable not only for its insight, but for how differently Maxwell framed electrodynamics from how we frame it today.

# Chapter 5

# Electromagnetic Waves

## 5.1 One-dimensional waves

A string is stretched taut. Call position along the straight string $x$. Now the string is pulled away from straightness in some fashion: the displacement from straightness is called $\eta(x,t)$. How does this displacement change with time?

To high accuracy, this displacement obeys "the wave equation"

$$\frac{\partial^2 \eta}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 \eta}{\partial t^2}. \tag{5.1}$$

(The wave equation does not describe *exactly* the motion of a disturbed spring. In fact, the wave equation is not *exactly* correct even for electromagnetic waves in vacuum, because of quantum mechanical effects. But it is highly accurate for both situations.) I won't derive the wave equation for a taut string (see W.C. Elmore and M.A. Heald, *Physics of Waves* (1969) section 1.1), but instead focus on solving it.

I genuinely enjoy solving ordinary differential equations, but partial differential equations tend to make my stomach churn. Can we make any progress on this equation? Jean-Baptiste le Rond d'Alembert led the way. Inspired by the algebraic factorization $a^2 - b^2 = (a-b)(a+b)$, he decided to write the wave equation as

$$0 = \frac{\partial^2 \eta}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 \eta}{\partial t^2} = \left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right)\eta(x,t).$$

This suggests a change of variable to

$$\begin{aligned} u &= x - ct \\ v &= x + ct \end{aligned}$$

or, what is the same thing,

$$\begin{aligned} x &= \frac{1}{2}(v+u) \\ t &= \frac{1}{2c}(v-u). \end{aligned}$$

In terms of these variables, the equation above is

$$0 = 4\frac{\partial^2 \eta(u,v)}{\partial u\, \partial v}$$

and the solution of this equation is clear from inspection: If $f(\ )$ and $g(\ )$ are arbitrary one-variable functions, then the general solution is

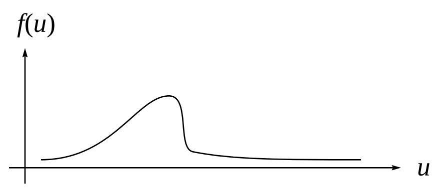
$$\eta(u,v) = f(u) + g(v).$$

In short, d'Alembert's solution shows that the general solution to the wave equation (5.1) above is

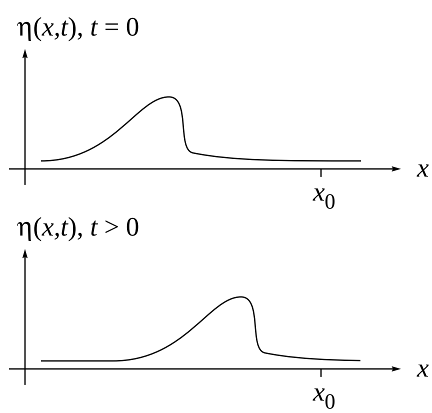
$$\eta(x,t) = f(x-ct) + g(x+ct). \tag{5.2}$$

Wow, that was easy. Three cheers for d'Alembert!

**Character of d'Alembert's solution.** But after the cheering stops we need to dissect this general solution to see what it's telling us about nature. Begin with the case $g(v) = 0$. If $f(u)$ looks like

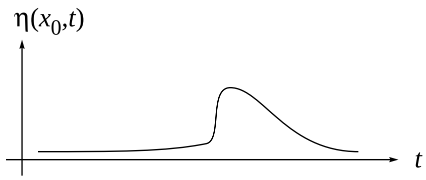



then $\eta(x,t) = f(x-ct)$ looks like

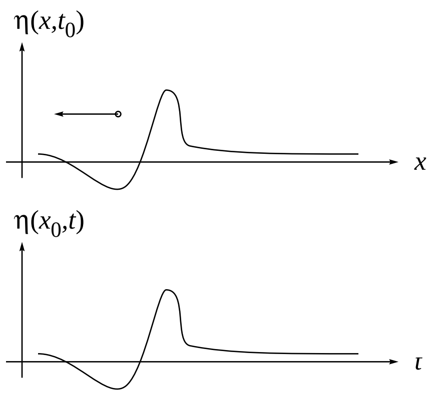



This is a shape-preserving pulse wave moving to the right at speed $c$!

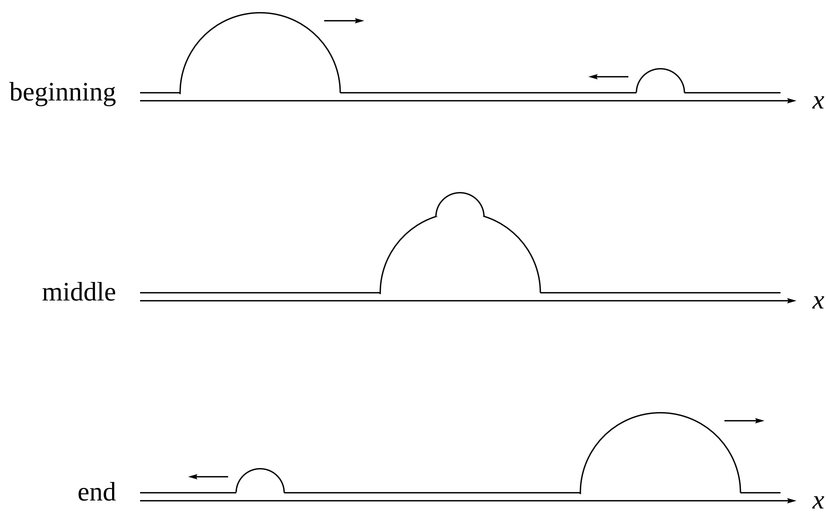
Meanwhile, what does this wave look like as a function of $t$? Suppose I stand at point $x_0$ and let the wave wash over me. The function $\eta(x_0, t)$ looks like the mirror image of the above.



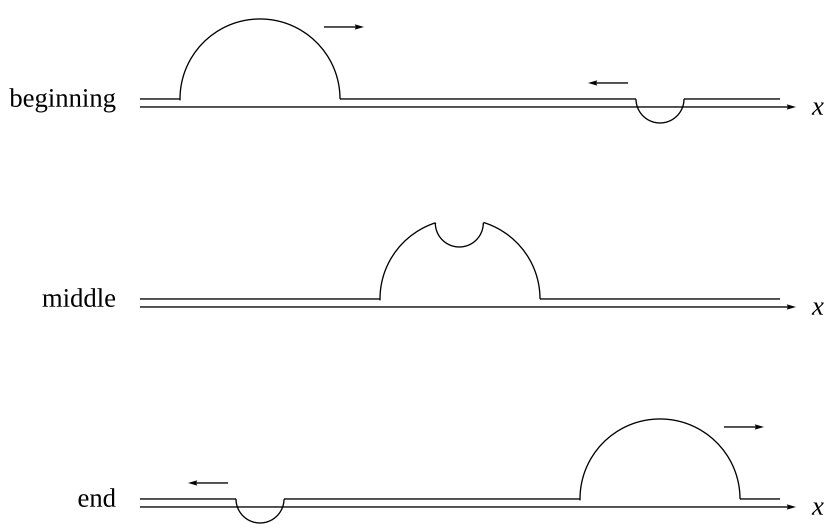

Move on to the case $f(u) = 0$, so $\eta(x, t) = g(x - ct)$. This is a shape-preserving pulse wave moving to the left at speed $c$:



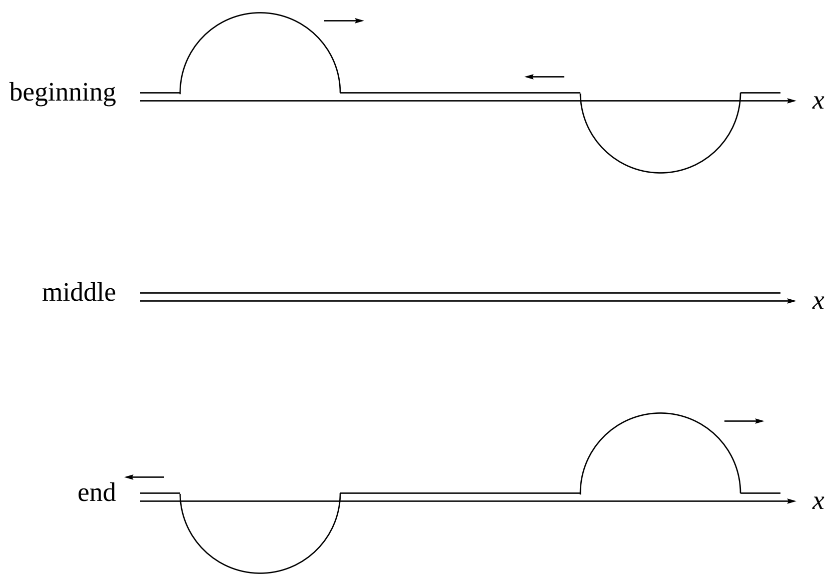
What if neither $f(u)$ nor $g(v)$ vanish? This is called "superposition"! In the example below, a big semicircular wave moves right, a small one moves left. When they cross over each other, the two waves add. Then each continues independently on its own way as if they had never known each other: "two ships that pass at night".




The same holds if the small wave moving left happens to have a downward rather than an upward displacement.



What if the two waves are the same size? In this case their displacements cancel out completely as they pass over each other.

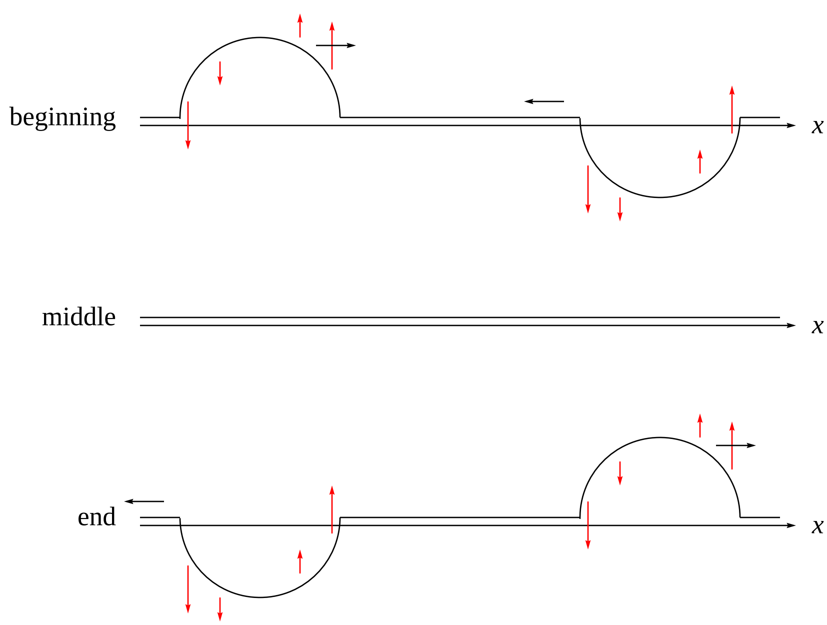



It would be very funny if this were done as a lecture demonstration and you happened to walk into class late just at the time marked "middle". You would see a straight string with no displacement at all, then two semicircular waves would pop into being on the straight string, the "up" wave moving right and the "down" wave moving left! This is a puzzle. How can a straight, unstretched spring just pop two semicircular waves into existence? Where does the energy for those two waves come from?

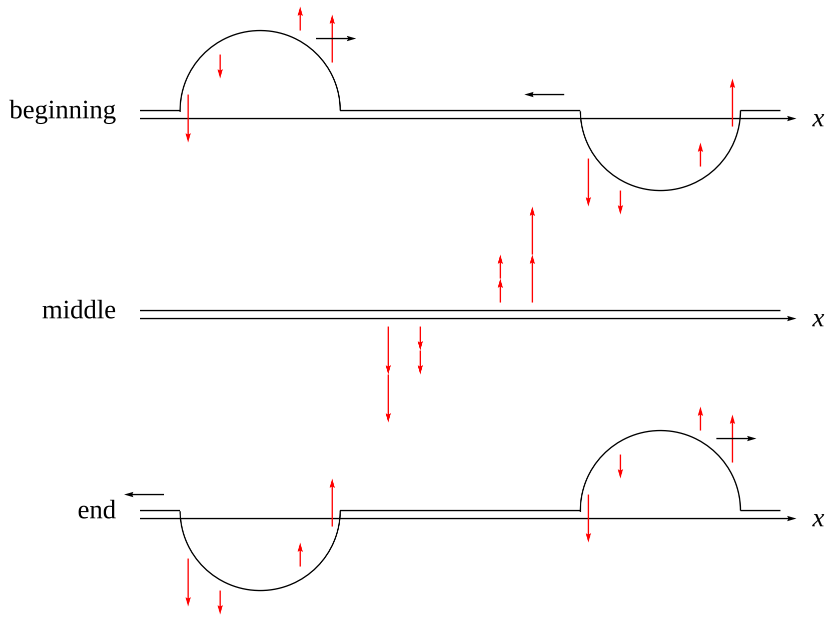
I'll resolve this puzzle on the next page. But think about it for a few moments before turning the page, both to make sure you understand why it's puzzling and to see if you can't resolve the puzzle yourself.

You know from introductory mechanics that to specify the state of a particle you must specify both its position and its velocity. The wave pictures on the previous pages show only the positions of the particles that make up the string, and not their velocities, so they don't specify the state. (There are black velocity arrows, but they signify the velocity of the waveform, not the velocity of the particles on the string.)

Paint a green dot on a single bit of string, and think about the motion of that dot as the rightward moving, "up" wave washes over it. That dot moves first up, then down. When the leftward moving, "down" wave washes over the green dot, it moves first down, then up. For the "beginning" and "end" situations, when the two waves are well-separated, the velocities of representative dots are shown using red arrows in the figure below.




Now think about what happens to a single string element at the "middle" situation. The total displacement of a dot on the string is the sum of the displacement due to the two superposing waves. And the same is true of the velocities. But at the "middle" time, when the string element *displacements* sum they cancel out to zero, while when the string element *velocities* sum they actually increase.




When you walked late into class, you saw the string at an instant, as in a snapshot, and of course a snapshot can't show the velocities. The situation at the middle is indeed a straight, unstretched string, but it's not a straight, unstretched spring at rest. It's the motion of the string (invisible in the snapshot) that enables it to pop two semicircular waves into existence.

*Challenge:* Can you show that if the waveform is $y(x,t) = f(x - ct)$, and if

$$f'(x) = \frac{df(x)}{dx},$$

then the velocity of the string element at $(x,t)$ is $-cf'(x - ct)$?

**Sine waves.** Suppose $f(u) = A\sin(ku)$ so that the wave is

$$f(x - ct) = A\sin(kx - \omega t).$$

Here $kc = \omega$ ("Kansas cows eat wheat"). Also the wavelength is $\lambda = 2\pi/k$ and the period is $T = 2\pi/\omega$. This is a solution to the wave equation, but it doesn't correspond to any physical wave, because the function $\sin(kx - \omega t)$ extends over all space and all time. This wave started infinitely far in the past and will keep going for ever and ever, amen. Real waves are finite in space (waves on the ocean end when they hit the beach) and of course finite in time.[1] *However*, the pure sine wave can

[1] Following songwriters Nickolas Ashford and Valerie Simpson, ("Ain't No Mountain High Enough," 1966, sung most famously by Marvin Gaye and Tammi Terrell) I like to say that there "ain't no ocean wide enough, ain't no string long enough" to carry a pure sine wave.

be an accurate approximation to a wave that lasts for many periods. (And because the period of light is so small, most light waves fit this description.)

**Superposition of sine waves.** Suppose the two waves superposing are not semicircular pulses, but instead sine waves:

$$\begin{aligned} f(x-ct) &= A\sin(kx-\omega t) \\ g(x+ct) &= A\sin(kx+\omega t). \end{aligned}$$

Then the total wave is

$$\eta(x,t) = A\sin(kx-\omega t) + A\sin(kx+\omega t). \tag{5.3}$$

Okay, so that's the sum, but how can we understand the character of $\eta(x,t)$? If you knew the trigonometric sum and difference formulas, you might be able to make some progress. But I forgot those formulas the minute I left high school (if not before). Instead, I like to perform trigonometric manipulations using complex arithmetic and Euler's formula that, for $\theta$ real,

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

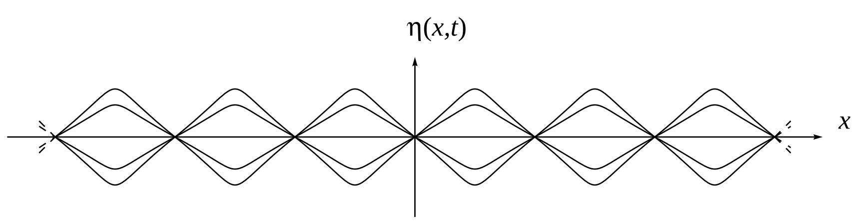
First, establish two consequences of Euler's formula:

$$\sin(\theta) = \Im m\left\{e^{i\theta}\right\} \text{ and } \cos(\theta) = \tfrac{1}{2}(e^{i\theta}+e^{-i\theta}).$$

Now go at it:

$$\begin{aligned} \eta(x,t) &= A\sin(kx-\omega t) + A\sin(kx+\omega t) && (5.4) \\ &= A\,\Im m\left\{e^{i(kx-\omega t)} + e^{i(kx+\omega t)}\right\} \\ &= A\,\Im m\left\{e^{ikx}[e^{-i\omega t}+e^{i\omega t}]\right\} \\ &= A\,\Im m\left\{e^{ikx}[2\cos(\omega t)]\right\} \\ &= 2A\cos(\omega t)\,\Im m\left\{e^{ikx}\right\} \\ &= 2A\cos(\omega t)\sin(kx). && (5.5) \end{aligned}$$

In other words, $y(x,t)$ is just a sine function of $x$, but with amplitude that varies with time: The amplitude is $2A$ at $t=0$, then diminishes to 0 at $t=\frac{1}{2}\pi/\omega$, becomes $-2A$ at $t=\pi/\omega$, 0 again at $t=\frac{3}{2}\pi/\omega$, and returns to $2A$ at $t=2\pi/\omega$.



I don't know about you, but I never would have guessed that this behavior is hidden within the equation (5.4). This does not look like "two ships that pass at night", but it is! These are called "standing waves".

It's tempting to **overgeneralize d'Alembert's solution.** It seems so natural: The general solution of a second-order linear ODE has two adjustable parameters, which are set by the initial or boundary conditions. A natural generalization is "A second-order linear PDE has two adjustable functions, which are set by the initial or boundary conditions." This is wrong.

For example a wave $\eta(x, y, t)$ moving in two dimensions (like the motion of a taut drumhead, or ripples on the surface of a puddle) satisfies, to high accuracy, the two-dimensional wave equation

$$\begin{aligned}\frac{\partial^2\eta}{\partial x^2}+\frac{\partial^2\eta}{\partial y^2} &= \frac{1}{c^2}\frac{\partial^2\eta}{\partial t^2},\\ \nabla^2\eta &= \frac{1}{c^2}\frac{\partial^2\eta}{\partial t^2}.\end{aligned}$$

What is the general solution to this equation? One is tempted to guess

$$f_1(x-ct)+g_1(x+ct)+f_2(y-ct)+g_2(y+ct). \tag{5.6}$$

This is indeed a solution, as you can readily check, but it's not the most general solution. Drop a pebble into the puddle: a circular wave emerges. This circular wave is *not* shape-preserving: it gets smaller as it spreads from the pebble. This waveform is *not* of the form (5.6), so form (5.6) can't be the most general solution.

## 5.2 From Maxwell equations to wave equations

In a region with no charges or currents, the Maxwell equations are

$$\begin{array}{ll}\vec\nabla\cdot\vec E=0 & \vec\nabla\cdot\vec B=0\\ \vec\nabla\times\vec E=-\dfrac{\partial\vec B}{\partial t} & \vec\nabla\times\vec B=\mu_0\epsilon_0\dfrac{\partial\vec E}{\partial t}\end{array}$$

An equation like Faraday's law on the bottom left,

$$\vec\nabla\times\vec E=-\frac{\partial\vec B}{\partial t},$$

doesn't tell you much, because you don't know what $\vec B$ is! But, because of the Ampere-Maxwell law on the bottom right, you *do* know what $\vec\nabla\times\vec B$ is. So let's take the curl of Faraday's law:

$$\vec\nabla\times\vec\nabla\times\vec E=-\frac{\partial}{\partial t}\vec\nabla\times\vec B=-\frac{\partial}{\partial t}\left(\mu_0\epsilon_0\frac{\partial\vec E}{\partial t}\right)=-\mu_0\epsilon_0\frac{\partial^2\vec E}{\partial t^2}.$$

The right hand side looks like it belongs in a wave equation, but the left hand side looks dramatically foreign. Remember the vector calculus identity

$$\vec{\nabla} \times \vec{\nabla} \times \vec{F} = \vec{\nabla}(\vec{\nabla} \cdot \vec{F}) - \vec{\nabla}^2 \vec{F}.$$

(Normally the Laplacian $\vec{\nabla}^2$ is applied to scalar functions. Here we apply it to each of the three Cartesian components in turn.) For our case $\vec{\nabla} \cdot \vec{E} = 0$, so

$$\vec{\nabla}^2 \vec{E} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}. \tag{5.7}$$

This is the wave equation with

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}. \tag{5.8}$$

You can apply exactly the same reasoning to the magnetic field to find

$$\vec{\nabla}^2 \vec{B} = \mu_0 \epsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}. \tag{5.9}$$

So much fuss is made of these two wave equations that you might forget that the Maxwell equations are more fundamental. The two wave equations can be derived from the Maxwell equations but not vice versa. That's because the Maxwell equations contain more information than the two wave equations. For example, Griffiths shows (section 9.2.2) that monochromatic plane electromagnetic waves are traverse waves: If the direction of wave propagation is $\hat{k}$, then $\vec{E}$ and $\vec{B}$ are perpendicular to $\hat{k}$. Furthermore $\vec{E}$ is perpendicular to $\vec{B}$. Still further, at any one place and time the magnitudes of the fields are related through $|\vec{E}| = c|\vec{B}|$. None of these facts can be derived from the two wave equations by themselves.

Finally, the analysis of this section shows that electromagnetic waves exist, but doesn't show how to use charges and currents to get them started. That's the task of chapter 7, "Resultant Potentials and Fields".

## 5.3 Waves in media

Electromagnetic waves in vacuum: Start with the wave equation

$$\frac{\partial^2 \eta}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 \eta}{\partial t^2},$$

this generates the "shape preserving" solutions

$$\eta(x,t) = f(x - ct) + g(x + ct)$$

which specializes to

$$\eta(x,t) = Ae^{i(kx - \omega t)} \quad \text{with} \quad kc = \omega.$$

(There is an understood "text" behind this equation: Take the real or the imaginary part. No wave lasts forever, but this applies approximately to any wavepacket long compared to $\lambda$.)

In a medium, the process goes almost the other way around: Think of waves like

$$\eta(x,t) = Ae^{i(kx-\omega t)}.$$

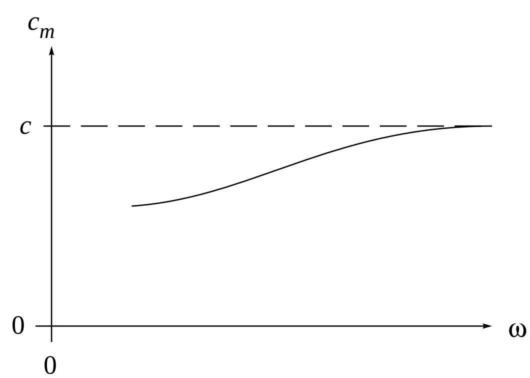
This sine wave moves with speed $c_m(\omega)$: In an infinite (or very long) sine wave, the crests move at speed $c_m(\omega)$.

⟦You will ask: Why does wave speed depend upon $\omega$ alone? The answer is, it might not. In an inhomogeneous medium, $c_m$ will depend upon location. In a nonlinear medium, $c_m$ will depend upon amplitude $A$. But it's an experimental result that for many uniform media, $c_m$ is a function of $\omega$ alone.⟧

There is not, and cannot be, an equation like

$$\frac{\partial^2\eta}{\partial x^2} = \frac{1}{c_m^2(\omega)}\frac{\partial^2\eta}{\partial t^2} \; !$$

One often sees experimental results like

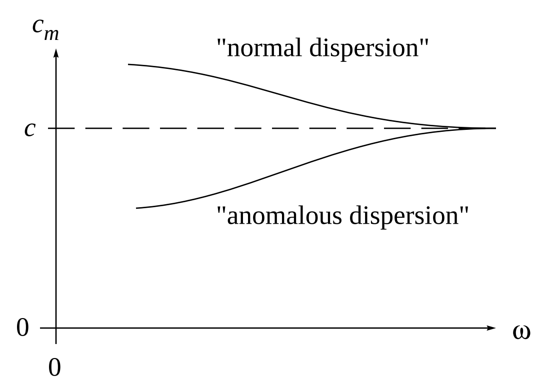



It makes sense that at high frequencies the speed of light in medium $c_m$ approaches the speed of light in vacuum $c$: When the jiggling is very fast the atoms in the medium cannot react to the light, so the light passes through as if the atoms were not there.

Okay, so this is how infinitely long sine waves behave. But what about real, finite waves? In this case you have to start with the initial finite wave, Fourier decompose it into sine components, move each component forward at its own speed $c_m(\omega)$, then Fourier recompose it to find the wave at the final time. [Very much like quantum mechanics, where, to solve time evolution problems, you start with the initial wavefunction, expand in energy eigenstates (with eigenvalue $E_n$), propagate each energy eigenstate forward in time through its phase factor $e^{-(i/\hbar)E_n t}$, then recombine them.]

So, to solve time evolution problems, you must know the details of $\eta(x, t_0)$ and of $c_m(\omega)$. The wave $\eta(x, t)$ might spread out, compact in, change shape, split in two, etc. Lots of strange behaviors are possible, showing the many, various, and diverse phenomena of physics. As such, there is no single "wave speed". In his book *Wave Propagation and Group Velocity*, Léon Brillouin discusses phase velocity, group velocity, velocity of energy transport, signal velocity, and forerunner velocity.

But most of the time the packet spreads out or "disperses". Graphs like the previous one are called "dispersion curves". For example




What?! Speeds $c_m$ greater than $c$?!?!

Yes. This is the speed of the crests in an infinite sine wave. When you watch the wave come in, it's been known for a long time when the crests are going to arrive. An infinite sine wave in fact transmits no information whatsoever!

By the way, anomalous dispersion is in fact more common than normal dispersion.

**Phase velocity and group velocity.** To get an idea of the variety of behaviors possible through this decompose-recompose technique, consider a simple initial wave, composed of only two Fourier components:

$$\begin{aligned}\eta_1(x,t) &= A\cos(k_1 x - \omega_1 t) = A\,\Re e\left\{e^{i(k_1 x - \omega_1 t)}\right\}\\ \eta_2(x,t) &= A\cos(k_2 x - \omega_2 t) = A\,\Re e\left\{e^{i(k_2 x - \omega_2 t)}\right\}.\end{aligned} \tag{5.10}$$

If the two components are close in frequency, it makes sense to define the average wavenumber $k_0$ and the average frequency $\omega_0$ such that

$$\begin{aligned}k_1 &= k_0 - \tfrac{1}{2}\Delta k \qquad \omega_1 = \omega_0 - \tfrac{1}{2}\Delta\omega\\ k_2 &= k_0 + \tfrac{1}{2}\Delta k \qquad \omega_2 = \omega_0 + \tfrac{1}{2}\Delta\omega.\end{aligned} \tag{5.11}$$

In terms of these new variables,

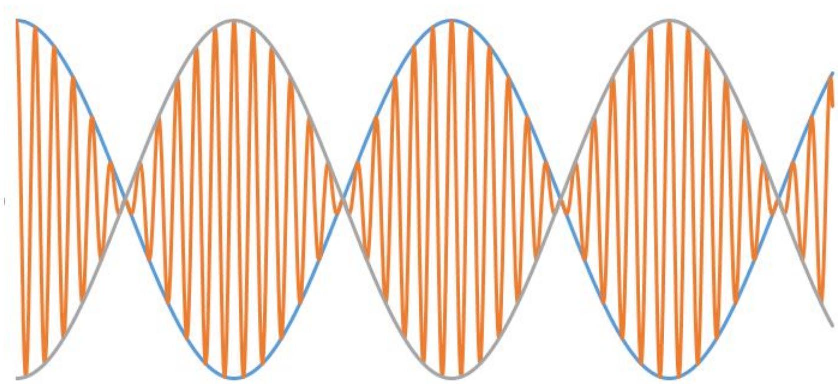
$$\begin{aligned}\eta_1(x,t) &= A\,\Re e\left\{e^{i(k_0x-\omega_0 t)-\frac{1}{2}i(\Delta k\,x-\Delta\omega\,t)}\right\}\\ \eta_2(x,t) &= A\,\Re e\left\{e^{i(k_0x-\omega_0 t)+\frac{1}{2}i(\Delta k\,x-\Delta\omega\,t)}\right\}.\end{aligned}\tag{5.12}$$

The total wave is then

$$\begin{aligned}\eta_1+\eta_2 &= A\,\Re e\left\{e^{i(k_0x-\omega_0 t)}\left[e^{-\frac{1}{2}i(\Delta k\,x-\Delta\omega\,t)} - e^{+\frac{1}{2}i(\Delta k\,x-\Delta\omega\,t)}\right]\right\}\\ &= A\,\Re e\left\{e^{i(k_0x-\omega_0 t)}\left[2\cos\left(\tfrac{1}{2}(\Delta k\,x-\Delta\omega\,t)\right)\right]\right\}\\ &= 2A\cos\left(k_0x-\omega_0 t\right)\cos\left(\tfrac{1}{2}\Delta k\,x-\tfrac{1}{2}\Delta\omega\,t\right),\end{aligned}\tag{5.13}$$

the product of two cosine waves!

So far we have not made use of our idea that $\omega_1$ and $\omega_2$ are close. Adopting this idea, we see that $\Delta k \ll k_0$, so $\Delta k$ corresponds to a long wavelength, so the right-hand factor in equation (5.13) varies slowly compared to the left-hand factor. We can plot the sum wave $\eta_1+\eta_2$ using the "envelope trick". In blue I plot the slowly varying envelope $2A\cos\left(\frac{1}{2}\Delta k\,x-\frac{1}{2}\Delta\omega\,t\right)$. In gray I plot the negative of this envelope. In brown I plot the sum wave $\eta_1+\eta_2$:



The envelope wave — the humps — has a wavenumber of $\frac{1}{2}\Delta k$ and a frequency of $\frac{1}{2}\Delta\omega$, so it moves at the so-called "group velocity"

$$v_g = \frac{\frac{1}{2}\Delta\omega}{\frac{1}{2}\Delta k} \approx \frac{d\omega}{dk},\tag{5.14}$$

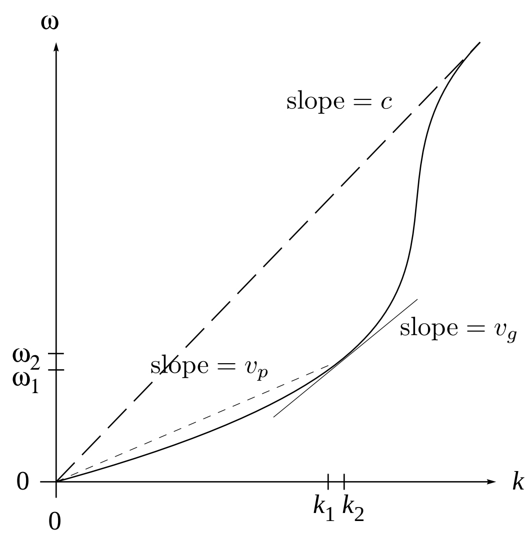
whereas the short-wavelength wave within this envelope — the wiggles — moves at the so-called "phase velocity"

$$v_p = \frac{\omega_0}{k_0}.\tag{5.15}$$

This argument is based on a simplification (only two Fourier components) but it contains the seeds of general truths:

group velocity: speed of hump
phase velocity: speed of wiggles within hump
usually information and energy moves at the group velocity, not always!
usually $v_g < c$, not always!
when $v_g > c$, the hump breaks up too fast to carry information

This graph shows the curve of $\omega(k)$ for a hypothetical medium. For $k$ large, wavelength and period small, this medium behaves like vacuum, where $\omega = ck$.




I show the wave numbers $k_1$ and $k_2$ that appear in equations (5.10). We've discussed what happens when these are the sole Fourier components, but I think you can see that if you had a spread of wave numbers the qualitative behavior would be much the same.

**Wave speed faster than infinity!** Here's an example of fantastic wave behavior told to me by Steven Wong.[2]

Everyone knows that a light wave in a medium might make that medium more absorptive or, in lay terms, darker. This is why people lie in the sun to get a tan. Steven developed a material that would switch from pure transparent to pure absorptive in about half a nanosecond, the time required for light to travel about 10 centimeters.

Steven flashed a light pulse two meters long (6 nanoseconds worth) into a wafer of his material. The first 10 centimeters of this light pulse made it through, but the

[2]S. Wong and D. Styer, "Answer to Question #52: Group Velocity and Energy Propagation," *Am. J. Phys.* **66**, 659–661 (1998).

rest was blocked. The entire emerging pulse departed while 1.9 meters of incoming pulse had yet to encounter the medium. The "center of the light pulse", emerged from the wafer before the "center of the light pulse" had reached the wafer!

But information in this case is carried by the forerunner that makes it through, not by the mathematical "center of the light pulse". The information travels safely at or below light speed.

**The wide variety of wave behavior.** We have only scratched the surface. If you examine quantal waves, deep-water waves, shallow-water waves, rogue waves, seismic waves, and electromagnetic waves within waveguides, or within plasmas, or within magnetic materials, you will encounter behaviors[3] beyond your wildest imagination.

You will understand rainbows, sun dogs, rings around the moon, the green flash the precedes sunrise, and mirages. You will understand the erie sounds made by frozen lakes. You will see with a new eye boat wakes and ocean breakers. Both your everyday life and your scientific life will be enriched.

[3]See, for example, Anatoly Patsyk, Uri Sivan, Mordechai Segev, and Miguel A. Bandres, "Observation of branched flow of light" *Nature* **583** (2 July 2020) 60–65.

# Chapter 6

# Potentials and Gauges

## 6.1 Math

This chapter relies on two mathematical theorems:

**Vanishing curl.** If a vector function $\vec{F}(\vec{r})$ has vanishing curl everywhere, then that function is the gradient of some scalar function $f(\vec{r})$. In other words $\vec{\nabla}\times\vec{F}(\vec{r}) = 0$ implies $\vec{F}(\vec{r}) = \vec{\nabla}f(\vec{r})$.

**Vanishing divergence.** If a vector function $\vec{F}(\vec{r})$ has vanishing divergence everywhere, then that function is the curl of some other vector function $\vec{A}(\vec{r})$. In other words $\vec{\nabla} \cdot \vec{F}(\vec{r}) = 0$ implies $\vec{F}(\vec{r}) = \vec{\nabla} \times \vec{A}(\vec{r})$.

These theorems hold for any reasonable function. If you are interested in unreasonable functions, look up "Helmholtz decomposition".

## 6.2 Potentials

| Electrostatics | Magnetostatics |
| --- | --- |
| $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_0$ | $\vec{\nabla} \cdot \vec{B} = 0$ |
| $\vec{\nabla} \times \vec{E} = 0$ | $\vec{\nabla} \times \vec{B} = \mu_0 \vec{J}$ |

On the electrostatic side, start with $\vec{\nabla} \times \vec{E} = 0$. This means that $\vec{E} = -\vec{\nabla} V$, so Gauss's law becomes Poisson's equation

$$\vec{\nabla}^2 V = -\rho/\epsilon_0.$$

It is much easier to work with electrostatic potential than with electric field. For example the potential due to source charge $q$ at the origin is

$$V(r) = \frac{1}{4\pi\epsilon_0}\frac{q}{r}.$$

We don't have to mess with vectors, this is a scalar!

On the magnetostatic side, start with $\vec{\nabla} \cdot \vec{B} = 0$. This means that $\vec{B} = \vec{\nabla} \times \vec{A}$, so Ampere's law becomes

$$\begin{aligned} \vec{\nabla} \times \vec{\nabla} \times \vec{A} &= \mu_0 \vec{J} \\ \vec{\nabla}^2 \vec{A} - \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) &= -\mu_0 \vec{J} \end{aligned}$$

which is almost Poisson's equation. It is somewhat easier to work with this vector potential than with magnetic field. For example the vector potential due to source charge $q$ moving with velocity $\vec{v}$ at the origin is

$$\vec{A}(r) = \frac{\mu_0}{4\pi}\frac{q\vec{v}}{r}.$$

This is still a vector, but at least it doesn't involve a cross product.

Electrodynamics

$$\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_0 \qquad \vec{\nabla} \cdot \vec{B} = 0$$
$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad \vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t}$$

In this case, start with $\vec{\nabla} \cdot \vec{B} = 0$. This still means that

$$\vec{B}(\vec{r}, t) = \vec{\nabla} \times \vec{A}(\vec{r}, t), \tag{6.1}$$

so Faraday's law becomes

$$\begin{aligned}
\vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\
\vec{\nabla} \times \vec{E} &= -\frac{\partial}{\partial t} \vec{\nabla} \times \vec{A} \\
\vec{\nabla} \times \left( \vec{E} + \frac{\partial \vec{A}}{\partial t} \right) &= 0 \\
\vec{E} + \frac{\partial \vec{A}}{\partial t} &= -\vec{\nabla} V
\end{aligned}$$

or

$$\vec{E}(\vec{r}, t) = -\vec{\nabla} V(\vec{r}, t) - \frac{\partial \vec{A}(\vec{r}, t)}{\partial t}. \tag{6.2}$$

What happens with the other two Maxwell equations? Gauss's law becomes

$$\begin{aligned}
\vec{\nabla} \cdot \vec{E} &= \rho/\epsilon_0 \\
\vec{\nabla} \cdot \left( -\vec{\nabla} V - \frac{\partial \vec{A}}{\partial t} \right) &= \rho/\epsilon_0 \\
\vec{\nabla}^2 V + \frac{\partial}{\partial t} \vec{\nabla} \cdot \vec{A} &= -\rho/\epsilon_0 \qquad (6.3)
\end{aligned}$$

which is a little more involved than Poisson's equation.

Meanwhile the Ampere-Maxwell equation becomes

$$\begin{aligned}
\vec{\nabla} \times \vec{B} &= \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \\
\vec{\nabla} \times \vec{\nabla} \times \vec{A} &= \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \left( -\vec{\nabla} V - \frac{\partial \vec{A}}{\partial t} \right) \\
\vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \vec{\nabla}^2 \vec{A} + \mu_0 \epsilon_0 \vec{\nabla} \frac{\partial V}{\partial t} + \mu_0 \epsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} &= \mu_0 \vec{J} \\
\vec{\nabla}^2 \vec{A} - \mu_0 \epsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} - \vec{\nabla} \left( \vec{\nabla} \cdot \vec{A} + \mu_0 \epsilon_0 \frac{\partial V}{\partial t} \right) &= -\mu_0 \vec{J}. \qquad (6.4)
\end{aligned}$$

Notice that if magnetic monopoles exist, none of these steps would be possible.

Thinking **mathematically**, we have gone from a description of electromagnetism in terms of $(\vec{E}, \vec{B})$ (six components) to a description in terms of $(V, \vec{A})$ (four components). There are eight component Maxwell equations. (Gauss's law is one equation, Faraday's law is a vector equation whence three components, that's four, and then there are four on the magnetism side as well.) There are four component equations (equations (6.3) and (6.4)) for the scalar and vector potentials.

| | |
|---|---|
| Maxwell approach: | 6 functions determined by 8 equations |
| Potential approach: | 4 functions determined by 4 equations |

Thinking **physically**, it's pretty simple to have an idea of scalar potential — indeed "voltage" is known to little kids who don't know about electrostatic field and who couldn't line integrate to save their lives. But the idea of vector potential is more iffy. James Clerk Maxwell was interested in cases far from electrostatics, where the primary source of $\vec{E}$ was not $-\vec{\nabla}V$ but instead

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t}.$$

He compared that equation to

$$\vec{F} = \frac{d\vec{p}}{dt}$$

and called $\vec{A}$ "electromagnetic momentum". (He was not someone to let a minus sign get in his way.) This is not momentum in the sense of

$$\frac{1}{c^2}\int \vec{S}\, d^3r,$$

but momentum in the sense of "oomph". (When George H.W. Bush was running for president in 1980, he used the term "momentum" in this sense when he said of his campaign's "oomph" that "what we will have is momentum. We will look forward to big mo".)

When Maxwell gave names to electromagnetic quantities, he listed them in order of importance. This is how magnetic field got its strange name $\vec{B}$. And what did Maxwell consider the most important of all quantities? That's right, what we today call vector potential, $\vec{A}$.

**The principal vectors which we have to consider are :—**

| | Symbol of Vector. | Constituents. |
|---|---|---|
| The radius vector of a point .................. | $\rho$ | $x\ \ y\ \ z$ |
| The electromagnetic momentum at a point | $\mathfrak{A}$ | $F\ \ G\ \ H$ |
| The magnetic induction ...................... | $\mathfrak{B}$ | $a\ \ b\ \ c$ |
| The (total) electric current .................. | $\mathfrak{C}$ | $u\ \ v\ \ w$ |
| The electric displacement ..................... | $\mathfrak{D}$ | $f\ \ g\ \ h$ |



619.] QUATERNION EXPRESSIONS. 237

| | Symbol of Vector. | Constituents. |
|---|---|---|
| The electromotive force ..................... | $\mathfrak{E}$ | $P\ \ Q\ \ R$ |
| The mechanical force ........................ | $\mathfrak{F}$ | $X\ \ Y\ \ Z$ |
| The velocity of a point ........................ | $\mathfrak{G}$ or $\dot{\rho}$ | $\dot{x}\ \ \dot{y}\ \ \dot{z}$ |
| The magnetic force ........................... | $\mathfrak{H}$ | $\alpha\ \ \beta\ \ \gamma$ |
| The intensity of magnetization ............ | $\mathfrak{I}$ | $A\ \ B\ \ C$ |
| The current of conduction .................. | $\mathfrak{K}$ | $p\ \ q\ \ r$ |

**We have also the following scalar functions :—**

**The electric potential $\Psi$.**

James Clerk Maxwell, *A Treatise on Electricity and Magnetism*, volume II (Clarendon Press, Oxford, 1873) pages 236–237.

(You see from this table that it is also Maxwell who gave the name $\vec{J}$ to current density.)

Maxwell, in fact, thought of $\vec{E}$, $\vec{B}$, $\vec{A}$, and $V$ as fields of comparable footing. It was Oliver Heaviside who, in 1888, made the distinction between fields and potentials that we know and love today. [The self-educated Heaviside also invented coaxial cable, the concept of impedance, the use of complex numbers to represent sinusoidal current, the Heaviside step function and, remarkably, the Lorentz force law ($\vec{F} = q\vec{v} \times \vec{B}$)! In 1902 Heaviside surmised the existence of the ionosphere to explain Guglielmo Marconi's 1901 observation that radio waves can propagate over the horizon. In Britain the ionosphere is sometimes called the "Heaviside layer". See Bruce J. Hunt, "Oliver Heaviside: A first-rate oddity" *Physics Today* **65** (November 2012) pages 48–54.]

Heaviside made the distinction between fields and potentials when he realized that he could change $\vec{A}$ and $V$ without changing $\vec{E}$ and $\vec{B}$, and hence without changing the force experienced by any particle. He asked this question: I know that in electrostatics I can add a constant to the potential $V(\vec{r})$ and obtain a potential just as good as the original. Can I do a similar thing in electrodynamics? This is, if I have potentials $\vec{A}(\vec{r},t)$ and $V(\vec{r},t)$ that produce particular fields $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$, then can I find different potentials

$$\vec{A}'(\vec{r},t) = \vec{A}(\vec{r},t) + \vec{\alpha}(\vec{r},t).$$

and

$$V'(\vec{r},t) = V(\vec{r},t) + \beta(\vec{r},t)$$

that produce the same fields?

If these are to generate the same $\vec{B} = \vec{\nabla} \times \vec{A}$, then $\vec{\nabla} \times \vec{A}' = \vec{\nabla} \times \vec{A}$, so

$$\vec{\nabla} \times \vec{\alpha}(\vec{r},t) = 0$$

which happens whenever

$$\vec{\alpha}(\vec{r},t) = \vec{\nabla}\lambda(\vec{r},t).$$

Meanwhile, if these two different potentials are to generate the same

$$\vec{E} = -\vec{\nabla}V - \frac{\partial \vec{A}}{\partial t}$$

they will have to have

$$\begin{aligned}
-\vec{\nabla}\beta - \frac{\partial \vec{\alpha}}{\partial t} &= 0 \\
\vec{\nabla}\beta + \frac{\partial \vec{\nabla}\lambda}{\partial t} &= 0 \\
\vec{\nabla}\left(\beta + \frac{\partial \lambda}{\partial t}\right) &= 0 \\
\beta(\vec{r},t) + \frac{\partial \lambda(\vec{r},t)}{\partial t} &= k(t).
\end{aligned}$$

That is, if the new potentials are

$$\vec{A}'(\vec{r},t) = \vec{A}(\vec{r},t) + \vec{\nabla}\lambda(\vec{r},t) \quad \text{and} \quad V'(\vec{r},t) = V(\vec{r},t) - \frac{\partial \lambda(\vec{r},t)}{\partial t} + k(t),$$

then the $\vec{E}$ and $\vec{B}$ generated will be exactly the same for both potentials. In fact, we can replace $\lambda(\vec{r},t)$ with the different function

$$\lambda'(\vec{r},t) = \lambda(\vec{r},t) - \int k(t)\,dt$$

and make the same transformation.

Thus, finally, for any function $\lambda(\vec{r},t)$, the potentials

$$\vec{A}'(\vec{r},t) = \vec{A}(\vec{r},t) + \vec{\nabla}\lambda(\vec{r},t) \quad \text{and} \quad V'(\vec{r},t) = V(\vec{r},t) - \frac{\partial\lambda(\vec{r},t)}{\partial t} \tag{6.5}$$

generate the same $\vec{E}$ and $\vec{B}$ fields. Changing from one set of potentials to another is called a "gauge transformation". The word "gauge" means "a standard or scale of measurement". For example, the choice of whether to measure elevations of building floors relative to sea level or relative to the elevation of the basement is a choice of gauge.

⟦I am not certain, but I suspect that the term "gauge" follows from the nineteenth century fascination with railroad technology. In this context "gauge" refers to the separation between the two rails. Today, most railroads are "standard gauge" (4 feet, $8\frac{1}{2}$ inches or 1435 mm) or, for railroads in mountainous regions, "narrow gauge". But formerly each railroad company used its own gauge. If you search the Internet for "track gauge" you will uncover information about the 1841–1892 "Gauge War" (also called the "Battle of the Gauges").⟧

So you see that in fact we don't need four functions ($V(\vec{r},t)$ and $\vec{A}(\vec{r},t)$) to describe electromagnetism in a particular situation. Because we can select a gauge at will, there are only three independent functions. This is great power! But with great power comes great responsibility. How will we use our newfound power?

## 6.3 Gauge freedom

### Weyl gauge

By hook or by crook, we have found a set of potentials $(V(\vec{r},t), \vec{A}(\vec{r},t))$ for our situation. Now set

$$\begin{aligned} \frac{\partial\lambda(\vec{r},t)}{\partial t} &= V(\vec{r},t) \\ \lambda(\vec{r},t) &= \int V(\vec{r},t)\,dt. \end{aligned}$$

The new potentials obtained through the gauge transformation equations (6.5) are

$$\begin{aligned} V'(\vec{r},t) &= V(\vec{r},t) - \frac{\partial\lambda(\vec{r},t)}{\partial t} = 0 \\ \vec{A}'(\vec{r},t) &= \vec{A}(\vec{r},t) + \vec{\nabla}\lambda(\vec{r},t) = \vec{A}(\vec{r},t) + \vec{\nabla}\int V(\vec{r},t)\,dt. \end{aligned}$$

In this gauge — called the Weyl gauge — the scalar potential is zero everywhere and there are only three component functions to find. Through a similar argument you could instead set $A_x = 0$ everywhere, or $A_y = 0$, or $A_\theta = 0$.

The Weyl gauge is amusing when applied to any electrostatic situation, where $\vec{E}(\vec{r})$ is given and $\vec{B}(\vec{r}) = 0$. In the Weyl gauge in this situation the potentials are

$$V(\vec{r}, t) = 0 \quad \text{and} \quad \vec{A}(\vec{r}, t) = -\vec{E}(\vec{r})t, \tag{6.6}$$

as you can readily check through equations (6.2) and (6.1):

$$\begin{aligned}
\vec{E}(\vec{r}, t) &= -\vec{\nabla}V(\vec{r}, t) - \frac{\partial \vec{A}(\vec{r}, t)}{\partial t} = -\vec{\nabla}0 - \frac{\partial(-\vec{E}(\vec{r})t)}{\partial t} = \vec{E}(\vec{r}) \\
\vec{B}(\vec{r}, t) &= \vec{\nabla} \times \vec{A}(\vec{r}, t) = -t\vec{\nabla} \times \vec{E}(\vec{r}) = 0.
\end{aligned}$$

I imagine walking into the office of the president of Eveready Battery Company: "I purchased a battery advertised as 9 volts. But the voltage is zero everywhere! Furthermore [pointing to the minus sign in equation (6.6)] the vector potential has been going down ever since I bought it!"

## Coulomb gauge

Well, the Weyl gauge is great for jokes, but can we do something more profitable with our new power? Recall from equations (6.3) and (6.4) that the equations for the potentials are

$$\vec{\nabla}^2 V + \frac{\partial}{\partial t}\vec{\nabla} \cdot \vec{A} = -\rho/\epsilon_0, \tag{6.7}$$

$$\vec{\nabla}^2 \vec{A} - \mu_0\epsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} - \vec{\nabla}\left(\vec{\nabla} \cdot \vec{A} + \mu_0\epsilon_0 \frac{\partial V}{\partial t}\right) = -\mu_0 \vec{J}. \tag{6.8}$$

When we wrote down the first equation we said it was "a little more involved than Poisson's equation", because of that $\vec{\nabla} \cdot \vec{A}$ piece. Is there some gauge in which we can get rid of it and turn it exactly into Poisson's equation?

Suppose that, by hook or by crook, we have found a set of potentials $(V(\vec{r}, t), \vec{A}(\vec{r}, t))$ for our situation. Now we look for a new set of potentials for which $\vec{\nabla} \cdot \vec{A}' = 0$.

$$\begin{aligned}
\vec{A}'(\vec{r}, t) &= \vec{A}(\vec{r}, t) + \vec{\nabla}\lambda(\vec{r}, t) \\
\vec{\nabla} \cdot \vec{A}'(\vec{r}, t) &= \vec{\nabla} \cdot \vec{A}(\vec{r}, t) + \vec{\nabla}^2\lambda(\vec{r}, t)
\end{aligned}$$

We desire for the left hand side to vanish, and it will do so when

$$\vec{\nabla}^2\lambda(\vec{r}, t) = -\vec{\nabla} \cdot \vec{A}(\vec{r}, t).$$

To find $\lambda(\vec{r}, t)$, we have to solve Poisson's equation using, as a "source", $-\vec{\nabla} \cdot \vec{A}(\vec{r}, t)$. This might be hard to do, but we have a theorem guaranteeing that such a solution exists.

In this new gauge (dropping the primes),

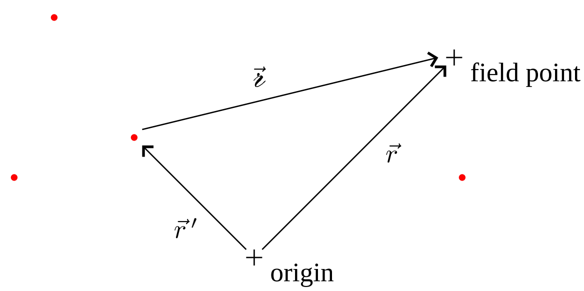
$$\vec{\nabla} \cdot \vec{A}(\vec{r}, t) = 0$$

so the scalar potential is generated exactly by Poisson's equation of electrostatics

$$\vec{\nabla}^2 V(\vec{r},t) = -\rho(\vec{r},t)/\epsilon_0.$$

But from electrostatics we know the solution for the potential:

$$V(\vec{r},t) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\vec{r}\,',t)}{\text{𝓇}}\, d^3r'.$$




On the other hand, the vector potential is particularly difficult to find in this gauge!

One peculiarity of this gauge demands investigation. If a source charge moves at $\vec{r}\,'$, the potential $V(\vec{r})$ at the field point $\vec{r}$ changes. . . instantly! The field point might be light-years away, but the potential changes the instant that the source charge moves. How can this be?

When the source charge moves, both the scalar potential and the vector potential at the field point change instantly, but they change in such a way that the resulting electric and magnetic fields at the field point do *not* change instantly. They don't change until the information reaches the field point, traveling at light speed. (This claim will be proven on page 64.) This underscores the role of potentials as mathematical tools that help enormously in making calculations but that are not themselves physically observable. (You know from way back in introductory circuits that a voltmeter doesn't measure absolute potential, it can measure only potential differences.) In the Coulomb gauge $V(\vec{r},t)$ and $\vec{A}(\vec{r},t)$ propagate instantly, but all gauges produce the same $\vec{E}(\vec{r},t)$ and $\vec{B}(\vec{r},t)$, and these fields propagate at light speed.

⟦This is my favorite analogy for wavefunction in quantum mechanics. When a measurement is made the wavefunction "collapses" instantaneously, but that's allowed because wavefunction is not itself physically observable: it is nothing more than a computational tool.⟧

This gauge is often used in quantum electrodynamics, and it's used far from charges and currents so that $V(\vec{r},t) = 0$.

It is called the "Coulomb gauge" despite the facts that (*i*) Charles-Augustin de Coulomb knew nothing about it (my limited historical research suggests it was introduced by J. Willard Gibbs) and that (*ii*) Coulomb's law

$$\vec{E}(\vec{r},t) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\vec{r}\,',t)}{\text{\char"0D}^2}\,\hat{\text{\char"0D}}\,d^3r'$$

doesn't hold!!

## Lorenz gauge

Let's look again at equations (6.7) and (6.8). Instead of trying to make the first equation simple at the expense of the second, can we make the second simple?

Yes we can, by selecting a gauge in which

$$\vec{\nabla}\cdot\vec{A} + \mu_0\epsilon_0\frac{\partial V}{\partial t} = 0.$$

(You can find the appropriate $\lambda(\vec{r},t)$ yourself, if you wish.) In this gauge equations (6.7) and (6.8) become

$$\vec{\nabla}^2 V - \mu_0\epsilon_0\frac{\partial^2 V}{\partial t^2} = -\rho/\epsilon_0, \tag{6.9}$$

$$\vec{\nabla}^2 \vec{A} - \mu_0\epsilon_0\frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0\vec{J}. \tag{6.10}$$

This pair has a very pleasing symmetry between $V$ and $\vec{A}$. Furthermore (and in contrast to the general equations (6.3) and (6.4)) the equation for $V$ does not involve $\vec{A}$, and vice versa. In our attempt to make the second equation simple, we have in fact made both of them simple!

Recall that

$$\vec{\nabla}^2 f(\vec{r},t) - \frac{1}{c^2}\frac{\partial^2 f(\vec{r},t)}{\partial t^2} = 0$$

was called "the wave equation". Equations like

$$\vec{\nabla}^2 f(\vec{r},t) - \frac{1}{c^2}\frac{\partial^2 f(\vec{r},t)}{\partial t^2} = \rho(\vec{r},t)$$

are called "inhomogenious wave equations" or "wave equations with the source $\rho(\vec{r},t)$".

This gauge was invented by the Dane Ludvig V. Lorenz and is called the Lorenz gauge. Unfortunately Lorenz's last name is so similar to that of the famed Dutch theorist Hendrik A. Lorentz that for many years it was misnamed the "Lorentz gauge". (This tendency to name things, not after the inventor, but after someone already famous is called "the Matthew effect" from the Biblical passage Matthew 25:29: "For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away.") You should not use this misnomer. I'm telling you because you might look up the gauge in an older book and find the wrong name.

This gauge is so convenient that for the rest of the course we will use it exclusively.

# Chapter 7

# Resultant Potentials and Fields

**Statics:** You have a bunch of charges $\rho(\vec{r}\,')$ and currents $\vec{J}(\vec{r}\,')$. What are the resulting potentials at point $\vec{r}$? (Define $\vec{\mathscr{r}} = \vec{r} - \vec{r}\,'$.) You know the answer:

| Governing equation | Resulting solution |
|---|---|
| $\vec{\nabla}^2 V(\vec{r}) = -\rho(\vec{r})/\epsilon_0$ | $V(\vec{r}) = \dfrac{1}{4\pi\epsilon_0}\displaystyle\int \dfrac{\rho(\vec{r}\,')}{\mathscr{r}}\, d^3r'$ |
| $\vec{\nabla}^2 \vec{A}(\vec{r}) = -\mu_0 \vec{J}(\vec{r})$ | $\vec{A}(\vec{r}) = \dfrac{\mu_0}{4\pi}\displaystyle\int \dfrac{\vec{J}(\vec{r}\,')}{\mathscr{r}}\, d^3r'$ |

**Dynamics:** What if these source charges $\rho(\vec{r}\,', t)$ and currents $\vec{J}(\vec{r}\,', t)$ are changing? A good guess is that the potentials are *retarded*... that is the effect of a charge or current propagates away from that source at the speed of light $c$. Thus the potential at point $\vec{r}$ at time $t$ will depend not upon the sources at $\vec{r}\,'$ now, but upon the sources that were there some time ago, at the "retarded" time

$$t_r = t - \frac{\mathscr{r}}{c}.$$

If that guess holds, then the answer would be:

| Governing equation | Resulting solution |
|---|---|
| $\vec{\nabla}^2 V(\vec{r}, t) - \dfrac{1}{c^2}\dfrac{\partial^2 V(\vec{r}, t)}{\partial^2 t} = -\rho(\vec{r}, t)/\epsilon_0$ | $V(\vec{r}, t) = \dfrac{1}{4\pi\epsilon_0}\displaystyle\int \dfrac{\rho(\vec{r}\,', t_r)}{\mathscr{r}}\, d^3r'$ |
| $\vec{\nabla}^2 \vec{A}(\vec{r}, t) - \dfrac{1}{c^2}\dfrac{\partial^2 \vec{A}(\vec{r}, t)}{\partial^2 t} = -\mu_0 \vec{J}(\vec{r}, t)$ | $\vec{A}(\vec{r}, t) = \dfrac{\mu_0}{4\pi}\displaystyle\int \dfrac{\vec{J}(\vec{r}\,', t_r)}{\mathscr{r}}\, d^3r'$ |

Does our guess hold true? Griffiths proves (pages 445–446) that it does.

There are two equivalent ways to look at these retarded solutions. The first ("source centric") way is that the sources make $V$ and $\vec{A}$ and those effects propagate outward from the sources at point $\vec{r}\,'$ at light speed. The second ("field point centric") way is to start at the field point $\vec{r}$. To calculate the field there, think of a spherical shell centered on $\vec{r}$, expanding at light speed but going backwards in time. When this shell touches a source, that's the source affecting the field point at time $t$. The first approach focuses on a spherical shell expanding from each source point going forward in time, the second focuses on a spherical shell expanding from each field point going backward in time.

You might think this is an obvious generalization, but (*i*) it holds only for potentials in the Lorenz gauge and (*ii*) it *doesn't* hold for fields. The fields are *not* simply retarded. In fact, to find the fields you need to use equations (6.1) and (6.2), namely

$$\vec{E}(\vec{r},t) = -\vec{\nabla}V(\vec{r},t) - \frac{\partial \vec{A}(\vec{r},t)}{\partial t} \quad \text{and} \quad \vec{B}(\vec{r},t) = \vec{\nabla} \times \vec{A}(\vec{r},t).$$

This is more difficult to do than you might think (Griffiths pages 449–450), but the answers are

$$\vec{E}(\vec{r},t) = \frac{1}{4\pi\epsilon_0}\int \left[ \frac{\rho(\vec{r}\,',t_r)}{\mathscr{r}^2}\,\hat{\mathscr{r}} + \frac{\dot{\rho}(\vec{r}\,',t_r)}{c\,\mathscr{r}}\,\hat{\mathscr{r}} - \frac{\dot{\vec{J}}(\vec{r}\,',t_r)}{c^2\,\mathscr{r}} \right] d^3r' \tag{7.1}$$

$$\vec{B}(\vec{r},t) = \frac{\mu_0}{4\pi}\int \left[ \frac{\vec{J}(\vec{r}\,',t_r)}{\mathscr{r}^2} + \frac{\dot{\vec{J}}(\vec{r}\,',t_r)}{c\,\mathscr{r}} \right] \times \hat{\mathscr{r}}\, d^3r'. \tag{7.2}$$

These are the Jefimenko equations (after Oleg Dmitrovich Jefimenko of West Virginia University in Morgantown).

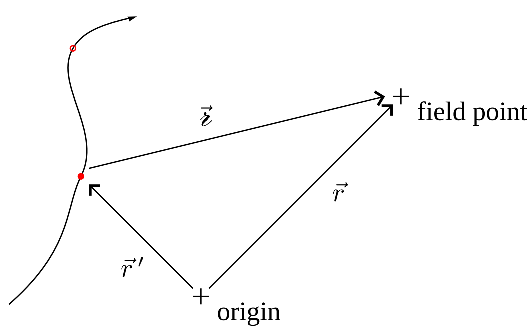
It is clear that these fields are "retarded" — that is the electromagnetic effects of a source propagate away from that source at light speed: no faster, no slower. These same fields would be derived from any gauge, but it's easiest to find them in the Lorenz gauge. This is the source of my claim on page 61 that in the Coulomb gauge the potentials propagate instantaneously, but the fields do not.

These equations exhibit surprising, perhaps shocking, behavior. From way back at the Coulomb and Biot-Savart laws, we've grown used to the idea that electric and magnetic fields fall off like $1/(\text{distance})^2$. And sure enough two terms in the Jefimenko equations do just that. If the situation is static ($\dot{\rho} = 0$, $\dot{\vec{J}} = 0$) these are the only terms that exist. But if the situation is dynamic, then there are three terms that fall off much more slowly, like $1/(\text{distance})$. These terms resulting from $\dot{\rho}$ and $\dot{\vec{J}}$ are the EM fields involved in light and radio propagation. Because they fall off like $1/(\text{distance})$, not $1/(\text{distance})^2$, they are much more effective in spreading over large distances. This is why we use electrodynamics, not electrostatics, to send messages over long distances (e.g., via radio).

Scientists who work in detecting gravity waves (LIGO, VIRGO, NANOGRAV, LISA) like to boast about gravity wave observatories: "Light falls off like $1/r^2$, but gravitational field from gravity waves falls off like $1/r$, so we can see farther into space than optical observatories can." Well in fact, the electric field due to light radiation and the gravitational field due to gravitational radiation both fall off like $1/r$. The difference is that optical observatories measure the *energy* carried by light, whereas gravity wave observatories measure the *field* carried by the gravity wave. If we detected the $\vec{E}$ field associated with light, rather than the energy associated with light, we would find that it falls off like $1/r$ too.

## 7.1 Single moving point charge

A single point particle with charge $q$ travels along a trajectory. What potentials and fields does it produce?




The open red dot represents the location of the charged particle *now*. More important is the solid red dot which represents the location of the charged particle back at the retarded time,

$$t_r = t - \frac{\mathscr{r}}{c}.$$

The retarded potential is

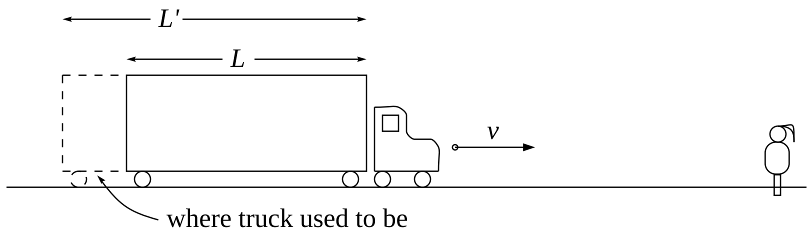
$$V(\vec{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\vec{r}\,', t_r)}{\mathscr{r}}\, d^3r'$$

and I'd think that this would evaluate to

$$\frac{1}{4\pi\epsilon_0} \frac{q}{\mathscr{r}}.$$

I'm wrong, for a subtle reason involving the spherical shell expanding as it moves backward in time.

To make things easier we consider, first treat the apparent volume of a truck cargo box, not of a spherical shell expanding backwards in time.




Jessica stands on the sidewalk as a truck travels down the road at speed $v$. The light from the front of the cargo box takes some time to reach Jessica's eye, but the light from the back takes still more time. Thus she sees the cargo box as longer than it really is. How much longer? If $L$ is the true length of the cargo box and $L'$ the apparent length, then the truck moves a distance $L' - L$ while the light moves a distance $L$:

$$\begin{aligned}
\frac{L' - L}{v} &= \frac{L'}{c} \\
L' &= \frac{L}{1 - v/c} \\
L' &= \frac{L}{1 - \hat{\mathscr{r}} \cdot \vec{v}/c} \\
\text{apparent volume of cargo box} &= \frac{\text{true volume of cargo box}}{1 - \hat{\mathscr{r}} \cdot \vec{v}/c} \\
d^3r' &= \frac{d^3r}{1 - \hat{\mathscr{r}} \cdot \vec{v}/c}
\end{aligned}$$

Thus

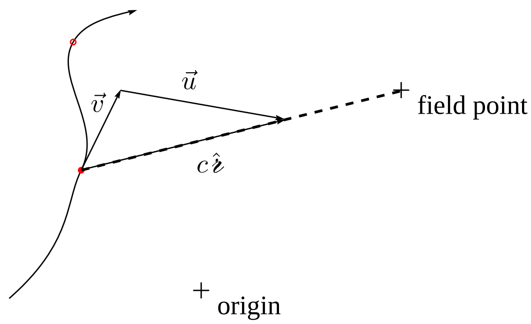
$$V(\vec{r}, t) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\vec{r}\,', t_r)}{\mathscr{r}} \, d^3r' = \frac{1}{4\pi\epsilon_0} \frac{q}{\mathscr{r}} \frac{1}{1 - \hat{\mathscr{r}} \cdot \vec{v}/c} \tag{7.3}$$

where all the quantities on the right are evaluated at the retarded time $t_r$. Similarly for the vector potential

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \frac{q\vec{v}}{\mathscr{r}} \frac{1}{1 - \hat{\mathscr{r}} \cdot \vec{v}/c} \tag{7.4}$$

where again all the quantities on the right are evaluated at the retarded time. These two equations are the Liénard–Wiechert potentials.

⟦The French mining engineer Alfred-Marie Liénard derived these potentials in 1898. The German geophysicist Emil Wiechert derived them independently in 1900. Liénard was not related to Philipp Lenard, who won the Nobel Prize for his work with electron beams and who was appointed "Chief of Aryan Physics" by Adolf Hitler.⟧




The fields associated with these potentials are, where $\vec{u} = c\hat{\mathscr{r}} - \vec{v}$,

$$\vec{E}(\vec{r},t) = \frac{q}{4\pi\epsilon_0}\frac{\mathscr{r}}{(\vec{\mathscr{r}}\cdot\vec{u})^3}[(c^2-v^2)\vec{u} + \vec{\mathscr{r}}\times(\vec{u}\times\vec{a})] \tag{7.5}$$

$$\vec{B}(\vec{r},t) = \frac{1}{c}\vec{\mathscr{r}}\times\vec{E}(\vec{r},t) \tag{7.6}$$

all quantities on the right of the first equation being evaluated at the retarded time. (I noted back on page 6 that the familiar Coulomb and Biot-Savart laws do *not* hold in electrodynamics. What holds instead? These two equations.)

The challenging thing about these two formulas is not in their derivation — which is straightforward if laborious — but in making sense of the result. What do these formulas tell us about nature?

First, make sure that they give the proper electrostatic result. If $\vec{a} = 0$ and $\vec{v} = 0$, then $\vec{u} = c\hat{\mathscr{r}}$ so

$$\begin{aligned}
\vec{E}(\vec{r},t) &= \frac{q}{4\pi\epsilon_0}\frac{\mathscr{r}}{(\vec{\mathscr{r}}\cdot(c\hat{\mathscr{r}}))^3}[(c^2)c\hat{\mathscr{r}}] = \frac{q}{4\pi\epsilon_0}\frac{1}{\mathscr{r}^2}\hat{\mathscr{r}} \\
\vec{B}(\vec{r},t) &= 0
\end{aligned}$$

So far so good.

I like to break equation (7.5) down into two parts: the electric field $\vec{E}^{(za)}(\vec{r},t)$ that would be at point $\vec{r}$ if there were zero acceleration, and the electric field $\vec{E}^{(a)}(\vec{r},t)$ due to acceleration:

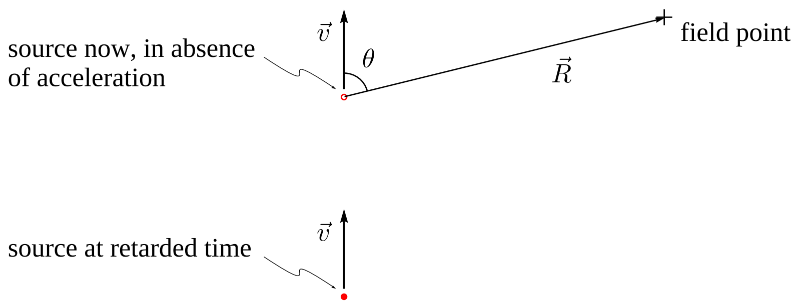
$$\vec{E}(\vec{r},t) = \vec{E}^{(za)}(\vec{r},t) + \vec{E}^{(a)}(\vec{r},t) \tag{7.7}$$

$$\vec{E}^{(za)}(\vec{r},t) = \frac{q}{4\pi\epsilon_0}\frac{\mathscr{r}}{(\vec{\mathscr{r}}\cdot\vec{u})^3}(c^2-v^2)\vec{u} \tag{7.8}$$

$$\vec{E}^{(a)}(\vec{r},t) = \frac{\vec{\mathscr{r}}\times(\vec{E}^{(za)}\times\vec{a})}{(c^2-v^2)}. \tag{7.9}$$

Griffiths investigates the constant velocity result $\vec{E}^{(za)}$ in his example 10.4. He finds that if $\vec{R}$ is the vector from the location where the charge is *now* (not its retarded location) to the field point, then

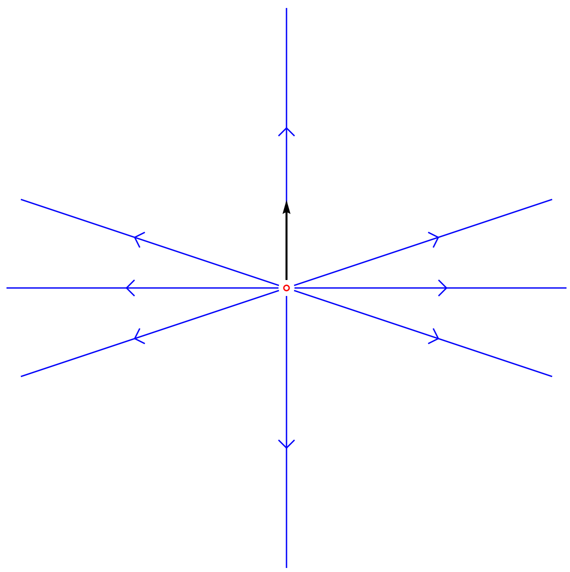
$$\vec{E}^{(za)} = \frac{q}{4\pi\epsilon_0}\frac{1-(v/c)^2}{[1-(v\sin\theta/c)^2]^{3/2}}\frac{\hat{R}}{R^2}. \tag{7.10}$$




What does this equation tell us? One surprising feature is that the field is radial from the source's *current* location, not from its retarded location. Once you're over this surprise, the equation looks remarkably like Coulomb's law. There is an obvious and necessary cylindrical symmetry. There is a less obvious front-back reflection symmetry: Because $\sin(\theta) = \sin(180° - \theta)$, the field of a particle with velocity $-\vec{v}$ is exactly the same as that of a particle with velocity $+\vec{v}$.

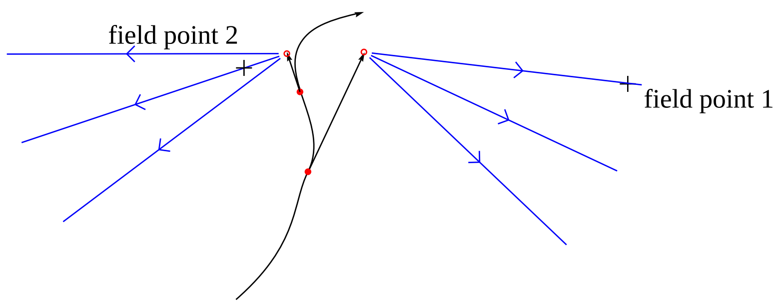
The field immediately in front of the particle ($\theta = 0$) is weaker than the Coulomb field by a factor of $1 - (v/c)^2$. As $\theta$ increases the field becomes stronger until, at $\theta = 90°$, the field is stronger than the Coulomb field by a factor of $1/\sqrt{1-(v/c)^2}$. As $\theta$ increases still further, the field declines until, immediately behind the particle, it is equal to the magnitude immediately in front.

All this information is summarized in a field line diagram that I call “cat whiskers”:



I think of it this way: Gauss’s Law says that the number of electric field lines launched by a charge $q$ is always the same, regardless of velocity. When the velocity is zero, the field lines necessarily spread out uniformly. When the velocity increases, they drift away from the bow and wake of the particle toward “midships” (left and right of the velocity).

So now we have a global picture for $\vec{E}^{(za)}$. Can we use this, accompanied by equation (7.9), to obtain a global picture for $\vec{E}^{(a)}$? In general no. The problem is that at each field point we would need a different set of cat whiskers. For example in the figure below, if we wanted to find the field at point 1, far from the source charge trajectory, we would need to use the cat whiskers from some time in the distant past. But at point 2, close to the source charge trajectory, we would need to use the cat whiskers from some time in the recent past. Every point needs to employ a different set of cat whiskers, and the task quickly degenerates into a morass for all but the simplest situations.

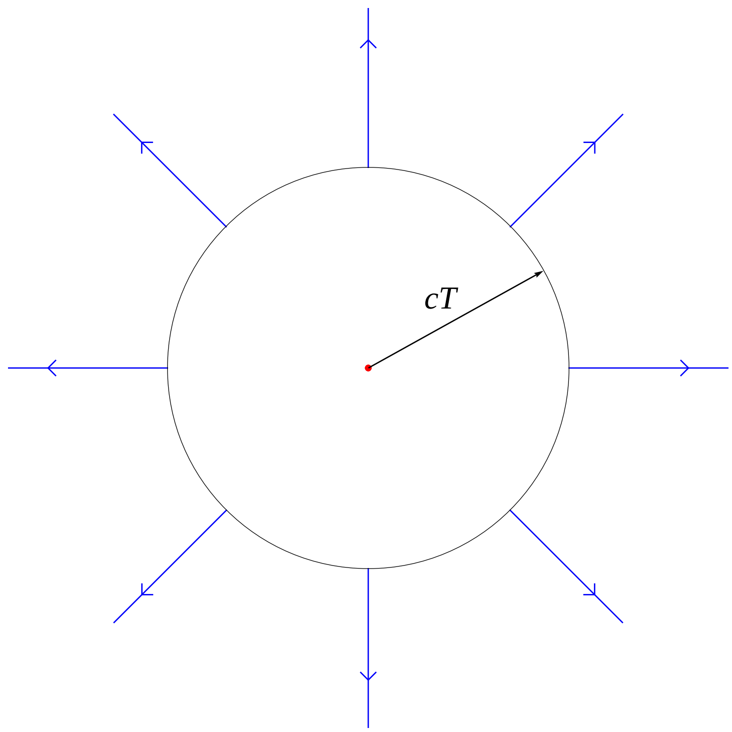



Before giving up completely, I do want to look at one qualitative feature. The cat whiskers equation (7.10) shows the electric field due to a constant velocity charge falling off like $1/(\text{distance})^2$. But the equation for $\vec{E}^{(a)}$ from an accelerating charge (7.9) multiplies this $\vec{E}^{(za)}$ by a distance $\vec{\mathscr{r}}$ in the numerator. Admittedly the distance $R^2$ in the denominator is not exactly the same as the distance $\mathscr{r}$ in the numerator, but this at least reinforces our arguments on page 64 that the electric field due to acceleration falls off like $1/(\text{distance})$.

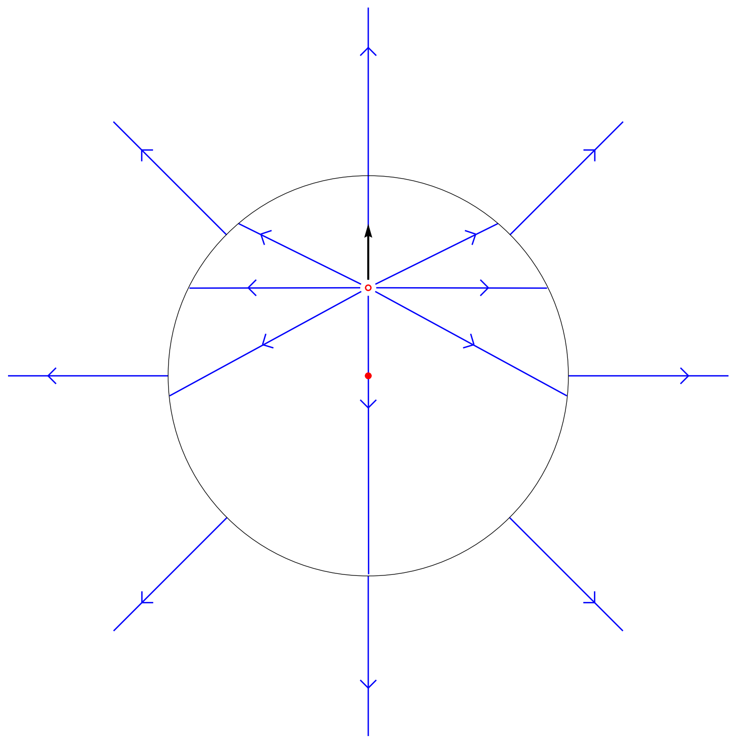
## 7.2 Fields from a good swift kick

The task of finding the electric field everywhere "quickly degenerates into a morass for all but the simplest situations." All right then, let's not give up, let's look at a simple situation.

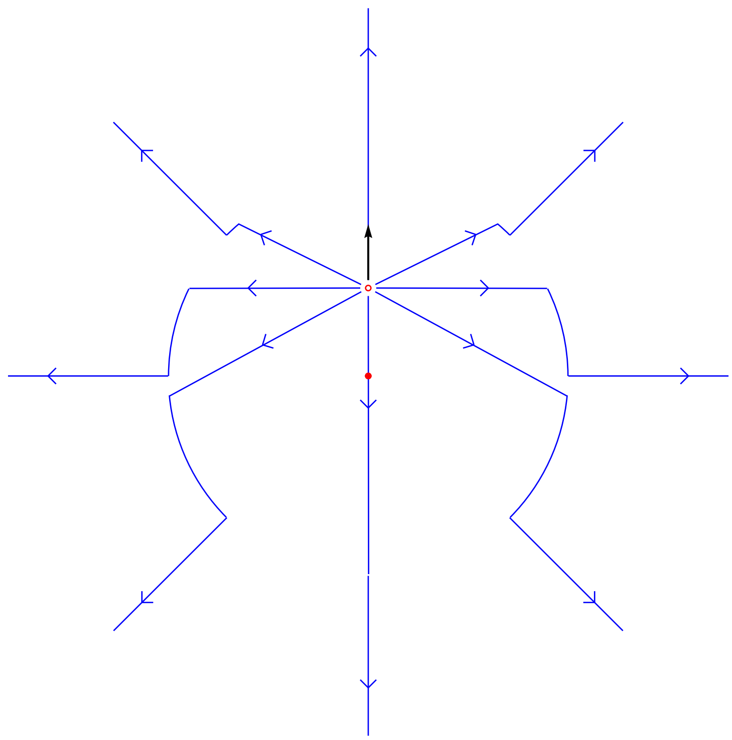
A charge is stationary until given a good swift kick that accelerates it instantly to velocity $\vec{v}$. Then it maintains that velocity without change for a time $T$. All points outside of sphere of radius $cT$ haven't yet gotten the message that the charge has gotten the kick. The electric field outside that sphere is just the Coulomb field pointing at where the charge had been before the kick.



What about points inside the sphere? It's resonable to suppose that the field lines inside are cat whiskers.



How do the field lines inside match up with those outside? They can only connect through field lines on the surface of the expanding sphere.



Can we make these observations quantitative? Look at equation (7.9),

$$\vec{E}^{(a)}(\vec{r},t) = \frac{\vec{\boldsymbol{\imath}} \times (\vec{E}^{(za)} \times \vec{a})}{c^2 - v^2}.$$

Here $\vec{a}$ and $v$ refer to the retarded source point, $\vec{E}^{(za)}$ to the field point, and $\vec{\boldsymbol{\imath}}$ the vector from the retarded source point to the field point.

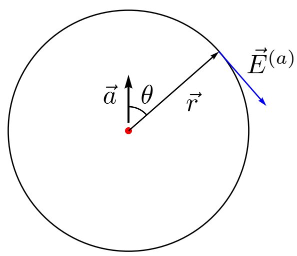
On the outer edge of the expanding shell, $\vec{\boldsymbol{\imath}} = \vec{r}$, $v = 0$, and

$$\vec{E}^{(za)} = \frac{q}{4\pi\epsilon_0}\frac{\hat{r}}{r^2}.$$

Thus

$$\vec{E}^{(a)} = \frac{q}{4\pi\epsilon_0}\frac{1}{c^2r^2}\vec{r} \times (\hat{r} \times \vec{a}).$$

The direction is tangential to the expanding shell,




and the magnitude is

$$|\vec{E}^{(a)}| = \frac{q}{4\pi\epsilon_0}\frac{a\sin\theta}{c^2 r}.$$

Everything about this formula makes sense. The prefactor $q/4\pi\epsilon_0$ is expected for an electric field. It has the $1/r$ character that we've been harping on for radiation fields. It is directly proportional to $a\sin\theta$, which is the amount of acceleration perpendicular to $\vec{r}$. (If you peer down $-\vec{r}$ from the field point toward the accelerating charge, you won't see the part of $\vec{a}$ parallel to $\vec{r}$... you only see the perpendicular part, with magnitude $a\sin\theta$). And if the equation with all those characteristics is to have the proper dimensions, there must be a constant factor with dimensions $1/(\text{speed})^2$.

This electric field is perpendicular to the direction of propagation. It's light!

## 7.3 Radiation simulations

I have solved enough electrostatics problems that I have a "gut" feel for them and I can usually guess the qualitative character of electrostatic field before I solve the problem mathematically. I'm usually but not always right. When I'm not right I have fun figuring out why my gut feel is wrong, and that figuring helps refine my feel.

But I have a harder time getting a gut feel for radiation problems. My best approach is to look at an arbitrary trajectory as a sum of good swift kicks, but I really need a computer simulation to get a good qualitative picture. I recommend you spend time looking at these two simulations:

1. The PhET simulation "Radiating Charge" at

   https://phet.colorado.edu/en/simulation/legacy/radiating-charge.

   (Try giving a stationary charge a good swift kick to check out the conclusions of the previous section.)

2. Bala Juluir's simulation of dipole radiation at

   https://juluribk.com/radiation-from-dipole.html.

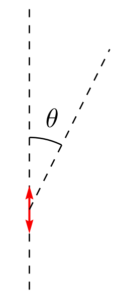
   (Note that the electric field lines always loop! We are a long distance from electrostatics.)

What do I take away from these simulations? That electric field lines act like furry rubber bands.

## 7.4 Radiation from an electric dipole

Why would anyone be interested in the artificial situation of a pure dipole oscillating with a single frequency? Surely any charge configuration encountered in nature will be more complicated than two equal point charges of opposite sign! The reason is that for many charge configurations the dominant portion of the radiated EM field is dipolar. (Furthermore, once we understand the math of finding the radiation from a dipole, it's (relatively) easy to go back and redo the calculation for quadrupoles, octupoles, etc.)

An oscillating electric dipole points always along the same axis and has dipole moment $p_0 \cos(\omega t)$. Call the axis $z$ and locate the coordinate origin at the center of the dipole.




At equation (11.20), Griffiths finds[1] that at detection point $\vec{r}$ the current of electromagnetic energy (power radiated/cross-sectional area) produced by this dipole is

$$\vec{S}(\vec{r}, t) = \frac{\mu_0}{16\pi^2} \frac{p_0^2 \omega^4}{cr^2} \sin^2\theta \cos^2[\omega(t - r/c)]\hat{r}.$$

I don't want to derive this result; instead I want to dissect it to uncover what it tells us about nature.

The radiation is outbound (in the direction $\hat{r}$) and has frequency $\omega$. As time proceeds, this energy current varies from 0 to some maximum and then back to zero and so forth. This is expected behavior and (for almost all frequencies) it will be hard to follow those rapid variations. Instead, focus on the time-averaged[2] energy current of

$$\langle S \rangle = \frac{\mu_0}{32\pi^2} \frac{p_0^2 \omega^4}{cr^2} \sin^2\theta.$$

**(1) Dependence on distance $r$.** It makes perfect sense that the power should decrease like $1/r^2$: For radiated fields the electric field behaves like $1/r$, the magnetic field behaves like $1/r$, so the Poynting vector $\vec{S} = (1/\mu_0)\vec{E} \times \vec{B}$ should behave like $1/r^2$.

Furthermore, the $1/r^2$ behavior is expected from energy conservation. Imagine two spherical shells surrounding the dipole. In a steady state the total power passing

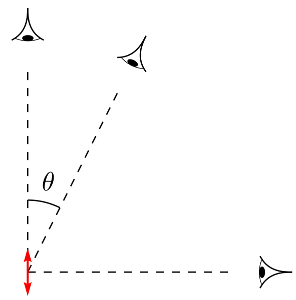
[1] After making three reasonable approximations saying that the detection point is far from a small dipole.

[2] The time average of $\cos(\omega t + \phi)$ over a period is zero. The time average of $\cos^2(\omega t + \phi)$ over a period is $\frac{1}{2}$. I know I always say (quoting from the syllabus) "In writing your solutions, do *not* just write down the final answer. Show your reasoning and your intermediate steps. Describe (in words) the thought that went into your work as well as describing (in equations) the mathematical manipulations involved." But by this point in your education you are suppose to know these facts without having to derive them.

through the inner shell must equal the total power passing through the outer shell. Thus the power per area must obey

$$S = \frac{\text{total power}}{4\pi r^2} \sim \frac{1}{r^2}.$$

**(2) Dependence on angle $\theta$.**




If you view the dipole from the right, you see a tall dipole of height $p_0$. But if you view it from above, you look down and see but a single point! If you view it looking down the axis at angle $\theta$, you see an intermediate dipole of height $p_0 \sin\theta$. So it makes sense that the electric field would vary as the dipole height, $E^{(a)} \sim p_0 \sin\theta$, and that the power would vary as $S \sim p_0^2 \sin^2\theta$.

**(3) Dependence on frequency $\omega$.** It certainly makes sense that higher frequencies, hence higher accelerations, would radiate more power. But can we understand why the power goes up like $\omega^4$ rather than some other power?

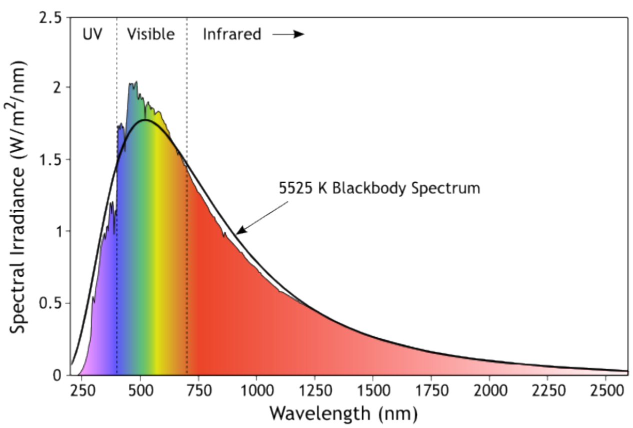
We can. Simple harmonic motion has

$$\begin{aligned} x(t) &= x_0 \cos(\omega t) \\ v(t) &= -\omega x_0 \sin(\omega t) \\ a(t) &= -\omega^2 x_0 \cos(\omega t) \end{aligned}$$

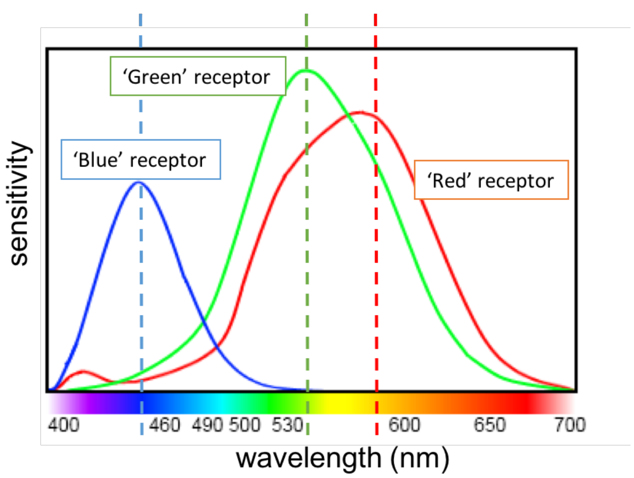
But we know (from equation 7.9) that the radiated electric field is proportional to acceleration $a$. Hence the power, proportional to electric field squared, is proportional to $\omega^4$.

This is a big difference. Even within the narrow band of visible light, in going from violet light ($\lambda = 380$ nm) to red light ($\lambda = 700$ nm), $\omega$ decreases by a factor of $380/700 = 0.54$ so $\omega^4$ decreases by a factor of 0.087.

**Here is one consequence:** On the Moon, the sky is dark. But on Earth, the sky is bright because the atmosphere scatters sunlight. Sunlight streams through the atmosphere. The oscillating electric field in that sunlight induces oscillating dipoles in the nitrogen, oxygen, and other molecules in the atmosphere. (In the absence of light, the nitrogen molecule has zero dipole moment.) Because the sunlight oscillates at frequencies ranging from fast (violet) to slow (red), the molecular dipoles oscillate in that same frequency range. The oscillating dipoles themselves radiate, explaining why the sky on Earth is bright. We've just seen that the fast oscillators (short wavelength, violet) radiate far more than the slow oscillators (long wavelength, red) so you might expect the sky to be violet. However the sun doesn't radiate strongly in the violet (380–450 nm)




and the human eye is not particularly sensitive in the violet



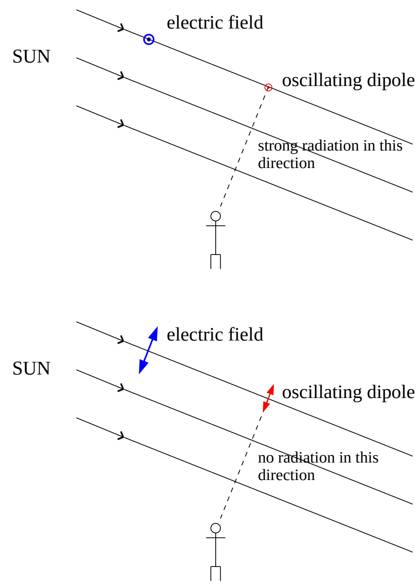

so in fact the sky is blue. Because the blue is removed from sunlight through this scattering process, the light that remains after passing through a lot of atmosphere, as during sunrise and sunset, is red.

*Challenge:* Make reasonable approximations to show that when the sun is directly overhead, sunlight passes through 10 miles of air to reach our eyes, whereas when the sun is setting, it passes through 300 miles of air.

This is one of the things I love about physics. We'll be pushing our way through a thorny patch of dense mathematics, tangled and rough, then suddenly emerge into an open grassy meadow to understand something about the everyday world that has puzzled us since childhood.

But another thing I love about physics is that it not only explains what we already know, it also brings out things we didn't know.

The light from the sun is randomly polarized, switching back and forth rapidly from $x$ polarized to $y$ polarized to 17° polarized (which can be regarded as a superposition of $x$ and of $y$ polarized). [Such light is often called "unpolarized", but it really deserves the name "randomly polarized".] At an instant when the light from the Sun is polarized in-and-out of the page (top of figure), it makes the dipole of a nitrogen molecule oscillate in-and-out of the page. In turn, that oscillating dipole radiates in the direction of the person shown.




At an instant when the light from the Sun is polarized within the page (bottom of figure), the dipole of the nitrogen molecule oscillates within the page. That oscillating dipole radiates, but, as seen above at "(2) Dependence on angle $\theta$", it sends none of that resulting radiation in the direction of the person shown.

Look at the blue sky through a Polaroid. (For safety's sake, don't look directly at the sun.) You will find that the sky is unpolarized near the sun, but highly polarized 90° from the sun. At sunset on a clear evening, this 90° belt is noticeable directly above your head if you look up while wearing Polaroid sunglasses.

Because the clear blue sky is polarized but the white clouds are unpolarized, landscape photographers often use polarizing filters to make white clouds stand out beautifully in a deep blue sky.

**There's a lot more to say about waves and optics in the everyday world.** On 19 July 2004, I was goatpacking with my brother on Grandfather Mountain in Saint Joe National Forest, Idaho. We had a beautiful hike up to a meadow. We were beginning to set up camp at the crest of the meadow when a thunderstorm rolled in. We scurried to finish setting up the tent. Once the tent was erect, one of the goats slipped inside, bringing a lot of water and mud with it. By the time we cleaned up from that goatscapade the cloudburst was almost over. A stunningly beautiful rainbow hung over the meadow to the east, and over the woods and mountains beyond. I grabbed my brother's camera to take a photo of him standing in the meadow with rainbow and mountains in the background. I peered through the camera's viewfinder. My brother was there, and the meadow and the mountains, but no rainbow! I looked over the camera instead of through the viewfinder. The rainbow was there! I looked back through the viewfinder. No rainbow! I remembered that light from a rainbow is polarized. "Bill, do you have a polarizing filter on this camera?" "I don't know," he replied, "I just put on whatever the salesman recommended." So I twisted the camera by 90°, and there was the rainbow in the viewfinder. I snapped the photo.

To give you a glimpse of more topics involved, I show one photo that I took while backpacking on the Na Pali coast of Kauai (part of my effort to go backpacking in each of the fifty states)

and a photo of so-called "supernumerary rainbows" (which I have never personally witnessed)

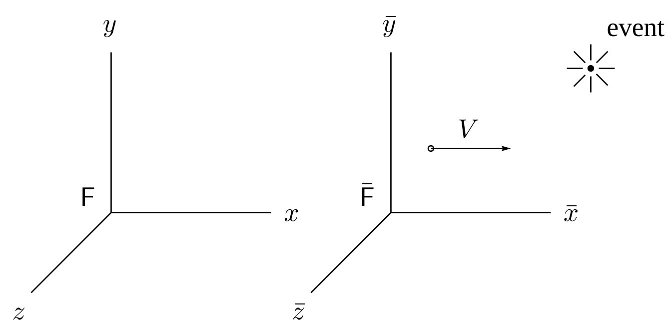
If you like these sorts of questions, I recommend these three books: *Waves* by Frank S. Crawford, Jr. (1968); *Light and Colour in the Open Air* by M.G.J. Minnaert (1940) [also published under titles *The Nature of Light and Colour in the Open Air* and *Light and Color in the Outdoors*]; and *Rainbows, Halos, and Glories* by Robert Greenler (1980). (Bob Greenler's daughter Robin attended Oberlin College, and Bob is the only person to have served twice as the OC Physics Department's Hays lecturer.)

# Chapter 8

# Relativistic Electrodynamics

## 8.1 Space and time in relativity

You know the setup: Reference frame $\bar{\mathsf{F}}$ moves uniformly at speed $V$ to the right past reference frame $\mathsf{F}$. The two reference frames coincide when $t = \bar{t} = 0$. Some event occurs.




How are the event coordinates $(t, x, y, z)$ in frame $\mathsf{F}$ related to the event coordinates $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$ in frame $\bar{\mathsf{F}}$? The answer is

$$\begin{aligned}
\bar{t} &= \frac{t - Vx/c^2}{\sqrt{1-(V/c)^2}} \\
\bar{x} &= \frac{x - Vt}{\sqrt{1-(V/c)^2}} \\
\bar{y} &= y \\
\bar{z} &= z
\end{aligned} \tag{8.1}$$

This is called the "Lorentz transformation".

You know that there are some extremely non-obvious consequences of these four simple equations: time dilation, length contraction, relativity of simultaneity, the transformation of velocity

$$\bar{v}_x = \frac{v_x - V}{1 - v_x V/c^2}, \tag{8.2}$$

the fact that interval

$$(ct)^2 - (x^2 + y^2 + z^2) \tag{8.3}$$

is "invariant" — the same in all reference frames, no causal signal can travel faster than light, no body is infinitely rigid, no body is infinitely strong, the twin paradox, the busted bus, the pole in the barn, etc., etc.[1] I am willing to discuss these consequences with you for as long as you are willing to ask questions.

But eventually we will stop taking about consequences and return to the Lorentz transformation equations themselves. It's pretty clear that these equations become more symmetric and easier to remember if you focus on the product $ct$ rather than on $t$. The variable $ct$ has the dimensions of length, putting it on an even footing with $x$, $y$, and $z$. In addition, we will use the common abbreviations

$$\beta = \frac{V}{c} \quad \text{and} \quad \gamma = \frac{1}{\sqrt{1 - (V/c)^2}}, \tag{8.4}$$

which save a lot of penstrokes. ⟦Use these abbreviations cautiously, however. I remember solving a long and intricate relativity problem and being puzzled by the fact that my solution differed from the solution in the back of the book by a factor of $\gamma^2(1 - \beta^2)$. Try as I might, I just couldn't find my error.⟧

With these cosmetic changes, the Lorentz transformation becomes

$$\begin{aligned} c\bar{t} &= \gamma(ct - \beta x) \\ \bar{x} &= \gamma(x - \beta ct) \\ \bar{y} &= y \\ \bar{z} &= z \end{aligned} \tag{8.5}$$

And this form suggests one last cosmetic change. Call the space-time coordinates of the event

$$(x^0, x^1, x^2, x^3) \equiv (ct, x, y, z) \tag{8.6}$$

and write

$$\begin{aligned} \bar{x}^0 &= \gamma(x^0 - \beta x^1) \\ \bar{x}^1 &= \gamma(x^1 - \beta x^0) \\ \bar{x}^2 &= x^2 \\ \bar{x}^3 &= x^3 \end{aligned} \tag{8.7}$$

[1] The appendix to G.P. Sastry, "Is length contraction really paradoxical?" *American Journal of Physics* **55** (October 1987) 943–946 presents an extensive yet necessarily incomplete catalog of paradoxes in relativity.

*Challenge:* Because of the many apparent paradoxes of relativity, it's not clear that these equations are consistent. Prove to yourself: (*i*) If you transform coordinates from frame $\mathsf{F}$ to frame $\bar{\mathsf{F}}$ moving at velocity $V$ relative to $\mathsf{F}$, and then from frame $\bar{\mathsf{F}}$ to frame $\mathsf{F}$ moving at velocity $-V$ relative to $\bar{\mathsf{F}}$, you get back to your original coordinates. (*ii*) If you transform coordinates from frame $\mathsf{F}_0$ to frame $\mathsf{F}_1$ moving at speed $V_1$ relative to $\mathsf{F}_0$, and then transform those coordinates from frame $\mathsf{F}_1$ to frame $\mathsf{F}_2$ moving at speed $V_2$ relative to $\mathsf{F}_1$, you get the same result as transforming directly from frame $\mathsf{F}_0$ to frame $\mathsf{F}_2$, moving at speed

$$\frac{V_1 + V_2}{1 + V_1 V_2/c^2}$$

relative to $\mathsf{F}_0$. These two items are sufficient to prove that the set of Lorentz transformations constitute a mathematical group, and then group theory assures us that the equations are consistent.

## Four-scalars and four-vectors

A four-scalar is the same in all reference frames. Examples are the mass of a particle, the charge of a particle, and the invariant interval between two events. The time separation between two events is *not* a four-scalar, but the time ticked off by a watch attached to a particle — the so-called "proper time"[2] $\tau$ — is a four-scalar: different reference frames disagree on the time elapsed, but all agree that the watch attached to a particle has ticked off a certain amount of time. (The proper time squared is just the invariant interval between two events, so of course it is a four-scalar.) According to time dilation, the relation between lab time and proper time is

$$dt = \frac{d\tau}{\sqrt{1-(v/c)^2}}. \tag{8.8}$$

The coordinates of a four-vector transform from frame to frame like

$$\begin{aligned}
\bar{a}^0 &= \gamma(a^0 - \beta a^1)\\
\bar{a}^1 &= \gamma(a^1 - \beta a^0)\\
\bar{a}^2 &= a^2\\
\bar{a}^3 &= a^3
\end{aligned}$$

Or in other words

$$\bar{a}^\mu = \sum_{\alpha=0}^{3} \Lambda^\mu{}_\alpha a^\alpha,$$

[2]The word "proper" is irritating. Any inertial frame is as good as any other inertial frame, so why should the time in one of those frames be considered more "proper" than any other time? The word origin is that the "proper" in "proper time" derives not from the English "proper" meaning "respectable, genteel", but from the French "propre" meaning "own". The particle's "proper time" means the particle's "own time".

where the matrix $\Lambda$ with components $\Lambda^{\mu}{}_{\alpha}$ ($\mu$th row; $\alpha$th column) is

$$\begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The most familiar four-vector is the location of an event in space-time $\mathbb{r}$ which is represented in a particular reference frame through

$$\mathbb{r} \doteq [ct, x, y, z] = [ct, \vec{r}]. \tag{8.9}$$

(I like to represent four-vectors using open face symbols. It would be nice if this habit were universal, but it isn't.)

## 8.2 Energy and momentum in relativity

Another four-vector is the four-momentum

$$\mathbb{p} = m\frac{d\mathbb{r}}{d\tau}. \tag{8.10}$$

In some particular inertial frame

$$\mathbb{p} = m\frac{d\mathbb{r}}{d\tau} = m\frac{dt}{d\tau}\frac{d\mathbb{r}}{dt} \doteq \frac{m}{\sqrt{1-(v/c)^2}}[c, v_x, v_y, v_z] = [E/c, p_x, p_y, p_z]. \tag{8.11}$$

## 8.3 Current density in relativity

A charged wind moves through the laboratory. A particular dot of charge has velocity $\vec{v}$. Transform into the reference frame moving along with that dot of charge. (In general, this will involve a rotation as well as a Lorentz boost — a so-called Poincaré transformation.) In this so-called comoving (or proper) reference frame erect a tiny box of volume $V_0$ about the dot. The proper charge density is

$$\rho_0 = \frac{Q}{V_0} \tag{8.12}$$

and of course in this reference frame the current vanishes.

In the laboratory reference frame the tiny box has contracted volume

$$V = \sqrt{1-(v/c)^2}\, V_0,$$

so the lab charge density is

$$\rho = \frac{\rho_0}{\sqrt{1-(v/c)^2}}$$

while the lab current density is

$$\vec{J} = \rho\vec{v} = \frac{\rho_0\vec{v}}{\sqrt{1-(v/c)^2}}.$$

Comparison to the four-momentum

$$\mathbb{p} \doteq [E/c, \vec{p}] = \left[\frac{mc}{\sqrt{1-(v/c)^2}}, \frac{m\vec{v}}{\sqrt{1-(v/c)^2}}\right]$$

leaves no doubt that the four-current

$$\mathbb{J} \doteq [\rho c, \vec{J}] = \left[\frac{\rho_0 c}{\sqrt{1-(v/c)^2}}, \frac{\rho_0\vec{v}}{\sqrt{1-(v/c)^2}}\right] \tag{8.13}$$

transforms as a four-vector.

The continuity equation

$$\vec{\nabla}\cdot\vec{J} = -\frac{\partial\rho}{\partial t}$$

becomes in four-vector notation

$$\sum_{\mu=0}^{3}\frac{\partial J^\mu}{\partial x^\mu} = 0. \tag{8.14}$$

## 8.4 Four-tensors

Just as we generalized ordinary vectors to tensors (page 36), so we can generalize four-vectors to four-tensors. A four-tensor has 16 components

$$\begin{pmatrix} t^{00} & t^{01} & t^{02} & t^{03} \\ t^{10} & t^{11} & t^{12} & t^{13} \\ t^{20} & t^{21} & t^{22} & t^{23} \\ t^{30} & t^{31} & t^{32} & t^{33} \end{pmatrix}$$

that transform from one reference frame to another through

$$\bar{t}^{\mu\nu} = \sum_{\alpha=0}^{3}\sum_{\beta=0}^{3}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta t^{\alpha\beta}. \tag{8.15}$$

Normally statements about tensor components, as opposed to statements about tensors, are frame-dependent. For example if $t^{13} = 7$, then in general $\bar{t}^{13} \neq 7$. But not always. For example the tensor with components $\delta^{\mu\nu}$ has the same components in all frames. Another example is that if the tensor components are symmetric ($t^{\mu\nu} = t^{\nu\mu}$ for all $\mu$, $\nu$) in one reference frame then they are symmetric in all reference frames. Similarly if they are antisymmetric ($t^{\mu\nu} = -t^{\nu\mu}$ for all $\mu$, $\nu$).

The proof starts with equation (8.15). Write it down again, except swap the indices $\mu$ and $\nu$:

$$\bar{t}^{\nu\mu} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \Lambda^{\nu}{}_{\alpha} \Lambda^{\mu}{}_{\beta} t^{\alpha\beta}. \tag{8.16}$$

Now, write this second equation again, but swapping the dummy summation indices $\alpha$ and $\beta$:

$$\bar{t}^{\nu\mu} = \sum_{\beta=0}^{3} \sum_{\alpha=0}^{3} \Lambda^{\nu}{}_{\beta} \Lambda^{\mu}{}_{\alpha} t^{\beta\alpha}. \tag{8.17}$$

Finally, make use of the fact that addition and multiplication of real numbers is commutative:

$$\bar{t}^{\nu\mu} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \Lambda^{\mu}{}_{\alpha} \Lambda^{\nu}{}_{\beta} t^{\beta\alpha}. \tag{8.18}$$

Finally write equations (8.15) and (8.18) right on top of each other:

$$\bar{t}^{\mu\nu} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \Lambda^{\mu}{}_{\alpha} \Lambda^{\nu}{}_{\beta} t^{\alpha\beta}, \tag{8.19}$$

$$\bar{t}^{\nu\mu} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \Lambda^{\mu}{}_{\alpha} \Lambda^{\nu}{}_{\beta} t^{\beta\alpha}. \tag{8.20}$$

It's now plain to see that if $t^{\alpha\beta} = t^{\beta\alpha}$ for all $\alpha$, $\beta$ on the right, then it's also true that $\bar{t}^{\mu\nu} = \bar{t}^{\nu\mu}$ for all $\mu$, $\nu$ on the left. Furthermore, if $t^{\alpha\beta} = -t^{\beta\alpha}$ for all $\alpha$, $\beta$ then $\bar{t}^{\mu\nu} = -\bar{t}^{\nu\mu}$ for all $\mu$, $\nu$.

*Challenge:* The proof also tells us that if $t^{\mu\nu} = 5t^{\nu\mu}$ for all $\mu$, $\nu$ in one reference frame, that this property holds in all reference frames. Does this mean that we should talk about a class of "five-fold tensors"?

A general four-tensor has 16 independent components. But a symmetric four-tensor has 10 independent components. And an antisymmetric four-tensor has four zero components on the diagonal so it has only 6 independent components.

Two things about this theorem: (*i*) It works for ordinary tensors in three-dimensions as well as for four-tensors. (*ii*) It's telling us something about geometry. It's telling us that the property of symmetry or antisymmetry doesn't depend merely on coordinates, but on something coordinate independent. I wish I understood what this something was; I wish I could give you a picture of what a symmetric four-tensor "looks like". For that matter I wish I could give you a picture of what a symmetric three-tensor "looks like". Or what a four-vector "looks like". Alas, I have spend much of my life trying to develop a pictorial sense of tensors and of spacetime, yet I cannot give you any of these things.

## 8.5 Magnetism as a relativistic phenomenon

Griffiths section 12.3.1 shows that magnetism can be regarded as a relativistic phenomenon due to electric charges in motion (currents).

**You will immediately object:** But relativity is important only for objects moving near light speed! Magnetic effects are produced by currents moving at very modest speeds.

Indeed. But the electrostatic force is so huge (on a human scale) that even these tiny relativistic effects are noticeable (on a human scale).

⟦Are electrostatic forces huge on a human scale? Whenever I teach Physics 111, I assign this problem on the first week:

> Two students, Ivan and Veronica, stand about 100 feet apart. Ivan weighs about 200 pounds and Veronica weighs about 100 pounds. Suppose each student has a 0.01% excess in his or her amount of positive and negative charge, one student being positive and the other negative. Estimate the force of attraction between the two.

In my model solution I estimate the charge on the protons in 100 pounds of water (and people are mostly water) as about $25 \times 10^8$ C. Thus

> Force at 100 ft $\approx$ 30 m is
>
> $$(9 \times 10^9 \text{ Nm}^2/\text{C}^2)\frac{(25 \times 10^4 \text{ C})(50 \times 10^4 \text{ C})}{(30 \text{ m})^2} \approx 1.2 \times 10^{18} \text{ N}.$$
>
> Wow! This is, of course, a huge force. If Ivan and Veronica really were attracted to each other with a force like this, they would be pulled toward each other and both would die in the resulting crash. The moral of the story is that Ivan and Veronica do *not* suffer charge excesses of 0.01% ... real life charge excesses are much smaller than this.⟧

**"Derivation" of Maxwell Equations from Coulomb's law plus relativity.** It was once popular[3] to say

$$\text{Coulomb's Law of Electrostatics} \ + \ \text{Relativity} \ \Longrightarrow \ \text{Maxwell Equations}$$

That's not exactly true. If it were true, then

$$\text{Newtons's Law of Gravity} \ + \ \text{Relativity} \ \Longrightarrow \ \text{Maxwell Equations}$$

would hold. And it doesn't. Such discussions shine light on neglected corners of electrodynamics, but in the end they are suggestive. They don't "derive" anything.

Ben Lemberger (in 2014) pointed out that Coulomb's Law is exactly true in classical electrostatics (as far as we know). But Newton's Law of Gravity is not exactly true — it doesn't hold in high field situations, and field becomes high as $r \to 0$.

**Total charge.** Katie Rigdon (in 2018) asked: Why is it that

$$\begin{aligned} \text{total charge} \ &= \ \text{sum of constituent charges} \\ \text{but} \ & \\ \text{total mass} \ &\neq \ \text{sum of constituent masses} \quad ? \end{aligned}$$

It is because the source of EM field is charge, a four-scalar, whereas the source of gravitational field is not mass, a four-scalar, but instead the stress-energy four-tensor, which involves mass, and energy, and momentum, and the flow of mass, energy, and momentum. (This also explains why massless photons fall in gravity.)

[3] Jack R. Tessman, "Maxwell — Out of Newton, Coulomb, and Einstein" *American Journal of Physics* **34**, 1048–1055 (1966). W.G.V. Rosser, *Classical electromagnetism via relativity: An alternative approach to Maxwell's equations* (Butterworth, 1968). A. Nicolaide, "Derivation of the electromagnetic field equations by applying Coulomb's formula and the special theory of relativity" *Archiv für Elektrotechnik* **56**, 156–160 (August 1974). B. Konorski, "Coulomb's law and the theory of relativity — a basis of the entire electrodynamics (Horak's method)" *Przeglad elektrotechniczny* **53**, 428–432 (1 October 1977) (in Polish).

## 8.6 Electromagnetic fields in relativity

Griffiths equation (12.109) shows how fields transform from one reference frame to another:

$$\begin{array}{lll} \bar{E}_x = E_x & \bar{E}_y = \gamma\,(E_y - VB_z) & \bar{E}_z = \gamma\,(E_z + VB_y) \\ \bar{B}_x = B_x & \bar{B}_y = \gamma\left(B_y + \dfrac{V}{c^2}E_z\right) & \bar{B}_z = \gamma\left(B_z - \dfrac{V}{c^2}E_y\right) \end{array}$$

I like to write these with an emphasis not on $\vec{E}$, but on $\vec{E}/c$ (which has the same dimensions as $\vec{B}$), giving

$$\begin{array}{lll} \bar{E}_x/c = E_x/c & \bar{E}_y/c = \gamma\,(E_y/c - \beta B_z) & \bar{E}_z/c = \gamma\,(E_z/c + \beta B_y) \\ \bar{B}_x = B_x & \bar{B}_y = \gamma\,(B_y + \beta E_z/c) & \bar{B}_z = \gamma\,(B_z - \beta E_y/c) \end{array}$$

In the cases of four-vectors for spacetime $\mathbb{r}$ and energy-momentum $\mathbb{p}$ and current density $\mathbb{J}$, we started with a three-vector ($\vec{r}$ and $\vec{p}$ and $\vec{J}$) and tacked on front a three-scalar ($ct$ and $E/c$ and $\rho c$) to make a four-vector. This strategy will not work for $\vec{E}$ and $\vec{B}$. Just look at the transformation properties for $E_x$ and $B_x$!

Instead, the trick is to look at $\vec{E}$ and $\vec{B}$ together. There are six components. And an antisymmetric four-tensor also has six components! Is there any way to shoehorn the six components of $\vec{E}$ and $\vec{B}$ into an antisymmetric four-tensor?

Yes! You have to do some fiddling around, but with sufficient fiddling you find not one but two different ways to perform this shoehorning! The first is the "electromagnetic field four-tensor" $\mathbb{F}$, with components

$$\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ & 0 & B_z & -B_y \\ & & 0 & B_x \\ & & & 0 \end{pmatrix}. \tag{8.21}$$

(I leave blank the elements below the diagonal because they're just the negatives of the elements above the diagonal.) The second is the "dual four-tensor" $\mathbb{G}$, with components

$$\begin{pmatrix} 0 & B_x & B_y & B_z \\ & 0 & -E_z/c & E_y/c \\ & & 0 & -E_x/c \\ & & & 0 \end{pmatrix}. \tag{8.22}$$

Let me inject a few words about general relativity here. (C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation*, pages 220–221.) There is a field four-tensor for gravitational field just as there is for electromagnetic field. It is the rank-4 "Riemann tensor" $R^{\mu}{}_{\nu\sigma\lambda}$. This tensor has $4^4 = 256$ components, but they're not all independent. Because of various symmetries and antisymmetries, there are "only" 20 independent components. (By comparison the electromagnetic field four-tensor has 16 components, of which 6 are independent.)

Just as the 6 independent components of the electromagnetic field four-tensor can be usefully dissected into two three-vector fields with 3 components each (electric and magnetic), so the 20 independent components of the Riemann tensor can be usefully dissected. This dissection produces, among other things, two symmetric, tracefree, three-tensor fields with 5 independent components each. One of the people who figured out how to do this is our own Rob Owen, so if you have questions you should ask him.

## 8.7 Maxwell equations in relativistic notation

You have to fiddle around a lot to write the Maxwell equations in terms of the field and dual four-tensors. When you do, you find:

| The four source equations | The four homogeneous equations |
|---|---|
| $\vec{\nabla} \cdot \vec{E} = \rho/\epsilon_0$ | $\vec{\nabla} \cdot \vec{B} = 0$ |
| $\vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t}$ | $\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ |
| are | are |
| $\sum_{\nu=0}^{3} \frac{\partial F^{\mu\nu}}{\partial x^\nu} = \mu_0 J^\mu$ | $\sum_{\nu=0}^{3} \frac{\partial G^{\mu\nu}}{\partial x^\nu} = 0$ |

All those funny curls and weird derivatives that show up in the traditional form of the Maxwell equations... what were they there for? To make this relativistic result come out so cleanly!

⟦If you don't like using the dual vector, the bottom right equation can be recast as

$$\frac{\partial F_{\mu\nu}}{\partial x^\lambda} + \frac{\partial F_{\nu\lambda}}{\partial x^\mu} + \frac{\partial F_{\lambda\mu}}{\partial x^\nu} = 0.$$

I don't like this form because: ($i$) These look like $4 \times 4 \times 4 = 64$ equations but in fact there are only 4 independent equations. (For example, if $\mu = 2$, $\nu = 2$, and $\lambda = 2$ then this equation reads $0 + 0 + 0 = 0$, which is true but not informative.) ($ii$) To use it, you have to appreciate the distinction between $F^{\mu\nu}$ and $F_{\mu\nu}$ ("covariant components" vs. "contravariant components"). I have learned this distinction a few dozen times and each time forgotten it a few weeks later.[4] I suspect you're the same way. Learn this distinction when you need to use it.⟧

[4]Some people (including, I'm sorry to say, Griffiths on page 526) refer to a "covariant four-vector" and a "contravariant four-vector" instead of to the "covariant components" and "contravariant components" of a single four-vector. This is very bad. An arrow has one set of components in one

So, this is elegant! Physics 111, 311, and 411 is the study of these two equations. But as Einstein pointed out

> "Matters of elegance ought to be left to the tailor and to the cobbler."[5]

and

> "Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore."[6]

In fact, it represents more than elegance. If you write the Maxwell equations in the traditional form shown on page 6, then transform them to the barred frame using the equations on page 91, and if you also remember to transform all the positions and times and charge densities and current densities correctly, you will be surprised to find at the end that all the $\gamma$s and $\beta$s exactly cancel out: the Maxwell equations in the barred frame are exactly the same as the Maxwell equations on page 6. But this way of doing the transformation is clearly a slog. In the four-tensor formulation, it's obvious that the Maxwell equations have the same form in all reference frames.

Maxwell himself didn't understand this. He knew that his equations weren't Galilean invariant, but he thought that meant his equations applied only in the reference frame of the ether.

*Challenge:* The scalar potential $V(\vec{r},t)$ and the vector potential $\vec{A}(\vec{r},t)$ combine to make a four-vector potential $\mathbb{A}(\mathbb{r})$ through

$$\mathbb{A} = [V/c, \vec{A}].$$

Show that the fields are related to the potentials through

$$F^{\mu\nu} = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu}.$$

For example, this equation with $\mu = 0$, $\nu = 1$ reads

$$\frac{E_x}{c} = \frac{\partial A_x}{\partial(-ct)} - \frac{\partial(V/c)}{\partial x} = \frac{1}{c}\left(-\frac{\partial A_x}{\partial t} - \frac{\partial V}{\partial x}\right).$$

And now that it's in this light, the strange combination

$$\vec{E} = -\vec{\nabla}V - \frac{\partial \vec{A}}{\partial t}$$

really does look sort of like a curl.

---

coordinate system and a different set of components in a rotated coordinate system. The arrow (the vector $\vec{r}$) doesn't change when the coordinate system is rotated — only the coordinates change (from $(r_x, r_y, r_z)$ to $(r'_x, r'_y, r'_z)$). That's the whole point of "vector" (see section 2.1, "What is a vector?"). Similarly for the distinction between convariant and contravariant components: The change in components does *not* mean there's a change in four-vector.

[5] Albert Einstein, *Relativity: The Special and the General Theory* (1916) page v.

[6] Statement by Einstein as recalled by Arnold Sommerfeld, "To Albert Einstein's Seventieth Birthday" in *Albert Einstein, Philosopher–Scientist*, edited by Paul A. Schilpp (Library of Living Philosophers, Evanston, Illinois, 1949) page 102.

## 8.8 The stress-energy four-tensor

We've previously discussed the flow of momentum in space, and our discussion resulted in the Maxwell stress tensor. In relativity we have to discuss the flow of four-momentum in space-time. The result will be the stress-energy four-tensor.

As a technical appetizer, I just want to mention that the time-space four-vector expressed in the laboratory frame

$$\mathbb{r} \doteq [ct, \vec{r}] \tag{8.23}$$

has time derivative

$$\frac{d\mathbb{r}}{dt} \doteq [c, \vec{v}]. \tag{8.24}$$

This is *not* a four-vector, because we're taking the derivative with respect to laboratory time $t$ rather than with respect to the four-scaler proper time $\tau$. But we'll soon find the result useful anyway. Now on to the main course.

We build up the stress-energy four-tensor representation in the laboratory frame. Remember that the negative of the Maxwell stress tensor described the flow of electromagnetic momentum. We called the momentum density $\vec{g}$, and the velocity of a plug of fields $\vec{v}$, and described that momentum flow through the tensor

$$\begin{array}{cl} & \text{current density of} \ldots \\ \begin{pmatrix} \phantom{g_xc} & \phantom{u\vec{v}/c} \\ & g_x\vec{v} \\ & g_y\vec{v} \\ & g_z\vec{v} \end{pmatrix} & \begin{array}{l} \\ \leftarrow x\text{-momentum} \\ \leftarrow y\text{-momentum} \\ \leftarrow z\text{-momentum} \end{array} \\ \begin{array}{cc} \phantom{\text{direction}} & \begin{array}{c} \uparrow\uparrow\uparrow \\ \text{move in} \\ \text{direction} \\ x,\ y,\ z \end{array} \end{array} & \end{array}$$

One lesson of relativity is that we can't think of momentum in isolation: we have to consider also the zero component of the energy-momentum four-vector, namely energy/$c$. But we've already talked about the energy density $u$, so we fill in the zero row of this four-tensor as

$$\begin{array}{cl} & \text{current density of} \ldots \\ \begin{pmatrix} \phantom{g_xc} & u\vec{v}/c \\ & g_x\vec{v} \\ & g_y\vec{v} \\ & g_z\vec{v} \end{pmatrix} & \begin{array}{l} \leftarrow \text{energy}/c \\ \leftarrow x\text{-momentum} \\ \leftarrow y\text{-momentum} \\ \leftarrow z\text{-momentum} \end{array} \\ \begin{array}{cc} \phantom{\text{direction}} & \begin{array}{c} \uparrow\uparrow\uparrow \\ \text{move in} \\ \text{direction} \\ x,\ y,\ z \end{array} \end{array} & \end{array}$$

An even earlier lesson of relativity is that we can't think of space in isolation: we have to consider the zero component of the time-space four-vector, namely $ct$. Using equation (8.24), we extend the bottom three rows of this four-tensor into the zero column:

$$\begin{array}{cl} & \text{current density of} \ldots \\ \begin{pmatrix} & u\vec{v}/c \\ g_xc & g_x\vec{v} \\ g_yc & g_y\vec{v} \\ g_zc & g_z\vec{v} \end{pmatrix} & \begin{array}{l} \leftarrow \text{energy}/c \\ \leftarrow x\text{-momentum} \\ \leftarrow y\text{-momentum} \\ \leftarrow z\text{-momentum} \end{array} \\ \begin{array}{cc} \begin{array}{c} \uparrow \\ \text{move in} \\ \text{direction} \\ ct \end{array} & \begin{array}{c} \uparrow\uparrow\uparrow \\ \text{move in} \\ \text{direction} \\ x,\ y,\ z \end{array} \end{array} & \end{array}$$

Finally we fill in the zero-zero component of the four-tensor

$$\begin{array}{cl}
 & \text{current density of} \ldots \\
\begin{pmatrix} u & u\vec{v}/c \\ g_x c & g_x\vec{v} \\ g_y c & g_y\vec{v} \\ g_z c & g_z\vec{v} \end{pmatrix} & \begin{matrix} \leftarrow \text{energy}/c \\ \leftarrow x\text{-momentum} \\ \leftarrow y\text{-momentum} \\ \leftarrow z\text{-momentum} \end{matrix} \\
\begin{matrix} \uparrow & \uparrow\uparrow\uparrow \\ \text{move in} & \text{move in} \\ \text{direction} & \text{direction} \\ ct & x,\ y,\ z \end{matrix} &
\end{array}$$

Recognizing that the flow of energy is related to the Poynting vector $\vec{S} = u\vec{v}$ and that the momentum density is $\vec{g} = \vec{S}/c^2$, this four-tensor can be written

$$\begin{pmatrix} u & \vec{S}/c \\ S_x/c & g_x\vec{v} \\ S_y/c & g_y\vec{v} \\ S_z/c & g_z\vec{v} \end{pmatrix}$$

or, remembering the Maxwell stress tensor $\overset{\leftrightarrow}{T}$

$$\begin{pmatrix} u & \vec{S}/c \\ S_x/c & \\ S_y/c & -\overset{\leftrightarrow}{T} \\ S_z/c & \end{pmatrix}$$

and in this form, the four-tensor is clearly symmetric. It is called the "stress-energy four-tensor".

The stress-energy four-tensor resolves the conundrum raised in the problem assignment concerning the transformation of electromagnetic energy-momentum, but the concept goes beyond electromagnetism: *any* flow of energy-momentum has an associated stress-energy four-tensor. For example, a particle of mass $m$ follows the trajectory $\vec{r}(t)$ in the laboratory. Using $\vec{v} = d\vec{r}/dt$ and $\gamma = 1/\sqrt{1-(v/c)^2}$, I think you can see for yourself that the stress-energy four-tensor is

$$\begin{pmatrix} \gamma mc^2 & \gamma mc\vec{v} \\ \gamma mv_x c & \gamma mv_x\vec{v} \\ \gamma mv_y c & \gamma mv_y\vec{v} \\ \gamma mv_z c & \gamma mv_z\vec{v} \end{pmatrix} \times \delta^{(3)}(\vec{x} - \vec{r}(t)).$$

In Newton's theory of gravity, mass is the source of gravity. But in Einstein's theory of gravity, general relativity, the stress-energy four-tensor is the source of gravity. The zero-zero component above shows that mass is a source of gravity, but the other components show that energy and momentum are also sources, and that the flow of energy and momentum is also a source. This explains how light can be bent by gravity: it has zero mass but it doesn't have zero energy or zero momentum.

## 8.9 What to do with relativistic electrodynamics

We have set up relativistic electrodynamics. The famous graduate-level textbook by J.D. Jackson, *Classical Electrodynamics*, uses that setup to solve problems: he poses a problem, finds the easiest frame in which to solve it, and then transforms back to the lab frame in order to make testable predictions.

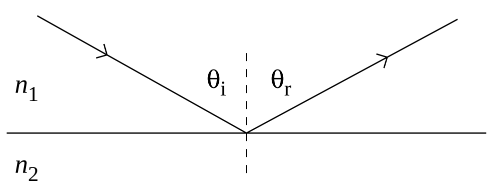
# Chapter 9

# Atmospheric Optics

We've been working at a high level of abstraction and generality. Nothing wrong with that — it's been stimulating and enlightening. But we started the course with electrostatics: rubbing your hair with a balloon and then watching it cling to a wall. Let's conclude the course by coming back to everyday phenomena.

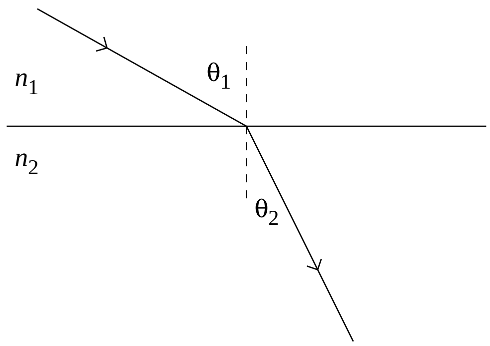
I want to think about light scattering from a transparent sphere, at the level of geometric (ray) optics. Think of a sphere of radius $R$ and index of refraction $n$ (for water, $n = 1.333$, for glass, $n = 1.52$), immersed in air ($n = 1.000$). Let me remind you of three facts:

1. If light reflects off an interface, the angle of incidence equals the angle of reflection: $\theta_i = \theta_r$.



2. If light transmits through the interface, the angle of incidence $\theta_1$ in medium $n_1$ is related to the angle of refraction $\theta_2$ in medium $n_2$ through Snell's Law

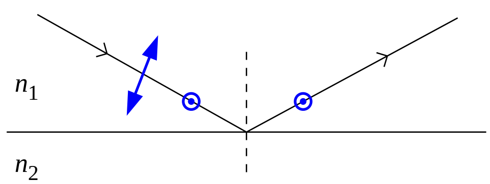
$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$



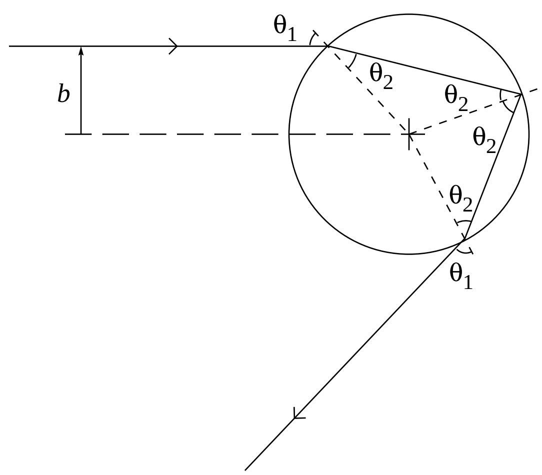

3. If light reflects off an interface at Brewster's angle where

$$\tan\theta_B = n_2/n_1,$$

then the reflected light is polarized as shown.



With this background let's tackle the problem of light scattering from a sphere. A ray of light enters the sphere a distance $b$ (the "impact parameter") above the sphere's center. (After finding the behavior of a ray entering above the axis indicated by the dashed line, we will rotate about this axis to find the behavior of any ray.)




This ray strikes the sphere with angle of incidence $\theta_1$. If you like geometry, you'll enjoy proving that

$$\sin\theta_1 = \frac{b}{R}.$$

Some of this light is reflected, but let's follow the ray that's transmitted. This ray strikes the surface of the sphere toward the right of the figure. Some of this light is transmitted, but let's follow the ray that's reflected. This ray strikes the surface of the sphere toward the bottom of the figure. Some of this light is reflected, but let's follow the ray that's transmitted and leaves the sphere for good.

What is the total angle of bending suffered by this ray? Put on your geometry hat again. At the first transmission it is bent by angle $\theta_1 - \theta_2$. At the reflection it is bent by angle $\pi - 2\theta_2$. At the second transmission it is bent by angle $\theta_1 - \theta_2$. The total bend is thus

$$\Theta = \pi + 2\theta_1 - 4\theta_2. \tag{9.1}$$

Remember that $\theta_1$ and $\theta_2$ are related through Snell's Law

$$\sin\theta_1 = n\sin\theta_2,$$

so the bend is of course a function of $b$.

An equation is not merely a jumble of symbols awaiting numbers to "plug in and chug through". An equation is a troubadour singing songs about nature. Let's examine this equation's song. As $b$ varies from 0 to $R$, other relevant parameters vary as follows:

$$\begin{array}{rcccl} b & : & 0 & \to & R \\ \theta_1 & : & 0 & \to & \pi/2 \\ \sin\theta_1 & : & 0 & \to & 1 \\ \sin\theta_2 & : & 0 & \to & 1/n \\ \theta_2 & : & 0 & \to & \operatorname{asin}(1/n) \\ \Theta & : & \pi & \to & 2\pi - 4\operatorname{asin}(1/n) \end{array}$$

To help us hear the song, we graph this dependence of $\Theta$ upon $\theta_1$. First, find the slope

$$\frac{d\Theta}{d\theta_1} = 2 - 4\frac{d\theta_2}{d\theta_1}.$$

But taking the derivative of both sides of Snell's Law

$$\cos\theta_1 = n\cos\theta_2\frac{d\theta_2}{d\theta_1},$$

so

$$\frac{d\Theta}{d\theta_1} = 2 - \frac{4}{n}\frac{\cos\theta_1}{\cos\theta_2}.$$

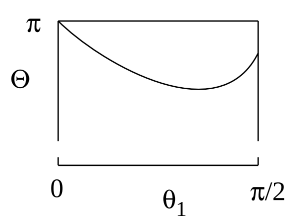
Thus at $\theta_1 = 0$,

$$\frac{d\Theta}{d\theta_1} = 2 - \frac{4}{n},$$

a negative number for typical values of $n$, while at $\theta_1 = \pi/2$,

$$\frac{d\Theta}{d\theta_1} = 2.$$

The function $\Theta(\theta_1)$ starts off sloping downward, but ends up sloping upward. It must have a minimum. For $n = 1.333$, the graph looks something like this.




The graph is, of course, flat near the minimum. Most of the outgoing rays will have this value of $\Theta$ — call it the most popular value $\Theta_p$. (*All* of the outgoing rays will have $\Theta \geq \Theta_p$.)

The location of this popular value is given through

$$0 = \frac{d\Theta}{d\theta_1} = 2 - \frac{4}{n}\frac{\cos\theta_1}{\cos\theta_2}.$$

That is

$$\begin{aligned} 2 &= \frac{4}{n}\frac{\cos\theta_1}{\cos\theta_2} \\ n\cos\theta_2 &= 2\cos\theta_1 \\ n^2\cos^2\theta_2 &= 4\cos^2\theta_1. \end{aligned}$$

But the square of Snell's law is

$$\begin{aligned} n^2\sin^2\theta_2 &= \sin^2\theta_1 \\ n^2(1-\cos^2\theta_2) &= 1-\cos^2\theta_1 \\ n^2-1+\cos^2\theta_1 &= n^2\cos^2\theta_2. \end{aligned}$$

We can eliminate $\theta_2$ by putting these two together to get

$$\begin{aligned} n^2-1+\cos^2\theta_1 &= 4\cos^2\theta_1 \\ n^2-1 &= 3\cos^2\theta_1 \\ \cos^2\theta_1 &= (n^2-1)/3 \\ \sin^2\theta_1 &= (4-n^2)/3. \end{aligned} \tag{9.2}$$

For any particular value of $n$, we can use this equation to find $\theta_1$, then use Snell's law to find $\theta_2$, and finally use the total bend equation (9.1) to find the most popular total bend $\Theta_p$.

⟦*Challenge:* Our math shows that $\Theta_p$ depends on index of refraction $n$ but not on radius $R$. Can you produce any simple argument showing that $\Theta_p$ should be independent of $R$?⟧

Executing this program for water ($n = 1.333$) gives $\Theta_p = 138°$ — most of the rays bend this much, that is they leave the sphere 42° from the incoming light beam. (Because the index of refraction varies with color, this most popular angle will vary somewhat with color.) Furthermore, in the $n = 1.333$ case the most popular reflection angle $\theta_2$ is 40°, and just by coincidence Brewster's angle for this reflection is 37°! The exiting light is largely polarized perpendicular to the plane of the paper.

Rotating this construction around the dashed axis we find that the sphere scatters a cone of light backwards, and this light is polarized parallel to the cone.

What would happen if there were a bunch of water spheres in the sky, perhaps very small ones, and you looked up at them? Here's a planar sketch of the situation:

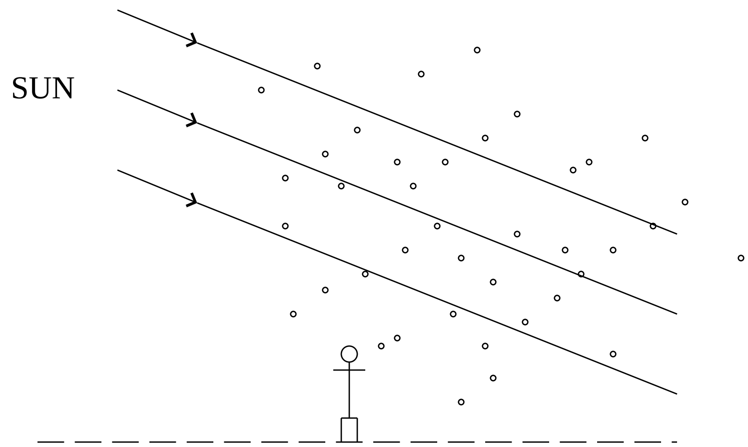



Figure out what's happening within this sketch, then rotate about the dashed line of symmetry to figure out what happens in three-dimensional space. What is the name of this phenomenon?

# Appendix A

# Pictorializing divergence and curl

The picture on the next page is the answer given by Purcell to the challenge poised on page 19.

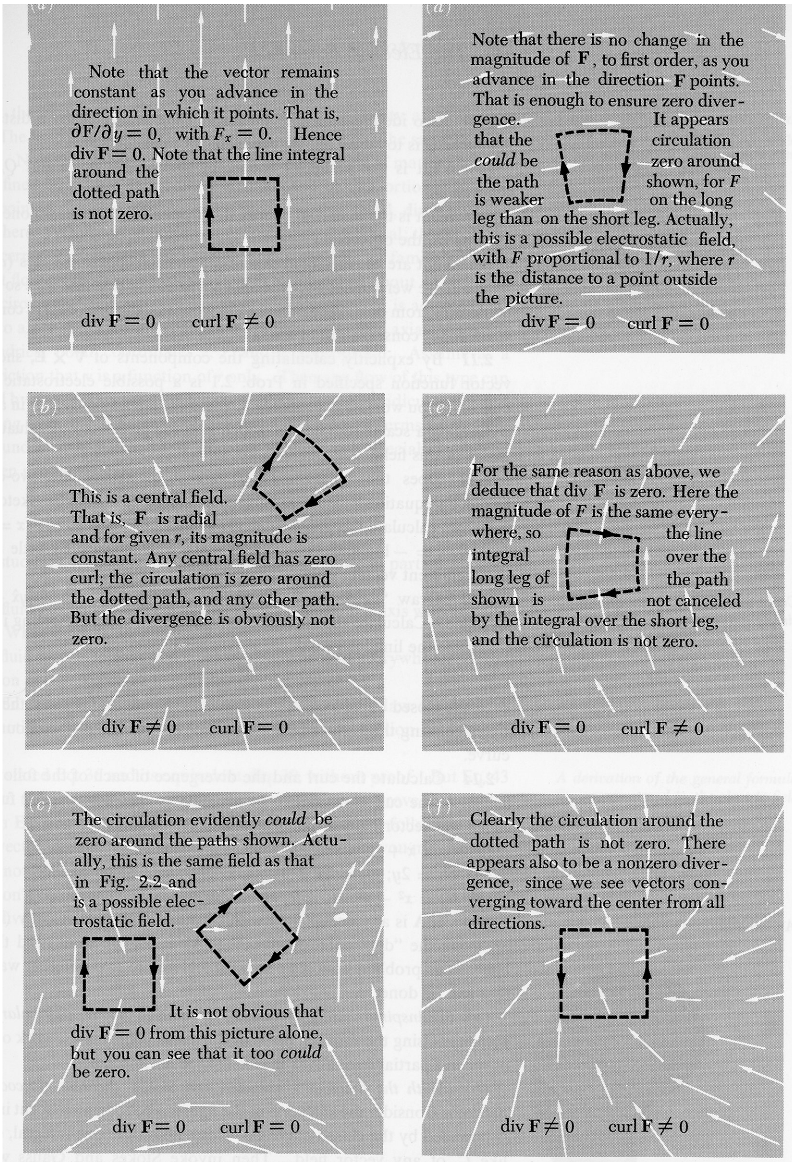



*Fig. 2.34* Discussion of Fig. 2.32.