Classical Field Theory

Classical field theory, which concerns the generation and interaction of fields, is a logical precursor to quantum field theory and can be used to describe phenomena such as gravity and electromagnetism.

Written for advanced undergraduates, and appropriate for graduate-level classes, this book provides a comprehensive introduction to field theories, with a focus on their relativistic structural elements. Such structural notions enable a deeper understanding of Maxwell's equations, which lie at the heart of electromagnetism, and can also be applied to modern variants such as Chern-Simons and Born-Infeld electricity and magnetism.

The structure of field theories and their physical predictions are illustrated with compelling examples, making this book perfect as a text in a dedicated field theory course, for self-study, or as a reference for those interested in classical field theory, advanced electromagnetism, or general relativity. Demonstrating a modern approach to model building, this text is also ideal for students of theoretical physics.

Joel Franklin is a professor in the Physics Department of Reed College. His work focuses on mathematical and computational methods with applications to classical mechanics, quantum mechanics, electrodynamics, general relativity, and modifications of general relativity. He is author of two previous titles: *Advanced Mechanics and General Relativity* and *Computational Methods for Physics*.

Classical Field Theory

JOEL FRANKLIN
Reed College, Oregon

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107189614
DOI: 10.1017/9781316995419

First published 2017

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalog record for this publication is available from the British Library

ISBN 978-1-107-18961-4 Hardback

For Lancaster, Lewis, and Oliver

Contents

Preface

This is a book on classical field theory, with a focus on the most studied one: electricity and magnetism (E&M). It was developed to fill a gap in the current undergraduate physics curriculum – most departments teach classical mechanics and then quantum mechanics, the idea being that one informs the other and logically precedes it. The same is true for the pair "classical field theory" and "quantum field theory," except that there are almost no dedicated classical field theory classes. Instead, the subject is reviewed briefly at the start of a quantum field theory course. There are a variety of reasons why this is so, most notably because quantum field theory is enormously successful and, as a language, can be used to describe three of the four forces of nature. The only classical field theory (of the four forces of nature) that requires the machinery developed in this book is general relativity, which is not typically taught at the undergraduate level. Other applications include fluid mechanics (also generally absent from the undergraduate course catalogue) and "continuum mechanics" applications, but these tend to be meant primarily for engineers.

Yet classical field theory provides a good way to think about modern physical model building, in a time where such models are relevant. In this book, we take the "bottom up" view of physics, that there are certain rules for constructing physical theories. Knowing what those rules are and what happens to a physical model when you break or modify them is important in developing physical models beyond the ones that currently exist. One of the main points of the book is that if you ask for a "natural" vector field theory, one that is linear (so superposition holds) and is already "relativistic," you get Maxwell's E&M almost uniquely. This idea is echoed in other areas of physics, notably in gravity, where if you similarly start with a second rank symmetric field that is linear and relativistic, and further require the universal coupling that is the hallmark of gravity, you get general relativity (almost uniquely). So for these two prime examples of classical field theories, the model building paradigm works beautifully, and you would naturally develop this pair even absent any sort of experimental observation (as indeed was done in the case of general relativity, and even Maxwell's correction to Faraday's law represents a similar brute force theoretical approach). But we should also be able to go beyond E&M and gravity. Using the same guidelines, we can develop a theory of E&M in which the photon has mass, for example, and probe the physics implied by that change.

Lagrangians and actions are a structure-revealing way to explore a physical theory, but they do not lend themselves to specific solutions. That's why E&M is the primary theory discussed in this book. By the time a student encounters the material presented here, they will have seen numerous examples of solutions to the Maxwell field equations, so that they know what statements like $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, $\nabla \times \mathbf{B} = 0$, mean physically, having introduced various types of charge sources and solved the field equations. What is less clear from Maxwell's equations are notions of gauge freedom and relativistic covariance. Moving the familiar **E** and **B** (or perhaps more appropriately, V and **A**) into a revealing scalar structure, like the E&M action, allows for a discussion beyond the specific solutions that are studied in most introductory E&M courses. As an example, one can easily develop the notion of a conserved stress tensor from the E&M action. Doing that without the action is much harder and less clear (in terms of conservation laws and symmetries).

I have used this text, or pieces of it, as the basis for a second, advanced semester of E&M, complementing a semester of E&M at the level of reference [19]. In such a course, the physical focus is on the acceleration fields that are necessary to describe radiation.[1] Those fields come directly from the Green's function for the wave equation in four-dimensional space-time, so a discussion of Green's functions is reasonable and also represents the beginning and end of the integral solution for field theories that are linear. The goal, in general, is to use advanced elements of E&M to motivate techniques that are useful for all field theories. But it is also possible to teach a dedicated classical field theory course from the book, without over-dependence on E&M as the primary example. There are plenty of additional physical ideas present in the book, including the action and conserved quantities for both Schrödinger's equation and the Klein–Gordon equation, in addition to the latter's interpretation as the scalar relativistic analogue to Schrödinger's equation.

The text is organized into four chapters and three appendixes. Chapter 1 is a relatively standard review of special relativity, with some in-depth discussion of transformation and invariants and a focus on the modified dynamics that comes from using Newton's second law with relativistic momentum rather than the pre-relativistic momentum. In Chapter 2, the focus is on Green's functions, with the main examples being first static problems in electricity, and then the full wave equation of E&M. The role of the Green's function as an integral building block is the focus, and the new radiation fields the physical beneficiary. Chapter 3 reviews Lagrangian mechanics and then introduces the notion of a field Lagrangian and action whose minimization yields familiar field equations (just as the extremization of classical actions yields familiar equations of motion, i.e., the Newtonian ones). The free particle field equations (and associated actions) for scalars, vectors, and tensors are developed, and then the free fields are coupled to both field sources and particle sources. One of the advantages of the scalar action is the automatic conservation of a stress tensor, a good example of the utility of Noether's theorem. The end of the chapter has a discussion of physical model building and introduces Born–Infeld E&M and Chern–Simons E&M. Finally, Chapter 4 is on gravity, another classical field theory. We establish that Newtonian gravity is insufficient (because it is not relativistically covariant) and explore the implications of universal coupling on gravity as a field theory (basically requiring gravity to be represented by a second rank symmetric tensor field to couple to the full stress tensor, with nonlinear field equations, to couple to itself).

[1] Those fields are sufficiently complicated as to be avoided in most undergraduate courses, except in passing. They are difficult for a number of reasons. First, structurally, since they bear little resemblance to the familiar Coulomb field that starts off all E&M investigation. Second, the acceleration fields are analytically intractable except for trivial cases. Even the first step in their evaluation, calculating the retarded time at a field point, requires numerical solution in general.

The appendixes are side notes, and fill in the details for some techniques that are useful for field theories (both classical and quantum). There is an appendix on mathematical methods (particularly complex contour integration, which is good for evaluating Green's functions) and one on numerical methods (that can be used to solve the actual field equations of E&M in the general setting, for example). Finally, there is a short essay that makes up the third appendix and is meant to demonstrate how you can take a compact action and develop from it all the physics that you know from E&M. When I teach the class, I ask the students to perform a similar analysis for a modified action (typically Proca, but one could use any of a number of interesting current ones).

There exist many excellent texts on classical field theory, classics such as [21] and [25], and the more modern [15]. I recommend them to the interested reader. I hope that my current contribution might complement these and perhaps extend some of the ideas in them. The focus on E&M as a model theory for thinking about special relativity (relativistic covariance) and field theoretic manifestations of it is common at the graduate level, in books such as [26] and [20]. What I have tried to do is split off that discussion from the rest of the E&M problem-solving found in those books and amplify the structural elements. This book could be used alongside [19] for a second, advanced semester of E&M or as a standalone text for a course on classical field theory, one that might precede a quantum field theory course and whose techniques could be used fairly quickly in the quantum setting (much of the field–field interaction that quantum field theory is built to handle has classical analogues that benefit from many of the same techniques, including both perturbative ones and ones having to do with finding Green's functions).

Acknowledgments

My own view of field theories in physics was developed under the mentorship of Stanley Deser, and much of the motivation for exploring more complicated field theories by comparison/contrast with E&M comes directly from his approach to physics. My students and colleagues at Reed have been wonderful sounding boards for this material. In particular, I'd like to thank Owen Gross and David Latimer for commentary on drafts of this book; they have helped improve the text immensely. Irena Swanson in the math department gave excellent technical suggestions for the bits of complex analysis needed in the first appendix, and I thank her for her notes. Finally, David Griffiths has given numerous suggestions, in his inimitable manner, since I started the project – my thanks to him, as always, for helping me clarify, expand, and contract the text.

1 Special Relativity

1.1 Geometry

Special relativity focuses our attention on the geometry of space-time, rather than the usual Euclidean geometry of space by itself. We'll start by reviewing the role of rotations as a transformation that preserves a specific definition of length, then introduce Lorentz boosts as the analogous transformation that preserves a new definition of length.

1.1.1 Rotations

In two dimensions, we know it makes little difference (physically) how we orient the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ axes – the laws of physics do not depend on the axis orientation. The description of those laws changes a little (what is "straight down" for one set of axes might be "off to the side" for another, as in Figure 1.1), but their fundamental predictions are independent of the details of these basis vectors.

A point in one coordinate system can be described in terms of a rotated coordinate system. For a vector that points from the origin to the point labeled by x and y in the $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ basis vectors: $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}}$, we want a description of the same point in the coordinate system with basis vectors $\hat{\bar{\mathbf{x}}}$ and $\hat{\bar{\mathbf{y}}}$ – i.e., what are $\bar{x}$ and $\bar{y}$ in $\bar{\mathbf{r}} = \bar{x}\,\hat{\bar{\mathbf{x}}} + \bar{y}\,\hat{\bar{\mathbf{y}}}$? Referring to Figure 1.2, we can use the fact that the length of the vector $\mathbf{r}$ is the same as the length of the vector $\bar{\mathbf{r}}$ – lengths are invariant under rotation. If we call the length r, then:

$$\begin{aligned} \bar{x} &= r\cos(\psi-\theta) = r\cos\psi\,\cos\theta + r\sin\psi\,\sin\theta = x\cos\theta + y\sin\theta \\ \bar{y} &= r\sin(\psi-\theta) = r\sin\psi\,\cos\theta - r\cos\psi\,\sin\theta = y\cos\theta - x\sin\theta. \end{aligned} \tag{1.1}$$

So we know how to go back and forth between the barred coordinates and the unbarred ones. The connection between the two is provided by the invariant[1] length of the vector. In fact, the very definition of a rotation "transformation" between two coordinate systems is tied to length invariance. We could start with the idea that two coordinate systems agree on the length of a vector and use that to generate the transformation (1.1). Let's see how that goes: the most general linear transformation connecting $\bar{x}$ and $\bar{y}$ to x and y is

$$\begin{aligned} \bar{x} &= A\,x + B\,y \\ \bar{y} &= C\,x + D\,y \end{aligned} \tag{1.2}$$

[1] Invariant here means "the same in all coordinate systems related by some transformation."

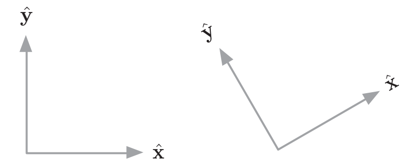

Fig. 1.1 Two different orientations for the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ basis vectors. In either case, the laws of physics are the same (e.g., $\mathbf{F} = m\,\mathbf{a}$ holds in both).

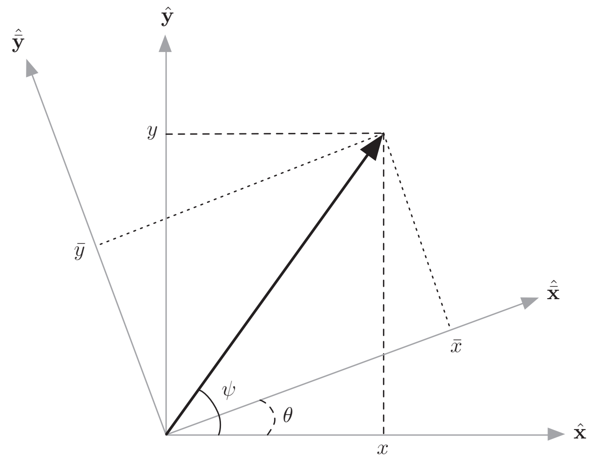

Fig. 1.2 The same point in two different bases (rotated by an angle θ with respect to one another). The angle ψ is the angle made by the vector with the $\hat{\mathbf{x}}$-axis.

for constants A, B, C, and D. Now, if we demand that $\bar{x}^2 + \bar{y}^2 = x^2 + y^2$, we have:

$$\left(A^2 + C^2\right) x^2 + \left(B^2 + D^2\right) y^2 + 2\,(A\,B + C\,D)\;x\,y = x^2 + y^2, \tag{1.3}$$

and then the requirement is: $A^2 + C^2 = 1$, $B^2 + D^2 = 1$, and $A\,B + C\,D = 0$. We can satisfy these with a one-parameter family of solutions by letting $A = \cos\theta$, $B = \sin\theta$, $C = -\sin\theta$, and $D = \cos\theta$ for arbitrary parameter θ. These choices reproduce (1.1) (other choices give back clockwise rotation).

1.1.2 Boosts

There is a fundamental invariant in special relativity, a quantity like length for rotations, that serves to define the transformation of interest. We'll start with the invariant and then build the transformation that preserves it. The length invariant used above for rotations comes from the Pythagorean notion of length. The invariant "length" in special relativity comes from the first postulate of the theory:

The speed of light is the same in all frames of reference traveling at constant velocity with respect to one another.

The idea, experimentally verified, is that no matter what you are doing, light has speed c. If you are at rest in a laboratory, and you measure the speed of light, you will get c. If you are running alongside a light beam in your laboratory, you will measure its speed to be c. If you are running toward a light beam in your laboratory, you will measure speed c. If the light is not traveling in some medium (like water), its speed will be c. This observation is very different from our everyday experience measuring relative speeds. If you are traveling at 25 mph to the right, and I am traveling next to you at 25 mph, then our relative speed is 0 mph. Not so with light.

What does the constancy of the speed of light tell us about lengths? Well, suppose I measure the position of light in my (one-dimensional) laboratory as a function of time: I flash a light on and off at time $t = 0$ at the origin of my coordinate system. Then the position of the light flash at time t is given by $x = c\,t$. Your laboratory, moving at constant speed with respect to mine, would call the position of the flash: $\bar{x} = c\,\bar{t}$ (assuming your lab lined up with mine at $t = \bar{t} = 0$ and $x = \bar{x} = 0$) at your time $\bar{t}$. We can make this look a lot like the Pythagorean length by squaring – our two coordinate systems must agree that:

$$-c^2\,\bar{t}^2 + \bar{x}^2 = -c^2\,t^2 + x^2 = 0. \tag{1.4}$$

Instead of x^2+y^2, the new invariant is $-c^2\,t^2+x^2$. There are two new elements here: (1) $c\,t$ is playing the role of a coordinate like x (the fact that c is the same for everyone makes it a safe quantity for setting units), and (2) the sign of the temporal portion is negative. Those aside, we have our invariant. While it was motivated by the constancy of the speed of light and applied in that setting, we'll now promote the invariant to a general rule (see Section 1.2.3 if this bothers you), and find the transformation that leaves the value $-c^2\,t^2 + x^2$ unchanged. Working from the most general linear transformation relating $c\,\bar{t}$ and $\bar{x}$ to $c\,t$ and x:

$$\begin{aligned} c\,\bar{t} &= A\,(c\,t) + B\,x \\ \bar{x} &= C\,(c\,t) + D\,x, \end{aligned} \tag{1.5}$$

and evaluating $-c^2\,\bar{t}^2 + \bar{x}^2$,

$$-(c\,\bar{t})^2 + \bar{x}^2 = -\left(A^2 - C^2\right)\,(c\,t)^2 + 2\,x\,(c\,t)\,(C\,D - A\,B) + \left(D^2 - B^2\right)\,x^2, \tag{1.6}$$

we can see that to make this equal to $-c^2\,t^2+x^2$, we must have $A^2 - C^2 = 1$, $D^2 - B^2 = 1$, and $C\,D - A\,B = 0$. This is a lot like the requirements for length invariance above, but with funny signs. Noting that the hyperbolic version of $\cos^2\theta + \sin^2\theta = 1$ is

$$\cosh^2\eta - \sinh^2\eta = 1, \tag{1.7}$$

we can write $A = \cosh\eta$, $C = \sinh\eta$, $D = \cosh\eta$, and $B = \sinh\eta$. The transformation analogous to (1.1) is then:

$$\begin{aligned} c\,\bar{t} &= (c\,t)\,\cosh\eta + x\,\sinh\eta \\ \bar{x} &= (c\,t)\,\sinh\eta + x\,\cosh\eta. \end{aligned} \tag{1.8}$$

That's the form of the so-called Lorentz transformation. The parameter η, playing the role of θ for rotations, is called the "rapidity" of the transformation.

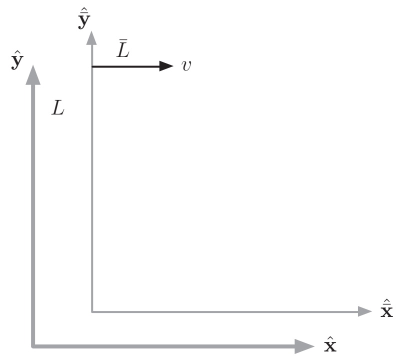

Fig. 1.3 A frame L and another frame, $\bar{L}$, moving to the right with speed v (in L). At time $t = 0$ in L, we have $\bar{t} = 0$ and $x = \bar{x} = 0$. Then at time t, the origin of $\bar{L}$ is at horizontal location $x = v\,t$ in L.

We can attach a physical picture to the transformation. Let the barred axes move with speed v to the right in the unbarred axes (the "lab frame," L). At time $t = 0$, we agree that $\bar{t} = 0$ and the spatial origins coincide ($x = \bar{x} = 0$, similar to keeping the origin fixed in rotations). Then the barred axes (the "moving frame," $\bar{L}$) have points $(c\,\bar{t}, \bar{x})$ related to the points in L by (1.8), and we need to find the relation between η and v. Referring to Figure 1.3, the origin of $\bar{L}$ is moving according to $x = v\,t$ in L. So we know to associate $\bar{x} = 0$ with $x = v\,t$, and using the second equation in (1.8) with this information gives:

$$0 = c\,t\,\sinh\eta + v\,t\,\cosh\eta \longrightarrow \tanh\eta = -\frac{v}{c}. \tag{1.9}$$

We can isolate both $\sinh\eta$ and $\cosh\eta$ given $\tanh\eta$ and the identity $\cosh^2\eta - \sinh^2\eta = 1$:

$$\sinh\eta = -\frac{\frac{v}{c}}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad \cosh\eta = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{1.10}$$

From these it is convenient to define the dimensionless quantities:

$$\gamma \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad \beta \equiv \frac{v}{c}, \tag{1.11}$$

and then (1.8) can be written:

$$\begin{aligned} c\,\bar{t} &= \gamma\,((c\,t) - x\,\beta) \\ \bar{x} &= \gamma\,(-(c\,t)\,\beta + x)\,. \end{aligned} \tag{1.12}$$

This form of the "Lorentz boost" is equivalent to the hyperbolic cosine and sine transformation in (1.8) (which connect nicely with rotations), but written with the physical configuration (one lab traveling at constant speed with respect to another) in mind.

Problem 1.1 The definition of hyperbolic cosine is $\cosh\eta = \frac{1}{2}\left(e^{\eta} + e^{-\eta}\right)$, and hyperbolic sine is $\sinh\eta = \frac{1}{2}\left(e^{\eta} - e^{-\eta}\right)$. Using these definitions, show that $\cosh^2\eta - \sinh^2\eta = 1$.

Problem 1.2 Using the configuration shown in Figure 1.3, develop the "inverse Lorentz transformation" relating $c\,t$ and x to $c\,\bar{t}$ and $\bar{x}$ (inverting (1.12)). Hint: don't do any algebra.

Problem 1.3 Starting from $\tanh\eta = -\frac{v}{c}$, use the properties of hyperbolic sine and cosine to establish (1.10).

Problem 1.4 Show that $-c^2\,\bar{t}^2 + \bar{x}^2 = -c^2\,t^2 + x^2$ using the Lorentz transformations in "boost" form (1.12). If we take x–y coordinates (in addition to time), we would write the invariant as $s^2 \equiv -c^2\,t^2 + x^2 + y^2$. Show that rotations (1.1) (with $\bar{t} = t$) also preserve this "length"; i.e., show that $-c^2\,\bar{t}^2 + \bar{x}^2 + \bar{y}^2 = -c^2\,t^2 + x^2 + y^2$ where the spatial coordinates x and y are related by a rotation.

1.2 Examples and Implications

The Lorentz transformation is easy to describe and understand in terms of redefined "length," but its geometric implications for space and time are less familiar than their rotational analogues. For rotations, statements like "one lab's north is another's northwest" are perfectly reasonable, with no problem of interpretation. The inclusion of time in the transformation, allowing for relations like "one lab's time is another's space-and-time" sound more interesting but are really of the same basic type.

1.2.1 Length Contraction and Time Dilation

We can think of two "labs," L and $\bar{L}$, and compare observations made in each. We'll refer to the setup in Figure 1.3, so that the lab $\bar{L}$ moves at speed v to the right within L (and $x = \bar{x} = 0$ at $t = \bar{t} = 0$). As an example of moving measurements from one lab to the other: suppose we have a rod at rest in $\bar{L}$. Its length is $\Delta\bar{x}$ as measured in $\bar{L}$. What is the length of the rod in L? In order to make a pure length measurement, we will find Δx with $\Delta t = 0$, i.e., an *instantaneous*[2] length measurement in L. From the Lorentz transformation (1.12), we have:

$$\Delta\bar{x} = \gamma\,(-c\,\Delta t\,\beta + \Delta x) = \gamma\,\Delta x \longrightarrow \Delta x = \frac{1}{\gamma}\,\Delta\bar{x} = \sqrt{1 - \frac{v^2}{c^2}}\,\Delta\bar{x}. \tag{1.13}$$

For[3] $v < c$, we have $\Delta x < \Delta\bar{x}$, and the moving rod has a length in L that is *less than* its length in $\bar{L}$. We cryptically refer to this phenomenon as length contraction and say that "moving objects are shortened." To be concrete, suppose $v = \frac{3}{5}\,c$, then a meter stick in $\bar{L}$ ($\Delta\bar{x} = 1$ m) has length $\Delta x = \frac{4}{5}$ m in L.

What happens to a pure temporal measurement? Take a clock in $\bar{L}$ that is at the origin $\bar{x} = 0$ (and remains there). Suppose we measure an elapsed time in $\bar{L}$ that is $\Delta\bar{t}$, how much

[2] If you measure the ends of a moving rod at two different times, you will, of course, get an artificially inflated or deflated length.

[3] If $v > c$, $\Delta\bar{x}$ is imaginary, a reminder that superluminal motion is no longer sensible.

time elapsed in L? This time, we'll use the inverse Lorentz transformation (since we know $\Delta\bar{x} = 0$, the clock remained at the origin) from the solution to Problem 1.2:

$$c\,\Delta t = \gamma\left(c\,\Delta\bar{t} + \beta\,\Delta\bar{x}\right) = \gamma\,c\,\Delta\bar{t} \longrightarrow \Delta t = \gamma\,\Delta\bar{t}. \tag{1.14}$$

Remember that $\gamma > 1$ for $v < c$, so this time we have $\Delta t > \Delta\bar{t}$: if 1 s passes in $\bar{L}$, traveling at $v = \frac{3}{5}\,c$, then $\Delta t = \frac{5}{4}$ s passes in L. The phrase we use to describe this "time dilation" is "moving clocks run slow."

Time dilation and length contraction are very real physical effects. The muon, an elementary particle, has a lifetime of 2×10^{-6} s when at rest. Muons produced in the upper atmosphere travel very fast. For $v = .999\,c$, we would expect a muon to decay in about 600 m, yet they make it down to the surface of the earth for detection. The reason is time dilation. From the earth's point of view, the muon's lifetime is $T = 1/\sqrt{1 - 0.999^2}\left(2 \times 10^{-6}\text{ s}\right) \approx 4 \times 10^{-5}$ s. In this amount of time, they can go a distance $\approx 0.999\,c\,T \approx 12$ km, far enough to reach the earth's surface.

1.2.2 Directions Perpendicular to Travel

The "first" postulate of special relativity is the following.

The laws of physics are the same in all inertial frames of reference

or, sloppily, in any frames moving at constant velocity with respect to each other. That doesn't have a lot of quantitative meaning just yet, but it can be used to establish that, for example, length contraction occurs only in the direction parallel to the motion. When we define the Lorentz boost in three dimensions for a lab $\bar{L}$ moving along the $\hat{\mathbf{x}}$-axis in L, we say that $(c\,t)$, x, y, and z become, in the "barred" coordinates:

$$\begin{aligned} c\,\bar{t} &= \gamma\,((c\,t) - x\,\beta) \\ \bar{x} &= \gamma\,(-(c\,t)\,\beta + x) \\ \bar{y} &= y \\ \bar{z} &= z. \end{aligned} \tag{1.15}$$

But how do we *know* that nothing interesting happens in the y- and z-directions (perpendicular to the relative motion)? We can use the first postulate – suppose I have a door that is exactly my height, so that I just barely fit under it at rest as in the top panel of Figure 1.4. Now assume that length contraction holds in the $\hat{\mathbf{y}}$-direction even for relative motion in the $\hat{\mathbf{x}}$-direction. I move toward the door at speed v. From the door's point of view, I am moving and by our (false) assumption, I appear shorter than normal (to the door) and easily make it through the door. From my point of view, the door is moving toward me; it appears shorter than when it is at rest, and I do not make it through. This conflict violates the notion that the door and I had better agree on whether I can get through. The easy fix: there is no length contraction perpendicular to the direction of travel.

Problem 1.5 A meter stick sits at rest in $\bar{L}$. The $\bar{L}$ frame is moving to the right at speed v in the lab (the same setup as shown in Figure 1.3). In L, the stick is measured to have length $12/13$ m. What is v?

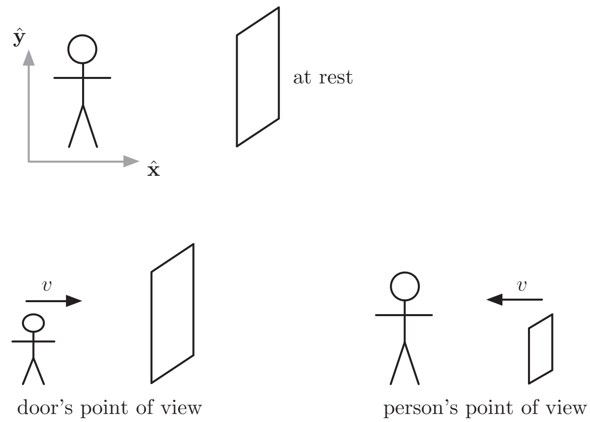

Fig. 1.4 A person and a door are the same height when at rest (top). If length contraction happened in directions perpendicular to travel, we'd have a shorter person from the door's point of view (bottom left) or a shorter door from the person's point of view (bottom right). These two points of view lead to different physical predictions: the person either makes it through the door or does not, and we can't have that. There must therefore be no length contraction in the direction perpendicular to the relative motion.

Problem 1.6 A clock moves through a lab with speed $v = (12/13)\,c$. It goes from $x = 0$ to $x = 5$ m. How long did its trip take in the lab? How long did its trip take in the rest frame of the clock (i.e., in $\bar{L}$ moving at speed v, where the clock remains at $\bar{x} = 0$)?

Problem 1.7 You take off in a rocket ship, headed for the stars, at a speed of $v = 3/5\,c$. After a year of traveling, you turn around and head back to earth at the same speed. How much time has passed on earth? (Make $\bar{L}$ the ship's frame.)

1.2.3 Alternative

We generated the Lorentz transformations by defining the quantity:

$$s^2 \equiv -c^2\,t^2 + x^2 + y^2 + z^2 \tag{1.16}$$

and demanding that it be unchanged by the transformation. But our only example, light traveling at c, has $s^2 = 0$, so it is still possible that we should allow, for $s^2 \neq 0$, a scaling. Maybe we require only that:

$$\bar{s}^2 = \alpha^2\,s^2, \tag{1.17}$$

where $\bar{s}^2$ is in the transformed frame and α is a scaling factor. If $\bar{s}^2 = s^2 = 0$, we wouldn't see the α at all. We will now show that the constancy of the speed of light requires $\alpha = \pm 1$, and we'll take the positive root to avoid switching the signs of intervals. First note that α cannot be a constant function of the speed v (other than 1) in the transformation. If it were, we would get $\bar{s}^2 \neq s^2$ in the $v = 0$ limit. So $\alpha(v)$ has the constraint that $\alpha(0) = 1$ in order to recover the same s^2 when the boost speed is zero.

Now the Lorentz transformation that puts an α^2 in front of the length s^2 upon transformation, without violating any of our other rules, is:

$$
\begin{aligned}
c\,\bar{t} &= \alpha\,\gamma\,((c\,t) - x\,\beta) \\
\bar{x} &= \alpha\,\gamma\,(-(c\,t)\,\beta + x) \\
\bar{y} &= y \\
\bar{z} &= z,
\end{aligned}
\tag{1.18}
$$

for a boost in the x-direction. The inverse transformation can be obtained by taking $\beta \to -\beta$ (as usual) and $\alpha \to 1/\alpha$.

In $\bar{L}$, we shoot a laser along the $\hat{\bar{\mathbf{y}}}$-axis. In a time $\Delta\bar{t}$, the light has traveled $\Delta\bar{x} = \Delta\bar{z} = 0$, $\Delta\bar{y} = c\,\Delta\bar{t}$. Using these in the inverse transformation associated with (1.18), we have

$$
\Delta t = \frac{\gamma}{\alpha}\,\Delta\bar{t}, \qquad \Delta x = \frac{\gamma}{\alpha}\,v\,\Delta\bar{t}, \qquad \Delta y = c\,\Delta\bar{t},
\tag{1.19}
$$

with $\Delta z = 0$. The distance traveled by the light in L is:

$$
\Delta \equiv \sqrt{\Delta x^2 + \Delta y^2} = \sqrt{\frac{\gamma^2}{\alpha^2}\,v^2 + c^2}\,\Delta\bar{t}
\tag{1.20}
$$

so that the speed of the light, as measured in L, is

$$
c_L = \frac{\Delta}{\Delta t} = \sqrt{v^2 + \alpha^2\,c^2\left(1 - \frac{v^2}{c^2}\right)} = \sqrt{v^2\left(1 - \alpha^2\right) + c^2},
\tag{1.21}
$$

and the only way for the speed of light to be c in L, $c_L = c$, is if $\alpha = \pm 1$, and we'll take the positive root.

1.2.4 Simultaneity

Suppose we have an "event" that occurs in the lab at position x_1, time t_1 (e.g., a firecracker goes off) and another event that occurs at x_2, time t_1. The two events are separated spatially but happen at the same time. For our moving $\bar{L}$ coordinate system, these points become:

$$
\begin{aligned}
c\,\bar{t}_1 &= \gamma\,((c\,t_1) - x_1\,\beta)\,, & \bar{x}_1 &= \gamma\,(x_1 - (c\,t_1)\,\beta) \\
c\,\bar{t}_2 &= \gamma\,((c\,t_1) - x_2\,\beta)\,, & \bar{x}_2 &= \gamma\,(x_2 - (c\,t_1)\,\beta)\,,
\end{aligned}
\tag{1.22}
$$

so that $\bar{t}_2 - \bar{t}_1 = -\frac{\gamma\,\beta}{c}\,(x_2 - x_1) \neq 0$. The two events did not occur at the same time in $\bar{L}$. Events simultaneous in one coordinate system are not simultaneous in another.

Related to simultaneity is the idea of "causality" – one event (at t_1, x_1, say) can cause another (at t_2, x_2) only if there is a way for a signal traveling at the speed of light (or less) to be sent between the two in a time less than (or equal to) $t_2 - t_1$. Think of E&M, where electric and magnetic fields propagate (in vacuum) at speed c. The only way for one charge to act on another electromagnetically is for the fields to travel from one location to the other at speed c. The requirement that two space-time points be "causally connected"[4] amounts to: $c^2\,(t_2 - t_1)^2 \geq (x_2 - x_1)^2$ (where we squared both sides to avoid absolute values), or

$$-\,c^2\,(t_2 - t_1)^2 + (x_2 - x_1)^2 \leq 0. \tag{1.23}$$

Problem 1.8 Two events (t_1 at x_1 and t_2 at x_2) are causally related if $-c^2\,(t_2 - t_1)^2 + (x_2 - x_1)^2 \leq 0$. Show that if event one causes event two in L (i.e., the two events are causally related with $t_1 < t_2$), then event one causes event two in any $\bar{L}$ traveling at speed v in L (where L and $\bar{L}$ are related as usual by Figure 1.3).

1.2.5 Minkowksi Diagrams

Just as we can draw motion in the "usual" two-dimensional space spanned by $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, and relate that motion to rotated (or otherwise transformed) axes, we can think of motion occurring in the two-dimensional space spanned by time $c\,\hat{\mathbf{t}}$ and $\hat{\mathbf{x}}$.[5] Referring to Figure 1.5, we can display motion that takes the form $x(t)$ by inverting to get $t(x)$ and then plotting $c\,t(x)$ versus x. Light travels with $x = c\,t$, and the dashed line in Figure 1.5 shows light emitted from the origin. For a particle traveling with constant speed v, we have $x = v\,t$ and can invert to write

$$c\,t = \frac{c}{v}\,x, \tag{1.24}$$

describing a line with slope c/v. For light, the slope is 1, and for particle motion with $v < c$, the slope of the line is > 1. Superluminal motion occurs with lines of slope < 1. For the "fun" trajectory shown in Figure 1.5, if we look at the tangent to a point on the curve, we can estimate the speed of the particle at a given time – portions of that trajectory do have tangent lines with slope < 1, indicating that the particle is traveling faster than the speed of light.

For a particle traveling with constant speed through "the lab" L, we know that the particle is at rest in its own $\bar{L}$ (moving with constant speed through L). That means that the line of the particle's motion in L represents the particle's $c\,\hat{\bar{\mathbf{t}}}$-axis (purely temporal motion). What does the $\hat{\bar{\mathbf{x}}}$-axis look like? Well, those are points for which $\bar{t} = 0$. From the inverse of the transformation (1.12), we have

$$\begin{aligned} c\,t &= \gamma\,\beta\,\bar{x} \\ x &= \gamma\,\bar{x}, \end{aligned} \tag{1.25}$$

defining the line: $c\,t = \frac{v}{c}\,x$ in L. That's just the reflection of the line $c\,t = \frac{c}{v}\,x$ about the "light" line. In Figure 1.6, we have the picture relating the axes for boosts analogous to Figure 1.1 for rotations. By looking at the two coordinate systems, it should be clear that a pure temporal interval in the barred coordinates, $\Delta\bar{t}$, corresponds to a combination of Δt and Δx in the unbarred coordinate system (and vice versa).

It is interesting to note that our new "Minkowski length" definition: $s^2 = -c^2\,t^2 + x^2$ allows for three distinct options: s^2 can be greater than, less than, or equal to zero. We

[4] Meaning that light can travel between the two points in a time that is less than their temporal separation.
[5] The basis vector for the temporal direction is denoted $c\,\hat{\mathbf{t}}$ because writing $\hat{c\mathbf{t}}$ looks awkward.

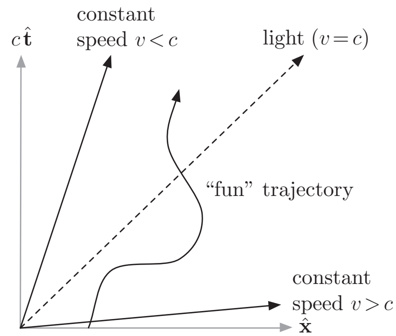

Fig. 1.5 A Minkowski diagram: we plot the motion of a particle or light with $c\,t$ as the vertical axis and x as the horizontal one.

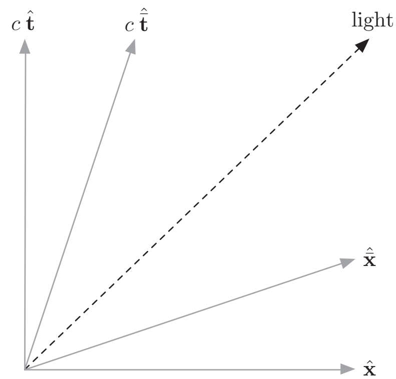

Fig. 1.6 Two coordinate systems related by a Lorentz boost (1.12).

generated interest in this particular combination of $c\,t$ and x by thinking about light, for which $-c^2\,t^2 + x^2 = 0$ (since light travels along the line $x = c\,t$). We call points separated by "zero" light-like since a light-like trajectory connects these points. But we also have motion with $v < c$: $x = v\,t$, and this gives a separation $s^2 = -c^2\,t^2 + v^2\,t^2 < 0$. Such intervals are called "time-like" because there is a rest frame (in which only temporal motion occurs) connecting the two points. Finally, for $v > c$, $-c^2\,t^2 + x^2 = -c^2\,t^2 + v^2\,t^2 > 0$ and these separations are "space-like" (like motion along the spatial $\hat{\mathbf{x}}$ and $\hat{\hat{\mathbf{x}}}$ axes). For material particles, only time-like separations are possible. It is clear that speeds greater than c cannot be achieved: the boost factor γ becomes imaginary. That's a mathematical warning sign, the physical barrier is the amount of energy that a massive particle moving at c would have (infinite, as we shall see).

Problem 1.9 We have $\bar{L}$ moving at constant speed v in L as in Figure 1.3 (with origins $x = \bar{x} = 0$ that overlap at $t = \bar{t} = 0$). In $\bar{L}$, we observe the following two events: $\bar{t}_1 = 0$, $\bar{x}_1 = 0$, and $\bar{t}_2 = 0$, $\bar{x}_2 = d$. The $c\,\hat{\hat{\mathbf{t}}}$–$\hat{\hat{\mathbf{x}}}$ picture is shown in Figure 1.7. Find the location and time of these two events in L and draw the $c\,\hat{\mathbf{t}} - \hat{\mathbf{x}}$ version of the picture using $d = 1$ m and $v = 4/5$ c.

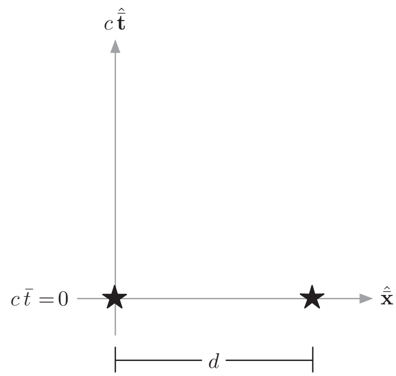

Fig. 1.7 Two events occur simultaneously in $\bar{L}$ (for Problem 1.9).

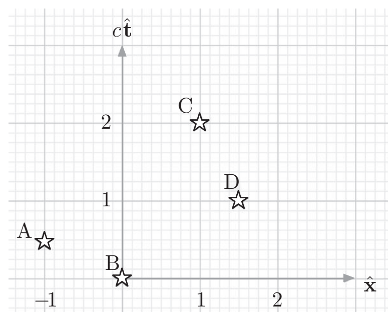

Fig. 1.8 Points for Problem 1.10.

Problem 1.10 Which of the points shown in Figure 1.8 could cause to the others? (Make a list.)

Problem 1.11 A particle moves according to $x = \alpha t^2$ for constant $\alpha > 0$. Make a Minkowski diagram (sketch) of this motion and mark the first point at which the particle moves faster than c.

Problem 1.12 A particle is at rest at $x_1 > 0$, $t_1 > 0$. Indicate, on a Minkowski diagram, all those points that could be causally influenced by the particle and all those points that could have (causally) influenced the particle.

1.3 Velocity Addition

Velocity addition is an operation that must be modified in special relativity. If two trains travel in the same direction at the same speed, "ordinary" velocity addition tells us that the relative speed between the two trains is zero. But if one of those trains is replaced by a light beam, then the relative speed between the train and the beam is c.

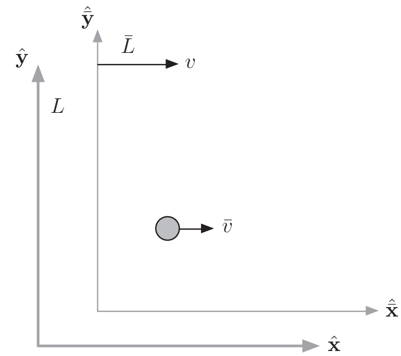

Fig. 1.9 A ball moves with speed $\bar{v}$ in $\bar{L}$. What is its speed in L?

1.3.1 Non-relativistic Velocity Addition

Let's review the non-relativistic form of velocity addition to set the scene. We'll use the "frames" language of special relativity for later comparison. Suppose I have $\bar{L}$ moving at speed v in L, and there is a ball in $\bar{L}$ that moves with speed $\bar{v}$ relative to $\bar{L}$. What is the speed of the ball relative to L? The setup is shown in Figure 1.9. In $\bar{L}$, the ball moves a distance $\Delta\bar{x}$ in time $\Delta\bar{t}$, with $\Delta\bar{x} = \bar{v}\,\Delta\bar{t}$. In L, the ball moves a distance $\Delta x = \Delta\bar{x} + v\,\Delta t$ in time Δt with $\Delta\bar{t} = \Delta t$ (non-relativistic, remember), so that the ball's speed in L is:

$$V \equiv \frac{\Delta x}{\Delta t} = \bar{v} + v. \tag{1.26}$$

In the more general three-dimensional case, where we have a velocity $\mathbf{v}_{B\bar{L}}$ (the velocity of the ball relative to $\bar{L}$) and $\mathbf{v}_{\bar{L}L}$ (the velocity of $\bar{L}$ relative to L), the pre-relativistic velocity addition formula for $\mathbf{v}_{BL}$ (the velocity of the ball relative to L) is

$$\mathbf{v}_{BL} = \mathbf{v}_{B\bar{L}} + \mathbf{v}_{\bar{L}L}. \tag{1.27}$$

1.3.2 Relativistic Velocity Addition

Now, we'll do the same analysis, but using the Lorentz transformation that relates L and $\bar{L}$. Again: in $\bar{L}$, the ball moves a distance $\Delta\bar{x}$ in time $\Delta\bar{t}$, with $\Delta\bar{x} = \bar{v}\,\Delta\bar{t}$. From the Lorentz boost, the Δx and Δt in L are

$$\begin{aligned}\Delta x &= \gamma\left(\beta\, c\,\Delta\bar{t} + \Delta\bar{x}\right)\\ \Delta t &= \gamma\left(\Delta\bar{t} + \frac{\beta}{c}\,\Delta\bar{x}\right),\end{aligned} \tag{1.28}$$

where $\beta = \frac{v}{c}$ as usual. The speed of the ball in L is:

$$V = \frac{\Delta x}{\Delta t} = \frac{\beta\, c\,\Delta\bar{t} + \Delta\bar{x}}{\Delta\bar{t} + \frac{\beta}{c}\,\Delta\bar{x}} = \frac{1 + \frac{v}{\bar{v}}}{\frac{1}{\bar{v}} + \frac{v}{c^2}} = \frac{v + \bar{v}}{1 + \frac{v\,\bar{v}}{c^2}}. \tag{1.29}$$

This is the Einstein velocity addition formula. It says that for an object moving with speed $\bar{v}$ in $\bar{L}$, which is moving with speed v in L, the speed V of the object in L is:

$$V = \frac{v + \bar{v}}{1 + \frac{v\bar{v}}{c^2}}. \tag{1.30}$$

Notice that if $v = c$ and $\bar{v} = c$, we get $V = c$, as desired. For small speeds, with $v \ll c$ and $\bar{v} \ll c$, we recover $V \approx v + \bar{v}$, the "usual" result (1.26).

Problem 1.13 A particle moves at $\bar{v} = c$ relative to $\bar{L}$, and $\bar{L}$ travels at speed $v = -w$ (to the left) in L. What is the speed of the particle relative to L? What happens if $w = c$? What would the speed be in this case ($w = c$) if we used non-relativistic velocity addition?

Problem 1.14 The velocity addition in (1.30) is for motion that is all in the same direction (parallel or anti-parallel). What are the components of velocity in L for a ball that moves with $\bar{\mathbf{v}} = \bar{v}\,\hat{\mathbf{y}}$ in $\bar{L}$? (Refer to the setup in Figure 1.9.)

Problem 1.15 We generated Lorentz transformations using the "rapidity" η, with $\tanh\eta = -v/c$ for a boost with speed v (see (1.8)). We want to express Einstein's velocity addition law in terms of rapidities. Write (1.30) in terms of the rapidities η, χ, and ψ where $\tanh\eta = -v/c$, $\tanh\chi = -\bar{v}/c$, and $\tanh\psi = -V/c$. Use hyperbolic trigonometry simplifications on the resulting expression to establish that "rapidities add."

1.4 Doppler Shift for Sound

We know that when an object emitting sound moves toward or away from us, our perception of the emitted frequency changes. This is a problem in relative motion, and we can calculate the frequency shift. Suppose a source emits frequency f, and we want to find the frequency $\bar{f}$ that an observer hears. Take a source with velocity[6] $\mathbf{v}_{sg} = v_s\,\hat{\mathbf{x}}$ emitting a sound wave of frequency f that travels at $\mathbf{v}_{wg} = v_w\,\hat{\mathbf{x}}$ (where v_w is the speed of sound in air), and an observer moving at $\mathbf{v}_{og} = v_o\,\hat{\mathbf{x}}$ as in Figure 1.10. We'll go to the rest frame of the wave. There, the wavelength λ of the wave will be perceived to travel at different speeds by the observer and source, and the time it takes each to go one wavelength defines the period (and hence frequency) that is perceived by each.

The velocity of the observer relative to the wave is (using the non-relativistic velocity addition rule, and noting that $\mathbf{v}_{gw} = -\mathbf{v}_{wg}$)

$$\mathbf{v}_{ow} = \mathbf{v}_{og} + \mathbf{v}_{gw} = \mathbf{v}_{og} - \mathbf{v}_{wg} \tag{1.31}$$

so that in the rest frame of the wave, the observer is moving at $\mathbf{v}_{ow} = (v_o - v_w)\,\hat{\mathbf{x}}$. It takes the observer $T_o = \frac{\lambda}{v_{ow}}$ to go one full wavelength, so the frequency heard is[7]

$$\bar{f} = \frac{1}{T_o} = \frac{v_o - v_w}{\lambda}. \tag{1.32}$$

[6] $\mathbf{v}_{sg}$ refers to the velocity of the source (s) relative to the ground (g).
[7] We assume that $v_o < v_w$; the observer is moving slower than the speed of sound in air.

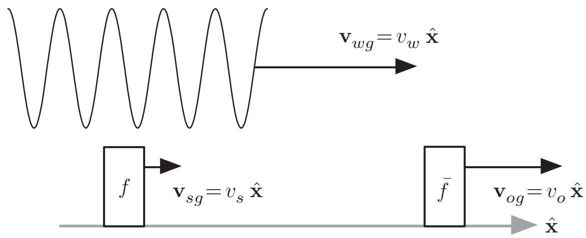

Fig. 1.10 Setup for calculating the Doppler shift associated with a moving source and observer.

Similarly, the source's motion relative to the wave has velocity: $\mathbf{v}_{sw} = (v_s - v_w)\,\hat{\mathbf{x}}$, and the frequency is

$$f = \frac{v_s - v_w}{\lambda}. \tag{1.33}$$

Since we imagine fixing f, we can rewrite $\bar{f}$ in terms of f by eliminating λ:

$$\bar{f} = \frac{v_o - v_w}{v_s - v_w} f. \tag{1.34}$$

This gives us back all the usual degenerate cases: for a stationary source ($v_s = 0$), with observer moving away ($v_o > 0$) from the source, we get

$$\bar{f} = \left(1 - \frac{v_o}{v_w}\right) f, \tag{1.35}$$

while if the observer is moving toward the source, $v_o < 0$, and the formula still holds. For a stationary observer, $v_o = 0$, we have

$$\bar{f} = \frac{1}{1 - \frac{v_s}{v_w}} f \tag{1.36}$$

where now "toward" means $v_s > 0$ (referring again to Figure 1.10), and "away" has $v_s < 0$.

Problem 1.16 A police siren emits $f = 440$ Hz at rest. Make a sketch of the frequency you hear as the police car approaches you at constant speed (assume the car passes your location at $t = 0$).

Problem 1.17 Re-do the Doppler shift for sound using the relativistic velocity addition formula. You can still go to the rest frame of the sound wave, since $v_w < c$.

Problem 1.18 The Doppler shift we worked out in this section is one-dimensional, with the motion of the source and observer occuring along a shared axis. In order to describe more realistic situations (where the police car doesn't run you over as you observe its siren's frequency shift), you can generalize (1.36) by using $v_s = \mathbf{v}_s \cdot \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the vector pointing from the source to the observer. Now we are picking off the one-dimensional direction connecting source and observer. Write down the version of (1.36) that is appropriate for a police car that moves with constant velocity $\mathbf{v} = v\,\hat{\mathbf{x}}$ but does so at $d\,\hat{\mathbf{y}}$ (so that the motion is a straight line parallel to the x-axis, but at a distance d along the y-axis). Sketch the frequency as a function of time assuming the car passes you at $t = 0$ (pick an arbitrary frequency).

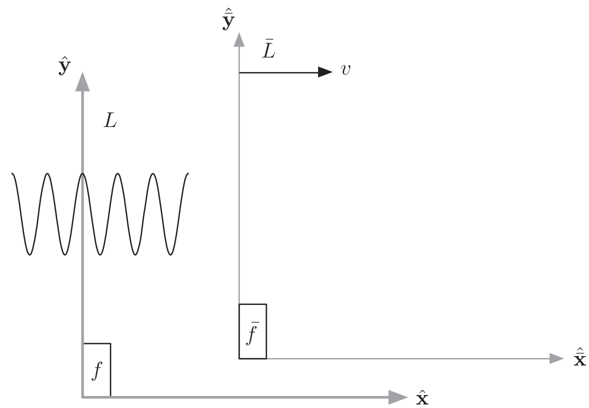

Fig. 1.11 A frequency source generates f; a moving observer receives $\bar{f}$.

1.5 Doppler Shift for Light

In developing the Doppler effect for sound, we moved to the rest frame of the sound wave and analyzed the perceived frequencies in that rest frame. In light[8] of special relativity, we cannot perform the same analysis on a light wave since it moves with speed c for all observers. There is no way to go to the rest frame of light. Instead, we will use our Lorentz transformations. Suppose at $x = 0$ in L we have a light source emitting light with frequency f (period $T = 1/f$). What is the frequency of light observed by someone moving to the right at speed v ($v > 0$ means the observer is going away)? Well, as is by now usual, we will imagine the observer sitting at rest at $\bar{x} = 0$ in $\bar{L}$, moving to the right in L with speed v as in Figure 1.3. The current setup is shown in Figure 1.11.

We'll align the two coordinate systems so that at $t = 0$, $\bar{t} = 0$, and the origins coincide, $x = \bar{x} = 0$ at that time. We'll also imagine that as the two origins pass each other, the wave (or, if you like, the electric field component of the wave) is at a maximum. The idea is to find the next time at which a maximum occurs at $x = 0$ and at $\bar{x} = 0$. Those times will give us the period in L and $\bar{L}$ from which we can get the frequencies.

In L, the next maximum occurs at T (see Figure 1.12), and that peak travels at speed c toward the origin of $\bar{L}$, which it will certainly catch ($v < c$). Working in L (so as to avoid mixing and matching), the location of the origin of $\bar{L}$ is $x_o = v\,t$, and the location of the peak is $x_p = c\,(t - T)$ (so that at $t = T$, the peak is at $x_p = 0$). We want the time t^* at which x_p is equal to x_o:

$$v\,t^* = c\,(t^* - T) \longrightarrow t^* = \frac{c\,T}{c - v}; \tag{1.37}$$

this occurs at $x^* = v\,t^*$. Now we take that event and transform, using the Lorentz transformations, to find the location and time of the event in $\bar{L}$:

[8] Pun intended.

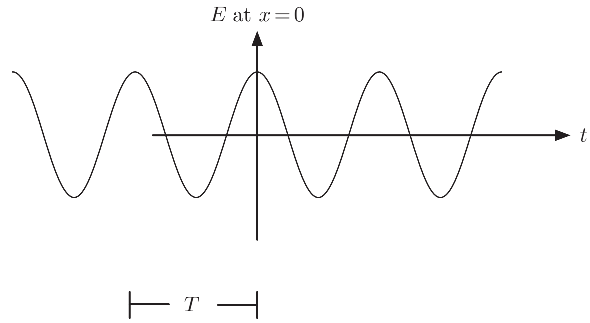

Fig. 1.12 Electric field at the origin of L ($x = 0$) as a function of time.

$$\begin{aligned} c\bar{t}^* &= \gamma\left(c\,t^* - \frac{v}{c}x^*\right) = c\,\gamma\left(t^* - \frac{v^2}{c^2}t^*\right) = c\,t^*\,\gamma\left(1 - \frac{v^2}{c^2}\right) \\ \bar{x}^* &= \gamma\left(-v\,t^* + x^*\right) = \gamma\left(-v\,t^* + v\,t^*\right) = 0 \end{aligned} \tag{1.38}$$

where $\bar{x}^* = 0$ is just a sense check – we were supposed to calculate when the peak arrived at the origin of $\bar{L}$, which is at $\bar{x} = 0$ in $\bar{L}$. Going back to the time relation, using the definition of γ, we have:

$$\bar{t}^* = \frac{t^*}{\gamma}. \tag{1.39}$$

Since $\bar{t}^*$ is the first time after $\bar{t} = 0$ at which a maximum arrived at the origin of $\bar{L}$, we would call the period of the light in $\bar{L}$: $\bar{T} = \bar{t}^* - 0 = \bar{t}^*$. From $t^* = \frac{cT}{c-v}$, we can relate $\bar{T}$ to T:

$$\bar{T} = \frac{1}{\gamma}\frac{c\,T}{c-v} \longrightarrow \frac{1}{\bar{T}} = \gamma\left(1 - \frac{v}{c}\right)\frac{1}{T}, \tag{1.40}$$

or, since $f = 1/T$ and $\bar{f} = 1/\bar{T}$:

$$\bar{f} = \frac{1 - \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}} f = \frac{\sqrt{c-v}}{\sqrt{c+v}} f. \tag{1.41}$$

This is the Doppler shift for an observer moving away from a source. For an observer moving toward a source, we have:

$$\bar{f} = \frac{\sqrt{c+v}}{\sqrt{c-v}} f \tag{1.42}$$

just taking $v \longrightarrow -v$.

Problem 1.19 A moving source of light emits frequency f. What frequency $\bar{f}$ does a stationary observer measure? (Calculate for both cases: source moving toward and away from the observer at speed v.) Suppose we have a star that sends out light with a frequency $f = 5 \times 10^{14}$ Hz (among others) moving away from the earth at a speed $v = 0.1\,c$. What is the frequency of the light detected on earth? In which direction (red or blue) is the shift? (Refer to a standard electromagnetic spectrum table.)

Problem 1.20 For an observer moving away from a stationary source, we can write (1.41) as:

$$\bar{f} = \frac{\sqrt{1-\beta}}{\sqrt{1+\beta}} f \tag{1.43}$$

where $\beta \equiv \frac{v}{c}$. Suppose β is small, i.e., $v \ll c$. What is $\bar{f}$ if you use the Taylor series expansion? Remember that for $\bar{f}(\beta)$, the expansion reads:

$$\bar{f}(\beta) \approx \bar{f}(0) + \bar{f}'(0)\,\beta \tag{1.44}$$

where $\bar{f}'(0)$ means "evaluate the derivative of $\bar{f}(\beta)$ at $\beta = 0$". Compare your result with the Doppler shift for sound using the same configuration (observer moving away from a stationary source).

1.6 Proper Time

For a particle traveling with constant speed $v < c$ in a lab L, we know that we can "boost to the rest frame" of the particle; i.e., we can define a coordinate transformation such that the particle is at rest (at the origin, say) of the new coordinate system $\bar{L}$. The boost is defined precisely by (1.12). Take $x = v\,t$ (describing the particle's location in L), then $\bar{x} = 0$ and $\bar{t} = \sqrt{1 - v^2/c^2}\,t$. For a particle that moves along a particular path in the lab, $x(t)$, we can define the "instantaneous" rest frame of the particle by boosting to the rest frame at time t. To do this, we just perform a Lorentz boost with parameter $v = \dot{x}(t) \equiv \frac{dx}{dt}$.

In general, we know that for infinitesimal motion in the t- and x-directions, $ds^2 = -c^2\,dt^2 + dx^2$ is the same for all coordinate systems related by a Lorentz transformation. Since there is a Lorentz transformation that puts us in the instantaneous rest frame of the particle, we know that in that rest frame, the particle moves only in time. Call the particle's time, τ, the "proper time." An interval with dt and dx in L can be written in the instantaneous rest frame in terms of the purely temporal $d\tau$:

$$-c^2\,d\tau^2 = -c^2\,dt^2 + dx^2. \tag{1.45}$$

If we have parametrized motion, $x(t)$, then $dx = \dot{x}(t)\,dt$ and the above becomes:

$$-c^2\,d\tau^2 = \left(-c^2 + \dot{x}(t)^2\right)\,dt^2. \tag{1.46}$$

From this, we can write the defining relationship between the proper time and the "coordinate" time (in L):

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{\dot{x}(t)^2}{c^2}}}. \tag{1.47}$$

Problem 1.21 A particle moves along the $\hat{\mathbf{x}}$-axis with constant speed v_0 in the lab so that the particle's location is given by $x(t) = v_0\,t$. Find the proper time τ of the particle in terms of t, and write $x(\tau)$, i.e., the position in the lab with proper time instead of coordinate time as the parameter.

Problem 1.22 A particle starts from rest and accelerates to the right in a lab. It is moving relativistically and has the following relation between proper time and coordinate time:

$$\tau(t) = \alpha \sinh^{-1}\left(\frac{t}{\alpha}\right) \tag{1.48}$$

where α is a constant with temporal dimension. Find the motion of the particle in the lab (i.e., what is $x(t)$?). What is the force acting on this particle? (Calculate the force using the familiar form of Newton's second law, but come back and redo the problem after Newton's second law has been updated in Section 1.7.)

Problem 1.23 In the "twin paradox," two people of the same age start on earth, and one heads out into space on a rocket. After a time t_f on earth, the space twin returns. The question is: how old is the space twin? As a concrete model, suppose the rocket travels according to $x(t) = a/2\,(1 - \cos(\omega t))$. In terms of ω, what is t_f? Find the proper time τ_f elapsed in the rest frame of the rocket during this earth-based time interval. This is the time elapsed for the space twin. The following integral identity will be useful:

$$\int_0^{2\pi} \sqrt{1 - k^2 \sin^2\theta}\, d\theta = 4\,\mathrm{E}(k) \tag{1.49}$$

where the (complete) elliptic integral (of the second kind) on the right is defined by

$$\mathrm{E}(k) \equiv \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2\theta}\, d\theta. \tag{1.50}$$

The maximum speed of the rocket twin is $v_{\max} = a\,\omega/2$. Define $\beta_{\max} \equiv v_{\max}/c$, and make a plot of the total time elapsed on earth (t_f) and on the rocket (τ_f) in units of $1/\omega$ for $\beta_{\max} = 0 \to 1$.

Proper time is a new object in special relativity. For your "usual" classical mechanics problems, there is only one time, t. In special relativity, there are two natural times: (1) t the coordinate time in the lab, and now (2) the proper time of the particle. Which one is the correct one for dynamics? We normally write $p = m\,\frac{dx}{dt}$ to define momentum, but we could define a new "relativistic" momentum:

$$p \equiv m\,\frac{dx}{d\tau} = m\,\frac{dx}{dt}\,\frac{dt}{d\tau} = \frac{m\,v(t)}{\sqrt{1 - \frac{v(t)^2}{c^2}}}, \tag{1.51}$$

where $v \equiv \frac{dx(t)}{dt}$. Notice that while this relativistic momentum is inspired by the presence of τ, it is written entirely in terms of x and t, the coordinates in our stationary laboratory L; that's where measurements are made, after all, so that's the relevant place to write everything.

Now think about Newton's second law: $\frac{dp}{dt} = F$. Which p should we use here?[9] Take $F = F_0$, a constant. If we use $p = m\,v(t)$, then $v(t) = \frac{F_0}{m}\,t$ for a particle that starts from

[9] Incidentally, you might wonder if Newton's second law should read $\frac{dp}{dt} = F$ or $\frac{dp}{d\tau} = F$. Good question!

rest, and the particle's speed will reach and exceed the speed of light in a finite time. Now take p to be the relativistic momentum, then

$$\frac{d}{dt}\left[\frac{m\,v(t)}{\sqrt{1-\frac{v(t)^2}{c^2}}}\right] = F \longrightarrow \frac{m\,v(t)}{\sqrt{1-\frac{v(t)^2}{c^2}}} = F_0\,t, \tag{1.52}$$

and the speed, as a function of time, in this case, is

$$v(t) = \frac{c\,\frac{F_0\,t}{m\,c}}{\sqrt{1+\left(\frac{F_0\,t}{m\,c}\right)^2}}. \tag{1.53}$$

This speed is bounded; it will never reach c (except at $t = \infty$), suggesting that Newton's second law is correct provided we use the relativistic momentum.

1.7 Energy and Momentum

The new relativistic momentum, in vector form, is

$$\mathbf{p} = \frac{m\,\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{1.54}$$

where $\mathbf{v}(t)$ is the velocity of the particle moving through the lab L, and $v^2 \equiv \mathbf{v}\cdot\mathbf{v}$.

One way to view the definition of energy in classical mechanics is to "integrate" Newton's second law once. Suppose we have a force $F(x) = -\frac{dU}{dx}$ for potential energy $U(x)$, then

$$m\,\frac{dv}{dt} = -\frac{dU}{dx} \tag{1.55}$$

and if we multiply both sides by v, we get:

$$m\,v\,\frac{dv}{dt} = -v\,\frac{dU}{dx}. \tag{1.56}$$

The left-hand side can be written as the total time derivative: $m\,v\,\frac{dv}{dt} = \frac{d}{dt}\left(\frac{1}{2}\,m\,v^2\right)$ and the right-hand side is the total time derivative: $-v\,\frac{dU}{dx} = -\frac{dU(x)}{dt}$ (remember that $v = \frac{dx}{dt}$, so we are using the chain rule here). Then we integrate both sides in time to get

$$\frac{d}{dt}\left(\frac{1}{2}\,m\,v^2\right) = -\frac{dU(x)}{dt} \longrightarrow \frac{1}{2}\,m\,v^2 = -U + H \tag{1.57}$$

where H (the Hamiltonian) is a constant of integration that we interpret as the total energy, and write

$$H = \frac{1}{2}\,m\,v^2 + U. \tag{1.58}$$

The analogous procedure, for our reinterpreted Newton's second law (with conservative force $F = -\frac{dU}{dx}$),

$$\frac{d}{dt}\left[\frac{m\,v}{\sqrt{1-\frac{v^2}{c^2}}}\right] = -\frac{dU}{dx} \tag{1.59}$$

gives

$$H = \frac{m\,c^2}{\sqrt{1-\frac{v^2}{c^2}}} + U. \tag{1.60}$$

As in the non-relativistic case, H is a constant with the interpretation of total energy.

Notice that in the absence of a potential energy function, we have

$$E = H = \frac{m\,c^2}{\sqrt{1-\frac{v^2}{c^2}}} \tag{1.61}$$

(where E is the energy of the particle), and this tells us that for a particle that is at rest, with $\mathbf{v} = 0$,

$$E = m\,c^2. \tag{1.62}$$

We can also use the free particle expression (1.61) to establish that massive particles cannot travel at c: the energy required to accelerate a particle to speed c is clearly infinite (put $v = c$ into (1.61); there are also problems with the particle's momentum in that case). Finally, if we assume that the energy of a massless particle is *not* zero, then (1.61) also implies that massless particles *must* travel at speed c (and the same is implied if a massless particle's momentum is to be non-zero).

Example

Let's take $U(x) = \frac{1}{2}\,k\,x^2$, the harmonic oscillator potential energy (for a spring with spring constant k and equilibrium location at zero). If we start a mass m from rest at an initial extension of a, then we can identify the total energy of the system. For the classical (pre-special-relativistic) energy, we get:

$$H = \frac{1}{2}\,k\,a^2 = \frac{1}{2}\,m\,v^2 + \frac{1}{2}\,k\,x^2, \tag{1.63}$$

so that

$$v^2 = \frac{k}{m}\left(a^2 - x^2\right), \tag{1.64}$$

and the maximum speed of the mass is $v_{\max} = \sqrt{\frac{k}{m}}\,a$, which can be greater than c for (large) initial extension a.

If we instead use the relativistic energy (1.60), we have, at time $t = 0$:

$$H = m\,c^2 + \frac{1}{2}\,k\,a^2, \tag{1.65}$$

and since energy is conserved, this value is maintained throughout the motion of the mass:

$$H = m c^2 + \frac{1}{2} k a^2 = \frac{m c^2}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{1}{2} k x^2. \tag{1.66}$$

We can solve for v and identify the maximum speed in this case:

$$v^2 = c^2 \left(1 - \frac{1}{\left(1 + \frac{1}{2} k \left(\frac{a^2 - x^2}{m c^2}\right)\right)^2} \right). \tag{1.67}$$

It is clear that $v^2 < c^2$ for all $|x| < a$, so there is again a speed limit.

We can get dynamics information out of the speed here. Think about what happens in the "very" relativistic limit (high energies, in this case achieved by taking a very large). We can make the second term in parentheses in (1.67) small by taking a large, so that the speed gets close to c. The value of x ranges from $-a$ to a, and if you look at the functional form of the speed v, as displayed in Figure 1.13 for example, it is clear that the speed achieves values close to c "abruptly." Referring to the right-most plot, for $a = 3$ (in units of $\sqrt{m/k}\, c$), we can see that for x relatively close to a, the speed is very close to c (1 represents c in these plots).

So we would predict that in the high-energy limit, the motion of a mass connected to a spring looks like motion at speeds $\pm c$, where the sign changes abruptly at $\pm a$. An example of this ultra-relativistic motion is shown, together with the non-relativistic motion, in Figure 1.14. It is interesting to note that the non-relativistic solution can be written as:

$$x(t) = a \cos\left(\sqrt{\frac{k}{m}}\, t\right) \tag{1.68}$$

and the period of this motion is $T = 2\pi \sqrt{\frac{m}{k}}$. That does not depend on the initial extension at all, only on the spring constant k and mass m. Conversely, the high-energy motion has period $T = \frac{4a}{c}$, which depends *only* on the initial extension and is independent of k and m, precisely the opposite behavior. In between the two regimes, the period depends on both $\sqrt{\frac{m}{k}}$ and a in some complicated manner (see [7] for details).

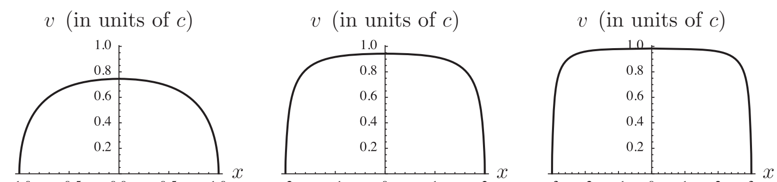

Fig. 1.13 $v(x)$ from (1.67) for $a = 1, 2$, and 3 (left to right, in units of $\sqrt{m/k}\, c$). Notice that for "large a" ($a = 3$) the speed $v \sim c$ even for x close to $\pm a$.

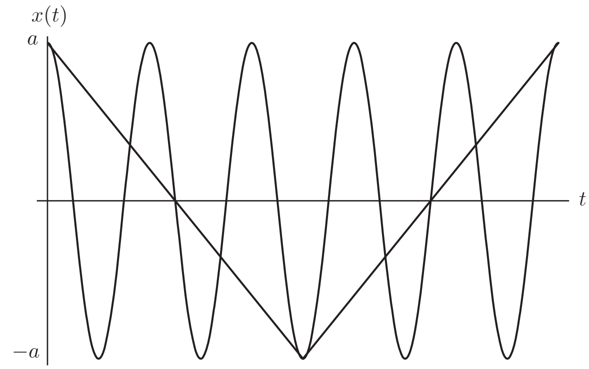

Fig. 1.14 The cosine is associated with non-relativistic harmonic motion, while the straight line represents the motion of the mass moving in the ultra-relativistic regime.

Problem 1.24 Rewrite the relativistic energy of a free particle:

$$E = \frac{m\,c^2}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{1.69}$$

in terms of relativistic momentum (i.e., solve $p = \frac{m\,v}{\sqrt{1-\frac{v^2}{c^2}}}$ for v in terms of p and insert in E).

Problem 1.25 If you take the derivative on the left in Newton's second law:

$$\frac{d}{dt}\left(\frac{m\,v(t)}{\sqrt{1 - \frac{v(t)^2}{c^2}}}\right) = F, \tag{1.70}$$

show that you get

$$m\,\frac{d^2x(t)}{dt^2} = \left(1 - \frac{v(t)^2}{c^2}\right)^{3/2} F. \tag{1.71}$$

We can gain some (special) relativistic intuition by thinking of the right-hand-side as an "effective" force (notice that as $v \to \pm c$, the effective force goes to zero) for the "usual" form of Newton's second law.

Problem 1.26 Take the time derivative of

$$H = \frac{m\,c^2}{\sqrt{1 - \frac{v(t)^2}{c^2}}} + U(x(t)) \tag{1.72}$$

and show that you recover Newton's second law (i.e., $\frac{dp}{dt} = -\frac{dU}{dx}$ where p is the relativistic momentum).

Problem 1.27 For a positive charge q pinned at the origin and a negative charge, $-q$, with mass m that starts from rest a distance a away, use conservation of energy to find $v(r)$

(r is the distance to the origin for the negative charge) and sketch $|v(r)|$ to establish that $v < c$ for all locations as the negative charge falls toward the positive one.

1.7.1 Energy–Momentum Transformation

The momentum of a particle moving with constant speed v in the $\hat{\mathbf{x}}$-direction is

$$p = \frac{m\,v}{\sqrt{1-\frac{v^2}{c^2}}} = m\,v\,\gamma, \tag{1.73}$$

and its energy is[10] $E = m\,c^2\,\gamma$. In the particle's rest frame, we have $\bar{p} = 0$ and $\bar{E} = m\,c^2$, so we can relate the two frames:

$$\begin{aligned} \frac{E}{c} &= \gamma\left(\frac{\bar{E}}{c} + \beta\,\bar{p}\right) \\ p &= \gamma\left(\beta\,\frac{\bar{E}}{c} + \bar{p}\right), \end{aligned} \tag{1.74}$$

where the factor of c has been inserted to get the units to work out correctly. The $\bar{p}$ is in place (even though it is zero) to establish the additive structure of the transformation (although we can only pin it down up to sign – we expect, in the presence of a non-zero $\bar{p}$, that if $\bar{p} > 0$, $p > 0$ as well). Inverting gives us the same transformation as (1.12), with E/c playing the role of $c\,t$, and p acting like x. You will establish the correctness of (1.74) in Problem 1.28.

Just as the quantity $-c^2\,t^2 + x^2$ is the same in all coordinate systems related by Lorentz transformation, we expect:

$$-\frac{E^2}{c^2} + p^2 = \text{ constant.} \tag{1.75}$$

Evaluating this expression for a particle at rest sets the constant to $-E^2/c^2 = -m^2\,c^2$. If we expand to the full three-dimensional case, then we have

$$-\frac{E^2}{c^2} + \mathbf{p}\cdot\mathbf{p} = -m^2\,c^2. \tag{1.76}$$

It will be useful to have the time derivative of this expression for later:

$$-2\,\frac{E}{c^2}\,\frac{dE}{dt} + 2\,\frac{d\mathbf{p}}{dt}\cdot\mathbf{p} = 0, \tag{1.77}$$

or, dividing by $E/c^2 = m/\sqrt{1-\frac{v^2}{c^2}}$, we have

$$\frac{dE}{dt} = \frac{d\mathbf{p}}{dt}\cdot\frac{d\mathbf{x}}{dt} = \mathbf{F}\cdot\frac{d\mathbf{x}}{dt}. \tag{1.78}$$

[10] Note that γ is now shorthand for

$$\gamma = \frac{1}{\sqrt{1-\frac{\dot{x}^2}{c^2}}}$$

not just a constant boost factor (unless $\dot{x}$ is constant).

This result is familiar in the work–energy context. There we expect the change in energy to be due to an applied force: $dE = \mathbf{F} \cdot d\boldsymbol{\ell} = \mathbf{F} \cdot \mathbf{v}\, dt$, which is precisely the above.

Problem 1.28 We can fill in the gaps in the above transformation argument leading to (1.74). For L and $\bar{L}$ in the "standard" setup of Figure 1.3, we have a particle of mass m moving with speed $\bar{v}$ in the $\hat{\bar{\mathbf{x}}}$-direction in $\bar{L}$. What is its energy $\bar{E}$ and momentum $\bar{p}$ in $\bar{L}$? Using velocity addition, find the speed of the particle in L and write down the energy E and momentum p in L. Now show (by algebraically manipulating these expressions) that the two are related by

$$\begin{aligned} \frac{E}{c} &= \gamma \left[\frac{\bar{E}}{c} + \beta \bar{p} \right] \\ p &= \gamma \left[\beta \frac{\bar{E}}{c} + \bar{p} \right], \end{aligned} \tag{1.79}$$

from which

$$\begin{aligned} \frac{\bar{E}}{c} &= \gamma \left[\frac{E}{c} - \beta p \right] \\ \bar{p} &= \gamma \left[-\beta \frac{E}{c} + p \right] \end{aligned} \tag{1.80}$$

follows via the usual inverse transformation.

1.8 Force

In classical mechanics (relativistic or not) a force causes a change in momentum: $\mathbf{F} = \frac{d\mathbf{p}}{dt}$. How do forces transform under Lorentz boosts? You can tell it's going to be tricky, since $\mathbf{p}$ transforms into a combination of energy and momentum, and t becomes a combination of time and spatial coordinates under Lorentz transformation. Let's work through the transformation for $\mathbf{F}$ using these ingredients. Take a boost in the $\hat{\mathbf{x}}$-direction, with speed v:

$$\bar{\mathbf{F}} = \frac{d\bar{\mathbf{p}}}{d\bar{t}} = \frac{d\bar{\mathbf{p}}}{\gamma \left(dt - dx \frac{v}{c^2} \right)} = \frac{d\bar{\mathbf{p}}}{dt} \frac{1}{\gamma \left(1 - \frac{dx}{dt} \frac{v}{c^2} \right)}. \tag{1.81}$$

For the component of $\mathbf{p}$ that is parallel to the boost, the x-component here, we have $\bar{p}_\parallel = \gamma \left(-E\beta/c + p_\parallel \right)$ from the transformation (1.74), so that

$$\bar{F}_\parallel = \frac{\frac{dp_\parallel}{dt} - \frac{dE}{dt} \frac{v}{c^2}}{1 - \frac{dx}{dt} \frac{v}{c^2}} = \frac{F_\parallel - \frac{dE}{dt} \frac{v}{c^2}}{1 - \frac{dx}{dt} \frac{v}{c^2}} \tag{1.82}$$

while the components of $\bar{\mathbf{F}}$ that are perpendicular to the boost direction do not transform: $\bar{\mathbf{p}}_\perp = \mathbf{p}_\perp$, then

$$\bar{\mathbf{F}}_\perp = \frac{\mathbf{F}_\perp}{\gamma \left(1 - \frac{dx}{dt} \frac{v}{c^2} \right)}. \tag{1.83}$$

We can summarize and "simplify" the transformation, using the relation for $\frac{dE}{dt}$ from (1.78), with $\frac{d\mathbf{p}}{dt} = \mathbf{F}$:

$$\bar{F}_{\parallel} = \frac{F_{\parallel} - \frac{v}{c^2}\,\mathbf{F}\cdot\frac{d\mathbf{x}}{dt}}{1 - \frac{dx}{dt}\frac{v}{c^2}} \qquad \bar{\mathbf{F}}_{\perp} = \frac{\mathbf{F}_{\perp}}{\gamma\left(1 - \frac{dx}{dt}\frac{v}{c^2}\right)}. \tag{1.84}$$

It is a, to use a technical term, "messed up" transformation. Evidently, force is not an obviously relativistic quantity (if it were, it would transform like $c\,t$ and $\mathbf{x}$ or E/c and $\mathbf{p}$).

The inverse transformation, in which L moves to the left with speed $\bar{v}$ in $\bar{L}$, can be obtained from (1.84) by sending $v \to -v$,

$$F_{\parallel} = \frac{\bar{F}_{\parallel} + \frac{v}{c^2}\,\bar{\mathbf{F}}\cdot\frac{d\bar{\mathbf{x}}}{d\bar{t}}}{1 + \frac{d\bar{x}}{d\bar{t}}\frac{v}{c^2}}, \qquad \mathbf{F}_{\perp} = \frac{\bar{\mathbf{F}}_{\perp}}{\gamma\left(1 + \frac{d\bar{x}}{d\bar{t}}\frac{v}{c^2}\right)}. \tag{1.85}$$

If we think of the $\bar{L}$ frame as the one in which the particle is at rest (even just instantaneously at rest, which is a pretty weak requirement), then $\frac{d\bar{\mathbf{x}}}{d\bar{t}} = 0$, and (1.85) becomes

$$F_{\parallel} = \bar{F}_{\parallel}, \qquad \mathbf{F}_{\perp} = \frac{1}{\gamma}\,\bar{\mathbf{F}}_{\perp}. \tag{1.86}$$

You should be thinking about the form of Newton's second law in the context of special relativity. When we move from non-relativistic expressions to relativistic ones, we originally set $\frac{d\mathbf{p}}{dt} = \mathbf{F}$ where $\mathbf{p}$ is the relativistic momentum, and $\mathbf{F}$ is the force: $-k\,x$, $\frac{q^2}{4\pi\,\epsilon_0\,r^2}\hat{\mathbf{r}}$, etc. But given that $t = \tau$, "coordinate time" equals "proper time" in non-relativistic classical mechanics; we could just as easily have updated Newton's second law, for special relativity, with $\frac{d\mathbf{p}}{d\tau} = \mathbf{F}$, i.e., the derivative of the relativistic momentum with respect to *proper* time. In special relativity, t and τ *are* different, and one has to be careful where to put the force.

1.9 Special Relativity Requires Magnetic Fields

A line of charge has constant charge-per-unit-length λ_0 when at rest. It is moving through a lab at speed v (to the right). A charge q is a distance d from the line of charge and moves to the right with speed u; the setup is shown in Figure 1.15. We'll find the force on the charge in the lab frame and in its rest frame and relate them using the force transformation rules (1.86). The punchline is that without the magnetic field/force, the force prediction in each frame is different, which is not allowed by special relativity.

Let's start by finding the charge-per-unit-length in the lab. When the line of charge is at rest, there is a charge $\Delta q = \lambda_0\,\Delta x_0$ in a window of width Δx_0. Now the amount of charge doesn't change when the line of charge is moving, but the window size is length-contracted, so that $\Delta q = \lambda\,\Delta x$ with $\Delta x = \frac{1}{\gamma_v}\,\Delta x_0$ (that's length contraction: when the line of charge is moving, the length is less) where we define $\gamma_v \equiv 1/\sqrt{1 - v^2/c^2}$, so the moving line of charge has constant charge-per-unit-length

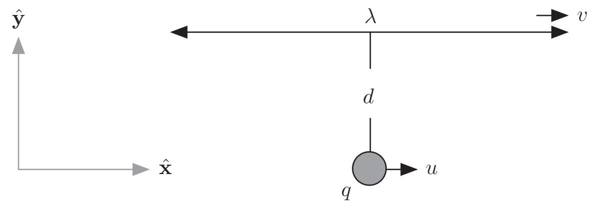

Fig. 1.15 An infinite line of charge with constant λ_0 at rest is pulled to the right at speed v (at which point it has constant charge density λ). A particle with charge q is a distance d from the line of charge and moves with speed u to the right.

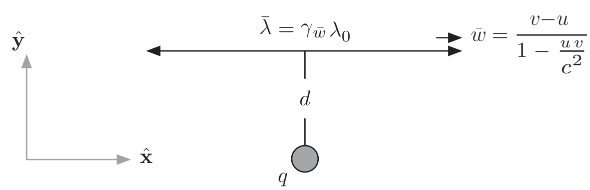

Fig. 1.16 The charge q is at rest and sees a line of charge with constant $\bar{\lambda} = \gamma_{\bar{w}} \lambda_0$.

$$\lambda = \frac{\Delta q}{\Delta x} = \frac{\lambda_0 \, \Delta x_0}{\frac{1}{\gamma_v} \, \Delta x_0} = \gamma_v \, \lambda_0. \tag{1.87}$$

In the lab, we have two forces on the charge q: the electric force

$$\mathbf{F}_E = -\frac{q\,\lambda}{2\,\pi\,\epsilon_0\,d}\,\hat{\mathbf{y}}, \tag{1.88}$$

and the magnetic force

$$\mathbf{F}_B = \frac{q\,\mu_0\,\lambda\,v}{2\,\pi\,d}\,u\,\hat{\mathbf{y}}, \tag{1.89}$$

so that the total force, in the lab, is

$$\begin{aligned} \mathbf{F} = \mathbf{F}_E + \mathbf{F}_B &= \frac{q\,\lambda_0}{2\,\pi\,d}\left[\mu_0\,u\,v - \frac{1}{\epsilon_0}\right]\gamma_v\,\hat{\mathbf{y}} \\ &= -\frac{q\,\lambda_0}{2\,\pi\,\epsilon_0\,d}\left[1 - \frac{u\,v}{c^2}\right]\gamma_v\,\hat{\mathbf{y}}. \end{aligned} \tag{1.90}$$

Now we'll analyze the force in the rest frame of the charge q. From its point of view, the scenario is shown in Figure 1.16. The line of charge is moving at speed $\bar{w}$ given by velocity addition (1.30):

$$\bar{w} = \frac{v-u}{1-\frac{u\,v}{c^2}}, \tag{1.91}$$

and the line of charge has density $\bar{\lambda} = \gamma_{\bar{w}}\,\lambda_0$ with $\gamma_{\bar{w}} \equiv 1/\sqrt{1-\bar{w}^2/c^2}$.

Since the charge is not moving in its rest frame, there is no magnetic force there; the only force is electric:

$$\bar{\mathbf{F}} = -\frac{q\,\bar{\lambda}}{2\,\pi\,\epsilon_0\,d}\,\hat{\mathbf{y}} = -\frac{q\,\gamma_{\bar{w}}\,\lambda_0}{2\,\pi\,\epsilon_0\,d}\,\hat{\mathbf{y}}. \tag{1.92}$$

In the rest frame of the charge, it is instantaneously at rest (of course), so that we would use (1.86) to predict the lab frame's force (for comparison with **F**) to be:

$$\hat{\mathbf{F}} = \frac{1}{\gamma_u} \bar{\mathbf{F}} \tag{1.93}$$

since the force is perpendicular to the boost direction.

Let's compare $\hat{\mathbf{F}}$ with **F** (they should be equal). Putting the expression for $\bar{\mathbf{F}}$ into (1.93), and writing everything in terms of λ_0,

$$\hat{\mathbf{F}} = -\frac{q\,\lambda_0}{2\,\pi\,\epsilon_0\,d}\,\frac{\gamma_{\bar{w}}}{\gamma_u}\,\hat{\mathbf{y}}. \tag{1.94}$$

As an algebraic identity, we have

$$\begin{aligned} \frac{1}{\gamma_{\bar{w}}^2} &= 1 - \frac{(v-u)^2}{c^2\left(1-\frac{u\,v}{c^2}\right)^2} \\ &= \frac{c^2 - 2\,u\,v + \frac{u^2\,v^2}{c^2} - v^2 + 2\,u\,v - u^2}{c^2\left(1-\frac{u\,v}{c^2}\right)^2} \\ &= \frac{c^2\left(1-\frac{v^2}{c^2}\right)\left(1-\frac{u^2}{c^2}\right)}{c^2\left(1-\frac{u\,v}{c^2}\right)^2} \\ &= \frac{1}{\gamma_v^2}\,\frac{1}{\gamma_u^2}\,\frac{1}{\left(1-\frac{u\,v}{c^2}\right)^2}, \end{aligned} \tag{1.95}$$

and then

$$\frac{\gamma_{\bar{w}}}{\gamma_u} = \gamma_v\left[1 - \frac{u\,v}{c^2}\right], \tag{1.96}$$

from which we conclude that

$$\hat{\mathbf{F}} = -\frac{q\,\lambda_0}{2\,\pi\,\epsilon_0\,d}\left[1-\frac{u\,v}{c^2}\right]\gamma_v\,\hat{\mathbf{y}}, \tag{1.97}$$

which is precisely (1.90): the prediction matches the force as calculated in the lab.

Notice that if we did *not* have the magnetic field/force, our original evaluation of the force in the lab, replacing (1.90), would have been:

$$\mathbf{F} = -\frac{q\,\lambda_0}{2\,\pi\,\epsilon_0\,d}\,\gamma_v\,\hat{\mathbf{y}}, \tag{1.98}$$

which does not match the prediction $\hat{\mathbf{F}}$ made in the rest frame of the charge. Without the magnetic field, then, we would predict different forces in different inertial frames, and that is forbidden. So one way of telling the story of the magnetic field is to start with the electric force, posit the existence of special relativity, and then require a new force to make inertial frames agree. That new force, associated with *moving* charge, is precisely the magnetic force.

It is interesting to note that we could, in theory, have started with Newtonian gravity, which has a "field" $\mathbf{g}$ (force per unit mass) related to massive sources ρ (mass density) by:

$$\nabla \cdot \mathbf{g} = -4\pi G \rho. \tag{1.99}$$

For a moving infinite line of mass, the $\mathbf{g}$ solving (1.99) is $\mathbf{g} = -\frac{2\lambda G}{s}\hat{\mathbf{s}}$. The force on a mass m is then $\mathbf{F}_N = -\frac{2m\lambda G}{s}\hat{\mathbf{s}}$. We could run an argument similar to the above to arrive at the conclusion that for Newtonian gravity to yield the same force predictions in all inertial frames, we'd better have a "gravitomagnetic" force, one playing the role of magnetism for gravity. That gravitomagnetic force would be sourced by moving mass and would act on moving massive bodies just as the magnetic force acts on moving charges. Of course, once you have that, you get the rest of the E&M structure pretty quickly, so that, for example, you could reasonably expect to find gravitational waves by analogy with electromagnetic waves, all coming just from Newtonian gravity modified by the demands of special relativity. We'll return to this notion in Section 4.5.1.

1.10 Four-Vectors

We'll close with some notation that will prove useful for our description of relativistic field theories like E&M and gravity.

The boost defined by the transformation (1.15) can be written in matrix-vector form, provided we expand our notion of a vector to include a "0-component." Let

$$x^\mu \doteq \begin{pmatrix} c\,t \\ x \\ y \\ z \end{pmatrix}, \qquad \text{for } \mu = 0, 1, 2, 3, \tag{1.100}$$

so $x^0 = c\,t$ and $x^i \doteq \mathbf{r}^i$ (for $i = 1, 2, 3$). Define the matrix

$$\mathbb{L} \doteq \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.101}$$

where we denote the entries of the matrix as $L^\nu_{\ \mu}$ for ν and μ both going from $0 \to 3$ (then $L^0_{\ 0} = \gamma$, $L^0_{\ 1} = -\beta\gamma$, etc.). The Lorentz boost transformation (for relative motion along the x-axis) is given by the matrix-vector multiplication

$$\bar{x}^\nu = \sum_{\mu=0}^{3} L^\nu_{\ \mu}\, x^\mu, \qquad \text{for } \nu = 0, 1, 2, 3. \tag{1.102}$$

We can boost in other directions, too. A boost in the y-direction has associated matrix:

$$\mathbb{L} \doteq \begin{pmatrix} \gamma & 0 & -\beta\gamma & 0 \\ 0 & 1 & 0 & 0 \\ -\beta\gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.103}$$

Any set of four quantities A^{μ} that are related, under a Lorentz boost (or spatial rotation, for that matter), by

$$\bar{A}^{\nu} = \sum_{\mu=0}^{3} L^{\nu}_{\ \mu} A^{\mu}, \qquad \text{for } \nu = 0, 1, 2, 3, \tag{1.104}$$

is called a "four-vector." The most common examples are: (1) the space-time coordinate vector, (2) the energy-momentum vector, and (3) the combination of the electric scalar and magnetic vector potentials into a singly indexed object. There are also physical objects that do *not* form part of a four-vector, like the spatial force vectors **F** from Section 1.8.

1.10.1 Einstein Summation Notation

The sums (1.102) and (1.104), representing the matrix-vector multiplication (and a variety of other more complicated sums), benefit from the "Einstein summation" notation. In any term of an expression, if there is a repeated index, appearing once up and once down, then there is an implicit sum. Repeated indices are called "dummy" or "closed" indices. Indices appearing up or down that are not repeated define vector components. These "open" indices are meant to run from $0 \to d$ (where d is the number of spatial dimensions, three for the most part). Open indices must appear with the same "variance" (up or down) in all expressions.

For the Lorentz boost in (1.102), we would write

$$\bar{x}^{\nu} = L^{\nu}_{\ \mu} x^{\mu} \tag{1.105}$$

in Einstein summation notation. The savings should be clear: we removed the summation symbol and baggage associated with the μ index on the right, since it is "closed." The verbiage associated with the ν index: "for $\nu = 0, 1, 2, 3$" has also been removed and is implicitly defined since ν appears in every term in the equation only once (and note that in each term, it appears in the "up" position).

The notation is powerful and can provide a nice check of correctness when complicated algebraic manipulation is required. For practice, take a look at the following.

1. $M^{\mu\nu}_{\ \ \ \gamma} T^{\gamma} = F^{\mu} G^{\nu} = M^{\mu\nu}_{\ \ \ \sigma} T^{\sigma}$. Here, we have a valid expression with two open indices, μ and ν, appearing everywhere in the "up" position. The index γ on the far left is a "closed" index and is automatically summed from $0 \to 3$. In the expression on the right, this closed index has been renamed σ, but it still represents a summation index that goes from $0 \to 3$. How many equations are encoded in the expression (i.e., μ and ν each take on the values from 0 to 3, for a total of how many equations)?

2. $A^\alpha B_\alpha C^{\alpha\beta} = P^\beta$. This expression is invalid because it has *three* αs on the left-hand side. The open index β is correctly matched on the left and right. One of the perqs of the notation is that error-correction is simplified. In this case, it may be that A^α should have been A (a scalar), for example, or perhaps it shouldn't have been there at all; $B_\alpha C^{\alpha\beta} = P^\beta$ is a valid expression.
3. $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R^{\alpha\beta} g_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\mu\nu}$. This is Einstein's equation for general relativity. It has a pair of closed indices, α and β, each of which runs from 0 to 3 independently. The objects $R_{\mu\nu}$ and $T_{\mu\nu}$ are symmetric: $R_{\mu\nu} = R_{\nu\mu}$, $T_{\mu\nu} = T_{\nu\mu}$ and $g_{\mu\nu} = g_{\nu\mu}$. Given that symmetry, how many independent equations are there in Einstein's equation?
4. $\bar{x}_\mu = L^\mu_{\ \nu} x^\nu$. In this expression, we have a closed ν-index. The open index μ is incorrect; it appears in the lower position on the left, but the upper position on the right.

We can define the dot-product of two four-vectors – index notation helps – the "length" associated with the four-vector x^μ is the square root of

$$x^\mu x_\mu \equiv -\left(x^0\right)^2 + \left(x^1\right)^2 + \left(x^2\right)^2 + \left(x^3\right)^2. \tag{1.106}$$

As the notation implies, $x_0 = -x^0$ and $x_i = x^i$ for $i = 1, 2, 3$ (these expression are not in indexed notation and are meant to express numerical equality). We can relate the up and down indices by letting

$$x_\mu \equiv \eta_{\mu\nu} x^\nu \tag{1.107}$$

for the "matrix" with entries $\eta_{\mu\nu}$:

$$\mathbb{N} \doteq \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.108}$$

The length (squared) can also be expressed as

$$x^\mu x_\mu = x^\mu x^\nu \eta_{\mu\nu}. \tag{1.109}$$

The matrix $\mathbb{N}$ has entries $\eta_{\mu\nu}$, and that collection of entries is called the "metric" and is associated with the definition of length.[11] We also define the upper form of the metric, $\eta^{\mu\nu}$, to be the matrix inverse of the lower form (in this case, with Cartesian coordinates, the matrix and its inverse are the same, but that won't be true in general). The defining property is

$$\eta^{\alpha\sigma} \eta_{\sigma\beta} = \delta^\alpha_\beta \tag{1.110}$$

where δ^α_β is the usual Kronecker delta (zero unless $\alpha = \beta$, in which case it is 1). You can use the upper form to move a lower index up: $x^\mu = \eta^{\mu\nu} x_\nu$.

Because the metric has a lower 3×3 block identity matrix, we can separate the temporal and spatial sections. We sometimes write the spatial portion of a four-vector using the usual bold notation, so you'll see things like:

[11] The metric is usually written as $g_{\mu\nu}$, with $\eta_{\mu\nu}$ the special symbol given to the Minkowski metric represented by (1.108) in Cartesian coordinates.

$$x^\mu\, x_\mu = -(c\,t)^2 + \mathbf{x}\cdot\mathbf{x} \tag{1.111}$$

or, for the velocity $\dot{x}^\mu = \frac{dx^\mu}{dt}$,

$$\dot{x}^\mu\, \dot{x}_\mu = -c^2 + \frac{d\mathbf{x}}{dt}\cdot\frac{d\mathbf{x}}{dt}. \tag{1.112}$$

Problem 1.29 Which of the following are valid expressions in Einstein summation notation? Explain what is wrong with the expressions that are not and generate a correct expression (that process will not be unique).

1. $A^\mu\, B^\nu\, C_\nu = F$.
2. $A^{\mu\nu\gamma}\, B_\nu = C^\gamma\, D^\mu$.
3. $A^\mu\, B^\mu\, C_\gamma = D_\gamma$.
4. $\epsilon^{\alpha\beta\gamma\delta}\, A_\alpha\, B_\gamma = F^{\beta\delta}$.
5. $A^\mu\, B_\mu\, C_\mu = D_\nu$.
6. $A_{\mu\nu}\, B^\nu\, F^\gamma = C^\gamma_\mu$.

Problem 1.30 Suppose we have $S^{\alpha\beta} = S^{\beta\alpha}$ and $A_{\alpha\beta} = -A_{\beta\alpha}$. Show that $S^{\alpha\beta}\, A_{\alpha\beta} = 0$.

Problem 1.31 Show that $A^\mu\, B_\mu = A_\mu\, B^\mu$.

1.10.2 Energy–Momentum Four-Vector

We saw that E/c and $\mathbf{p}$ transform like $c\,t$ and $\mathbf{x}$. This means that energy and momentum can be combined into a four-vector, just like the spatial one (and once again, c can be used unambiguously since it takes on the same value for all reference frames):

$$p^\mu \doteq \begin{pmatrix} E/c \\ p^x \\ p^y \\ p^z \end{pmatrix} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}. \tag{1.113}$$

The length (squared) of this vector is (letting $p^2 \equiv \mathbf{p}\cdot\mathbf{p} = (p^x)^2 + (p^y)^2 + (p^z)^2$):

$$p^\mu\, p_\mu = p_\gamma\, p^\gamma = p^\alpha\, \eta_{\alpha\gamma}\, p^\gamma = -\frac{E^2}{c^2} + p^2 = -\frac{m^2\, c^2}{\left(1-\frac{v^2}{c^2}\right)} + \frac{m^2\, v^2}{\left(1-\frac{v^2}{c^2}\right)} = -m^2\, c^2 \tag{1.114}$$

as we established back in Section 1.7.1. This expression gives a relation between energy and momentum for both massive and massless particles (if $m = 0$, $p = \pm E/c$).

Problem 1.32 For a generic four-vector A^ν with components measured in a lab frame L, we have the transformed form for components measured in $\bar{L}$: $\bar{A}^\nu = L^\nu{}_\mu\, A^\mu$ where $L^\nu{}_\mu$ is the matrix associated with a Lorentz boost in the x-direction (the elements of (1.101)). Show that the "length" (squared) is the same in either L or $\bar{L}$; i.e., show that $\bar{A}^\nu\, \bar{A}_\nu = A^\nu\, A_\nu$ (where we lower the index using the metric as usual).

Problem 1.33 For Newton's second law in one spatial dimension, if we have a constant force F_0 (like gravity near the surface of the earth, for example), the equation reads: $\frac{dp}{dt} = F_0$ with p the relativistic momentum. Solve this equation for $t(v)$, the time as a function of velocity assuming a particle of mass m starts at the origin from rest (at $t = 0$). Suppose Newton's second law instead read: $\frac{dp}{d\tau} = F_0$ where τ is the proper time of the particle. Solve for $t(v)$ in this case, again assuming the particle starts from the origin at rest.

Problem 1.34

a. Write the components of the (manifestly) four-vector $\frac{dp^\mu}{d\tau}$, i.e., the derivative of the four-momentum with respect to the particle's proper time. Using the definition of proper time, write the components of this four-vector entirely in terms of $\frac{d\mathbf{p}}{dt}$ and $\mathbf{v}$ (where $\mathbf{p}$ contains the spatial components of the relativistic momentum, and $\mathbf{v}$ is the particle's velocity "in the lab").

b. Given a force from classical mechanics, $\mathbf{F}_{pr}$ (for "pre-relativity"), we updated Newton's second law to read $\frac{d\mathbf{p}}{dt} = \mathbf{F}_{pr}$ where we used the coordinate time, but the relativistic momentum. Given the equivalence of t and τ in pre-relativistic physics, we could just as well have had $\frac{d\mathbf{p}}{d\tau} = \mathbf{F}_{pr}$ for the update (again with relativistic momentum $\mathbf{p}$). Write out the equations of motion for this latter case "in the lab" (meaning in terms of t and $\mathbf{v} = \frac{d\mathbf{x}}{dt}$). Which one of these two, $\frac{d\mathbf{p}}{dt} = \mathbf{F}_{pr}$ or $\frac{d\mathbf{p}}{d\tau} = \mathbf{F}_{pr}$, is appropriate for the Lorentz force $\mathbf{F}_{pr} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})$?

Problem 1.35 The location of a mass in uniform circular motion of radius R with angular frequency ω is $\mathbf{r}(t) = R\,\cos(\omega\, t)\,\hat{\mathbf{x}} + R\,\sin(\omega\, t)\,\hat{\mathbf{y}}$ (in the x–y plane). If the circular motion is sustained by an inward directed force (could be E&M or gravity, for example) $\mathbf{F} = -F\,\hat{\mathbf{r}}$, find the relativistic analogue of $\frac{m\,v^2}{R} = F$ (hint: start from the relativistic form of Newton's second law, insert the circular motion, and algebrize).

1.10.3 Contravariant versus Covariant

In Einstein summation notation, we have up-down pairs of matched indices for summation. We use the metric to relate up and down indices, but what is the fundamental difference between them? When we think about vectors like $\mathbf{f} = f^x\,\hat{\mathbf{x}} + f^y\,\hat{\mathbf{y}} + f^z\,\hat{\mathbf{z}}$ (and you can include a temporal coordinate as well), there are two elements: the entries $\{f^x, f^y, f^z\}$ (or "components") and the basis vectors $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$. Both of these elements respond to coordinate transformations, and they do so differently. When we write f^α, we are focusing on the components and their transformation.

A contravariant (index up) first rank (one index) tensor responds to a coordinate transformation $x^\mu \longrightarrow \bar{x}^\mu$ (from old to new coordinates) like this:

$$\bar{f}^\mu = \frac{\partial \bar{x}^\mu}{\partial x^\nu} f^\nu. \tag{1.115}$$

As an example of this type of transformation, consider the vector $d\mathbf{r} = dx\,\hat{\mathbf{x}} + dy\,\hat{\mathbf{y}} + dz\,\hat{\mathbf{z}}$, with components dr^α. The transformed vector is $d\bar{\mathbf{r}} = d\bar{x}\,\hat{\bar{\mathbf{x}}} + d\bar{y}\,\hat{\bar{\mathbf{y}}} + d\bar{z}\,\hat{\bar{\mathbf{z}}}$ with components $d\bar{r}^\alpha$. Take $\alpha = 1$ of the new components, $d\bar{x}^1 = d\bar{x}$, and let's write it in terms of the old components. From calculus, viewing $\bar{x}$ as a function of x, y, z, we have

$$d\bar{x} = \frac{\partial \bar{x}}{\partial x}\, dx + \frac{\partial \bar{x}}{\partial y}\, dy + \frac{\partial \bar{x}}{\partial z}\, dz = \frac{\partial \bar{x}}{\partial x^\mu}\, dr^\mu, \tag{1.116}$$

and more generally (applying this to each of the components of $d\bar{\mathbf{r}}$)

$$d\bar{r}^\alpha = \frac{\partial \bar{x}^\alpha}{\partial x^\mu}\, dr^\mu \tag{1.117}$$

so that the components of the "coordinate differential," dr^μ, form a contravariant tensor (usually denoted dx^μ; I wanted to avoid strange statements like $d\mathbf{x} = dx\,\hat{\mathbf{x}} + \cdots$ which are difficult to parse).

A covariant (index down, "co goes below") first rank (one index) tensor responds to a coordinate transformation via

$$\bar{f}_\mu = \frac{\partial x^\nu}{\partial \bar{x}^\mu} f_\nu. \tag{1.118}$$

The classic example of a covariant tensor is the gradient of a scalar function: $\phi_{,\mu} \equiv \frac{\partial \phi}{\partial x^\mu}$. If we change from x to $\bar{x}$ coordinates, the scalar function ϕ is unchanged: $\bar{\phi}(\bar{x}) = \phi(\bar{x})$; we just evaluate ϕ in terms of the new variables. Then the chain rule of calculus gives

$$\bar{\phi}_{,\mu} \equiv \frac{\partial \bar{\phi}}{\partial \bar{x}^\mu} = \frac{\partial \phi}{\partial \bar{x}^\mu} = \frac{\partial \phi}{\partial x^\nu}\frac{\partial x^\nu}{\partial \bar{x}^\mu} = \frac{\partial x^\nu}{\partial \bar{x}^\mu}\,\phi_{,\nu}, \tag{1.119}$$

which is of the appropriate form. This type of transformation depends on the inverse of the $\frac{\partial \bar{x}^\mu}{\partial x^\nu}$ appearing in the contravariant definition (1.115).

Going back to our original vector $\mathbf{f} = f^x\,\hat{\mathbf{x}} + f^y\,\hat{\mathbf{y}} + f^z\,\hat{\mathbf{z}}$, we can think of this as a product of components and basis vectors. Let $\mathbf{e}_\mu$ be $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ for $\mu = 1, 2, 3$. We could write $\mathbf{f}$ as

$$\mathbf{f} = f^\alpha\,\mathbf{e}_\alpha, \tag{1.120}$$

where the components are contravariant, and the basis vectors are covariant. That's nice, because it is clear that transforming $\mathbf{f}$ under $x \to \bar{x}$ gives[12]

$$\bar{\mathbf{f}} = \bar{f}^\alpha\,\bar{\mathbf{e}}_\alpha = \frac{\partial \bar{x}^\alpha}{\partial x^\nu} f^\nu \frac{\partial x^\mu}{\partial \bar{x}^\alpha}\,\mathbf{e}_\mu = f^\nu\,\mathbf{e}_\mu\,\delta^\mu_\nu = f^\nu\,\mathbf{e}_\nu = \mathbf{f}. \tag{1.121}$$

The change in the components undoes the change in the basis vectors, and then the geometric object, **f**, is the same in both coordinate systems, as must be the case (a vector that points toward a particle, say, must point toward that particle in all coordinate systems no matter what the coordinates or basis vectors).

The coordinate transformations we have highlighted in this chapter are Lorentz boosts and rotations. In those cases, we have $\bar{x}^\mu = L^\mu{}_\nu\, x^\nu$ so that $\frac{\partial \bar{x}^\mu}{\partial x^\nu} = L^\mu{}_\nu$, a set of constant entries. But our definitions of covariant and contravariant are independent of the details of the transformation we have in mind. As an example, let's take as the old coordinates Cartesian $x = x^1$, $y = x^2$, and $z = x^3$, and for the new coordinates, spherical $r = \bar{x}^1$,

[12] Using the identity

$$\frac{\partial \bar{x}^\alpha}{\partial x^\nu}\frac{\partial x^\mu}{\partial \bar{x}^\alpha} = \frac{\partial x^\mu}{\partial x^\nu} = \delta^\mu_\nu.$$

$\theta = \bar{x}^2$, $\phi = \bar{x}^3$. As our vector, we'll evaluate the transformation of $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$ (both bases and components). From the coordinate transformation,

$$\begin{aligned}\bar{x}^1 &= \sqrt{x^2+y^2+z^2}\\ \bar{x}^2 &= \tan^{-1}\left(\frac{\sqrt{x^2+y^2}}{z}\right)\\ \bar{x}^3 &= \tan^{-1}\left(\frac{y}{x}\right),\end{aligned} \tag{1.122}$$

and its inverse:

$$\begin{aligned}x &= \bar{x}^1\,\sin\bar{x}^2\,\cos\bar{x}^3\\ y &= \bar{x}^1\,\sin\bar{x}^2\,\sin\bar{x}^3\\ z &= \bar{x}^1\,\cos\bar{x}^2,\end{aligned} \tag{1.123}$$

we have

$$\begin{aligned}\frac{\partial\bar{x}^\alpha}{\partial x^\nu} &\doteq \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2+z^2}} & \frac{y}{\sqrt{x^2+y^2+z^2}} & \frac{z}{\sqrt{x^2+y^2+z^2}}\\ \frac{x\,z}{\sqrt{x^2+y^2}\,(x^2+y^2+z^2)} & \frac{y\,z}{\sqrt{x^2+y^2}\,(x^2+y^2+z^2)} & -\frac{\sqrt{x^2+y^2}}{(x^2+y^2+z^2)}\\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} & 0\end{pmatrix}\\ \frac{\partial x^\mu}{\partial\bar{x}^\alpha} &\doteq \begin{pmatrix}\sin\theta\,\cos\phi & r\,\cos\theta\,\cos\phi & -r\,\sin\theta\,\sin\phi\\ \sin\theta\,\sin\phi & r\,\cos\theta\,\sin\phi & r\,\sin\theta\,\cos\phi\\ \cos\theta & -r\,\sin\theta & 0\end{pmatrix}\end{aligned} \tag{1.124}$$

and either matrix expression could be written in either set of coordinates.

The transformation rule for the components gives

$$\bar{r}^\alpha = \frac{\partial\bar{x}^\alpha}{\partial x^\nu}\,r^\nu \doteq \begin{pmatrix}\sqrt{x^2+y^2+z^2}\\ 0\\ 0\end{pmatrix} = \begin{pmatrix} r\\ 0\\ 0\end{pmatrix} \tag{1.125}$$

while the basis vectors have

$$\bar{\mathbf{e}}_\alpha = \frac{\partial x^\mu}{\partial\bar{x}^\alpha}\,\mathbf{e}_\mu \doteq \begin{pmatrix}\sin\theta\,\cos\phi\,\hat{\mathbf{x}} + \sin\theta\,\sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}}\\ r\left(\cos\theta\,\cos\phi\,\hat{\mathbf{x}} + \cos\theta\,\sin\phi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}}\right)\\ r\,\sin\theta\left(-\sin\phi\,\hat{\mathbf{x}} + \cos\phi\,\hat{\mathbf{y}}\right).\end{pmatrix}. \tag{1.126}$$

These new basis vectors are not normalized (no guarantees of that, of course), but we can see the familiar $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ lying inside $\bar{\mathbf{e}}_\alpha$. Putting the components together with the basis vectors gives

$$\bar{\mathbf{r}} = \bar{r}^\alpha\,\bar{\mathbf{e}}_\alpha = \bar{r}\,\bar{\mathbf{e}}_1 = r\,\hat{\mathbf{r}}, \tag{1.127}$$

which is what we expect, of course.

Problem 1.36 Verify that $\frac{\partial\bar{x}^\alpha}{\partial x^\nu}\frac{\partial x^\mu}{\partial\bar{x}^\alpha} = \delta^\mu_\nu$ using the expressions from (1.124).

Problem 1.37 Show, using the transformation rules for contravariant and covariant first rank tensors, that $A^\mu\,B_\mu$ is a scalar if A^μ is a contravariant first-rank tensor and B_μ is a covariant first-rank tensor.

Problem 1.38 Tensors of higher rank transform with additional factors of $\frac{\partial \bar{x}^\alpha}{\partial x^\beta}$ (or its inverse), so that, for example, the second-rank covariant metric tensor transforms like

$$\bar{g}_{\mu\nu} = \frac{\partial x^\alpha}{\partial \bar{x}^\mu} \frac{\partial x^\beta}{\partial \bar{x}^\nu} g_{\alpha\beta}. \tag{1.128}$$

From this, show that the matrix inverse of the metric, denoted $g^{\alpha\beta}$, is indeed a second-rank contravariant tensor by showing that it transforms like

$$\bar{g}^{\mu\nu} = \frac{\partial \bar{x}^\mu}{\partial x^\alpha} \frac{\partial \bar{x}^\nu}{\partial x^\beta} g^{\alpha\beta}. \tag{1.129}$$

We use the metric to raise and lower indices. If f_μ is a covariant first-rank tensor, show that $f^\mu \equiv g^{\mu\nu} f_\nu$ is a contravariant first-rank tensor.

Let's think about the "gradient vector" that you use in E&M all the time. What do we make of

$$\nabla s = \frac{\partial s}{\partial x}\,\hat{\mathbf{x}} + \frac{\partial s}{\partial y}\,\hat{\mathbf{y}} + \frac{\partial s}{\partial z}\,\hat{\mathbf{z}}? \tag{1.130}$$

The components form a covariant tensor (as we showed above), and the basis vectors are part of a covariant tensor as well. How can we write ∇s in component-basis form? There must be a factor of the metric inverse $g^{\mu\nu}$ in place:

$$\nabla s = s_{,\mu}\, \mathbf{e}_\nu\, g^{\mu\nu}. \tag{1.131}$$

If we again use the Cartesian to spherical transformation, we can write out the transformed version, $\bar{\nabla}\bar{s}$, noting that the metric, in spherical coordinates, is

$$\bar{g}_{\mu\nu} \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \tag{1.132}$$

and then

$$\bar{\nabla}\bar{s} = \bar{s}_{,\mu}\, \bar{\mathbf{e}}_\nu\, \bar{g}^{\mu\nu} \doteq \frac{\partial s}{\partial r}\,\bar{\mathbf{e}}_1 + \frac{\partial s}{\partial \theta}\frac{1}{r^2}\,\bar{\mathbf{e}}_2 + \frac{\partial s}{\partial \phi}\frac{1}{r^2 \sin^2\theta}\,\bar{\mathbf{e}}_3. \tag{1.133}$$

Using $\bar{\mathbf{e}}_1 = \hat{\mathbf{r}}$, $\bar{\mathbf{e}}_2 = r\,\hat{\boldsymbol{\theta}}$ and $\bar{\mathbf{e}}_3 = r\sin\theta\,\hat{\boldsymbol{\phi}}$ from (1.126), we can write

$$\bar{\nabla}\bar{s} = \frac{\partial s}{\partial r}\,\hat{\mathbf{r}} + \frac{1}{r}\frac{\partial s}{\partial \theta}\,\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial s}{\partial \phi}\,\hat{\boldsymbol{\phi}}, \tag{1.134}$$

the correct result.

This example highlights the importance of knowing where the objects you are working with live. If you assumed the components of $\bar{\nabla}\bar{s}$ were $\bar{s}_{,\mu}$ (instead of $\bar{s}_{,}{}^{\mu}$), you'd have:

$$\begin{aligned} \bar{\nabla}\bar{s} &= \frac{\partial s}{\partial r}\,\bar{\mathbf{e}}_1 + \frac{\partial s}{\partial \theta}\,\bar{\mathbf{e}}_2 + \frac{\partial s}{\partial \phi}\,\bar{\mathbf{e}}_3 \\ &= \frac{\partial s}{\partial r}\,\hat{\mathbf{r}} + \frac{\partial s}{\partial \theta}\,r\,\hat{\boldsymbol{\theta}} + \frac{\partial s}{\partial \phi}\,r\sin\theta\,\hat{\boldsymbol{\phi}}, \end{aligned} \tag{1.135}$$

which is incorrect (the individual terms don't even match in dimension).

As a final note: because of the vector form $\mathbf{f} = f^\alpha \, \mathbf{e}_\alpha$, most of the quantities you work with have contravariant components. Think of the electric field, $\mathbf{E} = E^\alpha \, \mathbf{e}_\alpha$, with components E^α. Sometimes, as with the gradient vector, that contravariance is at odds with the fundamental place of an object, $s_{,}{}^{\nu} \equiv g^{\mu\nu} \, s_{,\mu}$ so that while you are most familiar with what we call $s_{,}{}^{\nu}$ (the components appearing in (1.131)), that is a derived quantity, the actual derivatives are covariant.

Problem 1.39 Provide an "external" check of (1.134) – start with ∇s in Cartesian coordinates, then rewrite the derivatives using the chain rule, (so, for example, $\frac{\partial s}{\partial x} = \frac{\partial s}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial s}{\partial \theta}\frac{\partial \theta}{\partial x} + \frac{\partial s}{\partial \phi}\frac{\partial \phi}{\partial x}$) and then rewrite the basis vectors using the standard relation between Cartesian and spherical basis vectors (from the back cover of [19], for example).

Problem 1.40 Working from the transformation law (1.128), and using the specific transformation from Cartesian coordinates to spherical, show that you get (1.132) as the form of the metric in spherical coordinates.

Problem 1.41 The electric and magnetic fields are not clearly elements of a four vector. In this problem we'll try to cook up a plausible relativistic form for them. The fields are related to the potentials by

$$\begin{aligned} \mathbf{E} &= -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \\ \mathbf{B} &= \nabla \times \mathbf{A}. \end{aligned} \tag{1.136}$$

We want to use four-vector building blocks to construct $\mathbf{E}$ and $\mathbf{B}$. The building blocks are the potential four-vector, and the gradient:

$$A^\mu \doteq \begin{pmatrix} V/c \\ A^x \\ A^y \\ A^z \end{pmatrix}, \qquad \partial_\nu \doteq \begin{pmatrix} \frac{1}{c}\frac{\partial}{\partial t} \\ \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}. \tag{1.137}$$

We know that we must combine ∂^ν (note the upper form, that's for convenience, but don't forget to raise the index on the gradient) and A^μ somehow, since $\mathbf{E}$ and $\mathbf{B}$ involve the derivatives of V and $\mathbf{A}$. We also need six separate equations, so we can't just make, for example,[13] $\partial^\mu A_\mu = 0$ since that is only one equation. Write down the entries of $\partial^\mu A^\nu$ for $\mu = 0 \to 3$, $\nu = 0 \to 3$ (go ahead and think of matrix entries, here, using μ to index rows, ν to index columns). See anything familiar? Try swapping the indices, and writing the matrix form of $\partial^\nu A^\mu$ (retain μ as the row index, ν as the column index) – is there any way to combine these matrices to obtain the right-hand side of (1.136) (up to $\pm$ and factors of c)?

Problem 1.42

a. In a lab, there is a constant magnetic field pointing into the page (everywhere), $\mathbf{B} = B_0 \, \hat{\mathbf{x}}$. A charge sits at rest in the lab on the $\hat{\mathbf{y}}$-axis. What is the force on the charge?

[13] Which is a statement of the Lorenz gauge condition.

b. Suppose you move along the $\hat{\mathbf{y}}$-axis with constant speed v. From your point of view, the charge is moving in the $-\hat{\mathbf{y}}$-direction with speed v. What is the total force on the charge? (Transform the fields using your tensor expression from Problem 1.41 and find the force in the new frame.)

Problem 1.43 A pair of infinite sheets of charge lies in the x–y plane. The top one has charge density $\sigma_0 > 0$ (when at rest) and the bottom one has $-\sigma_0$ (at rest). The top sheet moves along the $\hat{\mathbf{y}}$-axis with speed v, and the bottom sheet moves in the $-\hat{\mathbf{y}}$-direction with speed v.

a. Find the electric and magnetic fields above the top sheet.

b. Suppose you move along the $\hat{\mathbf{y}}$-axis with speed v. Find the electric and magnetic fields in this moving reference frame (using the physical characteristics of the sheets that would be observed in this frame). Compare with your result from part (b) of Problem 1.42.

2 Point Particle Fields

In this chapter, we will explore the fields (meaning both the electromagnetic fields and their potentials) of point particles. The static fields are familiar to us from E&M, and we'll use them to introduce the notion of a Green's function. Moving point particles have Green's functions that are more complicated, and we will spend some time thinking about both "pure" point particle[1] fields and the more realistic fields of moving charges. When charges move, they radiate, and the end of the chapter explores this new component to the fields generated by charge motion.

2.1 Definition

For the Poisson problem $\nabla^2 V = -\frac{\rho}{\epsilon_0}$ (using the E&M form, with a given, finite charge density ρ), we know that the general integral solution can be written as

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{\mathscr{r}}\,d\tau', \tag{2.1}$$

where $\boldsymbol{\mathscr{r}} \equiv \mathbf{r} - \mathbf{r}'$. The physical intuition here comes from the point source solution

$$V_\circ = \frac{q}{4\pi\epsilon_0\,\mathscr{r}} \tag{2.2}$$

applied to a continuous distribution of charge. We use superposition:[2] chop up the distribution into pieces, figure out the contribution due to each piece, then add them up. Referring to Figure 2.1, if we are given a charge density ρ, then in a small volume $d\tau'$ located at $\mathbf{r}'$, there is charge $dq = \rho(\mathbf{r}')\,d\tau'$. That charge contributes to the potential at $\mathbf{r}$, using (2.2),

$$dV = \frac{dq}{4\pi\epsilon_0\,\mathscr{r}} = \frac{\rho(\mathbf{r}')\,d\tau'}{4\pi\epsilon_0\,\mathscr{r}}, \tag{2.3}$$

and superposition tells us we can add up these contributions to arrive at (2.1).

For the generalized Poisson problem $\nabla^2\phi = -s$, where s is a "source" of some sort, the integral solution can again be written in terms of the "point source" solution. In

[1] The Green's function source is a point charge that "flashes" on and off instantaneously. This type of source does not conserve charge and is not meant to be taken as a real physical configuration.

[2] Superposition is a property that must be experimentally observed/verified. There are two inputs in, for example, static E&M: (1) forces fall off like $1/r^2$ and act along the line connecting two charged particles, and (2) the forces associated with multiple particles add as vectors.

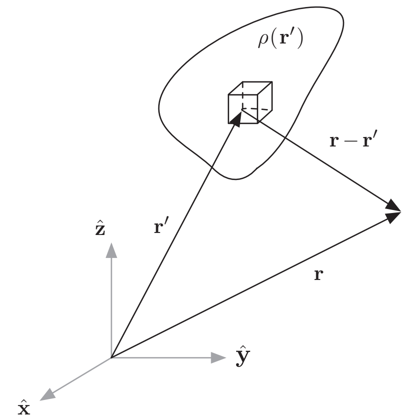

Fig. 2.1 Given a charge distribution $\rho(\mathbf{r}')$ (charge per unit volume), we consider a small volume $d\tau'$ at $\mathbf{r}'$ that has infinitesimal charge $dq = \rho(\mathbf{r}')\, d\tau'$ in it. This charge contributes $dV = dq/(4\pi\,\epsilon_0\,|\mathbf{r}-\mathbf{r}'|)$ to the potential at "field" location $\mathbf{r}$.

this context, the point source solution is called a "Green's function," and we'll denote it $G(\mathbf{r}, \mathbf{r}')$. Formally, the Green's function solves the particular Poisson problem

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = -\delta^3(\mathbf{r}-\mathbf{r}'), \tag{2.4}$$

where the right-hand side represents a point source (a density) at $\mathbf{r}'$. We don't usually specify the boundary conditions (although we should), but we'll take the implicit condition that $G(\mathbf{r}, \mathbf{r}') \rightarrow 0$ as $\mathbf{r}$ goes to infinity. The Green's function is sensitive to boundary conditions (different boundary conditions, different Green's functions). In addition, the Green's function is generally dependent on dimension.

Once you have solved (2.4), the utility lies in providing an integral solution to the Poisson problem, as above:

$$\phi(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}')\, s(\mathbf{r}')\, d\tau', \tag{2.5}$$

which almost seems like a tautology – if we act on both sides with ∇^2, it slips through the integral on the right (the $\mathbf{r}'$ variables are "dummy" integration variables) and hits G, which reproduces a delta function that can be used to do the integral, recovering $\nabla^2\phi(\mathbf{r}) = -s(\mathbf{r})$.

2.1.1 Example: Poisson $D = 3$

Let's start with the most familiar case, the Poisson problem in $D = 3$. We want the Green's function solving (2.4) in three dimensions (where does dimension show up in the problem?):

$$\nabla^2\, G(\mathbf{r}, \mathbf{r}') = -\delta^3(\mathbf{r}-\mathbf{r}'). \tag{2.6}$$

We find $G(\mathbf{r}, \mathbf{r}')$ via a four-step process:

1. Set the source at the origin, so that $\mathbf{r}' = 0$.
   Done, let $\nabla^2 G(\mathbf{r}) = -\delta^3(\mathbf{r})$.

2. Note that for all points in space that are not at the source (i.e., for all $\mathbf{r} \neq 0$), the equation of interest is $\nabla^2 G = 0$. We can solve that using spherical symmetry. A true point source has no structure, so the source is spherically symmetric, and it is reasonable to assume that its field will be, too.

   Spherical symmetry for $G(\mathbf{r})$ means that $G(\mathbf{r}) = G(r)$; the Green's function at a generic location $\mathbf{r}$ depends only on the distance to the source (which is at the origin, so the distance to the source is just r). With no θ- or ϕ-dependence in G, we have $\nabla^2 G = 1/r\,(r\,G)'' = 0$ where primes denote derivatives with respect to r, the radial coordinate. But this equation can be integrated twice to give

$$(r\,G)' = \beta \longrightarrow r\,G = \beta\,r + \alpha \longrightarrow G = \beta + \frac{\alpha}{r}. \tag{2.7}$$

   There are two constants of integration, α and β, that have shown up in the general solution to this second-order differential equation.
3. Use the boundary conditions (both at spatial infinity and at the origin) to fix any constants of integration.

   At spatial infinity, we'd like the point source solution to vanish,[3] so set $\beta = 0$. The other "boundary" is at $r = 0$, where the source is. To make it actionable, we'll integrate the defining equation (2.6) (with $\mathbf{r}' = 0$) over a ball of radius ϵ centered at the origin (to get rid of the delta function):

$$\int_{B(\epsilon)} \nabla \cdot \nabla G(\mathbf{r})\, d\tau' = -1, \tag{2.8}$$

   and we can use the divergence theorem on the left to integrate over the surface of the sphere of radius ϵ, denoted $\partial B(\epsilon)$,[4]

$$\oint_{\partial B(\epsilon)} \nabla G(\mathbf{r}) \cdot \left(\epsilon^2 \sin^2 d\theta\, d\phi\, \hat{\mathbf{r}}\right) = -1, \tag{2.9}$$

   or

$$-4\,\pi\,\epsilon^2\,\frac{\alpha}{\epsilon^2} = -1 \longrightarrow \alpha = \frac{1}{4\,\pi}. \tag{2.10}$$

   Our Green's function is

$$G(\mathbf{r}) = \frac{1}{4\,\pi\,r}, \tag{2.11}$$

   which should look very familiar from E&M.
4. You now have the Green's function for a source at the origin. To move the source around, just replace $\mathbf{r} \longrightarrow \boldsymbol{\mathscr{r}} \equiv \mathbf{r} - \mathbf{r}'$.

   That's easy:

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\,\pi\,\mathscr{r}}. \tag{2.12}$$

[3] Infinitely far away from sources, we should not be able to detect the source. That really means that the physical field (for example, the electric field with $\mathbf{E} = -\nabla G/\epsilon_0$) should vanish. But in $D = 3$, for this problem, we can also choose to zero the potential at spatial infinity.

[4] In the more general setting, we would take the limit as $\epsilon \to 0$, but for the current case, the result is ϵ-independent.

2.1.2 Example: Helmholtz $D = 3$

Suppose we have a modified equation, like the Helmholtz equation:

$$\nabla^2 \phi + \mu^2 \phi = -s, \tag{2.13}$$

for source s and constant μ. The Green's function problem is defined as usual by

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') + \mu^2 G(\mathbf{r}, \mathbf{r}') = -\delta^3(\mathbf{r} - \mathbf{r}') \tag{2.14}$$

(what are the dimensions of μ here?). Using the first step above, we set $\mathbf{r}' = 0$, and assume spherical symmetry, then solve

$$\nabla^2 G + \mu^2 G = 0 \tag{2.15}$$

(away from the source):

$$1/r\,(r\,G)'' = -\mu^2\,G \longrightarrow (r\,G)'' = -\mu^2\,(r\,G)\,. \tag{2.16}$$

In the second form, it's clear that the solution for $r\,G$ is a sum of exponentials:

$$r\,G = A\,e^{i\,\mu\,r} + B\,e^{-i\,\mu\,r} \longrightarrow G = \frac{A\,e^{i\,\mu\,r} + B\,e^{-i\,\mu\,r}}{r}. \tag{2.17}$$

To set the boundary conditions, we need to know more about the problem. If μ is real, then both terms might contribute, If μ is purely imaginary, we'd get rid of the growing exponential portion. For now, we know that when $\mu = 0$, we must recover (2.11), so given two new constants a and b, we have

$$G = \frac{a\,e^{i\,\mu\,r} + b\,e^{-i\,\mu\,r}}{4\,\pi\,r} \tag{2.18}$$

with $a + b = 1$. Moving the source to $\mathbf{r}'$, we have

$$G(\mathbf{r}, \mathbf{r}') = \frac{a\,e^{i\,\mu\,\mathscr{r}} + b\,e^{-i\,\mu\,\mathscr{r}}}{4\,\pi\,\mathscr{r}}. \tag{2.19}$$

Problem 2.1 Find the Green's function for the Poisson problem $\nabla^2 V = -k\,\rho$ in two dimensions, for constant k (what are its units assuming $\mathbf{E} = -\nabla V$ has its usual units and interpretation?) and given ρ, a charge-per-unit-area (the $D = 2$ version of a charge density). You won't be able to set the Green's function to zero at spatial infinity, so set it to zero at some arbitrary distance, s_0, from the point charge.

Problem 2.2 In three dimensions, the Green's function for the problem $\nabla^2\phi - \alpha^2\,\phi = -s$ is, pre-normalization (and for source at the origin), $G(r) = \frac{A\,e^{-\alpha\,r}}{r}$. Find the normalization constant by integrating the partial differential equation (PDE) over a sphere of radius ϵ, then taking the limit as $\epsilon \to 0$ to recover the correct value for A.

Problem 2.3 Find the Green's function for $\nabla^2\phi - \alpha^2\,\phi = -s$ in two dimensions.

Problem 2.4 For a uniformly charged sphere of radius R with charge Q (so constant charge density $\rho_0 = Q/(4/3\pi R^3)$), solve $\nabla^2 V = -\frac{\rho}{\epsilon_0}$ inside the sphere directly (no

Green's functions; just solve the ordinary differential equation (ODE) that comes from assuming that $V(\mathbf{r}) = V(r)$). Solve outside as well, and match the solutions and their derivatives at $r = R$; i.e., use $V_{\text{in}}(R) = V_{\text{out}}(R)$ and $V'_{\text{in}}(R) = V'_{\text{out}}(R)$ to fix constants (you will also need to use reasonable boundary conditions at $r = 0$ and $r \to \infty$). Find the associated electric fields inside and out.

Problem 2.5 If Maxwell's equation for V read $\nabla^2 V - \alpha^2 V = -\frac{\rho}{\epsilon_0}$, find the V inside and outside a uniformly charged solid sphere (with total charge Q, radius R). Use continuity and derivative continuity for V to set constants (in addition to requiring that the solution be finite at the origin and zero at spatial infinity). From V, find the electric field $\mathbf{E} = -\nabla V$ both inside and out (it is algebraically involved to verify that these fields limit to the correct form as $\alpha \to 0$, but that is a good check if you're up for it). Note that while you can develop and use the Green's function for the PDE here, it is probably easier to solve the PDE directly with the constant ρ in place as in the previous problem.

2.2 Four-Dimensional Poisson Problem

The natural relativistic scalar equation in the $3 + 1$-dimensional space-time of ordinary existence (and most other places) is[5]

$$\Box\phi \equiv -\frac{1}{c^2}\frac{\partial^2\phi(\mathbf{r}, t)}{\partial t^2} + \nabla^2\phi(\mathbf{r}, t) = -s(\mathbf{r}, t), \tag{2.20}$$

and of course, we have numerous examples of this type of equation. The electric and magnetic vector potential both solve a four-dimensional Poisson problem (in Lorenz gauge):

$$\begin{aligned} \Box V &= -\frac{\rho}{\epsilon_0} \\ \Box \mathbf{A} &= -\mu_0 \, \mathbf{J}. \end{aligned} \tag{2.21}$$

The equation is linear, so we expect to be able to find the Green's function and then use superposition to write the general solution to (2.20) as

$$\phi(\mathbf{r}, t) = \int G(\mathbf{r}, \mathbf{r}', t, t') \, s(\mathbf{r}', t') \, d\tau' \, dt'. \tag{2.22}$$

This is the formal solution written with spatial and temporal variables separated out. A more natural way to write the integral is to think of the position four-vector x^μ that combines both space and time, then the "volume element" is $d^4x = d\tau' \, (c \, dt')$. To keep the solution clear, it is useful to retain the separation for now.[6]

[5] The operator $\Box \equiv \partial_\mu \partial^\mu$, extending ∇^2 to space-time, is called the D'Alembertian.

[6] The Green's function can have dimension of either $1/\ell^2$ (inverse length squared) or $1/(\ell \, t)$ (inverse length times time). The choice depends on whether you are integrating over space-time using $d\tau' \, dt'$ or $d\tau' \, (c \, dt')$.

2.2.1 Fourier Transform

We're going to use the Fourier transform to solve for the Green's function in the space-time setting. There are many conventions for the Fourier transform, and we'll set the ones for this book here.

Given a function of time $p(t)$, the Fourier transform is

$$\tilde{p}(f) = \int_{-\infty}^{\infty} p(t)\, e^{2\pi i f t}\, dt \tag{2.23}$$

with inverse transformation

$$p(t) = \int_{-\infty}^{\infty} \tilde{p}(f)\, e^{-2\pi i f t}\, df. \tag{2.24}$$

In this temporal setting, the Fourier transform depends on f, a frequency with dimension of inverse time.

We can also define the Fourier transform of a spatial function, $q(\mathbf{r})$:

$$\tilde{q}(\mathbf{k}) = \int q(\mathbf{r})\, e^{2\pi i \mathbf{r}\cdot\mathbf{k}}\, d\tau \tag{2.25}$$

where the integral is over all space (a volume integral). For three-dimensional $\mathbf{r}$, the Fourier transform depends on a three-dimensional "wave vector" $\mathbf{k}$, with dimension of inverse length. The inverse Fourier transform here is

$$q(\mathbf{r}) = \int \tilde{q}(\mathbf{k})\, e^{-2\pi i \mathbf{r}\cdot\mathbf{k}}\, d\tau_k \tag{2.26}$$

where $d\tau_k \equiv dk^x\, dk^y\, dk^z$.

Problem 2.6 Show that for (2.24) to be self-consistent (meaning that if we take $\tilde{p}(f)$ from (2.23) and put it into (2.24), we do indeed get $p(t)$ out), we must have:

$$\int_{-\infty}^{\infty} e^{2\pi i f (\bar{t}-t)}\, df = \delta(\bar{t}-t). \tag{2.27}$$

Write out the real and imaginary pieces of this equation. Do they make "sense"?

Problem 2.7 In this problem, we'll find the Fourier transform of the step function by considering the $\alpha \to 0$ limit of:

$$p(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha t} & t > 0. \end{cases} \tag{2.28}$$

a. Write the integral

$$\int_0^{\infty} \cos(2\pi i f t)\, dt \tag{2.29}$$

in terms of the delta function.

b. Find the Fourier transform of $p(t)$ from (2.28) for constant $\alpha > 0$. Isolate the real and imaginary parts of the Fourier transform.

c. Take the $\alpha \to 0$ limit of the Fourier transform from part (b). Use your result from part (a) to help evaluate the real part in the $\alpha \to 0$ limit. This is the Fourier transform of the step function, $\theta(t)$.

Problem 2.8 What is the Fourier transform of $p(t) = p_0 \sin(2\pi \bar{f}\, t)$ (for constant $\bar{f}$)?

Problem 2.9 Find the two-dimensional Fourier transform of $f(x, y) = f_0 \cos(2\pi (\rho\, x + \sigma\, y))$ (with constants ρ and σ).

2.2.2 Green's Function

The point source for developing the Green's function associated with (2.20) is a spatial delta function at $\mathbf{r}'$ times a temporal delta function at t', representing an instantaneous "flash" at $\mathbf{r}'$. We want $G(\mathbf{r}, \mathbf{r}', t, t')$ solving[7]

$$-\frac{1}{c^2}\frac{\partial^2 G}{\partial t^2} + \nabla^2 G = -\delta^3(\mathbf{r} - \mathbf{r}')\,\delta(t - t'). \tag{2.30}$$

As usual, we'll start off with $\mathbf{r}' = 0$ and $t' = 0$, find the spherically symmetric solution to the above, and then replace $\mathbf{r}$ with $\boldsymbol{\mathscr{r}}$ and t with $t - t'$. The approach will be to Fourier transform in time, which will yield an equation that we already know how to solve.

Take

$$-\frac{1}{c^2}\frac{\partial^2 G}{\partial t^2} + \nabla^2 G = -\delta^3(\mathbf{r})\,\delta(t) \tag{2.31}$$

and multiply both sides by $e^{2\pi i f t}$, then integrate in t,

$$-\frac{1}{c^2}\int_{-\infty}^{\infty}\frac{\partial^2 G}{\partial t^2}\, e^{2\pi i f t}\, dt + \nabla^2 \int_{-\infty}^{\infty} G\, e^{2\pi i f t}\, dt = -\delta^3(\mathbf{r}). \tag{2.32}$$

Let $\tilde{G}(\mathbf{r}, f)$ be the Fourier transform of $G(\mathbf{r}, t)$; if we integrate the first term by parts twice, we get back $(2\pi f/c)^2\,\tilde{G} \equiv \mu^2\,\tilde{G}$. The second term is just $\nabla^2\,\tilde{G}$, so (2.32) is

$$\mu^2\,\tilde{G} + \nabla^2\,\tilde{G} = -\delta^3(\mathbf{r}), \tag{2.33}$$

which is precisely (2.14), the only difference being that it is now the temporal Fourier transform of G that is governed by the Helmholtz equation. We know the solution, from (2.18):

$$\tilde{G}(\mathbf{r}, f) = \frac{a\, e^{i\mu r} + b\, e^{-i\mu r}}{4\pi r}, \tag{2.34}$$

where we have a real $\mu \equiv \frac{2\pi f}{c}$, so that the numerator is oscillatory. You'll see in a moment what each term corresponds to, so let's just take one of them for now (set $a = 1$, $b = 0$, and we'll put $\pm$ in the *exponand*). In order to recover $G(\mathbf{r}, t)$, we must Fourier transform back:

[7] In order to obtain natural units for the left and right sides of (2.30), we should use $\delta(c\,(t - t'))$, an inverse length (and we will, later on). For our current purposes, omitting the c is fine as long as we also omit c from the integration element $d\tau'$ $(c\, dt')$. That omission is motivated by the definition of the Fourier transform.

$$
\begin{aligned}
G(\mathbf{r},t) &= \int_{-\infty}^{\infty} \tilde{G}(\mathbf{r},f)\, e^{-2\pi i f t}\, df \\
&= \frac{1}{4\pi r}\int_{-\infty}^{\infty} e^{2\pi i f(-t\pm\frac{r}{c})}\, df \\
&= \frac{1}{4\pi r}\,\delta\Big(t\pm\frac{r}{c}\Big).
\end{aligned}
\tag{2.35}
$$

Now the role of the $\pm$ is clear – if we are at location $\mathbf{r}$ at time t, and there is a source that flashes instantaneously at time $t'=0$ from the origin at $\mathbf{r}'=0$, then the flash "hits" us at $t=\pm\frac{r}{c}$; i.e., we feel the effect of the source only if our time t is the time of flight from the origin (imagine an expanding spherical shell of influence) of a signal traveling at speed c. Interestingly, we recover both the positive and negative solution, so that in theory, we are influenced by the source at time $t=-\frac{r}{c}$, which occurs before $t'=0$, the time at which the source flashed. This is the so-called advanced potential, and we generally get rid of it because of its acausal implication.[8] Then we keep the "retarded" potential, the one that occurs *after* the flash, so that you normally have (moving the source to position $\mathbf{r}'$ at time t' in the usual way):

$$
G(\mathbf{r},\mathbf{r}',t,t') = \frac{1}{4\pi\,\mathscr{r}}\,\delta\bigg((t-t')-\frac{\mathscr{r}}{c}\bigg).
\tag{2.36}
$$

Problem 2.10 Write the Lorenz gauge condition $\frac{1}{c^2}\frac{\partial V(\mathbf{r},t)}{\partial t}+\nabla\cdot\mathbf{A}(\mathbf{r},t)=0$ in terms of the Fourier transforms (in time only): $\tilde{V}(\mathbf{r},f)$ and $\tilde{\mathbf{A}}(\mathbf{r},f)$. You can see that finding the Fourier transform of $\mathbf{A}$ then algebraically determines the Fourier transform of V, so that giving $\mathbf{A}$ is enough to pin down V.

Problem 2.11 Charge conservation has the form: $\frac{\partial\rho(\mathbf{r},t)}{\partial t}+\nabla\cdot\mathbf{J}(\mathbf{r},t)=0$. Given $\mathbf{J}=J_0\,\hat{\mathbf{r}}$ for constant J_0, find the total charge Q contained in a sphere of radius r (centered at the origin) at time t.

2.3 Liénard–Wiechert Potentials

The point potential we used to generate the Green's function is problematic in the context of charge conservation in E&M (a charge cannot flash on and off instantaneously), so the more natural description of a point charge is one that moves along some prescribed trajectory, $\mathbf{w}(t)$, as shown in Figure 2.2 (here $\mathbf{w}(t)$ is a vector that points from the origin to the particle's location at time t). For the particle moving along the path, charge conservation holds.

We want to solve (2.21) for the charge distribution $\rho(\mathbf{r},t)=q\,\delta^3(\mathbf{r}-\mathbf{w}(t))$, with $\mathbf{J}=\rho\,\dot{\mathbf{w}}(t)$. Taking the electric potential first, the integral equation (2.22) becomes

[8] Although see Problem 2.24 – you could keep both advanced and retarded contributions to the potential, but then you need to figure out how to "weight" them since only the sum needs to reduce to the Coulomb case for the static limit (similar to picking a and b with $a+b=1$ in (2.18)).

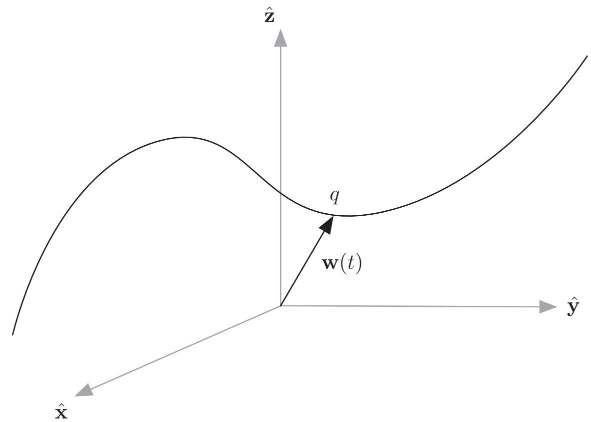

Fig. 2.2 A particle's path is given by the vector $\mathbf{w}(t)$ that points from the origin to the particle's location at time t.

$$V(\mathbf{r}, t) = \frac{q}{4\pi\epsilon_0} \int \int_{-\infty}^{\infty} \frac{\delta^3(\mathbf{r}' - \mathbf{w}(t'))}{|\mathbf{r} - \mathbf{r}'|} \delta\left(t - t' - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dt'\, d\tau', \tag{2.37}$$

and we can use the spatial delta function to replace $\mathbf{r}'$ with $\mathbf{w}(t')$, giving

$$V(\mathbf{r}, t) = \frac{q}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{1}{|\mathbf{r} - \mathbf{w}(t')|} \delta\left(t' - \left(t - \frac{|\mathbf{r} - \mathbf{w}(t')|}{c}\right)\right) dt'. \tag{2.38}$$

To make further progress, we need to evaluate a delta function with a functional argument. Let's work out the general case by evaluating

$$I = \int_{-\infty}^{\infty} g(t')\, \delta(h(t'))\, dt' \tag{2.39}$$

for functions $g(t')$ and $h(t')$. If we view $h(t')$ as defining h as a function of t', then $dh = \frac{dh}{dt'} dt'$, and we can invert the relation to write $t'(h)$. The integral is now

$$I = \int_{h(-\infty)}^{h(\infty)} g(t'(h)) \frac{\delta(h)}{\frac{dh}{dt'}}\, dh. \tag{2.40}$$

Whenever $h = 0$, we have a contribution to the integral. The function $h(t')$ could have a number of zeroes, and we have to count all of them (the ones lying between $h(-\infty)$ and $h(\infty)$). Let $\{t_i'\}_{i=1}^{n}$ be the values of t' for which $h(t_i') = 0$, then the integral I is

$$I = \sum_{i=1}^{n} \frac{g(t_i')}{\left|\frac{dh}{dt'}\right|_{t'=t_i'}} \tag{2.41}$$

where the absolute value in the denominator has crept in for reasons you will explore in Problem 2.12.

To apply this identity in (2.38), with $h(t') = t' - \left(t - \frac{|\mathbf{r} - \mathbf{w}(t')|}{c}\right)$, we need the derivative

$$\frac{dh}{dt'} = 1 - \frac{(\mathbf{r} - \mathbf{w}(t')) \cdot \dot{\mathbf{w}}(t')}{|\mathbf{r} - \mathbf{w}(t')|\, c} = \frac{|\mathbf{r} - \mathbf{w}(t')|\, c - (\mathbf{r} - \mathbf{w}(t')) \cdot \dot{\mathbf{w}}(t')}{|\mathbf{r} - \mathbf{w}(t')|\, c}, \tag{2.42}$$

and inserting this into (2.38), with the delta identity, we have

$$V(\mathbf{r}, t) = \frac{q\,c}{4\,\pi\,\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{w}(t')|\,c - (\mathbf{r} - \mathbf{w}(t')) \cdot \dot{\mathbf{w}}(t')}, \tag{2.43}$$

where we understand that we must evaluate t' according to

$$h(t') = t' - \left(t - \frac{|\mathbf{r} - \mathbf{w}(t')|}{c}\right) = 0; \tag{2.44}$$

that's what the delta function does, albeit implicitly. We sometimes remind ourselves of this requirement by referring to t' as t_r (for retarded time) – then (2.43) and (2.44) read:

$$\begin{aligned} V(\mathbf{r}, t) &= \frac{q\,c}{4\,\pi\,\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{w}(t_r)|\,c - (\mathbf{r} - \mathbf{w}(t_r)) \cdot \dot{\mathbf{w}}(t_r)} \\ c\,(t - t_r) &= |\mathbf{r} - \mathbf{w}(t_r)|. \end{aligned} \tag{2.45}$$

Physically, the potential at $\mathbf{r}$ and time t depends on the source charge's position *and* velocity at the retarded time t_r, defined by the time it takes light to travel from the retarded location to the field point. Note that the potential is not just Coulomb evaluated at a retarded time; that would look like:

$$\bar{V}(\mathbf{r}, t) = \frac{q}{4\,\pi\,\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{w}(t_r)|}. \tag{2.46}$$

It is not surprising that the Coulomb potential lacks the velocity term; it was valid only for stationary sources.

There is a final issue we need to think about. We assumed there was only one root in (2.44), hence only one term that contributed to the delta function. If there were additional roots, we would need to include them (as in (2.41)). Suppose you have two points in causal contact with the field point, so that t_1 and t_2 both satisfy (2.44), with $t_1 < t_2$. Then

$$c\,(t_2 - t_1) = |\mathbf{r} - \mathbf{w}(t_1)| - |\mathbf{r} - \mathbf{w}(t_2)|. \tag{2.47}$$

Take $t_1 = t_2 - \Delta t$ for small Δt; we can expand

$$\begin{aligned} c\,\Delta t &= |\mathbf{r} - \mathbf{w}(t_2 - \Delta t)| - |\mathbf{r} - \mathbf{w}(t_2)| \approx |\mathbf{r} - \mathbf{w}(t_2) + \dot{\mathbf{w}}(t_2)\,\Delta t| - |\mathbf{r} - \mathbf{w}(t_2)| \\ &\le |\mathbf{r} - \mathbf{w}(t_2)| + |\dot{\mathbf{w}}(t_2)|\,\Delta t - |\mathbf{r} - \mathbf{w}(t_2)| \\ &= \dot{w}(t_2)\,\Delta t, \end{aligned} \tag{2.48}$$

but then $c \le \dot{w}(t_2)$, meaning that the particle would have to travel at a speed of at least c, which is impossible.

Problem 2.12

a. Evaluate the integral $I_1 = \int_{-\infty}^{0} \delta(x^2 - 1)\,dx$ using explicit change of variables.

b. Do the same for $I_2 = \int_{0}^{\infty} \delta(x^2 - 1)\,dx$.

c. Put them together to find $I = \int_{-\infty}^{\infty} \delta(x^2 - 1)\,dx$. Finally, show that for $f(x) = x^2 - 1$ (as we have here), the correct expression (once boundaries are included) is $I = \sum_{i=1}^{2} \frac{1}{|f'(x_i)|}$ where $x_1 = 1$, $x_2 = -1$, or in general, the roots of $f(x)$.

Problem 2.13 What happened to the absolute values that appear in (2.41) when we used that identity in (2.43)?

Problem 2.14 We'll need the derivatives of V and $\mathbf{A}$ to calculate $\mathbf{E}$ and $\mathbf{B}$, and en route, we need to be able to express the derivatives of t_r. Evaluate the derivatives: $\frac{dt_r}{dt}$ and ∇t_r given $c\,(t - t_r) = |\mathbf{r} - \mathbf{w}(t_r)|$ for arbitrary $\mathbf{w}(t)$.

Another View: Geometric Charge Evaluation

There is another way to obtain the potential associated with a moving charge q with known trajectory $\mathbf{w}(t)$. The idea is to take the usual integral solution, where the charge density is evaluated at the retarded time,

$$V = \frac{1}{4\pi\,\epsilon_0} \int \frac{\rho(\mathbf{r}', t_r)}{|\mathbf{r} - \mathbf{r}'|}\, d\tau', \tag{2.49}$$

and note that $\mathbf{r}'$ is just $\mathbf{w}(t_r)$, so that the denominator comes outside the integral,

$$V = \frac{1}{4\pi\,\epsilon_0\,|\mathbf{r} - \mathbf{w}(t_r)|} \int \rho(\mathbf{r}', t_r)\, d\tau'. \tag{2.50}$$

The remaining integral would appear to be q, since we are integrating over all space. But we have to evaluate the different "pieces" of the point charge at different times (t_r is still lurking inside ρ). It's a hard sell for point charges and is easier to see in an extended setting. The result is that

$$\int \rho(\mathbf{r}', t_r)\, d\tau' = \frac{q}{1 - \frac{(\mathbf{r}-\mathbf{w})\cdot\dot{\mathbf{w}}}{c\,|\mathbf{r}-\mathbf{w}|}}, \tag{2.51}$$

all evaluated at the retarded time. Inserting this into (2.50) gives back

$$V(\mathbf{r}, t) = \frac{q\,c}{4\pi\,\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{w}(t')|\,c - (\mathbf{r} - \mathbf{w}(t'))\cdot\dot{\mathbf{w}}(t')}. \tag{2.52}$$

We recover the correct expression, so the association (2.51) must be correct (and it can be demonstrated geometrically; see [19]). Still, a simple example clarifies its role.

Example

Here we will see how evaluating the charge density at t_r, the retarded time, rather than t, changes what we would call the total charge. We'll do this by setting up a known charge density $\rho(\mathbf{r}, t)$ and evaluate both $\int \rho(\mathbf{r}', t)\, d\tau'$ and $\int \rho(\mathbf{r}', t_r)\, d\tau$ to see how they compare.

First we'll generate the charge distribution ρ at $t = 0$. Take a charge q and spread it uniformly in a sphere of radius R with a hole of radius ϵ cut out of the middle, so that there is a constant charge density

$$\rho_0 = \frac{q}{\frac{4}{3}\pi\left(R^3 - \epsilon^3\right)}. \tag{2.53}$$

The initial distribution, $\rho(\mathbf{r}, 0)$, can be written in terms of ρ_0 as

$$\rho(\mathbf{r}, 0) = \rho_0\,\left[\theta(R - r) + \theta(r - \epsilon) - 1\right], \tag{2.54}$$

where $\theta(x)$ is the step function:

$$\theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases} \tag{2.55}$$

To generate the temporal evolution of $\rho(\mathbf{r}, t)$, we'll start by assuming a form for the motion of the charge and ensure that that motion obeys charge conservation. At $t = 0$, turn on the constant velocity field $\mathbf{v} = v\,\hat{\mathbf{r}}$. We need to find $\rho(\mathbf{r}, t)$ with current density $\mathbf{J} = \rho\,\mathbf{v}$ that reduces to (2.54) at $t = 0$. From charge conservation, $\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}$, we get the PDE

$$\frac{\partial \rho}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\,\rho\,v\right) = 0. \tag{2.56}$$

Multiply by r^2 and define $u \equiv r^2\,\rho$ to get back a familiar (half) wave equation

$$\frac{\partial u}{\partial t} + v\,\frac{\partial u}{\partial r} = 0. \tag{2.57}$$

The solution to this equation is $u(r, t) = u(r - v\,t, 0)$, and because $u = r^2\,\rho$, this gives

$$\rho(r, t) = \frac{(r - v\,t)^2}{r^2}\,\rho_0\,[\theta(R - (r - v\,t)) + \theta((r - v\,t) - \epsilon) - 1]\,. \tag{2.58}$$

Now we'll integrate this density over all space using the time t, a snapshot of the distribution taken at a single time:

$$\begin{aligned} I &= \int \rho(r', t)\,d\tau' = \int_0^\infty 4\,\pi\,\rho(r', t)\,r'^2\,dr' \\ &= 4\,\pi \int_0^\infty (r' - v\,t)^2\,\rho_0\,\big[\theta(R - (r' - v\,t)) + \theta((r' - v\,t) - \epsilon) - 1\big]\,dr' \\ &= 4\,\pi \int_{\epsilon + v\,t}^{R + v\,t} \rho_0\,\left(r' - v\,t\right)^2\,dr'. \end{aligned} \tag{2.59}$$

Define $y \equiv r' - v\,t$ and you get

$$I = 4\,\pi \int_\epsilon^R \rho_0\,y^2\,dy = q \tag{2.60}$$

so that the integral over all space gives back the total charge. But that integral was performed at time t using $\rho(\mathbf{r}', t)$, an instantaneous evaluation of all the pieces of the charge distribution, not $\rho(r', t_r)$, which is what we want.

This time, we'll perform the correct integration of $\rho(\mathbf{r}', t_r)$. Take the observation point to be the origin, $\mathbf{r} = 0$, then $t_r = t - r'/c$, and

$$\begin{aligned} I &= \int \rho(r', t_r)\,d\tau' = \int_0^\infty 4\,\pi\,\rho(r', t_r)\,r'^2\,dr' \\ &= 4\,\pi \int_0^\infty (r' - v\,t_r)^2\,\rho_0\,\big[\theta(R - (r' - v\,t_r)) + \theta((r' - v\,t_r) - \epsilon) - 1\big]\,dr' \\ &= 4\,\pi \int_0^\infty (r' - v\,(t - r'/c))^2\,\rho_0\,\big[\theta(R - (r' - v\,(t - r'/c))) \\ &\quad + \theta((r' - v\,(t - r'/c)) - \epsilon) - 1\big]\,dr'. \end{aligned} \tag{2.61}$$

Here you can see the confounding effect of the retarded time: it adds additional r'-dependence to the integral. Now instead of integrating over the natural extent of the distribution at time t (from $\epsilon + v\,t$ to $R + v\,t$), we must start the integration at $(\epsilon + v\,t)\,(1 + v/c)^{-1}$, and we end at $(R + v\,t)\,(1 + v/c)^{-1}$ (from the step functions). The final expression is

$$
\begin{aligned}
I &= 4\,\pi\,\rho_0 \int_{\frac{\epsilon+v\,t}{1+v/c}}^{\frac{R+v\,t}{1+v/c}} \left(r' - v\,(t - r'/c)\right)^2\,dr' \\
&= \frac{q}{1+\frac{v}{c}}.
\end{aligned}
\tag{2.62}
$$

Our observation point is at the origin, so $\boldsymbol{\mathscr{r}} = 0 - \mathbf{r}$, since $\mathbf{w} = \mathbf{r}$ here. Then (2.62) is precisely $q/(1 - \hat{\boldsymbol{\mathscr{r}}}(t_r) \cdot \mathbf{v}/c)$ as in (2.51).

Problem 2.15 Why didn't we set $\epsilon = 0$ in this example? That would simplify the integrals and algebra significantly.

2.3.1 Magnetic Vector Potential

The magnetic vector potential, with source $\mathbf{J} = \rho\,\dot{\mathbf{w}}$, follows a similar pattern, just three copies of the electric potential case, and we can write down the result:

$$
\begin{aligned}
\mathbf{A}(\mathbf{r}, t) &= \frac{q\,c\,\mu_0}{4\,\pi} \frac{\dot{\mathbf{w}}(t_r)}{|\mathbf{r} - \mathbf{w}(t_r)|\,c - (\mathbf{r} - \mathbf{w}(t_r)) \cdot \dot{\mathbf{w}}(t_r)}, \text{ with} \\
c\,(t - t_r) &= |\mathbf{r} - \mathbf{w}(t_r)|.
\end{aligned}
\tag{2.63}
$$

Using $\mu_0\,\epsilon_0 = 1/c^2$, the magnetic vector potential can be expressed in terms of the electric one:

$$
\mathbf{A}(\mathbf{r}, t) = \frac{\dot{\mathbf{w}}(t_r)}{c^2}\,V(\mathbf{r}, t).
\tag{2.64}
$$

The full set of "Liénard–Wiechert" potentials, for a point charge moving along a given trajectory $\mathbf{w}(t)$, is

$$
\begin{aligned}
V(\mathbf{r}, t) &= \frac{q\,c}{4\,\pi\,\epsilon_0} \frac{1}{|\mathbf{r} - \mathbf{w}(t_r)|\,c - (\mathbf{r} - \mathbf{w}(t_r)) \cdot \dot{\mathbf{w}}(t_r)} \\
\mathbf{A}(\mathbf{r}, t) &= \frac{\dot{\mathbf{w}}(t_r)}{c^2}\,V(\mathbf{r}, t) \\
c\,(t - t_r) &= |\mathbf{r} - \mathbf{w}(t_r)|.
\end{aligned}
\tag{2.65}
$$

These look a little nicer if we define $\boldsymbol{\mathscr{r}}(t_r) \equiv \mathbf{r} - \mathbf{w}(t_r)$ with associated unit vector $\hat{\boldsymbol{\mathscr{r}}}(t_r)$, then

$$
\begin{aligned}
V(\mathbf{r}, t) &= \frac{q\,c}{4\,\pi\,\epsilon_0} \frac{1}{\mathscr{r}(t_r)\,c - \boldsymbol{\mathscr{r}}(t_r) \cdot \dot{\mathbf{w}}(t_r)} \\
\mathbf{A}(\mathbf{r}, t) &= \frac{\dot{\mathbf{w}}(t_r)}{c^2}\,V(\mathbf{r}, t) \\
c\,(t - t_r) &= |\mathbf{r} - \mathbf{w}(t_r)| = \mathscr{r}(t_r).
\end{aligned}
\tag{2.66}
$$

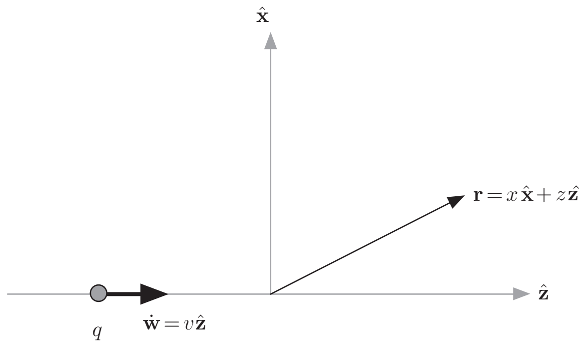

Fig. 2.3 A charge q travels along the $\hat{\mathbf{z}}$-axis with constant velocity. The field point of interest is in the x–z plane.

2.3.2 Example: Constant Velocity

The simplest example to which we can apply (2.66) is particle motion with constant velocity. Suppose a particle of charge q moves at constant speed down the $\hat{\mathbf{z}}$-axis, with $\mathbf{w}(t) = v\,t\,\hat{\mathbf{z}}$ as shown in Figure 2.3. We want to know the potentials at the point $\mathbf{r} = x\,\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}$ (working in the two-dimensional plane, without loss of generality) at time t. First, we need to solve the retarded time condition to find t_r, then we can evaluate the potential. The retarded time is given by solving

$$c^2\,(t - t_r)^2 = (z - v\,t_r)^2 + x^2, \tag{2.67}$$

a quadratic,[9] with

$$t_r = \frac{c^2\,t - v\,z \pm \sqrt{-v^2\,x^2 + c^2\left((z - v\,t)^2 + x^2\right)}}{c^2 - v^2}. \tag{2.68}$$

How should we pick the root here? Well, suppose $t = 0$ and we are at $z = 0$, then $t_r = \pm x/\sqrt{(c^2 - v^2)}$, and we'd take the negative root to enforce causality. It's pretty clear generally that the negative sign will give us a t_r that's less than taking the positive sign. Choosing the negative sign, let

$$t_r = \frac{c^2\,t - v\,z - \sqrt{-v^2\,x^2 + c^2\left((z - v\,t)^2 + x^2\right)}}{c^2 - v^2}. \tag{2.69}$$

Now we have to evaluate the denominator of the potential in (2.66),

$$c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r) = c^2\,(t - t_r) - (z - v\,t_r)\,v = (v^2 - c^2)\,t_r + c^2\,t - v\,z, \tag{2.70}$$

where we used the formal replacement $\mathscr{r}(t_r) = c\,(t - t_r)$ from its definition to simplify the right-hand side. Inserting t_r gives

$$c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r) = \sqrt{c^2\left((z - v\,t)^2 + x^2\right) - v^2\,x^2}. \tag{2.71}$$

[9] Note that by squaring the retarded time condition, we have *re*introduced the advanced solution, which we will have to then omit later on.

The potential is

$$V(x,z,t) = \frac{q}{4\pi\epsilon_0}\frac{1}{\sqrt{\left((z-v\,t)^2+x^2\right)-v^2\,x^2/c^2}}, \tag{2.72}$$

and we have $(z-v\,t)^2+x^2 = |\mathbf{r}-\mathbf{w}(t)|^2$ here. The difference vector, $\mathbf{R}\equiv\mathbf{r}-\mathbf{v}\,t$, is the vector that points from the *current* (at time t) location of the particle to the field point. We can write

$$V(x,z,t) = \frac{q}{4\pi\epsilon_0}\frac{1}{R\sqrt{1-\frac{v^2}{c^2}\frac{x^2}{R^2}}}, \tag{2.73}$$

and one can further simplify by noting that $x/R = \sin\theta$ where θ is the angle between $\mathbf{R}$ and $\mathbf{v}$. Obtaining the magnetic vector potential from (2.66) gives

$$\mathbf{A} = \frac{v\,\hat{\mathbf{z}}}{c^2}\,V(x,z,t). \tag{2.74}$$

We can use this result to establish an interesting property of the set $\{V,\mathbf{A}\}$. Notice that there are four quantities here, and from a numerological point of view, we are tempted to combine these into a four-vector like the coordinate one. That's silly: just because there are four of something doesn't mean that they form the components of a four-vector. What's the difference between four quantities and the components of a four-vector, anyway? Well, if V/c (sneaking in the factor of $1/c$ to make all the dimensions the same) and $\mathbf{A}$ are the 0 and 1, 2, and 3 components of a four-vector, then they transform like the position–time one, so we would expect, for a Lorentz boost in the z-direction:

$$\begin{aligned}\frac{\bar{V}}{c} &= \gamma\left(\frac{V}{c}-A^z\,\beta\right)\\ \bar{A}^z &= \gamma\left(-\frac{V}{c}\,\beta+A^z\right)\end{aligned} \tag{2.75}$$

(think of V/c as $c\,t$ and A^z as z from a Lorentz boost of position and time for relative motion in the z-direction). We will now show, from (2.73) and (2.74), that $A^\mu \doteq (V/c, A^x, A^y, A^z)$ does indeed transform like this.

In the rest frame of the charge q, we know that

$$\bar{V} = \frac{q}{4\pi\epsilon_0\sqrt{\bar{z}^2+\bar{x}^2}}, \tag{2.76}$$

just the Coulomb potential, and $\bar{A}^z = 0$ (these come from (2.73) and (2.74) written in the barred coordinates of the charge's rest frame). From the assumed transformation, then, we have

$$\bar{A}^z = 0 = \gamma\left(-\frac{V}{c}\,\beta+A^z\right) \tag{2.77}$$

from which we learn that $A^z = \frac{v\,V}{c^2}$, precisely the content of (2.74). The other piece of the transformation gives

$$\frac{\bar{V}}{c} = \gamma\left(\frac{V}{c}-A^z\,\beta\right) = \gamma\,\frac{V}{c}\left(1-\beta^2\right) = \frac{1}{\gamma}\frac{V}{c} \tag{2.78}$$

so that $V = \gamma \bar{V}$. Let's write V in terms of the unbarred coordinates using $\bar{z} = \gamma (z - v t)$; again from the Lorentz transformation, we have

$$V = \gamma \bar{V} = \frac{q}{4 \pi \epsilon_0 \sqrt{\gamma^2 (z - v t)^2 + x^2} \sqrt{1 - \frac{v^2}{c^2}}}, \tag{2.79}$$

and finally noting that the denominator is really the square root of

$$\begin{aligned} \left(\gamma^2 (z - v t)^2 + x^2\right)\left(1 - \frac{v^2}{c^2}\right) &= (z - v t)^2 + x^2 - \frac{v^2}{c^2} x^2 \\ &= R^2 - \frac{v^2}{c^2} x^2 \\ &= R^2 \left(1 - \frac{v^2}{c^2} \frac{x^2}{R^2}\right) \end{aligned} \tag{2.80}$$

up to constants. We can write the potential as

$$V = \frac{q}{4 \pi \epsilon_0} \frac{1}{R \sqrt{1 - \frac{v^2}{c^2} \frac{x^2}{R^2}}}, \tag{2.81}$$

in agreement with (2.73).

Conclusion: the combination

$$A^\mu \doteq \begin{pmatrix} \frac{V}{c} \\ A^x \\ A^y \\ A^z \end{pmatrix} \tag{2.82}$$

is a four-vector, by virtue of its response to Lorentz transformation.

Problem 2.16 Find $A^\mu A_\mu$ for the four-potential of a particle moving with $\mathbf{w}(t) = v t \hat{\mathbf{z}}$ for constant v. Does it match the value of $\bar{A}^\mu \bar{A}_\mu$ in the particle's rest frame?

Problem 2.17 An infinite sheet of charge lying in the x–y plane has constant charge-per-unit-area $\bar{\sigma}$ when at rest. Find the electric and magnetic vector potentials everywhere (use Lorenz gauge). Now the sheet moves with constant velocity $\mathbf{v} = v \hat{\mathbf{x}}$. Find the electric and magnetic vector potentials everywhere (do not use the transformation of the four-vector potential to do this), and write them in terms of the potentials at rest. From your solutions, verify that the potentials transform appropriately as elements of the potential four-vector: A^μ with $A^0 = V/c$ and spatial components given by $\mathbf{A}$.

Problem 2.18 Generalize the potential fields (2.72) and (2.74) of a point charge moving with constant velocity along the $\hat{\mathbf{z}}$-axis to field points in three dimensions (so that $\mathbf{r} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}}$), and write the potentials in cylindrical coordinates. Find the electric and magnetic fields from the potentials (stay in cylindrical coordinates). Finally, check the $\mathbf{v} = 0$ case. Do you recover the correct $\mathbf{E}$ and $\mathbf{B}$?

Problem 2.19 The "Lorenz gauge" condition relating V and $\mathbf{A}$ reads: $-\frac{1}{c^2} \frac{\partial V}{\partial t} + \nabla \cdot \mathbf{A} = 0$. Show that if you choose this gauge in one frame, you will remain in this gauge for all inertial frames related via a Lorentz boost. Hint: the easiest way to establish this is to write the condition in indexed notation and use your result from Problem 1.37.

Problem 2.20 For a charged particle that moves according to $\mathbf{w}(t) = d\,(\cos(\omega\,t)\,\hat{\mathbf{x}} + \sin(\omega\,t)\,\hat{\mathbf{y}})$ (with constant ω, d), find the full Liénard–Wiechert potentials (V and $\mathbf{A}$) for field points along the z-axis.

Another View: Destiny

There is another way to obtain the Liénard–Wiechert potentials,[10] one that relies on our belief in special relativity and the observation that E&M appears to support special relativity (it was, of course, one of the motivating pieces of evidence for special relativity). The idea is that we want to build V and $\mathbf{A}$ for a point charge out of Lorentz scalars and four-vectors, and we will require that our expression reduce to $\sim q/r$ for a source at rest. In addition to this manifestly Lorentz covariant form and correct static limit, we want some natural way to express the notion that the field information travels at constant speed c, so we'll be on the lookout for building blocks that make it easy to enforce the retarded time condition.

Start with the static case: for a charge at rest at $\mathbf{r}'$,

$$V = \frac{q}{4\,\pi\,\epsilon_0\,|\mathbf{r} - \mathbf{r}'|}, \tag{2.83}$$

so the main dependence is on the spatial separation between the source at $\mathbf{r}'$ and the observer at $\mathbf{r}$. In special relativity, the separation is a mix of temporal and spatial – take the separation vector R^μ to be the four-dimensional generalization of $\mathbf{r} - \mathbf{r}'$. For a source at space-time point x_s^μ (with $x_s^0 = c\,t'$, $\mathbf{x}_s = \mathbf{w}(t')$ for source trajectory $\mathbf{w}$) and an observer at space-time point x_o^μ (with $x_o^0 = c\,t$, $\mathbf{x}_o = \mathbf{r}$),

$$R^\mu \equiv x_o^\mu - x_s^\mu \doteq \begin{pmatrix} c\,(t - t') \\ \mathbf{r} - \mathbf{w}(t') \end{pmatrix}. \tag{2.84}$$

We can enforce the retarded time condition by requiring that $R^\mu\,R_\mu = 0$, i.e., that the separation is "light"-like. This gives us precisely the retarded time equation

$$R^\mu\,R_\mu = 0 = -c^2\,(t - t')^2 + (\mathbf{r} - \mathbf{w}(t')) \cdot (\mathbf{r} - \mathbf{w}(t')), \tag{2.85}$$

although the advanced time is also available here (because of the square).

We've hit a snag, though – the easiest replacement for $|\mathbf{r} - \mathbf{r}'|$ in the Coulomb potential would be $\sqrt{R^\mu\,R_\mu}$, appearing as:

$$V = \frac{q}{4\,\pi\,\epsilon_0\,\sqrt{R^\mu\,R_\mu}}, \tag{2.86}$$

but this replacement is now unavailable since $R^\mu\,R_\mu = 0$. Is there another four-vector lying around for us to use? Sure: the four-velocity[11] associated with x_s^μ, that's

[10] This is done in [21] with a nice discussion of the process, motivating the one here, from [26].

[11] It is clear that the object $\frac{dx_s^\mu}{d\tau}$ is a four-vector, since dx_s^μ is a four-vector and $d\tau$ is a scalar.

$$\dot{x}_s^\mu \equiv \frac{dx_s^\mu}{d\tau} \doteq \begin{pmatrix} c \\ \dot{\mathbf{w}}(t') \end{pmatrix} \frac{dt'}{d\tau}. \tag{2.87}$$

Now the most natural non-zero scalar available is

$$\dot{x}_s^\mu R_\mu = \left(-c^2 (t-t') + \dot{\mathbf{w}}(t') \cdot \left(\mathbf{r} - \mathbf{w}(t')\right)\right) \frac{dt'}{d\tau}, \tag{2.88}$$

which has the unknown factor $\frac{dt'}{d\tau}$ attached to it.

We know that V/c and $\mathbf{A}$ are the components of a four-vector, and the only four-vectors we have are R^μ and $\dot{x}_s^\mu$, so we expect that A^μ is proportional to either R^μ or $\dot{x}_s^\mu$. If we choose the latter, and stick a factor of $\dot{x}_s^\mu R_\mu$ in the denominator, then the factors of $\frac{dt'}{d\tau}$ will cancel. Let's take

$$A^\mu = K \frac{\dot{x}_s^\mu}{\dot{x}_s^\alpha R_\alpha} \doteq \frac{K}{\left(-c^2 (t-t') + \dot{\mathbf{w}}(t') \cdot (\mathbf{r} - \mathbf{w}(t'))\right)} \begin{pmatrix} c \\ \dot{\mathbf{w}}(t') \end{pmatrix} \tag{2.89}$$

for constant K that we'll set by comparison with the static limit. Let $\mathbf{w}(t) = \mathbf{r}'$, a constant source location with $\dot{\mathbf{w}}(t) = 0$. From the retarded time condition (2.85), we know that $c\,(t-t') = |\mathbf{r} - \mathbf{r}'|$ and can eliminate time entirely in our static expression for A^μ:

$$A^\mu \doteq \frac{K}{-c\,|\mathbf{r} - \mathbf{r}'|} \begin{pmatrix} c \\ 0 \end{pmatrix}. \tag{2.90}$$

If we want the zero-component here to match the Coulomb, $A^0 = V/c$, we must take $K = -q/(4\pi\, c\, \epsilon_0)$, and we can use this constant in the full case above to recover the Liénard–Wiechert form:

$$\begin{aligned} A^\mu &= -\frac{q}{4\pi\,\epsilon_0\, c} \frac{\dot{x}_s^\mu}{\dot{x}_s^\alpha R_\alpha} \doteq \frac{q}{4\pi\,\epsilon_0\,(c\,|\mathbf{r} - \mathbf{w}(t')| - \dot{\mathbf{w}}(t') \cdot (\mathbf{r} - \mathbf{w}(t')))} \begin{pmatrix} 1 \\ \frac{\dot{\mathbf{w}}(t')}{c} \end{pmatrix} \\ R^\mu R_\mu &= 0 = -c^2 (t-t')^2 + (\mathbf{r} - \mathbf{w}(t')) \cdot (\mathbf{r} - \mathbf{w}(t')), \end{aligned} \tag{2.91}$$

where the second equation serves to define the retarded time, as usual.

2.4 Particle Electric and Magnetic Fields

From the potentials, we are ready to compute the electric and magnetic fields using $\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$ and $\mathbf{B} = \nabla \times \mathbf{A}$. The tricky part lies in evaluating the derivatives of t_r, since it depends on both time and space. From the defining property of the retarded time, we can take the relevant derivatives. Take the t-derivative of the third equation in (2.66):

$$c\left(1 - \frac{dt_r}{dt}\right) = -\frac{(\mathbf{r} - \mathbf{w}(t_r)) \cdot \dot{\mathbf{w}}(t_r)}{|\mathbf{r} - \mathbf{w}(t_r)|} \frac{dt_r}{dt} \tag{2.92}$$

and we can solve for $\frac{dt_r}{dt}$

$$\frac{dt_r}{dt} = \left(1 - \frac{(\mathbf{r} - \mathbf{w}(t_r)) \cdot \dot{\mathbf{w}}(t_r)}{c\,|\mathbf{r} - \mathbf{w}(t_r)|}\right)^{-1} = \left(1 - \frac{1}{c}\, \hat{\boldsymbol{\imath}}(t_r) \cdot \dot{\mathbf{w}}(t_r)\right)^{-1}. \tag{2.93}$$

Similarly, we need the spatial derivatives of t_r, specifically the gradient. Working again from (2.66), this time taking the gradient of both sides:

$$
-\,c\,\nabla t_r = \frac{1}{|\mathbf{r}-\mathbf{w}(t_r)|}\left(\mathbf{r}-\mathbf{w}(t_r)-(\mathbf{r}-\mathbf{w}(t_r))\cdot\dot{\mathbf{w}}\,\nabla t_r\right), \tag{2.94}
$$

or using $\boldsymbol{\mathscr{r}}(t_r)$ notation:

$$
-\,c\,\nabla t_r = \hat{\boldsymbol{\mathscr{r}}}(t_r) - \hat{\boldsymbol{\mathscr{r}}}(t_r)\cdot\dot{\mathbf{w}}\,\nabla t_r, \tag{2.95}
$$

and we can isolate the gradient:

$$
\nabla t_r = -\frac{\hat{\boldsymbol{\mathscr{r}}}(t_r)}{c-\hat{\boldsymbol{\mathscr{r}}}(t_r)\cdot\dot{\mathbf{w}}}. \tag{2.96}
$$

Assembling the individual pieces is not too bad. We'll need $\nabla\mathscr{r}$, which, from (2.66), is just $-c\,\nabla t_r$. In addition, we'll want $\nabla(\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))$, and this can be evaluated using indexed notation:

$$
\begin{aligned}
\partial_i\left(\mathscr{r}_j\,\dot{w}^j\right) &= \left(\delta_{ij}-\dot{w}_j\,\partial_i\,t_r\right)\,\dot{w}^j + \left(r_j-w_j\right)\,\ddot{w}^j\,\partial_i\,t_r \\
&= \dot{w}_i + \partial_i t_r\left(-\dot{w}_j\,\dot{w}^j + (r_j-w_j)\,\ddot{w}^j\right),
\end{aligned} \tag{2.97}
$$

or, back in vector notation (for these spatial components, contravariant and covariant forms are numerically identical):

$$
\nabla(\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) = \dot{\mathbf{w}}(t_r) + (\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)-\dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r))\;\nabla t_r. \tag{2.98}
$$

Working our way along, we want $\mathbf{E}=-\nabla V-\frac{\partial\mathbf{A}}{\partial t}$. Start with the gradient of V:

$$
\begin{aligned}
-\nabla V &= \frac{q\,c}{4\,\pi\,\epsilon_0}\,\frac{1}{(\mathscr{r}(t_r)\,c-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^2}\left(\nabla\mathscr{r}(t_r)\,c-\nabla(\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))\right) \\
&= \frac{q\,c}{4\,\pi\,\epsilon_0}\,\frac{1}{(\mathscr{r}(t_r)\,c-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^2}\Big(-c^2\nabla t_r-\dot{\mathbf{w}}(t_r) \\
&\quad -(\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)-\dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r))\;\nabla t_r\Big) \\
&= \frac{q\,c}{4\,\pi\,\epsilon_0}\,\frac{1}{(\mathscr{r}(t_r)\,c-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^2}\Big(-\dot{\mathbf{w}}(t_r) \\
&\quad +\Big(\dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r)-c^2-\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\Big)\;\nabla t_r\Big).
\end{aligned} \tag{2.99}
$$

Finally, we can use the expression for ∇t_r from (2.96) to write

$$
-\nabla V = \frac{q\,c}{4\pi\epsilon_0}\,\frac{\left(-\dot{\mathbf{w}}(t_r)\,(\mathscr{r}(t_r)\,c-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))-\left(\dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r)-c^2-\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\right)\;\boldsymbol{\mathscr{r}}(t_r)\right)}{(\mathscr{r}(t_r)\,c-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^3}. \tag{2.100}
$$

To get the time derivative of the magnetic vector potential, we just take:

$$
\begin{aligned}
\frac{\partial \mathbf{A}}{\partial t} &= \frac{\ddot{\mathbf{w}}(t_r)}{c^2} V \frac{dt_r}{dt} - \frac{\dot{\mathbf{w}}(t_r)}{c^2} \frac{q\, c}{4\,\pi\,\epsilon_0} \\
&\quad \times \frac{1}{(\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^2} \frac{dt_r}{dt} \frac{d}{dt_r} (\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) \\
&= \left[\frac{\ddot{\mathbf{w}}(t_r)}{c^2} V - \frac{\dot{\mathbf{w}}(t_r)}{c^2} \frac{q\, c}{4\,\pi\,\epsilon_0} \frac{\left(-\frac{\dot{\mathbf{w}}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)\, c}{\mathscr{r}(t_r)} + \dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\right)}{(\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^2} \right] \frac{dt_r}{dt}.
\end{aligned}
\tag{2.101}
$$

Using $\frac{dt_r}{dt} = c\,\mathscr{r}\,/(c\,\mathscr{r}\; - \boldsymbol{\mathscr{r}}\; \cdot \dot{\mathbf{w}})$ from (2.93) (all evaluated at t_r, of course), we have

$$
\begin{aligned}
\frac{\partial \mathbf{A}}{\partial t} &= \frac{q\, c}{4\,\pi\,\epsilon_0} \frac{1}{(\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^3} \left[\frac{\mathscr{r}(t_r)\,\ddot{\mathbf{w}}(t_r)}{c} (c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) \right. \\
&\quad \left. + \dot{\mathbf{w}}(t_r) \left(\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r) - \frac{\dot{w}^2(t_r)\,\mathscr{r}(t_r)}{c} + \frac{\mathscr{r}(t_r)\,\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)}{c} \right) \right] \\
&= \frac{q\, c}{4\,\pi\,\epsilon_0} \frac{1}{(\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^3} \left[\frac{\mathscr{r}(t_r)\,\ddot{\mathbf{w}}(t_r)}{c} (c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) \right. \\
&\quad + \dot{\mathbf{w}}(t_r) \Bigg(-(c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) + c\,\mathscr{r}(t_r) \\
&\quad \left. - \frac{\dot{w}^2(t_r)\,\mathscr{r}(t_r)}{c} + \frac{\mathscr{r}(t_r)\,\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)}{c} \Bigg) \right] \\
&= \frac{q\, c}{4\,\pi\,\epsilon_0} \frac{1}{(\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r))^3} \left[\left(\frac{\mathscr{r}(t_r)\,\ddot{\mathbf{w}}(t_r)}{c} - \dot{\mathbf{w}}(t_r) \right) \right. \\
&\quad \left. \times (c\,\mathscr{r}(t_r) - \boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)) + \dot{\mathbf{w}}(t_r) \frac{\mathscr{r}(t_r)}{c} \left(c^2 - \dot{w}^2(t_r) + \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r) \right) \right].
\end{aligned}
\tag{2.102}
$$

Let $\mathbf{u}(t_r) \equiv c\;\hat{\boldsymbol{\mathscr{r}}}(t_r) - \dot{\mathbf{w}}(t_r)$, so that we can write the derivatives as

$$
\begin{aligned}
-\nabla V &= \frac{q\, c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \Big(-\dot{\mathbf{w}}(t_r)\; (\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) + \big(c^2 - \dot{w}^2(t_r) \\
&\quad + \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r) \big)\; \boldsymbol{\mathscr{r}}(t_r) \Big) \\
-\frac{\partial \mathbf{A}}{\partial t} &= \frac{q\, c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[\left(\dot{\mathbf{w}}(t_r) - \frac{\mathscr{r}(t_r)\,\ddot{\mathbf{w}}(t_r)}{c} \right) (\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) \right. \\
&\quad \left. - \dot{\mathbf{w}}(t_r) \frac{\mathscr{r}(t_r)}{c} \left(c^2 - \dot{w}^2(t_r) + \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r) \right) \right].
\end{aligned}
\tag{2.103}
$$

Adding these two pieces together gives the electric field:

$$
\begin{aligned}
\mathbf{E} &= \frac{q\, c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[-\frac{\mathscr{r}(t_r)\,\ddot{\mathbf{w}}(t_r)}{c} (\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) \right. \\
&\quad \left. + \left(-\dot{\mathbf{w}}(t_r) \frac{\mathscr{r}(t_r)}{c} + \boldsymbol{\mathscr{r}}(t_r) \right) \left(c^2 - \dot{w}^2(t_r) + \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r) \right) \right] \\
&= \frac{q\,\mathscr{r}(t_r)}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[-\ddot{\mathbf{w}}(t_r)\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) \right. \\
&\quad \left. + \mathbf{u}(t_r) \left(c^2 - \dot{w}^2(t_r) + \boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r) \right) \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{q\,\mathscr{r}(t_r)}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\Big[(c^2-\dot{w}^2(t_r))\,\mathbf{u}(t_r) \\
&\quad + \mathbf{u}(t_r)\,(\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)) - \ddot{\mathbf{w}}(t_r)\,(\boldsymbol{\mathscr{r}}(t_r)\cdot\mathbf{u}(t_r))\Big] \\
&= \frac{q\,\mathscr{r}(t_r)}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\Big[(c^2-\dot{w}^2(t_r))\,\mathbf{u}(t_r)+\boldsymbol{\mathscr{r}}(t_r)\times(\mathbf{u}(t_r)\times\ddot{\mathbf{w}}(t_r))\Big],
\end{aligned}
\tag{2.104}
$$

where the final line comes from the "BAC-CAB" identity for cross products.

From here, the magnetic field, $\mathbf{B}=\nabla\times\mathbf{A}$, is more of the same. In component form:

$$
\begin{aligned}
B^i &= \epsilon^{ijk}\partial_j A_k = \epsilon^{ijk}\,\partial_j\left(\frac{\dot{w}_k(t_r)}{c^2}\,V\right) \\
&= \epsilon^{ijk}\left(\frac{\ddot{w}_k(t_r)}{c^2}\,\frac{\partial t_r}{\partial x^j}\,V+\frac{\dot{w}_k(t_r)}{c^2}\,\frac{\partial V}{\partial x^j}\right) \\
&= \epsilon^{ijk}\left(\frac{\partial t_r}{\partial x^j}\,\frac{\ddot{w}_k(t_r)}{c^2}\,V+\frac{\partial V}{\partial x^j}\,\frac{\dot{w}_k(t_r)}{c^2}\right),
\end{aligned}
\tag{2.105}
$$

or back in vector form

$$
\mathbf{B}=\frac{1}{c^2}\left(V\,\nabla t_r\times\ddot{\mathbf{w}}(t_r)+\nabla V\times\dot{\mathbf{w}}(t_r)\right).
\tag{2.106}
$$

The two terms are

$$
\begin{aligned}
V\,\nabla t_r\times\ddot{\mathbf{w}}(t_r) &= -\frac{V\,\boldsymbol{\mathscr{r}}(t_r)\times\ddot{\mathbf{w}}(t_r)}{c\,\mathscr{r}(t_r)-\boldsymbol{\mathscr{r}}(t_r)\cdot\dot{\mathbf{w}}(t_r)} = -\frac{\boldsymbol{\mathscr{r}}(t_r)\times\ddot{\mathbf{w}}(t_r)}{\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)}\,\frac{q\,c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))} \\
\nabla V\times\dot{\mathbf{w}}(t_r) &= \frac{q\,c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\,\Big(-\Big(c^2-\dot{w}^2(t_r)+\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\Big) \\
&\quad \times\boldsymbol{\mathscr{r}}(t_r)\times\dot{\mathbf{w}}(t_r)\Big),
\end{aligned}
\tag{2.107}
$$

then

$$
\begin{aligned}
\mathbf{B} &= -\frac{1}{c^2}\,\frac{q\,c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\Big[\boldsymbol{\mathscr{r}}(t_r)\times\ddot{\mathbf{w}}(t_r)\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) \\
&\quad +\Big(c^2-\dot{w}^2(t_r)+\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\Big)\,\boldsymbol{\mathscr{r}}(t_r)\times\dot{\mathbf{w}}(t_r)\Big] \\
&= \frac{1}{c^2}\,\boldsymbol{\mathscr{r}}(t_r)\times\Bigg(\frac{q\,c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\big[-\ddot{\mathbf{w}}(t_r)\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)) \\
&\quad +\Big(c^2-\dot{w}^2(t_r)+\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)\Big)\,\big(c\,\hat{\boldsymbol{\mathscr{r}}}(t_r)-\dot{\mathbf{w}}(t_r)\big)\big]\Bigg),
\end{aligned}
\tag{2.108}
$$

where we have added $c\,\hat{\boldsymbol{\mathscr{r}}}(t_r)$ in the second line – that goes away when the cross product with $\boldsymbol{\mathscr{r}}(t_r)$ is taken (out front), and we can further simplify:

$$
\begin{aligned}
\mathbf{B} &= \frac{1}{c^2}\, \boldsymbol{\mathscr{r}}(t_r) \times \left(\frac{q\,c}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[\left(c^2 - \dot{w}^2(t_r)\right)\, \mathbf{u}(t_r) \right.\right. \\
&\quad \left.\left. + \, \mathbf{u}(t_r)\,(\boldsymbol{\mathscr{r}}(t_r)\cdot\ddot{\mathbf{w}}(t_r)) - \ddot{\mathbf{w}}(t_r)\,(\boldsymbol{\mathscr{r}}(t_r)\cdot\mathbf{u}(t_r)) \right] \right) \\
&= \frac{1}{c}\, \hat{\boldsymbol{\mathscr{r}}}(t_r) \times \left(\frac{q\,\mathscr{r}(t_r)}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[\left(c^2 - \dot{w}^2(t_r)\right)\, \mathbf{u}(t_r) + \boldsymbol{\mathscr{r}}(t_r) \right.\right. \\
&\quad \left.\left. \times\, (\mathbf{u}(t_r) \times \ddot{\mathbf{w}}(t_r)) \right] \right),
\end{aligned}
\tag{2.109}
$$

but the term in parentheses is just $\mathbf{E}$ from above, so we conclude:

$$
\mathbf{B} = \frac{1}{c}\, \hat{\boldsymbol{\mathscr{r}}}(t_r) \times \mathbf{E}. \tag{2.110}
$$

2.4.1 Example: Constant Velocity

From these general expressions, we can find the electric and magnetic fields for a particle traveling at constant velocity along the $\hat{\mathbf{z}}$-axis: $\mathbf{w}(t) = \mathbf{v}\,t\,\hat{\mathbf{z}}$. We'll take $\mathbf{r} = x\,\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}$ and find the field at time t. In terms of the components of the field:

$$
\mathbf{E} = \frac{q\,\mathscr{r}(t_r)}{4\,\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3} \left[(c^2 - \dot{w}^2(t_r))\,\mathbf{u}(t_r) + \boldsymbol{\mathscr{r}}(t_r) \times (\mathbf{u}(t_r) \times \ddot{\mathbf{w}}(t_r)) \right], \tag{2.111}
$$

we need to evaluate $\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r)$ and $\mathscr{r}(t_r)\,\mathbf{u}(t_r)$ (the acceleration term goes away since this is the constant velocity case). Thinking of the dot-product

$$
\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r) = c\,\mathscr{r}(t_r) - (\mathbf{v}\,t_r)\cdot\boldsymbol{\mathscr{r}}(t_r) = \sqrt{c^2\left((z - v\,t)^2 + x^2\right) - v^2\,x^2}, \tag{2.112}
$$

which we knew already from (2.71). For the $\mathbf{u}(t_r)$ term:

$$
\mathscr{r}(t_r)\,\mathbf{u}(t_r) = c\,\boldsymbol{\mathscr{r}}(t_r) - \mathscr{r}(t_r)\,\dot{\mathbf{w}}(t_r) = c\,(\mathbf{r} - \mathbf{v}\,t_r) - c\,(t - t_r)\mathbf{v} = c\,(\mathbf{r} - \mathbf{v}\,t) = c\,\mathbf{R}(t), \tag{2.113}
$$

where $\mathbf{R}(t)$ is the vector pointing from the current location of the charge to the field point (as before).

Then we can write the electric field as

$$
\mathbf{E} = \frac{q}{4\,\pi\,\epsilon_0} \frac{1 - \frac{v^2}{c^2}}{\left((z - v\,t)^2 + x^2 - \frac{v^2\,x^2}{c^2}\right)^{3/2}}\,(\mathbf{r} - \mathbf{v}\,t)\,. \tag{2.114}
$$

This expression can also be written in terms of $\mathbf{R}$ and θ (the angle between $\mathbf{R}$ and $\mathbf{v}$, here $\sin\theta = \frac{x}{R}$)

$$
\mathbf{E} = \frac{q}{4\,\pi\,\epsilon_0} \frac{1 - \frac{v^2}{c^2}}{R^2\left(1 - \frac{v^2}{c^2}\,\sin^2\theta\right)^{3/2}}\,\hat{\mathbf{R}}. \tag{2.115}
$$

The electric field points from the current location of the particle to the field point, even though physically it is a retarded time (with corresponding "earlier" location) that sets the magnitude of the electric field at different points. Notice that in the direction of travel ($\hat{\mathbf{v}} = \hat{\mathbf{z}}$ here, so $\theta = 0$), the field is smaller than in the directions perpendicular to $\hat{\mathbf{v}}$ – that's due to the $\sin\theta$ factor in the denominator. We also recover the Coulomb field for $v \ll c$, as is easy to see in (2.115).

For the magnetic field, we can start from $\mathbf{B} = \frac{1}{c}\,\hat{\boldsymbol{\mathscr{r}}}(t_r) \times \mathbf{E}$, and note that

$$\boldsymbol{\mathscr{r}}(t_r) = \mathbf{r} - \mathbf{v}\,t_r = \mathbf{r} - \mathbf{v}\,t + \mathbf{v}\,(t - t_r) = \mathbf{R} + \mathbf{v}\,\mathscr{r}(t_r)/c \tag{2.116}$$

using $c\,(t - t_r) = \mathscr{r}(t_r)$ as always. Since the electric field points in the $\mathbf{R}$-direction, that term does not contribute to the cross product, and we get

$$\mathbf{B} = \frac{1}{c^2}\,\mathbf{v} \times \mathbf{E}. \tag{2.117}$$

Particles Traveling at Speed c

From the form of (2.115), we see that as v gets larger and larger, the electric field becomes more concentrated in the $\theta = \pi/2$ direction. What would happen to (2.115) if we could take $v = c$?[12] It seems clear in a vague way that we'll get zero field except at $\theta = \pi/2$ in the $v \to c$ limit, implying a $\delta(z - c\,t)$ to localize the field to the plane perpendicular to the direction of travel and at the particle's location. In that plane, we expect the field to point "radially," so in the $\hat{\mathbf{s}}$-direction, and its magnitude should fall off with s, the distance to the charge (again, for points in the plane). Given the presence of the delta function (with dimension of inverse length), we expect:

$$\mathbf{E}_c \sim \frac{q}{\epsilon_0}\,\frac{\delta(z - c\,t)}{s}\,\hat{\mathbf{s}}, \tag{2.118}$$

but beyond that, we cannot fix dimensionless factors of 2, π, e, etc.

To pin down the correct limiting form, go back to (2.114), which can be written in terms of $\gamma \equiv \left(1 - v^2/c^2\right)^{-1/2}$:

$$\begin{aligned}\mathbf{E} &= \frac{q}{4\,\pi\,\epsilon_0}\,\frac{1/\gamma^2}{\left((z - v\,t)^2 + x^2/\gamma^2\right)^{3/2}}\left(x\,\hat{\mathbf{x}} + (z - c\,t)\,\hat{\mathbf{z}}\right)\\ &= \frac{q}{4\,\pi\,\epsilon_0}\,\frac{\gamma\,\left(x\,\hat{\mathbf{x}} + (z - c\,t)\,\hat{\mathbf{z}}\right)}{\left((z - c\,t)^2\,\gamma^2 + x^2\right)^{3/2}}.\end{aligned} \tag{2.119}$$

To find the field for $v = c$, we want:

$$\mathbf{E}_c = \lim_{v\to c}\mathbf{E} = \lim_{\gamma\to\infty}\frac{q}{4\,\pi\,\epsilon_0}\left[\frac{\gamma}{\left((z - c\,t)^2\,\gamma^2 + x^2\right)^{3/2}}\right]\left(x\,\hat{\mathbf{x}} + (z - c\,t)\,\hat{\mathbf{z}}\right). \tag{2.120}$$

The limit is determined by the limit of the term in square brackets, and there is a clever identity that we can use to correctly evaluate that limit. Given a function $f(p)$, we'll show that[13]

$$\lim_{\alpha\to\infty}\left[\alpha\,f(\alpha p)\right] = \delta(p)\int_{-\infty}^{\infty} f(q)\,dq. \tag{2.121}$$

[12] Generally, we reserve the speed c for massless particles, and there are no (known) massless charged particles, so it is not clear that we have immediate physical use for these fields. There is nothing forbidding a massless charged particle, and as a limiting case, the field of a charged particle traveling at c is interesting.

[13] See [1] for discussion and alternate derivation.

To see that this is the case, start with:

$$\int_{-\infty}^{\infty} f(q)\, dq = \lim_{\alpha \to \infty} \int_{-\infty}^{\infty} f(z)\, e^{i\,2\,\pi\,v\,z/\alpha}\, dz, \tag{2.122}$$

where the exponential goes to $e^{i\,0} = 1$ as $\alpha \to \infty$. Let $z = \alpha\, u$, to get:

$$\int_{-\infty}^{\infty} f(q)\, dq = \lim_{\alpha \to \infty} \alpha \int_{-\infty}^{\infty} f(\alpha\, u)\, e^{i\,2\,\pi\,v\,u}\, du. \tag{2.123}$$

Now multiply both sides of this equation by $e^{-i\,2\,\pi\,p\,v}$ and integrate in v from $-\infty$ to ∞. On the left, we pick up the same integral of f, but with a $\delta(p)$ sitting out front, and on the right, we get a $\delta(u-p)$ under the u-integral, and we can use that delta to perform the u-integration:

$$\delta(p) \int_{-\infty}^{\infty} f(q)\, dq = \lim_{\alpha \to \infty} \alpha\, f(\alpha\, p), \tag{2.124}$$

the desired result.

This pertains to (2.120), where we want to evaluate

$$\lim_{\gamma \to \infty} \frac{\gamma}{\left(p^2\,\gamma^2 + x^2\right)^{3/2}}, \tag{2.125}$$

with $p \equiv z - c\,t$. The function of interest is:

$$f(p) = \frac{1}{\left(p^2 + x^2\right)^{3/2}}, \tag{2.126}$$

and then applying (2.121) with this $f(p)$ gives

$$\begin{aligned} \lim_{\gamma \to \infty} \frac{\gamma}{\left(p^2\,\gamma^2 + x^2\right)^{3/2}} &= \delta(p) \int_{-\infty}^{\infty} \frac{1}{\left(q^2 + x^2\right)^{3/2}}\, dq \\ &= \delta(p) \left.\frac{q}{x^2\,\sqrt{q^2 + x^2}}\right|_{-\infty}^{\infty} \\ &= \delta(p)\, \frac{2}{x^2}. \end{aligned} \tag{2.127}$$

Inserting this result into the limit in (2.120),

$$\mathbf{E}_c = \frac{q\,\delta(z - c\,t)}{4\,\pi\,\epsilon_0}\, \frac{2}{x}\, \hat{\mathbf{x}}, \tag{2.128}$$

where the $\hat{\mathbf{z}}$ component goes away because $\delta(z - c\,t)\,(z - c\,t) = 0$. Generalizing to an arbitrary point in the plane perpendicular to the motion of the charge:

$$\mathbf{E}_c = \frac{q\,\delta(z - c\,t)}{2\,\pi\,\epsilon_0\, s}\, \hat{\mathbf{s}}, \tag{2.129}$$

just as we predicted in (2.118), and with the correct constants out front.

Problem 2.21 We know the electric field of a charge moving at c along the $\hat{\mathbf{z}}$-axis from (2.129). Find the magnetic field and potentials V and $\mathbf{A}$ in Lorenz gauge.

Problem 2.22 Find the electric and magnetic fields associated with the potentials

$$\begin{aligned} V(\mathbf{r},t) &= -\frac{q\,\omega\,d\cos\theta}{4\,\pi\,\epsilon_0\,c\,r}\,\sin(\omega\,(t-r/c)) \\ \mathbf{A}(\mathbf{r},t) &= -\frac{\mu_0\,q\,\omega\,d}{4\,\pi\,r}\,\sin(\omega\,(t-r/c))\,\hat{\mathbf{z}}. \end{aligned} \tag{2.130}$$

Express your answer in spherical coordinates with spherical basis vectors.

Problem 2.23 Use the identity (2.121) to show that

$$\lim_{\alpha\to\infty}\int_{-\alpha}^{\alpha}\cos(2\,\pi\,f\,t)\,df = \delta(t), \tag{2.131}$$

an identity from (2.27) we established in the interest of self-consistency for Fourier transforms.

Problem 2.24 Argue, using simple geometry and the electric fields for point charges, that you cannot have uniform circular motion for a pair of equal and oppositely charged particles (of the same mass) if you include only retarded time. Show how including an equal amount of advanced time might solve that problem (it's a problem because "positronium," an orbiting electron–anti-electron pair, exists).

Problem 2.25 Take spiraling motion of the form $\mathbf{w} = R\left(\cos(\omega\,t)\,\hat{\mathbf{x}} + \sin(\omega\,t)\,\hat{\mathbf{y}}\right) + v_z\,t\,\hat{\mathbf{z}}$ (the z-component depends on the speed v_z). For the field point $\mathbf{r} = 2\,R\,\hat{\mathbf{x}}$ and time $t = 0$, we'll calculate the retarded location, velocity, and acceleration of the particle, and compare that with the location, velocity, and acceleration at $t = 0$ (this is to show that you would get very different values if you used the instantaneous values of a charge in calculating the electric field at a field point). Start by writing down the expressions for $\boldsymbol{\mathscr{r}}$, $\mathbf{v}$, and $\mathbf{a}$ for the given $\mathbf{w}$.

Take $R = 10^4$ m, $\omega = 1000\,\text{s}^{-1}$, and $v_z = 10$ m/s. Use the `Mathematica` command "`FindRoot`"[14] to find the retarded time. From this retarded time value, evaluate $\mathbf{w}(t_r)$, $\mathbf{v}(t_r)$, and $\mathbf{a}(t_r)$. Compare those with $\mathbf{w}(0)$, $\mathbf{v}(0)$, and $\mathbf{a}(0)$.

Problem 2.26 A particle in "hyperbolic motion" travels along the z-axis according to $\mathbf{w}(t) = \sqrt{b^2 + (c\,t)^2}\,\hat{\mathbf{z}}$. For a field point $\mathbf{r} = x\,\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}$ (in the x–z plane), find the retarded time at $t = 0$. What is the retarded time associated with $z = 0$ (at $t = 0$)?

Use `Mathematica` (or other) to generate $\boldsymbol{\mathscr{r}}$, $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{a}$ at the retarded time for our $\mathbf{r} = x\,\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}$ and $t = 0$. From these, form the electric field (2.104). Try sketching the field lines in the x–z plane. Be careful; what happens at $z = 0$ (given its retarded time from above)? If you showed your field line sketch to a student in introductory physics (first learning about Gauss's law), what would they say? (you don't really have to answer this last question; I just want you to think about the problems with this electric field). For a full discussion of the electric field of a charge in hyperbolic motion, see [14] and its references.

[14] Or implement and use the bisection routine from Section B.1.

2.5 Radiation: Point Particles

As a physical model for this section, suppose we move a charge q up and down according to $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$ as in Figure 2.4. We'll find the potentials and then the fields using some natural approximations, and these will serve to define the "radiation" fields. Put our field point at $\mathbf{r} = r\,\sin\theta\,\hat{\mathbf{y}} + r\,\cos\theta\,\hat{\mathbf{z}}$, and assume that

$$d \ll r \rightarrow \frac{d}{r} \equiv \epsilon \ll 1, \qquad \omega\,d \ll c \rightarrow \frac{\omega\,d}{c} \equiv \delta \ll 1, \tag{2.132}$$

so that the observation point is far from the charge (which moves between $-d$ and d on the z-axis) and the charge is moving up and down slowly compared with the speed of light (for the $\mathbf{w}(t)$ given here, the speed is $\omega\,d\,\sin(\omega\,t)$ so its maximum value is $\omega\,d$).

2.5.1 Prediction and the Radiation Field Definition

Before we start working through these approximations in the Liénard–Wiechert potentials, let's guess the form of the potential using our static intuition. This will turn out to be useful and focus our attention on the "new" piece that we could *not* guess from statics. We have a charge q that is at location $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$ at time t, so there is a time-dependent dipole moment, $\mathbf{p}(t) = q\,w(t)\,\hat{\mathbf{z}}$ in addition to the monopole term. We could then guess (using $r \gg d$ so that the separation distance is basically r):

$$V(\mathbf{r}, t) = \frac{q}{4\,\pi\,\epsilon_0\,r} + \frac{q\,w(t)\,\cos\theta}{4\,\pi\,\epsilon_0\,r^2}, \tag{2.133}$$

just the monopole and dipole potentials.

We know, from the speed of light, that it takes some time for the information about the particle's location to reach the field point at $\mathbf{r}$, and for $r \gg d$, the time of flight is dominated by $t_r \approx t - r/c$, so we might even guess

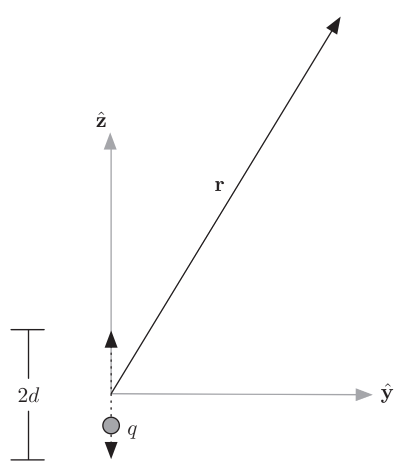

Fig. 2.4 A charge oscillates near the origin; the observation point for the fields is far away at $\mathbf{r}$.

$$V(\mathbf{r}, t) = \frac{q}{4\pi\epsilon_0 r} + \frac{q\, w(t - r/c)\cos\theta}{4\pi\epsilon_0 r^2}. \tag{2.134}$$

This is correct, as far as it goes, and it captures two of the pieces that will show up in the full case. But it is missing the most important piece, the one that serves to *define* radiation.

The defining feature of the "radiation fields" is that they carry energy away to infinity – that is, energy that will never return to the system. When you construct a sphere of charge, you do work bringing the charges in and putting them into a spherical arrangement. That work is then stored in the electric field, and you can recover the energy when you disassemble the configuration. By contrast, the energy that is needed to generate the radiation fields, our focus in this section, goes away and cannot be recovered.

What sort of electric field has this property? Think of the statement of conservation of energy: $\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{S}$ where u is the energy density associated with the electromagnetic fields ($u = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2\mu_0} B^2$) and $\mathbf{S}$ is the Poynting vector ($\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}$). In integral form, for a domain Ω with boundary $\partial\Omega$, we have

$$\frac{d}{dt}\int_\Omega u\, d\tau = -\oint_{\partial\Omega} \mathbf{S} \cdot d\mathbf{a}. \tag{2.135}$$

The integral on the left is the energy stored in the domain Ω, and the integral on the right is the energy flux through the surface of Ω. If we take our domain to be all space, then the radiation fields are ones for which the right-hand side is non-zero. Then the flux of energy through the surface at spatial infinity is not zero, so the energy has left the "universe."

To be concrete, take the domain to be a sphere of radius r (that we will send to infinity), and assume that $\mathbf{S}$ points in the $\hat{\mathbf{r}}$-direction; then the surface integral becomes

$$\oint_{\partial\Omega} \mathbf{S} \cdot d\mathbf{a} = \int_0^{2\pi}\int_0^{\pi} S\, r^2 \sin\theta\, d\theta\, d\phi \tag{2.136}$$

where $r \to \infty$. For electric and magnetic fields that go like $1/r^2$, we have $S \sim 1/r^4$ (from $\mathbf{E} \times \mathbf{B}$) and this integral is zero (the integrand goes like $1/r^2$ with $r \to \infty$). Referring back to our predicted potential (2.134), the first term will lead to precisely a $1/r^2$ electric field, and the second term gives a $1/r^3$ field, which is even more zero than the monopole term (for $1/r^3$ fields, $S \sim 1/r^6$ and the integrand in (2.136) goes like $1/r^4$ with $r \to \infty$).

Evidently, then, the radiation $\mathbf{E}$ and $\mathbf{B}$ fields must each go like $1/r$, and we have no example of this in our initial guess for the potential of our oscillating charge. As we move along with the full calculation, you'll see an obsession with the $1/r$ piece of the fields; these are the ones that carry new physics.

2.5.2 Using the Full Liénard–Wiechert Potentials

Go back to the problem description from the start of Section 2.5, and let's work out the content of the Liénard–Wiechert potentials. For starters, we'll record the separation vector and estimate the retarded time. The separation vector is

$$\boldsymbol{\mathscr{r}}(t_r) \equiv \mathbf{r} - \mathbf{w}(t_r) = r\sin\theta\,\hat{\mathbf{y}} + (r\cos\theta - d\cos(\omega t_r))\,\hat{\mathbf{z}}, \tag{2.137}$$

with magnitude

$$\mathscr{r}(t_r) = r\left(1 - 2\frac{d}{r}\cos\theta\,\cos(\omega\, t_r) + \frac{d^2}{r^2}\right)^{1/2} \approx r\,(1 - \epsilon\,\cos\theta\,\cos(\omega\, t_r))\,, \tag{2.138}$$

where we've kept the first-order term in $\epsilon \equiv d/r$. While we don't know t_r yet, we know that $\cos(\omega\, t_r)$ is bounded, so we are free to expand in small ϵ. Looking ahead, it is clear that quantities like $\mathscr{r}(t_r)$ will depend on t_r in terms with ϵ (and/or $\delta \equiv \omega\, d/c$) out front, so we only need to keep the leading order term in the retarded time definition:

$$c\,(t - t_r) = \mathscr{r}(t_r) \approx r + O(\epsilon) \longrightarrow t_r = t - \frac{r}{c} + O(\epsilon); \tag{2.139}$$

the retarded time is dominated by the time it takes light to travel from the origin to r. I'll leave t_r in the various expressions to keep things from getting cluttered, but remember that whenever you see t_r in this section, you'll replace it with the approximate $t - \frac{r}{c}$ at the end.

Next, we need to evaluate the denominator of (2.66). We'll start with the exact form, then keep only those terms that are constant or linear (in ϵ, δ)

$$\begin{aligned}
\mathscr{r}(t_r)\, c - \boldsymbol{\mathscr{r}}(t_r)\cdot \dot{\mathbf{w}}(t_r) &= c\,\mathscr{r}(t_r) + d\,\omega\,\sin(\omega\, t_r)\,(r\cos\theta - d\,\cos(\omega\, t_r)) \\
&\approx c\, r\left[1 - \epsilon\,\cos\theta\,\cos(\omega\, t_r)\right. \\
&\quad \left. + \frac{d\,\omega}{c}\,\sin(\omega\, t_r)\left(\cos\theta - \frac{d}{r}\,\cos\omega\, t_r\right)\right] \\
&\approx c\, r\,[1 - \cos\theta\,(\epsilon\,\cos(\omega\, t_r) - \delta\,\sin(\omega\, t_r))]\,,
\end{aligned} \tag{2.140}$$

where the $\epsilon\,\delta$ term has been dropped (it is quadratic in the small quantities of interest).

We're ready to put the pieces into (2.66):

$$\begin{aligned}
V(\mathbf{r}, t) &= \frac{q}{4\pi\,\epsilon_0}\,\frac{1}{r\,[1 - \cos\theta\,(\epsilon\,\cos(\omega\, t_r) - \delta\,\sin(\omega\, t_r))]} \\
&\approx \frac{1}{4\pi\,\epsilon_0}\,\frac{q}{r} + \frac{1}{4\pi\,\epsilon_0}\,\frac{q\,d\,\cos(\omega\, t_r)}{r^2}\,\cos\theta - \frac{1}{4\pi\,\epsilon_0}\,\frac{q\,d\,\omega\,\sin(\omega\, t_r)}{c\, r}\,\cos\theta.
\end{aligned} \tag{2.141}$$

The first term is just the Coulomb potential associated with the point charge q. As expected, it is time-independent (charge conservation holds here). The second term is precisely what we would expect from a time-varying dipole moment sitting at the origin: the numerator is what we would call $p(t_r)$, and we recover precisely the $\sim p(t_r)\,\cos\theta/r^2$ potential behavior we always get from a dipole moment. These two terms are the monopole and dipole contributions we predicted in (2.134).

The third term is new, and we would not have predicted its existence from a study of electrostatics. It has the interesting feature that it depends on $1/r$, like the monopole term, but carries a factor of $\delta \equiv \omega\, d/c$. We'll focus on this new term, and one way to differentiate between it and the other time-varying dipole term that goes like $\epsilon \equiv d/r$ is to assume that $\epsilon \ll \delta$. That limit translates to $\omega\, r \gg c$, or $\frac{\omega\, r}{c} \gg 1$. There is a nice physical interpretation for this assumption: if we think ahead to what all these oscillatory terms are going to mean for V and $\mathbf{A}$, it is clear there will be monochromatic fields with wavelength $\lambda \equiv \frac{2\pi\, c}{\omega}$ so that $\epsilon \ll \delta$ really means $r \gg \lambda$; the viewing distance is large compared with the wavelength of the fields.

Just to isolate it, we're thinking about a potential of the form:

$$V(\mathbf{r}, t) = -\frac{1}{4\,\pi\,\epsilon_0\,c}\,\frac{\cos\theta}{r}\,\omega\,q\,d\,\sin(\omega\,t_r). \tag{2.142}$$

This will be the first term that leads to an electric field with $1/r$-dependence (can you spot how?). We're focusing on it, and dropping the monopole and dipole terms. Those terms are there, of course, and if the omission bothers you, leave them in (or appeal to superposition and put them in at the end).

Next, we need the magnetic vector potential associated with the oscillating charge configuration, and again, we only keep the term linear in δ. If we think of the form of $\mathbf{A} = \frac{\dot{\mathbf{w}}(t_r)}{c^2}\,V$, then

$$\mathbf{A} = -\frac{\omega\,d}{c}\,\frac{V}{c}\,\sin(\omega\,t_r)\,\hat{\mathbf{z}}, \tag{2.143}$$

and there is a δ sitting out front, so we need only take the "Coulomb" (leading) portion of the electric potential; anything else would go like δ^2 or (even worse) $\epsilon\,\delta$. The new portion of the magnetic vector potential that goes along with (2.142) is

$$\mathbf{A} = -\frac{q}{4\,\pi\,\epsilon_0\,r\,c}\,\frac{\omega\,d}{c}\,\sin(\omega\,t_r)\,\hat{\mathbf{z}}. \tag{2.144}$$

For the fields, it is easiest to get $\mathbf{B} = \nabla \times \mathbf{A}$ and then work backward to $\mathbf{E}$. The vector potential depends on coordinates both directly, through r, and indirectly through $t_r = t - \frac{r}{c}$. The curl will give

$$\mathbf{B} = \nabla \times \mathbf{A} = \frac{\partial A_z}{\partial y}\,\hat{\mathbf{x}} - \frac{\partial A_z}{\partial x}\,\hat{\mathbf{y}}. \tag{2.145}$$

The derivative of r is straightforward: $\frac{\partial r}{\partial y} = \frac{y}{r}$ and $\frac{\partial r}{\partial x} = \frac{x}{r}$, so that

$$\frac{\partial A_z}{\partial y} = \frac{q}{4\,\pi\,\epsilon_0\,r^2\,c}\,\frac{\omega\,d}{c}\,\frac{y}{r}\left(\sin(\omega\,t_r) + \frac{\omega\,r}{c}\,\cos(\omega\,t_r)\right). \tag{2.146}$$

Keep in mind there is an overall δ out front. Inside the parentheses, though, is a factor of $\omega\,r/c$, and that term dominates, so we keep it, and evaluate y in spherical coordinates to get:

$$\frac{\partial A_z}{\partial y} \approx \frac{q}{4\,\pi\,\epsilon_0\,r\,c}\,\frac{\omega^2\,d}{c^2}\,\cos(\omega\,t_r)\,\sin\theta\,\sin\phi. \tag{2.147}$$

Similarly,

$$\frac{\partial A_z}{\partial x} \approx \frac{q}{4\,\pi\,\epsilon_0\,r\,c}\,\frac{\omega^2\,d}{c^2}\,\cos(\omega\,t_r)\,\sin\theta\,\cos\phi. \tag{2.148}$$

Putting these together in $\mathbf{B}$ gives

$$\mathbf{B} = \frac{\mu_0\,q}{4\,\pi\,r\,c}\,\omega^2\,d\,\sin\theta\,\cos(\omega\,t_r)\left(\sin\phi\,\hat{\mathbf{x}} - \cos\phi\,\hat{\mathbf{y}}\right), \tag{2.149}$$

and the direction is secretly $-\hat{\boldsymbol{\phi}}$.

To get $\mathbf{E}$, remember that $\mathbf{B} = \frac{1}{c}\,\hat{\boldsymbol{\mathscr{r}}}(t_r) \times \mathbf{E}$ from (2.110), and we can replace $\hat{\boldsymbol{\mathscr{r}}}(t_r)$ with $\hat{\mathbf{r}}$ at this order. We know the magnitude, E, will just be $c\,B$. For the direction, note that $\hat{\mathbf{r}} \times \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}}$ so that $\mathbf{E} \sim \hat{\boldsymbol{\theta}}$ to get the correct direction for $\mathbf{B}$. Then finally, we get the fields

$$\begin{aligned}\mathbf{E} &= -\frac{\mu_0\,q\,d\,\omega^2}{4\,\pi}\,\frac{\sin\theta}{r}\,\cos(\omega\,(t - r/c))\,\hat{\boldsymbol{\theta}}\\ \mathbf{B} &= -\frac{\mu_0\,q\,d\,\omega^2}{4\,\pi\,c}\,\frac{\sin\theta}{r}\,\cos(\omega\,(t - r/c))\,\hat{\boldsymbol{\phi}}.\end{aligned} \tag{2.150}$$

Notice that the electric and magnetic fields oscillate with the same frequency as the source; that's a characteristic of this dipole radiation. If you measure the frequency of the fields, you learn about the physical motion that generated those fields. The fields in (2.150) represent, locally, plane wave solutions. At a fixed field location, r and $\sin\theta$ are constant (so there is an overall constant in front of the cosine terms), and we have electric and magnetic fields that are perpendicular to each other and the direction of wave motion, with single-frequency oscillatory dependence, $\cos(\omega\,(t - r/c))$.

We can gain generality by noting that the dipole moment of this configuration is $\mathbf{p}(t) = q\,\mathbf{w}(t) = q\,d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$, so that the potentials could be written as

$$\begin{aligned}V(\mathbf{r},t) &= \frac{1}{4\,\pi\,\epsilon_0}\,\frac{\cos\theta}{r}\,\dot{p}(t - r/c)\\ \mathbf{A}(\mathbf{r},t) &= \frac{\dot{\mathbf{p}}(t - r/c)}{4\,\pi\,\epsilon_0\,c^2\,r},\end{aligned} \tag{2.151}$$

with fields

$$\begin{aligned}\mathbf{E} &= \frac{\mu_0}{4\pi}\,\frac{\sin\theta}{r}\,\ddot{p}(t - r/c)\,\hat{\boldsymbol{\theta}} = \frac{\mu_0}{4\,\pi\,r}\,\hat{\mathbf{r}} \times \big(\hat{\mathbf{r}} \times \ddot{\mathbf{p}}(t - r/c)\big)\\ \mathbf{B} &= \frac{\mu_0}{4\,\pi\,c}\,\frac{\sin\theta}{r}\,\ddot{p}(t - r/c)\,\hat{\boldsymbol{\phi}} = -\frac{\mu_0}{4\,\pi\,r\,c}\,\hat{\mathbf{r}} \times \ddot{\mathbf{p}}(t - r/c),\end{aligned} \tag{2.152}$$

which we'll see again in a more general context later on.

2.5.3 Radiated Power

The Poynting vector $\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E} \times \mathbf{B}$ tells us how the field energy is moving through space. In our initial definition of the radiation fields, we thought about a Poynting vector that went like $1/r^2$ and pointed in the $\hat{\mathbf{r}}$-direction. Going back to the fields in (2.150), the Poynting vector indeed has this form:

$$\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E} \times \mathbf{B} = \frac{1}{\mu_0\,c}\left(\frac{\mu_0\,q\,d\,\omega^2\,\sin\theta}{4\,\pi\,r}\right)^2\,\cos^2(\omega\,(t - r/c))\;\hat{\mathbf{r}}. \tag{2.153}$$

We can average the Poynting vector over time, to get the net energy change. Take an average over one full cycle (the period of the Poynting vector is $T = \omega/(2\,\pi)$),

$$\langle\mathbf{S}\rangle \equiv \frac{1}{T}\int_{t_0}^{t_0+T} \mathbf{S}(t)\,dt, \tag{2.154}$$

where t_0 represents the start of a cycle. For our Poynting vector, the relevant piece is the time average of $\cos^2(w\,(t - r/c))$, which is $1/2$ over a full cycle, giving

$$\langle \mathbf{S} \rangle = \frac{1}{2\,\mu_0\,c}\left(\frac{\mu_0\,q\,d\,\omega^2\,\sin\theta}{4\,\pi\,r}\right)^2 \hat{\mathbf{r}}. \tag{2.155}$$

The magnitude of the time-averaged Poynting vector is called the "intensity," and it has an interesting profile (owing to the $\sin^2\theta$-dependence): there is no energy flow in the $\theta = 0$ and π-directions. Instead, the intensity is sharply peaked around $\theta = \pi/2$, perpendicular to the axis of the motion of the particle.

Finally, we can calculate the total power, the energy lost per unit time, by integrating $\langle \mathbf{S} \rangle$:

$$P \equiv \oint_{\partial\Omega} \langle \mathbf{S} \rangle \cdot d\mathbf{a}, \tag{2.156}$$

where we take $\partial\Omega$ to be the sphere at spatial infinity (that bounds "all space"). The $1/r^2$ in the Poynting vector cancels the r^2 from da, and the ϕ integration throws in a $2\,\pi$, leaving just the polar angle integral of $\sin^3\theta$:

$$P = \left(\frac{\pi}{\mu_0\,c}\left(\frac{\mu_0\,q\,d\,\omega^2}{4\,\pi}\right)^2\right)\int_0^{\pi} \sin^3\theta\,d\theta = \frac{\mu_0\,(q\,d\,\omega^2)^2}{12\,\pi\,c}. \tag{2.157}$$

While we have used a specific configuration to generate this result, we can generalize by writing $q\,d\,\omega^2$ in terms of the dipole moment of the configuration: $p(t) = q\,d\,\cos(\omega\,t)$. The two factors of ω come from two time derivatives of $p(t)$, and we got rid of the explicit time-dependence by performing the time average. If we had not done the time-averaging, we would have just left in the $\cos^2(\omega\,(t - r/c))$ term in P (while losing the factor of $1/2$), and then

$$q\,d\,\omega^2\,\cos(\omega\,(t - r/c)) = \ddot{p}(t - r/c), \tag{2.158}$$

so that the power radiated by a generic dipole moment is

$$P(t) = \frac{\mu_0\,\ddot{p}(t - r/c)^2}{6\,\pi\,c}. \tag{2.159}$$

You might object that we used a sphere at spatial infinity to find the total power, so the appearance of r in this expression is questionable. But the whole point of the spatial integration was that the r-dependence drops out, so r could be anything. Since $\mathbf{S} \sim 1/r^2\,\hat{\mathbf{r}}$, the total power flowing through spheres of different radii is the same.

From our model, we have another interpretation for $q\,d\,\omega^2\,\cos(\omega\,(t - r/c))$: the particle's location at time t is $w(t) = d\,\cos(\omega\,t)$ so that

$$q\,d\,\omega^2\,\cos(\omega\,(t - r/c)) = q\,\ddot{w}(t - r/c) \tag{2.160}$$

or, more generally, $q\,a(t - r/c)$, where $a(t)$ is the acceleration of the particle at time t. Using this form in the power expression gives the Larmor formula for radiated power (again with time-dependence built in, you can average over a full cycle at the end if you like):

$$P(t) = \frac{\mu_0\,q^2\,a(t - r/c)^2}{6\,\pi\,c}. \tag{2.161}$$

It is the acceleration of a charge that leads to radiated power. That is no surprise – it couldn't be that a charge at rest radiates (how would it lose energy?) nor could a charge moving at constant velocity radiate (boosting to its rest frame, we again have no way to extract energy). So acceleration *has* to be the culprit (there could have been no culprit, but that describes a very different world from the one in which we live). To get a charge to go from rest to speed v requires that we dump energy in to make up for the radiative loss that the acceleration causes. Getting a neutral mass moving at speed v would not require this additional energy. Since it is a^2 that appears in P, the charge also loses energy when it *de*celerates – when you stop a charge, it radiates energy.

Our model of a charge moving up and down along the $\hat{\mathbf{z}}$-axis would require constant energy input on our part to keep the motion going, otherwise the charge would slow down as energy is carried away by the fields, a process known as radiation damping.

Problem 2.27 Show that

$$\frac{1}{T}\int_0^T \cos^2(2\pi\, t/T)\, dt = 1/2, \tag{2.162}$$

using (1) the exponential form of cos and (2) a geometric argument together with the identity $\cos^2\theta + \sin^2\theta = 1$.

Problem 2.28 Evaluate $\int_0^\pi \sin^3\theta\, d\theta$ using the exponential form of $\sin\theta$.

Problem 2.29 If you could harness all of a 100 watt bulb to move a charge up and down (oscillating with $\mathbf{w} = d\cos(\omega\, t)\,\hat{\mathbf{z}}$), and you took a $q = 1$ C charge moving with $\omega = 2\pi\, 60$ Hz, what would the amplitude of the motion, d, be?

Problem 2.30 You measure the maximum (during a day) power-per-unit-area of sunlight on the earth to be $I_e \approx 1400\ \mathrm{W/m^2}$. Assuming the sunlight is due to radiating dipoles, what is the power-per-unit-area on the surface of the sun, I_s?

Problem 2.31 A particle of charge q moves along a trajectory given by $\mathbf{w}(t) = f\cos(\omega\, t)\,\hat{\mathbf{x}} + g\sin(\omega\, t)\,\hat{\mathbf{y}}$ (for constants f, g, and ω). Sketch the trajectory of the particle. What is the total power (time-averaged) radiated? Which radiates more, a particle moving in a circular trajectory of radius f or a particle moving along a line of length $2f$ (each traversed with constant angular frequency ω)? What is the ratio of the power (time-averaged) radiated for the circle versus the line trajectories?

Problem 2.32 Here, we'll solve a problem exactly, then expand the solution and show we get the same thing out of a perturbative solution to the problem. We want to find y with:

$$\begin{aligned} 0 &= x^2 - \epsilon\, x - 1 \\ y &= \epsilon\, \cos(x). \end{aligned} \tag{2.163}$$

a. Solve the top equation for x (take the positive root), insert that x into the bottom equation, and then expand y in powers of $\epsilon \ll 1$ (use Taylor expansion, so $y(\epsilon) = y(0) + \epsilon\left(\frac{dy(\epsilon)}{d\epsilon}\Big|_{\epsilon=0}\right) + \cdots$) up through ϵ^1.

b. This time, expand your solution for x (take the positive root) through order ϵ^1, insert in y, and expand y in powers of ϵ (through ϵ^1). Do you need both the ϵ^0 and ϵ^1 term in x to get the same result for y that you had in part (a)?

2.6 Radiation: Continuous Sources

We used a special configuration to develop the radiation fields associated with dipoles, and those are the dominant fields. In a sense, then, the set (2.152) is the full story, and you just have to find the dipole moment of your more general configuration to use them. But we can get the same expressions directly from the general case, and it's worth seeing how our series of approximations works out there.

Let's go back to an arbitrary source: given $\rho(\mathbf{r}, t)$ and $\mathbf{J}(\mathbf{r}, t)$ as in Figure 2.5, what are the potentials (and fields)? From the integral form of the solution, we have

$$\begin{aligned} V(\mathbf{r}, t) &= \int G(\mathbf{r}, \mathbf{r}', t, t')\, \rho(\mathbf{r}', t')\, d\tau' dt' = \int \frac{\rho(\mathbf{r}', t')}{4\pi\epsilon_0 \mathscr{r}}\, \delta\left((t - t') - \frac{\mathscr{r}}{c}\right) d\tau' dt' \\ &= \int \frac{\rho(\mathbf{r}', t - \frac{\mathscr{r}}{c})}{4\pi\epsilon_0 \mathscr{r}}\, d\tau' \end{aligned} \tag{2.164}$$

where we used the temporal delta function to get rid of the t' integration. We'll expand the integrand in powers of r', allowing us to focus on solutions that are far from isolated sources ($\mathbf{r}'$ contributes where ρ is non-zero, so our physical picture is a finite bounded

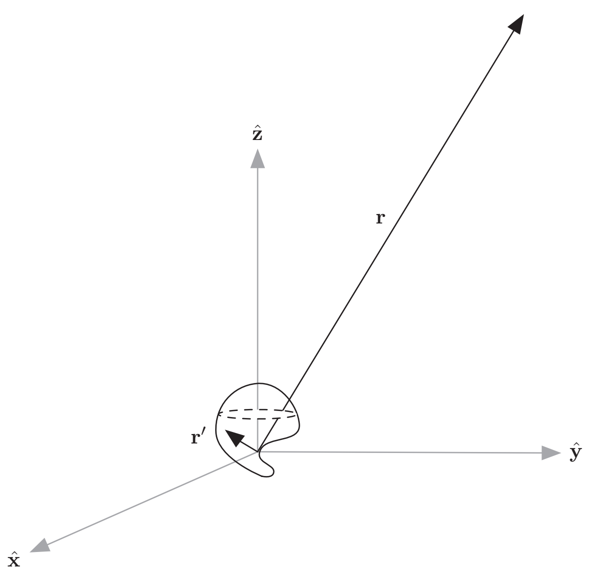

Fig. 2.5 A general distribution of charge, with $\rho(\mathbf{r}', t')$ and $\mathbf{J}(\mathbf{r}', t')$ given, and localized near the origin. The field point is at $\mathbf{r}$, time t.

source that is near the origin compared with our field-point location, **r**). Using the approximation, for constant α,

$$\mathscr{r}^{\alpha} \equiv \left[(\mathbf{r}-\mathbf{r}')\cdot(\mathbf{r}-\mathbf{r}')\right]^{\alpha/2} = r^{\alpha}\left(1-\frac{2\,\hat{\mathbf{r}}\cdot\mathbf{r}'}{r}+\frac{r'^2}{r^2}\right)^{\alpha/2} \approx r^{\alpha}-\frac{\alpha\,\hat{\mathbf{r}}\cdot\mathbf{r}'}{r^{1-\alpha}}, \tag{2.165}$$

the integrand in (2.164) is approximately

$$\begin{aligned}\frac{\rho(\mathbf{r}',t-\frac{\mathscr{r}}{c})}{4\pi\,\mathscr{r}} &\approx \frac{1}{4\pi}\,\mathscr{r}^{-1}\,\rho\left(\mathbf{r}',t-\frac{r}{c}\left(1-\frac{\hat{\mathbf{r}}\cdot\mathbf{r}'}{r}\right)\right)\\ &= \frac{1}{4\pi\,r}\left(1+\frac{\hat{\mathbf{r}}\cdot\mathbf{r}'}{r}\right)\left[\rho\left(\mathbf{r}',t-\frac{r}{c}\right)+\frac{r}{c}\,\dot{\rho}\left(\mathbf{r}',t-\frac{r}{c}\right)\frac{\hat{\mathbf{r}}\cdot\mathbf{r}'}{r}\right]\\ &= \frac{1}{4\pi\,r}\left(\rho\left(\mathbf{r}',t-\frac{r}{c}\right)+\frac{\hat{\mathbf{r}}\cdot\mathbf{r}'}{r}\left(\rho\left(\mathbf{r}',t-\frac{r}{c}\right)+\frac{r}{c}\,\dot{\rho}\left(\mathbf{r}',t-\frac{r}{c}\right)\right)\right),\end{aligned} \tag{2.166}$$

and then the integral solution becomes, letting $t_r \equiv t-\frac{r}{c}$ (again, this is only an approximation to the true t_r, but we keep the notation to make contact with our previous expressions),

$$\begin{aligned}V(\mathbf{r},t) &\approx \frac{q}{4\pi\,\epsilon_0\,r}+\frac{\hat{\mathbf{r}}}{4\pi\,\epsilon_0\,r^2}\cdot\int\left(\rho(\mathbf{r}',t_r)+\frac{r}{c}\,\dot{\rho}(\mathbf{r}',t_r)\right)\mathbf{r}'d\tau'\\ &= \frac{q}{4\pi\,\epsilon_0\,r}+\frac{\hat{\mathbf{r}}\cdot\mathbf{p}(t_r)}{4\pi\,\epsilon_0\,r^2}+\frac{\hat{\mathbf{r}}\cdot\dot{\mathbf{p}}(t_r)}{4\pi\,\epsilon_0\,r\,c}.\end{aligned} \tag{2.167}$$

The first term is constant in time, and represents the usual Coulomb term. The second term is also expected, that's just the dipole potential, with the dipole moment evaluated at $t_r = t - r/c$, i.e., at the retarded time for a dipole moment at the origin with the field point a distance r away. The third term is the first correction that comes from the time derivative and is familiar as the coordinate-free form of V from (2.151).

For the magnetic vector potential, we start with:

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\int\frac{\mathbf{J}\left(\mathbf{r}',t-\frac{\mathscr{r}}{c}\right)}{\mathscr{r}}\,d\tau', \tag{2.168}$$

and this time, we can drop the time-derivative term in the expansion of **J** (that, as it turns out, contributes at a higher order), so

$$\mathbf{A}(\mathbf{r},t) \approx \frac{\mu_0}{4\pi\,r}\int\mathbf{J}\left(\mathbf{r}',t-\frac{r}{c}\right)d\tau'. \tag{2.169}$$

Now use charge conservation to relate the integral of **J** to the time derivative of the dipole moment of the distribution. Take

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot\mathbf{J} = -\frac{\partial J^i}{\partial x^i}, \tag{2.170}$$

where we switch to index notation to make the result easier to develop. Multiply both sides of (2.170) by r^k (the kth component of **r**) and integrate over all space (assuming, of course, that the actual charge distribution is of finite extent)

$$\frac{d}{dt}\left(\int \rho\, r^k\, d\tau\right) = -\int \frac{\partial J^i}{\partial x^i}\, r^k\, d\tau. \tag{2.171}$$

The left-hand side is just $\dot{p}^k$, the time derivative of the dipole moment. To evaluate the right-hand side, note that

$$\int \frac{\partial}{\partial x^i}\left(J^i\, r^k\right)\, d\tau = \int \frac{\partial J^i}{\partial x^i}\, r^k\, d\tau + \int J^i\, \frac{\partial r^k}{\partial x^i}\, d\tau = \oint J^i\, r^k\, da_i = 0 \tag{2.172}$$

where the middle equality comes from the product rule and we have used the divergence theorem to establish that the integral is zero (by turning it into a boundary integral at spatial infinity). One of the terms in the middle is precisely the right-hand side of (2.171), and we can replace it with the negative of the other term (with $\frac{\partial r^k}{\partial x^i} = \delta_i^k$), then

$$\dot{p}^k = \int J^i\, \delta_i^k\, d\tau \longrightarrow \dot{\mathbf{p}} = \int \mathbf{J}\, d\tau. \tag{2.173}$$

We can use this to simplify the magnetic vector potential

$$\mathbf{A}(\mathbf{r}, t) \approx \frac{\mu_0}{4\,\pi\, r}\left(\dot{\mathbf{p}}(t_r)\right), \tag{2.174}$$

matching our earlier model-based form for $\mathbf{A}$ from (2.151).

To find the fields, we take the relevant derivatives, keeping in mind that (for $t_r = t - r/c$ here): $\nabla t_r = -\frac{1}{c}\nabla r = -\frac{1}{c}\hat{\mathbf{r}}$. We'll focus on the new piece, since we already know the electric and magnetic fields in the static case, and in keeping with the traditional definition of "radiation," it is only the portions of the field that go like $1/r$ that contribute (as discussed in Section 2.5.1)

$$\begin{aligned}
\mathbf{E}^{\text{rad}} &= -\nabla V - \frac{\partial \mathbf{A}}{\partial t} = \frac{\hat{\mathbf{r}} \cdot \ddot{\mathbf{p}}(t_r)}{4\,\pi\,\epsilon_0\, r\, c^2}\,\hat{\mathbf{r}} - \frac{\mu_0}{4\,\pi\, r}\,\ddot{\mathbf{p}}(t_r) \\
&= \frac{\mu_0}{4\,\pi\, r}\left(\left(\hat{\mathbf{r}} \cdot \ddot{\mathbf{p}}(t_r)\right)\hat{\mathbf{r}} - \ddot{\mathbf{p}}\right) \\
&= \frac{\mu_0}{4\,\pi\, r}\left(\hat{\mathbf{r}} \times \left(\hat{\mathbf{r}} \times \ddot{\mathbf{p}}(t_r)\right)\right).
\end{aligned} \tag{2.175}$$

Meanwhile, for $\mathbf{B} = \nabla \times \mathbf{A}$, we have, again dropping the contributions that go like $1/r^2$:

$$\mathbf{B}^{\text{rad}} = -\frac{\mu_0}{4\,\pi\, r\, c}\hat{\mathbf{r}} \times \ddot{\mathbf{p}}(t_r). \tag{2.176}$$

We have recovered the pair (2.152) that we got in a concrete setting, using the same approximations, but starting from the more general integral solutions for V and $\mathbf{A}$. Everything else (intensity, power radiated, etc.) pushes through similarly.

2.7 Exact Point Particle Radiation Fields

We've learned that the acceleration of a charge is what leads to radiation, and again, radiation *means* energy that is lost forever. In the previous section, we saw how to isolate the radiating contributions for an extended distribution of moving charge. Now let's do the special case of a point charge moving along the trajectory $\mathbf{w}(t)$, this time without making any simplifying assumptions (as we did in the case of an oscillating charge). We know the full electric and magnetic fields from (2.104) and (2.110); they are

$$\begin{aligned}
\mathbf{E} &= \frac{q\,\mathscr{r}(t_r)}{4\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\left[(c^2-\dot{w}^2(t_r))\,\mathbf{u}(t_r)+\boldsymbol{\mathscr{r}}(t_r)\times(\mathbf{u}(t_r)\times\ddot{\mathbf{w}}(t_r))\right]\\
\mathbf{B} &= \frac{1}{c}\;\hat{\boldsymbol{\mathscr{r}}}(t_r)\times\mathbf{E}\\
\mathbf{u} &\equiv c\;\hat{\boldsymbol{\mathscr{r}}}(t_r)-\dot{\mathbf{w}}(t_r)\\
c\,(t-t_r) &= |\mathbf{r}-\mathbf{w}(t_r)|.
\end{aligned} \tag{2.177}$$

The only term in $\mathbf{E}$ that depends on acceleration is the last, and so we expect it to be the term that contributes to the radiation, and similarly for $\mathbf{B}$:

$$\begin{aligned}
\mathbf{E}^{\text{rad}} &= \frac{q\,\mathscr{r}(t_r)}{4\pi\,\epsilon_0\,(\mathbf{u}(t_r)\cdot\boldsymbol{\mathscr{r}}(t_r))^3}\left[\boldsymbol{\mathscr{r}}(t_r)\times(\mathbf{u}(t_r)\times\ddot{\mathbf{w}}(t_r))\right]\\
\mathbf{B}^{\text{rad}} &= \frac{1}{c}\;\hat{\boldsymbol{\mathscr{r}}}(t_r)\times\mathbf{E}^{\text{rad}}.
\end{aligned} \tag{2.178}$$

It is relatively clear from a picture of an oscillating charge's electric field, for example, that some of the field stays nearby and some goes far away. We can compute the field of a particle moving along the $\hat{\mathbf{z}}$-axis, with $\mathbf{w}(t) = d\cos(\omega t)\,\hat{\mathbf{z}}$. We'll do this "exactly," meaning that we will not use the approximations that led us to dipole radiation, but rather compute all of the pieces in (2.177). We cannot demonstrate a fully closed form solution, of course, because the retarded time calculation must be done numerically (that procedure is described in Section B.1). In Figure 2.6, we can see a contour plot of the magnitude of the electric field in the y–z plane. The "wings" represent the acceleration fields, while the elliptical portion in the center comes from the "Coulomb" and velocity fields. In the panels, you can clearly see the acceleration fields leaving the domain.

2.7.1 Instantaneously at Rest

If the motion of the moving particle has $\dot{\mathbf{w}}(t_r) = 0$, the radiation fields simplify considerably. In that case

$$\mathbf{E}^{\text{rad}} = \frac{q}{4\pi\,\epsilon_0}\,\frac{1}{c^2\,\mathscr{r}(t_r)}\;\hat{\boldsymbol{\mathscr{r}}}(t_r)\times(\,\hat{\boldsymbol{\mathscr{r}}}(t_r)\times\mathbf{a}(t_r)) \tag{2.179}$$

and $\mathbf{B}^{\text{rad}} = \frac{1}{c}\;\hat{\boldsymbol{\mathscr{r}}}(t_r)\times\mathbf{E}^{\text{rad}}$. The Poynting vector is

$$\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E}^{\text{rad}}\times\mathbf{B}^{\text{rad}} = \frac{1}{\mu_0\,c}\,\mathbf{E}^{\text{rad}}\times(\,\hat{\boldsymbol{\mathscr{r}}}(t_r)\times\mathbf{E}^{\text{rad}}). \tag{2.180}$$

From (2.179), it is clear that $\mathbf{E}^{\text{rad}}\perp\hat{\boldsymbol{\mathscr{r}}}(t_r)$, and the cross products in the Poynting vector simplify using the BAC-CAB rule,

$$\mathbf{S} = \frac{1}{\mu_0\,c}\left(E^{\text{rad}}\right)^2\;\hat{\boldsymbol{\mathscr{r}}}(t_r). \tag{2.181}$$

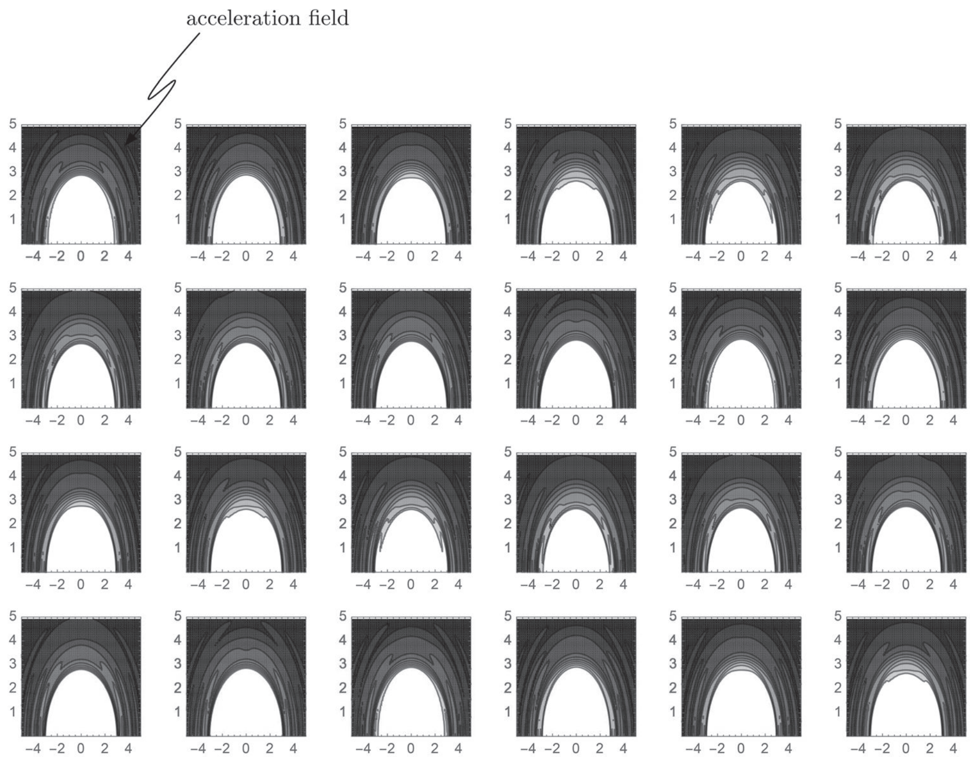

Fig. 2.6 The full electric field (magnitude) for a particle moving according to $\mathbf{w}(t) = d\ \cos(\omega\, t)\, \hat{\mathbf{z}}$. The acceleration portion of the field is marked, and you can watch multiple "wings" leaving the domain – time increases left to right within a row, top to bottom vertically.

The magnitude of the electric field, squared, is (dropping the t_r reminder, and using BAC-CAB in (2.179))

$$\begin{aligned}\left(E^{\text{rad}}\right)^2 &= \left(\frac{q}{4\,\pi\,\epsilon_0\,c^2\,\mathscr{r}}\right)^2 \left[\hat{\boldsymbol{\mathscr{r}}}\,(\hat{\boldsymbol{\mathscr{r}}}\cdot\mathbf{a}) - \mathbf{a}\right]\cdot\left[\hat{\boldsymbol{\mathscr{r}}}\,(\hat{\boldsymbol{\mathscr{r}}}\cdot\mathbf{a}) - \mathbf{a}\right] \\ &= \left(\frac{q}{4\,\pi\,\epsilon_0\,c^2\,\mathscr{r}}\right)^2 \underbrace{\left[a^2 - (\hat{\boldsymbol{\mathscr{r}}}\cdot\mathbf{a})^2\right]}_{=a^2(1-\cos^2\theta)},\end{aligned} \tag{2.182}$$

for θ the angle between $\hat{\boldsymbol{\mathscr{r}}}$ and $\mathbf{a}$. Putting this into the expression for the Poynting vector gives

$$\mathbf{S} = \frac{1}{\mu_0\,c}\left(\frac{q}{4\,\pi\,\epsilon_0\,c^2\,\mathscr{r}}\right)^2 a^2\ \sin^2\theta\ \hat{\boldsymbol{\mathscr{r}}}. \tag{2.183}$$

Integrating over a sphere centered at the instantaneously at-rest charge gives a total power of

$$P = \frac{\mu_0\,q^2\,a^2}{6\,\pi\,c}, \tag{2.184}$$

the Larmor formula (see (2.161)) again.

Problem 2.33 For $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$, and field point $\mathbf{r} = r\,\sin\theta\,\hat{\mathbf{y}} + r\,\cos\theta\,\hat{\mathbf{z}}$, take the three approximations $d/r \equiv \epsilon \ll 1$, $\omega\,d/c \equiv \delta \ll 1$, and $\omega\,r/c \gg 1$ (i.e., $\epsilon \ll \delta$) and find the radiation field (the part that goes like $1/r$; calculate electric only) directly from (2.177).

Problem 2.34 How much power is needed to get a charge with $q = 1\,\mu$C to move according to $\mathbf{w} = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$, with $d = 0.01$ mm and $\omega = 60$ Hz?

2.8 Radiation Reaction

The radiation emitted by an accelerating charge carries energy away from the charge. We know how much power is radiated from the Larmor formula, but how should we do the dynamical accounting for that lost energy? How, for example, does the particle's motion change because of the radiation? One way to address the issue is to find an effective force that would result in the same energy loss as that provided by the electromagnetic radiation. Then we use this force along with any others acting on a charged particle in Newton's second law.

The work done by a force acting on a particle as it moves from $\mathbf{b}$ to $\mathbf{c}$ is $W = \int_{\mathbf{b}}^{\mathbf{c}} \mathbf{F}\cdot d\boldsymbol{\ell}$. The energy associated with the Larmor power for a temporal interval $t_b \to t_c$ is $E = -\int_{t_b}^{t_c} P\,dt$ (negative because it is energy *loss*). If we set $W = E$, and take as our path the trajectory of the particle so that $d\boldsymbol{\ell} = \mathbf{v}\,dt$, then

$$\int_{t_b}^{t_c} \mathbf{F}\cdot\mathbf{v}\,dt = -\frac{\mu_0\,q^2}{6\,\pi\,c}\int_{t_b}^{t_c} a^2\,dt \tag{2.185}$$

where the particle is at $\mathbf{b}$ at time t_b and $\mathbf{c}$ at time t_c. The integral on the right can be written as a dot product with $\mathbf{v}$ as well, to match the dot product on the left:

$$\int_{t_b}^{t_c} a^2\,dt = \int_{t_b}^{t_c} \frac{d\mathbf{v}}{dt}\cdot\frac{d\mathbf{v}}{dt}\,dt = \mathbf{v}\cdot\left.\frac{d\mathbf{v}}{dt}\right|_{t_b}^{t_c} - \int_{t_b}^{t_c}\frac{d^2\mathbf{v}}{dt^2}\cdot\mathbf{v}\,dt \tag{2.186}$$

via integration by parts. If we take $\mathbf{v}(t_b) = \mathbf{v}(t_c)$ and $\mathbf{a}(t_b) = \mathbf{a}(t_c)$ (oscillatory motion has this property, for example), the boundary term vanishes, and we can use the integral piece in (2.185)

$$\int_{t_b}^{t_c} \mathbf{F}\cdot\mathbf{v}\,dt = \frac{\mu_0\,q^2}{6\,\pi\,c}\int_{t_b}^{t_c}\frac{d^2\mathbf{v}}{dt^2}\cdot\mathbf{v}\,dt \tag{2.187}$$

suggesting that we make the association

$$\mathbf{F} = \frac{\mu_0\,q^2}{6\,\pi\,c}\,\dot{\mathbf{a}}. \tag{2.188}$$

This is the "Abraham–Lorentz" force of radiation reaction.

As an example, suppose we have a neutral mass m attached to a spring with spring constant k, then the one-dimensional equation of motion is

$$m\,\ddot{x}(t) = -k\,x(t). \tag{2.189}$$

We would specify the initial position and velocity, two constants to set the two constants of integration we expect from Newton's second law.

If the particle carries charge q, the equation of motion is

$$m\ddot{x}(t) = -k\,x(t) + \frac{\mu_0\,q^2}{6\,\pi\,c}\,\dddot{x}(t). \tag{2.190}$$

This is very different and, in particular, requires an additional constant to start things off – we'd need to provide the initial position, velocity, and acceleration, and it's not entirely clear how to do that.

Even in the absence of an external force, the radiation reaction force is confusing. Let $\frac{\mu_0\,q^2}{6\,\pi\,m\,c} \equiv \tau$, a time scale, then

$$\ddot{x} = \tau\,\dddot{x} \longrightarrow \ddot{x}(t) = \ddot{x}(0)\,e^{t/\tau}, \tag{2.191}$$

so that if there is some initial acceleration, it grows in time. That would be fantastic, if it were true, since all you would have to do to solve the energy crisis is kick a charged particle, it would then take off and you could extract energy by slowing it down.

Landau–Lifschitz Form

There is a related but alternate formulation of the radiation reaction equation that eliminates some of the problematic behavior associated with the $\dot{\mathbf{a}}$ force above (but has its own difficulties). The so-called Landau–Lifschitz form of radiation reaction proceeds from (2.188). Take

$$m\,\mathbf{a}(t) = \mathbf{F}(t) + m\,\tau\,\dot{\mathbf{a}}(t), \tag{2.192}$$

where $\mathbf{F}(t)$ is an external (not radiation reaction) force. Moving the derivative of acceleration over to the left,

$$m\,\underbrace{(\mathbf{a}(t) - \tau\,\dot{\mathbf{a}}(t))}_{\approx\mathbf{a}(t-\tau)} = \mathbf{F}(t), \tag{2.193}$$

and if we let $s \equiv t - \tau$, we can associate the approximate evaluation $\mathbf{a}(t-\tau) = \mathbf{a}(s)$ with $\mathbf{F}(s+\tau)$. Relabeling $s \rightarrow t$, we have

$$m\,\mathbf{a}(t) \approx \mathbf{F}(t+\tau). \tag{2.194}$$

Finally, if we Taylor expand the force, we get, to first order in τ:

$$m\,\mathbf{a}(t) \approx \mathbf{F}(t) + \tau\,\frac{d\mathbf{F}(t)}{dt} \tag{2.195}$$

and this is the Landau–Lifschitz form of radiation reaction. It makes different predictions than the Abraham–Lorentz version. For example, in the Landau–Lifschitz setting, a constant force does not lead to any dynamical change due to radiation. That's not the same as the Abraham–Lorentz prediction, which includes (depending on initial conditions) exponential growth as in (2.190).

Problem 2.35 Find the value of $\tau \equiv \frac{\mu_0\,q^2}{6\,\pi\,m\,c}$ for an electron.

Problem 2.36 Show that Newton's second law, together with the Abraham–Lorentz force:

$$m\,a = F + m\,\tau\,\dot{a} \tag{2.196}$$

(working in one dimension for simplicity) leads to continuous acceleration (provided the force is not a delta function): $\lim_{\epsilon\to 0}(a(t+\epsilon) - a(t-\epsilon)) = 0$.

Problem 2.37 Take a constant force F_0 that turns on suddenly at $t = 0$ and find the motion of a charged particle (mass m, charge q) that starts at some initial position x_0 from rest using both the Abraham–Lorentz force, and the Landau–Lifschitz one (working in one dimension). For the Abraham–Lorentz case, pick either $\ddot{x}(0) = 0$ or $\ddot{x}(0) = F_0/m$. Both choices are problematic physically; why? (Use your result from the previous problem.)

2.8.1 Simple Model

The effective force in (2.188), while it cannot be the end of the story, is a fundamental prediction of special relativity in the sense that its physical mechanism is provided by the retarded time evaluation of the electric field. For an extended body, different pieces of the body can act on each other at different times because the retarded time puts different pieces in "contact" at different moments. It is this "self-force" (which exists even for point particles) that causes the radiation reaction force described in the previous section. To see how it works, take a pair of charges, each with charge $q/2$, separated by a distance d. This will be our model,[15] in the limit that $d \to 0$, for a point particle of charge q. Consider two different times, $t = 0$ and $t = T$, where the horizontal location of the charges at different times are $x(0) = 0$ and $x(T)$, respectively, as depicted in Figure 2.7. The electric field from the lower charge at $t = 0$ acts on the upper charge at time $t = T$ provided $t = 0$ is the retarded time for the upper charge at time T. We can ensure that this is the case by enforcing the retarded time condition

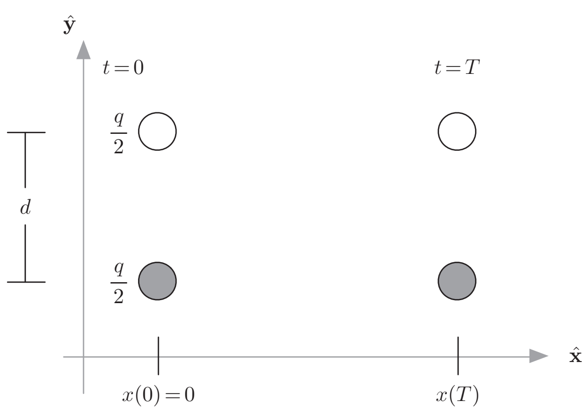

Fig. 2.7 Two charges separated by a distance d at time $t = 0$ and later at $t = T$.

[15] The dumbbell model used here comes from [19].

$$c\,T = \sqrt{d^2 + x(T)^2}. \tag{2.197}$$

We provide the trajectory, the $\mathbf{w}(t)$ to be used in (2.177), and in order to simplify life, take $x(0) = 0$, $\dot{x}(0) = 0$, $\ddot{x}(0) = 0$, and $\dddot{x}(0) = j_0$, so that the horizontal motion is governed by an initial jerk. Then the electric field at the upper charge, at time T, has $\mathbf{u} = c\,\hat{\boldsymbol{\mathscr{r}}}$, and no acceleration contribution – it looks like the Coulomb field,

$$\mathbf{E} = \frac{q/2}{4\,\pi\,\epsilon_0\,\mathscr{r}^2}\,\hat{\boldsymbol{\mathscr{r}}} = \frac{q/2}{4\,\pi\,\epsilon_0\,\mathscr{r}^3}\,\boldsymbol{\mathscr{r}}. \tag{2.198}$$

Using $\boldsymbol{\mathscr{r}} = x(T)\,\hat{\mathbf{x}} + d\,\hat{\mathbf{y}}$ with magnitude $\mathscr{r} = \sqrt{x(T)^2 + d^2} = c\,T$ from (2.197), the force on the upper charge is

$$\mathbf{F}_u = \frac{(q/2)^2}{4\,\pi\,\epsilon_0\,(c\,T)^3}\left(x(T)\,\hat{\mathbf{x}} + d\,\hat{\mathbf{y}}\right). \tag{2.199}$$

The same argument holds for the force on the lower charge at time T due to the upper charge at time 0 (all that changes is the sign of the $\hat{\mathbf{y}}$ component of the force), so the net force on the pair of charges is

$$\mathbf{F} = 2\,\frac{(q/2)^2}{4\,\pi\,\epsilon_0\,(c\,T)^3}\,x(T)\,\hat{\mathbf{x}}. \tag{2.200}$$

We're interested in the limit as $d \to 0$, and for small d, the time-of-flight for light is short, so that T will be small. For short times, we can approximate motion that starts with a constant jerk (and satisfies the rest of our initial conditions) by $x(T) = \frac{1}{6}\,j_0\,T^3$ (solving $\dddot{x}(t) = j_0$). But then the net force is just

$$\mathbf{F} = \frac{(q/2)^2}{2\,\pi\,\epsilon_0\,(c\,T)^3}\,\frac{1}{6}\,j_0\,T^3\,\hat{\mathbf{x}} = \frac{\mu_0\,q^2\,\dddot{x}(0)}{48\,\pi\,c}\,\hat{\mathbf{x}}. \tag{2.201}$$

The result stands even if we take the $d \to 0$ limit (d does not show up in the expression at all). This self-force is a factor of four smaller than the one we got in (2.188) (although most of that has to do with the evaluation of $a(t)$ at the current time instead of the retarded one), but it has the same basic form and is inspired by the same physics.

We chose a simplifying trajectory for the self-force evaluation, but that doesn't change its validity. It is interesting that the self-force on a point source is an immediate consequence of special relativity: the finite speed of field information is what drives the calculation. While radiation itself is a property of the acceleration portion of the electromagnetic fields, the self-force calculated here does not rely on the acceleration, or even velocity portion of the full electric field. Curiously, it is the Coulomb piece that is relevant, when evaluated at the retarded time.

3 Field Lagrangians

3.1 Review of Lagrangian Mechanics

In classical mechanics, Newton's second law is the primary quantitative tool. Given a potential energy $U(\mathbf{r})$, we have to solve

$$m\,\ddot{\mathbf{r}}(t) = -\nabla U(\mathbf{r}) \tag{3.1}$$

subject to some initial or boundary conditions. Then $\mathbf{r}(t) = x(t)\,\hat{\mathbf{x}} + y(t)\,\hat{\mathbf{y}} + z(t)\,\hat{\mathbf{z}}$ is a vector that tells us the location of a particle of mass m at time t moving under the influence of $U(\mathbf{r})$.

As written, Newton's second law refers to Cartesian coordinates, so the above is really three equations:

$$\begin{aligned} m\,\ddot{x}(t) &= -\frac{\partial U}{\partial x} \\ m\,\ddot{y}(t) &= -\frac{\partial U}{\partial y} \\ m\,\ddot{z}(t) &= -\frac{\partial U}{\partial z}. \end{aligned} \tag{3.2}$$

If we want to change coordinates, we have to manipulate these equations appropriately, and that can get unwieldy. For example, if we switch to cylindrical coordinates (a mild transformation, comparatively), in which $x = s\,\cos\phi$, $y = s\,\sin\phi$, and z remains the same, the equations in (3.2) become

$$\begin{aligned} m\left(\ddot{s}\,\cos\phi - 2\,\dot{s}\,\dot{\phi}\,\sin\phi - s\,\ddot{\phi}\,\sin\phi - s\,\dot{\phi}^2\,\cos\phi\right) &= -\frac{\partial U}{\partial s}\frac{\partial s}{\partial x} - \frac{\partial U}{\partial \phi}\frac{\partial \phi}{\partial x} \\ m\left(\ddot{s}\,\sin\phi + 2\,\dot{s}\,\dot{\phi}\,\cos\phi + s\,\ddot{\phi}\,\cos\phi - s\,\dot{\phi}^2\,\sin\phi\right) &= -\frac{\partial U}{\partial s}\frac{\partial s}{\partial y} - \frac{\partial U}{\partial \phi}\frac{\partial \phi}{\partial y} \\ m\,\ddot{z} &= -\frac{\partial U}{\partial z}. \end{aligned} \tag{3.3}$$

While Newton's second law can rarely be solved analytically, it is even harder to glean information when the physics is obscured by a different choice of coordinate system. Even the free particle solution is complicated in cylindrical coordinates. You would have to take linear combinations of the equations above to isolate, for example, equations for $\ddot{s}$ and $\ddot{\phi}$.

Enter the Lagrangian, and the Euler–Lagrange equations of motion that reproduce Newton's second law. The Euler–Lagrange equations of motion are coordinate invariant, meaning that they are structurally identical in all coordinate systems. We define the

Lagrangian, $L \equiv T - U$, the kinetic energy minus the potential energy for a particular physical configuration. Then we can integrate to get a functional (a function that takes a function and returns a number), the action:

$$S[\mathbf{r}(t)] = \int_{t_0}^{t_f} L(\mathbf{r}(t), \dot{\mathbf{r}}(t))\, dt = \int_{t_0}^{t_f} \left(\frac{1}{2} m\, \dot{\mathbf{r}} \cdot \dot{\mathbf{r}} - U(\mathbf{r}) \right) dt. \tag{3.4}$$

The action assigns a number to a given trajectory, described by the vector pointing from the origin to the location of the particle at time t: $\mathbf{r}(t)$ (defined for all $t \in [t_0, t_f]$). The "dynamical trajectory," the one taken by the particle according to Newton's second law, is the one that minimizes S.[1] That minimization occurs over the space of all possible paths connecting the initial and final locations of the particle.

When thinking about the minimum value for S, we have to compare apples with apples. Physical observation provides the relevant boundary conditions: we see the particle at location $\mathbf{r}_0$ at time t_0, and then we observe it later at location $\mathbf{r}_f$, time t_f. What happens in between is the question. In Figure 3.1, we see two different paths connecting the same endpoints; which one minimizes the action? In order to carry out the minimization, we perturb about the dynamical trajectory (which is at this point unknown). Let $\mathbf{r}(t)$ be the true trajectory, the solution to Newton's second law with appropriate boundary values. Then an arbitrary perturbation looks like: $\mathbf{r}(t) + \mathbf{u}(t)$ where $\mathbf{u}(t_0) = \mathbf{u}(t_f) = 0$ to respect the boundary conditions. How does the action respond to this arbitrary perturbation? If we take $\mathbf{u}(t)$ small, then we can expand the action,

$$\begin{aligned} S[\mathbf{r}(t) + \mathbf{u}(t)] &= \int_{t_0}^{t_f} L(\mathbf{r} + \mathbf{u}, \dot{\mathbf{r}} + \dot{\mathbf{u}})\, dt \\ &\approx \underbrace{\int_{t_0}^{t_f} L(\mathbf{r}, \dot{\mathbf{r}})\, dt}_{=S[\mathbf{r}(t)]} + \underbrace{\int_{t_0}^{t_f} \left[\frac{\partial L}{\partial \mathbf{r}} \cdot \mathbf{u} + \frac{\partial L}{\partial \dot{\mathbf{r}}} \cdot \dot{\mathbf{u}} \right] dt}_{\equiv \delta S}, \end{aligned} \tag{3.5}$$

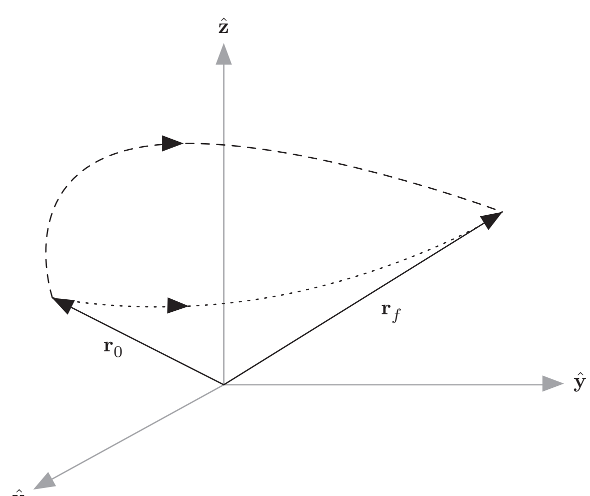

Fig. 3.1 Two trajectories connecting the initial location, $\mathbf{r}_0$, with the final location, $\mathbf{r}_f$.

[1] Technically, we'll be *extremizing* S, by setting its derivative to zero, but that extremum is, typically, a minimum.

where we obtain the second line by Taylor expansion in $\mathbf{u}(t)$ and $\dot{\mathbf{u}}(t)$ (assuming both are small). The notation here is convenient, but warrants definition. The derivative $\frac{\partial L}{\partial \mathbf{r}}$ is really just the gradient of L, what we would normally call ∇L. Similarly, $\frac{\partial L}{\partial \dot{\mathbf{r}}}$ is a vector whose components are the derivative of L with respect to $\dot{x}$, $\dot{y}$, and $\dot{z}$ (in Cartesian coordinates):

$$\frac{\partial L}{\partial \dot{\mathbf{r}}} \equiv \frac{\partial L}{\partial \dot{x}}\,\hat{\mathbf{x}} + \frac{\partial L}{\partial \dot{y}}\,\hat{\mathbf{y}} + \frac{\partial L}{\partial \dot{z}}\,\hat{\mathbf{z}}. \tag{3.6}$$

Now for the trajectory described by $\mathbf{r}(t)$ to be a minimum, we must have $\delta S = 0$ (δS is as close as we can get to a "derivative" for S, and we demand that it vanish) for all $\mathbf{u}$. To see what this implies about $\mathbf{r}(t)$ itself, note that $\mathbf{u}$ is arbitrary, but once it has been chosen, $\dot{\mathbf{u}}$ is determined. We'd like to write all of δS in terms of the arbitrary $\mathbf{u}$. We can do that by integrating the second term in δS by parts, noting that $\mathbf{u}$ vanishes at the end points of the integration (by construction):

$$\delta S = \int_{t_0}^{t_f} \left[\frac{\partial L}{\partial \mathbf{r}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}}\right] \cdot \mathbf{u}\, dt = 0 \tag{3.7}$$

and for this to hold for all excursions $\mathbf{u}$, we must have

$$\frac{\partial L}{\partial \mathbf{r}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}} = 0, \tag{3.8}$$

the Euler–Lagrange equations of motion.

It is important to remember that we have not assumed any particular coordinates for $\mathbf{r}(t)$. We can write L in terms of any coordinates we like; the equations of motion do not change. Sometimes, we highlight that by writing:

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = 0, \qquad \text{for } i = 1, 2, 3, \tag{3.9}$$

where $\{q^i\}_{i=1}^3$ are the "generalized coordinates."

This procedure, of extremizing an action, may seem obscure. Why should a peculiar combination of kinetic and potential energy lead to equations of motion like (3.9) that reproduce Newton's second law? And what, if any, physical insight does the story here contain? It is interesting that for $U = 0$, there is a natural geometric interpretation for both the action and the process of minimizing. The action is proportional to the length squared along the curve defined by $\mathbf{r}(t)$, and its extremization is really length minimization. For $U \neq 0$, we lose the length interpretation and rely on the correctness of the equations of motion (meaning the reproduction of Newton's second law) in Cartesian coordinates to justify the procedure.

Example

Of the two pieces that make up L, the kinetic energy is always the same. For a particle of mass m moving along the path $\mathbf{r}(t)$, $T = \frac{1}{2}\, m\, \dot{\mathbf{r}}(t) \cdot \dot{\mathbf{r}}(t)$. The potential energy depends on the physical system (the mass is attached to a spring, or it's charged and in the presence of an electric field, etc.); we'll leave it as $U(\mathbf{r})$ for now. In Cartesian coordinates, $T = \frac{1}{2}\, m \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right)$, and $U(x, y, z)$ is the potential energy. Then using (3.9), we recover precisely (3.2). If we instead use cylindrical coordinates, $T = \frac{1}{2}\, m \left(\dot{s}^2 + s^2\, \dot{\phi}^2 + \dot{z}^2\right)$ with $U(s, \phi, z)$ as the potential energy. Now our three equations of motion still come from (3.9) using $q^1 = s$, $q^2 = \phi$, and $q^3 = z$; we get

$$\begin{aligned} m\ddot{s} - m\,s\,\dot{\phi}^2 &= -\frac{\partial U}{\partial s} \\ m\,\frac{d}{dt}\left(m\,s^2\,\dot{\phi}\right) &= -\frac{\partial U}{\partial \phi} \\ m\ddot{z} &= -\frac{\partial U}{\partial z}. \end{aligned} \tag{3.10}$$

These are much better than (3.3), although they are still a coupled nonlinear set in general.

Problem 3.1 Suppose we started with a Lagrangian, in one dimension, that looked like: $L = \frac{1}{2}\,m\,\dot{x}^2 - U(x) + \frac{\partial F(x,t)}{\partial x}\,\dot{x} + \frac{\partial F(x,t)}{\partial t}$ for $F(x,t)$ an arbitrary function of x and t. What is the equation of motion in this case?

Problem 3.2 Using spherical coordinates, construct the Lagrangian for a central potential that depends only on r (the spherical radial coordinate). From the Lagrangian, find the three equations of motion.

Problem 3.3 In non-relativistic mechanics, you start with a particle action:

$$S = \int L\,dt, \tag{3.11}$$

with Lagrangian $L = \frac{1}{2}\,m\,v^2$ for free particles.

a. In relativistic mechanics, we'd also like to start with an action. The action must be a scalar (both the analogue of L and dt have to be scalars). Construct

$$S = \alpha \int \bar{L}\,d\bar{t}, \tag{3.12}$$

for relativistic dynamics (we'll work on free particles for now, no potentials). Pick an infinitesimal $d\bar{t}$ that is a scalar and an $\bar{L}$ that is also scalar (remember that there are ways of ensuring a quantity is a scalar having little to do with counting dangling indices!) with dimension of velocity (not velocity squared); α is a constant we'll set in a moment.

b. Show that your action is "reparametrization" invariant, meaning that you can reparametrize the motion in terms of anything you like (not just $\bar{t}$) without changing the form of the action. Pick the coordinate time t as the parameter and evaluate your $\bar{L}$. By requiring that you recover $\frac{1}{2}\,m\,v^2$ for the $v \ll c$ limit, set the constant α and write the final form of $\bar{L}$.

c. Using the usual Euler–Lagrange equations of motion applied to $\bar{L}$, find the equation of motion for a relativistic free particle in one spatial dimension. Does it look familiar?

d. Take the Legendre transform of $\bar{L}$ to find the Hamiltonian associated with this Lagrangian; is it what you expect? Remember that the Legendre transform is $\bar{H} = \mathbf{p}\cdot\dot{\mathbf{r}} - \bar{L}$ for canonical momentum $\mathbf{p} = \frac{\partial \bar{L}}{\partial \dot{\mathbf{r}}}$. The Hamiltonian is written entirely in terms of $\mathbf{p}$ and $\mathbf{r}$ (eliminating $\dot{\mathbf{r}}$ in favor of $\mathbf{p}$).

3.1.1 Length

For a particle moving along a trajectory given by $\mathbf{r}(t)$, as in Figure 3.2, we know that the distance traveled in a time interval dt is just

$$d\ell = \sqrt{\frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}}\, dt, \tag{3.13}$$

so that the length of the trajectory, as the particle goes from $\mathbf{a}$ at time t_0 to $\mathbf{b}$ at time t_f, is

$$\ell = \int_{\mathbf{a}}^{\mathbf{b}} d\ell = \int_{t_0}^{t_f} \sqrt{\frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}}\, dt. \tag{3.14}$$

This expression is "reparametrization invariant," meaning that the parameter t can be changed to anything else without changing the form of ℓ (you just rename t). That parameter freedom makes good geometrical sense: the length of a curve is independent of how you parametrize the curve (as long as you go from the correct starting point to the correct stopping point, of course). The integrand of the length is related to the free-particle Lagrangian; it looks like the square root of the kinetic energy (up to constants). The extremizing trajectories of the action are length-minimizing curves, i.e., straight lines.

In special relativity, we learn to measure lengths differently. For a particle moving in space-time, with $x^\mu(t)$ a vector pointing from the origin to its space-time location at time t (we could use a different parameter of course), we have:

$$d\ell = \sqrt{-\frac{dx^\mu}{dt} \frac{dx_\mu}{dt}}\, dt, \tag{3.15}$$

(the minus sign is just to make the length real, pure convention based on our choice of signs in the Minkowski metric), and again, we could use any parameter we like in $\ell = \int d\ell$. Sticking with time, we can compute the length of a curve in special relativity:

$$\ell = \int_{t_0}^{t_f} \sqrt{c^2 - \frac{d\mathbf{r}}{dt} \cdot \frac{d\mathbf{r}}{dt}}\, dt. \tag{3.16}$$

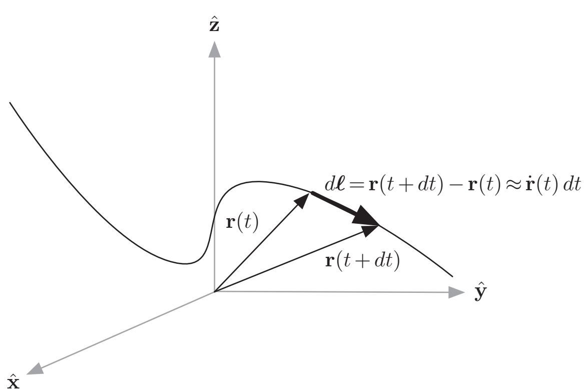

Fig. 3.2 A particle travels along the path shown, with its location at time t given by $\mathbf{r}(t)$.

The action for special relativity is more closely related to length than the one in non-relativistic mechanics. We start with an integrand proportional to $d\ell$ itself, not the length-squared-like starting point of the non-relativistic action (the kinetic energy term is missing the square root that would make it proportional to length). For an arbitrary curve parameter, λ, we have relativistic action

$$S[x(\lambda)] = \alpha \int_{\lambda_0}^{\lambda_f} \sqrt{-\frac{dx^\mu}{d\lambda}\frac{dx_\mu}{d\lambda}}\, d\lambda, \tag{3.17}$$

and reparametrization invariance tells us we can make λ anything, and the coordinate time is a good choice. Proper time is another natural candidate for curve parameter. In that case, you have the side constraint defining proper time, namely,

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{\mathbf{v}\cdot\mathbf{v}}{c^2}}}. \tag{3.18}$$

Problem 3.4 For hyperbolic motion we have: $x(t) = \sqrt{b^2 + (c\,t)^2}$. At time $t = 0$, the particle is at b. When the particle gets to $2\,b$ traveling back out along the x-axis (i.e., with $t > 0$), it has traveled a Pythagorean distance b. What is its *space-time* distance?

Problem 3.5 What is the space-time length of parabolic motion: $x(t) = \alpha\, t^2$ from $t = 0$ to $t = 1$?

Problem 3.6 Find the proper time for hyperbolic motion and write the position as a function of proper time.

Problem 3.7 Combining E&M with a relativistic particle action must be done via a scalar term of the form: $\int A^\mu\, \eta_{\mu\nu}\, \frac{dx^\nu}{d\tau}\, d\tau$ (this term has the advantage of being reparametrization invariant). Putting it all together in coordinate time parametrization:

$$S = -m\,c \int \sqrt{-\frac{dx^\mu}{dt}\frac{dx_\mu}{dt}}\, dt + \alpha \int \frac{dx^\mu}{dt} A_\mu\, dt \tag{3.19}$$

(where α is a constant that sets the coupling strength).

a. Use the Euler–Lagrange equations to find the equations of motion in Cartesian coordinates (use indexed notation[2]). Write these in terms of $\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$ and $\mathbf{B} = \nabla \times \mathbf{A}$ and set the value of α to ensure you recover the correct Lorentz force law (note also that the equation of motion here implies that the Lorentz force is an "ordinary" force and not a Minkowski force).

b. Find the canonical momentum $p_j = \frac{\partial L}{\partial \dot{x}^j}$ (i.e., evaluate the right-hand side).

c. Find the total energy, H (the Hamiltonian), by taking the Legendre transform of the Lagrangian using your momentum from part (b). You should write H in terms of the velocity $\mathbf{v}$ instead of the (usual) momentum, which makes it easier to see the result.

[2] In this expanded space-time setting, the Euler–Lagrange equations are

$$-\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) + \frac{\partial L}{\partial x^\mu} = 0.$$

3.1.2 Noether's Theorem

Aside from the uniformity of the equations of motion, the Lagrange approach makes it easy to see/prove Noether's theorem connecting symmetries to conservation laws. Suppose we have a free particle, so that there is no potential energy, and we work in Cartesian coordinates. Then $L = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right)$, and the Lagrangian (and hence the action) is insensitive to a shift in coordinates; i.e., introducing $\bar{x} = x + x_0$, $\bar{y} = y + y_0$, and $\bar{z} = z + z_0$ for arbitrary constants $\{x_0, y_0, z_0\}$ doesn't change the value or form of L, which is just $\bar{L} = \frac{1}{2} m \left(\dot{\bar{x}}^2 + \dot{\bar{y}}^2 + \dot{\bar{z}}^2\right)$. Noether's theorem says that this "symmetry" (or isometry) is associated with conserved quantities. Those quantities are easy to isolate and interpret given the Euler–Lagrange equations of motion, which look like:

$$\frac{d}{dt}(m\,\dot{x}) = 0, \qquad \frac{d}{dt}(m\,\dot{y}) = 0, \qquad \frac{d}{dt}(m\,\dot{z}) = 0. \tag{3.20}$$

Since there is no coordinate dependence in L, the quantities $m\,\dot{x}$, $m\,\dot{y}$, and $m\,\dot{z}$ all take on constant values, and we recognize these as the three components of momentum for the particle.

As another example, suppose we are working in spherical coordinates, for which

$$T = \frac{1}{2} m \left(\dot{r}^2 + r^2\,\dot{\theta}^2 + r^2 \sin^2\theta\,\dot{\phi}^2\right) \tag{3.21}$$

and we have a potential that is spherically symmetric, depending only on r: $U(r)$. Then think of the equation of motion for ϕ, a variable that does not appear in the Lagrangian at all:

$$\frac{d}{dt}\left(m\,r^2 \sin^2\theta\,\dot{\phi}\right) = 0 \tag{3.22}$$

which tells us that $m\,r^2 \sin^2\theta\,\dot{\phi}$ is a constant of the motion (an angular momentum, by the looks of it).

These are cases of "ignorable" coordinates – if a Lagrangian does not depend explicitly on one of the coordinates, call it q_c, then the associated "momentum" $\frac{\partial L}{\partial \dot{q}_c}$ is conserved (meaning constant along the trajectory). That's Noether's theorem in this context, and the symmetry is $q_c \rightarrow q_c + A$ for constant A (a coordinate shift).

3.2 Fields

Can we formulate an action for fields, as we did for particles? The advantages to such a formulation are the same for fields as for particles: compactness, structure-revelation, and access to conservation laws via Noether's theorem. We'll start with a field equation, the field version of an equation of motion, and develop an action that returns that field equation when extremized.

We know several field equations by now. The simplest one governs a scalar field ϕ in vacuum:

$$\Box\phi \equiv -\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} + \nabla^2\phi = 0, \tag{3.23}$$

just the wave equation with fundamental speed c. This field equation is analogous to Newton's second law – it tells us how to find ϕ for a source-free region of space-time (with boundary conditions specified, ensuring unique solution). The details of the coordinate system are hidden inside ∇^2.

Is there an underlying action minimization that gives this field equation? Define the following action (let $x^0 \equiv c\,t$ be the temporal coordinate, with dimension of length, and $d\tau$ is the spatial volume element, not proper time):

$$S[\phi] = \frac{1}{2}\int\left[-\left(\frac{\partial\phi}{\partial x^0}\right)^2 + \nabla\phi\cdot\nabla\phi\right]\underbrace{dx^0\,d\tau}_{\equiv d^4x}, \tag{3.24}$$

where we integrate in all four dimensions over some domain with a boundary (could be at infinity). If we write

$$\frac{\partial\phi}{\partial x^\mu} \doteq \begin{pmatrix} \frac{\partial\phi}{\partial x^0} \\ \frac{\partial\phi}{\partial x} \\ \frac{\partial\phi}{\partial y} \\ \frac{\partial\phi}{\partial z} \end{pmatrix} \equiv \partial_\mu\phi \equiv \phi_{,\mu} \tag{3.25}$$

working in Cartesian coordinates, then we can express the action neatly as

$$S[\phi] = \frac{1}{2}\int \partial_\mu\phi\,\eta^{\mu\nu}\,\partial_\nu\phi\,d^4x, \tag{3.26}$$

where $\eta^{\mu\nu}$ is the Minkowski metric,

$$\eta^{\mu\nu} \doteq \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.27}$$

in Cartesian coordinates. Notice that (3.26) is just the analogue of the free ($U = 0$) particle action (3.4) with t-derivatives generalized to coordinate derivatives, and the dot product generalized to the space-time inner product.

Now suppose we do the same thing as before: introduce a perturbation to the field that vanishes on the boundaries of the domain (again, the observables live on the boundary, and we don't want our additional arbitrary function to spoil those). Take arbitrary χ (a function of position and time) and expand the action:

$$\begin{aligned} S[\phi+\chi] &= \frac{1}{2}\int \partial_\mu(\phi+\chi)\;\eta^{\mu\nu}\;\partial_\nu(\phi+\chi)\;d^4x \\ &\approx \frac{1}{2}\int \partial_\mu\phi\,\eta^{\mu\nu}\,\partial_\nu\phi\,d^4x + \underbrace{\int \partial_\mu\phi\,\eta^{\mu\nu}\partial_\nu\chi\,d^4x}_{\equiv\delta S}. \end{aligned} \tag{3.28}$$

The first term is just the original action value, and it is δS that we will set equal to zero for all χ to establish a minimum. As in the particle case, the issue is that while χ is arbitrary, $\partial_\nu \chi$ is not. We want to write δS in terms of χ, and that means we have to "flip" a derivative from χ back onto ϕ.

The divergence theorem holds for our four-dimensional integrals just as it does for three-dimensional ones:

$$\int_\Omega \partial_\mu A^\mu d^4x = \oint_{\partial\Omega} A^\mu d^3a_\mu \tag{3.29}$$

where Ω is the domain of the integration with boundary $\partial\Omega$ and d^3a_μ is the surface "area" element.

Keeping in mind that χ vanishes on the boundary of the domain,

$$\int \partial_\nu \left(\partial_\mu\phi\, \eta^{\mu\nu}\, \chi\right) d^4x = \oint \left(\partial_\mu\phi\, \eta^{\mu\nu}\, \chi\right) d^3a_\nu = 0, \tag{3.30}$$

but using the product rule, we also have

$$\int \partial_\nu \left(\partial_\mu\phi\, \eta^{\mu\nu}\, \chi\right) d^4x = \int \chi \left(\partial_\nu\partial_\mu\phi\, \eta^{\mu\nu}\right) d^4x + \int \partial_\mu\phi \left(\eta^{\mu\nu}\, \partial_\nu\chi\right) d^4x. \tag{3.31}$$

Since the whole thing must be zero by (3.30), the two terms on the right are equal and opposite, and the second term is the one appearing in δS from (3.28). Substituting the (negative of the) first term there, we have the desired dependence on χ alone:

$$\delta S = -\int \left(\eta^{\mu\nu}\, \partial_\mu\partial_\nu\phi\right) \chi\, d^4x, \tag{3.32}$$

and for this to be zero for all χ, so that ϕ is really minimizing, we must have:

$$-\,\eta^{\mu\nu}\, \partial_\mu\partial_\nu\phi = 0 \tag{3.33}$$

which is precisely $\Box\phi = 0$, the field equation that was our original target.

The more general setting (although still in Cartesian coordinates) is:

$$S[\phi] = \int \mathcal{L}(\phi, \partial_\mu\phi)\, d^4x, \tag{3.34}$$

where $\mathcal{L}$ is a "Lagrangian" that takes, in theory, ϕ and $\partial_\mu\phi$ as arguments. The variation in this case yields the analogue of the Euler–Lagrange equation from above:

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial\left(\partial_\mu\phi\right)}\right) = 0, \tag{3.35}$$

which can be compared with (3.8).

Problem 3.8 Show that for the general action in (3.34), the minimizing ϕ is indeed given by solving (3.35): introduce arbitrary χ (vanishing on the boundaries) in this general setting, expand the action, etc. (you are generating a field-theoretic version of (3.5)).

Problem 3.9 What is the field equation for a Lagrangian that depends on ϕ, $\partial_\mu\phi$, and $\partial_\mu\partial_\nu\phi$: $\mathcal{L}(\phi, \partial_\mu\phi, \partial_\alpha\partial_\beta\phi)$? (Assume ϕ and $\partial\phi$ vanish on the boundary of the domain.) Use your result to find the field equation for $\mathcal{L} = \frac{1}{2}\,\phi_{,\alpha}\,\phi_{,}^{\;\alpha} + \frac{1}{2}\,\kappa\,\phi_{,\alpha\beta}\,\phi_{,}^{\;\alpha\beta}$ (κ is a constant used to set units).

Problem 3.10 What term could be added to $\mathcal{L} = \frac{1}{2}\,\partial_\mu\phi\,\eta^{\mu\nu}\,\partial_\nu\phi$ so that the field equation reads: $\Box\phi = -s$ where $s(t, x, y, z)$ is an arbitrary source function?

Problem 3.11 What term would you add to $\mathcal{L} = \frac{1}{2}\,\partial_\mu\phi\,\eta^{\mu\nu}\,\partial_\nu\phi$ to recover the field equation $\Box\phi = \mu^2\,\phi$ (which we've seen a few times now)?

Problem 3.12 For a pair of scalar fields, ψ and χ, we have the action:

$$S[\psi, \chi] = \int \left(\frac{1}{2}\,\psi_{,\mu}\,\eta^{\mu\nu}\,\psi_{,\nu} + \frac{1}{2}\,\chi_{,\mu}\,\eta^{\mu\nu}\,\chi_{,\nu} \right) d^4x \tag{3.36}$$

(working in Cartesian coordinates).

a. Write the action in terms of $\phi = \chi + i\,\psi$ and $\phi^* = \chi - i\,\psi$, allowing us to view a single complex field ϕ as a pair of independent fields.

b. Taking the integrand of your expression to be the Lagrangian, $\mathcal{L}(\phi_{,\mu}, \phi^*_{,\mu})$, find the field equations for ϕ and ϕ^*. Show that they are linear combinations of the field equations for χ and ψ (as is to be expected).

Problem 3.13 You can always add a total divergence to the action without changing the field equation. Show that this is the case by taking the Lagrangian $\mathcal{L} = \partial_\mu\,F^\mu(\phi)$ where $F^\mu(\phi)$ is an arbitrary function of ϕ and show that variation produces $0 = 0$ as the field equation.

Problem 3.14 Lagrangians provide a short-hand for the physical content of a theory's field equations and their solutions. Start with the following field action:

$$S[\phi] = -K^2 \int \sqrt{1 - \frac{\phi_{,\mu}\,\phi_{,\nu}\,\eta^{\mu\nu}}{K^2}}\, d^4x \tag{3.37}$$

where K is a constant.

a. Find the field equation governing ϕ.

b. Treating the field ϕ as an electrostatic potential, write the field equation in terms of $\mathbf{E}(x, y, z) = -\nabla\phi(x, y, z)$, assuming no temporal dependence.

c. Find the spherically symmetric point source solution. Your field equation is for vacuum, but you can insert the source $\frac{\rho}{\epsilon_0}$ on the right-hand side of your field equation by hand. Use a point charge Q at the origin for ρ (hint: take your field equation from part (b) and use the divergence theorem to generate the analogue of Gauss's law for this new electric field, then solve that). From your solution, what is an interpretation for K, the lone constant in the theory?

Problem 3.15 Suppose you have a Lagrangian that depends on a field Ψ and its complex conjugate Ψ^* and their temporal and spatial derivatives, i.e.,

$$S[\Psi] = \int \mathcal{L}(\Psi, \Psi^*, \dot{\Psi}, \dot{\Psi}^*, \Psi_{,j}, \Psi^*_{,j})\, d^4x, \tag{3.38}$$

where dots denote time derivatives, and $\Psi_{,j} \equiv \frac{\partial \Psi}{\partial x^j}$ for $j = 1, 2, 3$, the spatial derivatives.

a. Show that the field equations you get are given by:

$$-\frac{\partial}{\partial t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\Psi}^*}\right) - \frac{\partial}{\partial x^j}\left(\frac{\partial \mathcal{L}}{\partial \Psi^*_j}\right) + \frac{\partial \mathcal{L}}{\partial \Psi^*} = 0 \tag{3.39}$$

and the complex conjugate of this equation (varying with respect to Ψ instead of Ψ^*).

b. Take

$$\mathcal{L} = \frac{\hbar^2}{2m}\nabla\Psi^* \cdot \nabla\Psi + V\,\Psi^*\,\Psi - \frac{i\hbar}{2}\left(\Psi^*\,\dot{\Psi} - \Psi\,\dot{\Psi}^*\right), \tag{3.40}$$

with constants $\hbar$ and m, and real function V. Run it through your field equation (either one) and see what you get.

3.3 Noether's Theorem and Conservation

The action is a scalar quantity. That ensures that the field equations are generally covariant (meaning that they "look" the same in all coordinate systems). So coordinate transformations should leave the value of the action unchanged. A scalar function does not respond to coordinate transformation. For a scalar quantity $Q(x)$ and a transformation to new coordinates $\bar{x}$, we have $\bar{Q}(\bar{x}) = Q(\bar{x})$; i.e., the function is written in terms of the new coordinates. As an example, suppose we had, in two dimensions, $Q(x, y) = \sqrt{x^2 + y^2}$. If we switch to polar coordinates, $Q(s, \phi) = s$. The scalar nature of the action is the "symmetry" here, and in this section, we'll work out the associated conserved quantity, an expression of Noether's theorem.

In general, the integrand of the action is $\mathcal{L}(\phi, \partial_\mu\phi, \eta_{\mu\nu})\,d^4x$ (including the volume element now) where we have displayed the dependence of $\mathcal{L}$ on the metric $\eta_{\mu\nu}$ explicitly. For an arbitrary coordinate transformation, taking $x^\mu \rightarrow \bar{x}^\mu$, the action remains the same so that:

$$\mathcal{L}(\phi, \partial_\mu\phi, \eta_{\mu\nu})\,d^4x = \bar{\mathcal{L}}(\bar{\phi}, \bar{\partial}_\mu\bar{\phi}, \bar{\eta}_{\mu\nu})\,d^4\bar{x}. \tag{3.41}$$

Our goal now is to work out the implications of (3.41). A change of coordinates induces changes in the field ϕ, its derivative $\phi_{,\mu}$, and the metric $\eta_{\mu\nu}$. Those changes in turn generate a change δS in the action. We want $\delta S = 0$, since the action is a scalar. The requirement $\delta S = 0$, which is a statement of the coordinate invariance of the action, will define a set of conserved quantities.

3.3.1 Infinitesimal Transformation

Suppose we have a coordinate transformation $\bar{x}^\mu = x^\mu + \epsilon f^\mu$ where f^μ is a function of x^μ, and ϵ is small. There will be some linear (in ϵ) change in ϕ:

$$\bar{\phi}(\bar{x}) = \phi(\bar{x}) = \phi(x + \epsilon f) \approx \phi(x) + \epsilon\,\frac{\partial\phi}{\partial x^\mu}f^\mu, \tag{3.42}$$

which we can write as: $\phi(\bar{x}) = \phi(x) + \delta\phi(x)$, i.e., ϕ plus a small perturbation. Similarly,

$$\bar{\phi}_{,\mu}(\bar{x}) = \phi_{,\mu}(x) + \delta\phi_{,\mu}(x). \tag{3.43}$$

We also expect the metric to change, and we'll assume it retains a linearized form as well: $\bar{\eta}_{\mu\nu} \approx \eta_{\mu\nu} + \delta\eta_{\mu\nu}$ for small $\delta\eta_{\mu\nu}$. The transformation of the tensorial $\delta\phi_{,\mu}$ and $\eta_{\mu\nu}$ are more complicated than for the scalar case, so that $\delta\phi_{,\mu}$ and $\delta\eta_{\mu\nu}$ require more work to express in terms of f^μ than the $\delta\phi$ from (3.42).

With these changes to the field and metric, the action is[3]

$$\begin{aligned} S[\phi + \delta\phi] &= \int \mathcal{L}(\phi + \delta\phi, \phi_{,\alpha} + \delta\phi_{,\alpha}, \eta_{\mu\nu} + \delta\eta_{\mu\nu})\, d^4x \\ &\approx \underbrace{\int \mathcal{L}(\phi, \phi_{,\alpha}, \eta_{\mu\nu})\, d^4x}_{=S[\phi]} + \underbrace{\int \left[\frac{\partial \mathcal{L}}{\partial \phi}\,\delta\phi + \frac{\partial \mathcal{L}}{\partial \phi_{,\alpha}}\,\delta\phi_{,\alpha}\right] d^4x}_{=\delta S_\phi} \\ &\quad + \int \frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}}\,\delta\eta_{\mu\nu}\, d^4x. \end{aligned} \tag{3.44}$$

The first term is just the original action, and the fact that S is a scalar means that everything but this term should vanish at linear order. The second term is δS_ϕ, the response of the action to the variation $\phi \to \phi + \delta\phi$. But if ϕ satisfies its field equation, this term is zero (the extremum of the action holds for all $\delta\phi$ including this special one coming from coordinate transformation). That leaves the term on the third line, which must itself be zero. To evaluate the content of this term, we need to express $\delta\eta_{\mu\nu}$ in terms of the coordinate transformation.

How does the metric respond to coordinate transformation? We know that lengths must remain the same in both coordinate systems, so that

$$dx^\alpha\, \eta_{\alpha\beta}\, dx^\beta = d\bar{x}^\mu\, \bar{\eta}_{\mu\nu}\, d\bar{x}^\nu, \tag{3.45}$$

and inserting the linearized form on the right, using $d\bar{x}^\mu = dx^\mu + \epsilon\, \frac{\partial f^\mu}{\partial x^\rho}\, dx^\rho$ while dropping terms quadratic in ϵ, gives

$$\begin{aligned} d\bar{x}^\mu\, \bar{\eta}_{\mu\nu}\, d\bar{x}^\nu &= \left(dx^\mu + \epsilon\, \frac{\partial f^\mu}{\partial x^\rho}\, dx^\rho\right)(\eta_{\mu\nu} + \delta\eta_{\mu\nu})\left(dx^\nu + \epsilon\, \frac{\partial f^\nu}{\partial x^\sigma}\, dx^\sigma\right) \\ &= dx^\mu\, \eta_{\mu\nu}\, dx^\nu + dx^\mu\, \delta\eta_{\mu\nu}\, dx^\nu + \epsilon\, \eta_{\mu\nu}\left(dx^\mu\, \frac{\partial f^\nu}{\partial x^\sigma}\, dx^\sigma + dx^\nu\, \frac{\partial f^\mu}{\partial x^\rho}\, dx^\rho\right). \end{aligned} \tag{3.46}$$

Putting this on the right in (3.45), the first term cancels, and we get (relabeling closed indices):

$$dx^\alpha\, \delta\eta_{\alpha\beta}\, dx^\beta = -2\,\epsilon\, dx^\alpha\, \frac{\partial f_\alpha}{\partial x^\beta}\, dx^\beta \longrightarrow \delta\eta_{\mu\nu} = -2\,\epsilon\, \frac{\partial f_\mu}{\partial x^\nu}, \tag{3.47}$$

where this expression for $\delta\eta_{\mu\nu}$ suffices for our immediate purposes, although we should use the symmetric (in $\mu \leftrightarrow \nu$) form in general.

[3] We could write the action in terms of the new coordinates, then $d^4\bar{x}$ picks up a factor of the Jacobian, but we've elected to write the changes in field and metric back in the original coordinates, so we would immediately transform back to d^4x, obtaining the expression on the right in (3.44).

Going back to (3.44) with this $\delta\eta_{\mu\nu}$,

$$\int \frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}} \delta\eta_{\mu\nu} \, d^4x = \int \frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}} \left(-2\,\epsilon \frac{\partial f_\mu}{\partial x^\nu}\right) d^4x = -\epsilon \int \left[\partial_\nu \left(-2 \frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}}\right)\right] f_\mu \, d^4x, \tag{3.48}$$

where we have used integration by parts to flip the derivative in the last equality. Demanding that this integral be zero for all coordinate choices f^μ (and equally independent $f_\nu \equiv \eta_{\mu\nu} f^\mu$) means

$$\partial_\nu \mathcal{T}^{\mu\nu} = 0 \text{ for } \mathcal{T}^{\mu\nu} \equiv \left(-2 \frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}}\right). \tag{3.49}$$

We have a set of *four* conservation laws, one for each value of $\mu = 0, 1, 2, 3$. We will be probing the physical implications of these laws for the field theories of interest to us. First, though, stop and appreciate the symmetry–conservation law connection promised by Noether. The symmetry in this case is the coordinate choice. The action is independent of coordinate choice, since it is a scalar, and that automatically implies conservation of, in this case, four quantities (because of the four dimensions of space-time).

3.3.2 Other Isometries

For other "symmetries" of the action (and here we mean transformations that leave the action unchanged), there is a framework for extracting the conserved quantities and associated currents. Working with a scalar (complex or not) action to show the pattern, suppose $\phi \to \phi + \delta\phi$ leaves the action unchanged for some (specific) $\delta\phi$, so that

$$\begin{aligned} S[\phi + \delta\phi] &= \int \mathcal{L}(\phi + \delta\phi, \partial(\phi + \delta\phi))\, d^4x \\ &= \underbrace{\int \mathcal{L}(\phi, \partial\phi)\, d^4x}_{=S[\phi]} + \underbrace{\int \left(\frac{\partial \mathcal{L}}{\partial \phi} \delta\phi + \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \delta\phi_{,\mu}\right) d^4x}_{\equiv \delta S} \end{aligned} \tag{3.50}$$

has $\delta S = 0$ for this specific $\delta\phi$. We could be imagining an infinitesimal version of a more general transformation of ϕ, and then it makes sense to leave off the higher order terms in the Taylor expansion of the Lagrangian.

For *any* $\delta\phi$, we know that

$$-\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \phi_{,\mu}}\right) + \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{3.51}$$

as that is just the field equation for ϕ (the field equation comes from arbitrary variation of ϕ, and so holds for any specific $\delta\phi$ as well). Using the field equation (3.51), we can write

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \delta\phi\right) = \left(\partial_\mu \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}}\right) \delta\phi + \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \delta\phi_{,\mu} = \frac{\partial \mathcal{L}}{\partial \phi} \delta\phi + \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \delta\phi_{,\mu}. \tag{3.52}$$

The right-hand side is precisely what appears in the δS integrand of (3.50). Since $\delta S = 0$ for these $\delta\phi$, we have a conservation law; let $j^\mu \equiv \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \delta\phi$, then $\partial_\mu j^\mu = 0$ gives $\delta S = 0$. By analogy with charge conservation, the 0 component, j^0, is sometimes referred to as a conserved "charge" with "current" (both densities) $\mathbf{j}$, the spatial elements of j^μ.

Examples

Consider, first, an action that did not explicitly depend on ϕ; the scalar $\mathcal{L} = \frac{1}{2}\,\phi_{,\mu}\,\phi_{,}^{\ \mu}$ is a good example. Then we can take $\phi \rightarrow \phi + \delta\phi$ for $\delta\phi$ any constant; that doesn't change the action. The associated four-vector is:

$$j^\mu = \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}}\,\delta\phi = \phi_{,}^{\ \mu}\,\delta\phi. \tag{3.53}$$

This obviously has $\partial_\mu j^\mu \sim \Box\phi = 0$ by the field equation.

Suppose you have an action that consists of a pair of scalars, combined into a complex scalar as in Problem 3.12 (we'll even add a term), so start with

$$\mathcal{L} = \phi^*_{,\mu}\,\phi_{,}^{\ \mu} + \mu^2\,\phi^*\,\phi. \tag{3.54}$$

This Lagrangian is unchanged under $\phi \rightarrow e^{i\theta}\,\phi$ for constant θ. What is the associated conserved current? First we have to write this transformation in a form where we can identify $\delta\phi$. Since we are working with the Taylor expansion in (3.50), it makes sense to take $\theta \ll 1$, an infinitesimal transformation. Then

$$e^{i\theta}\,\phi \approx (1 + i\theta)\,\phi = \phi + \underbrace{i\,\theta\,\phi}_{\equiv\delta\phi}. \tag{3.55}$$

We automatically transform the conjugate: $\phi^* \rightarrow \phi^* - i\,\theta\,\phi^*$, so that $\delta\phi^* = -i\,\theta\,\phi^*$ as expected. The conserved current is

$$j^\mu = \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}}\,\delta\phi + \frac{\partial \mathcal{L}}{\partial \phi^*_{,\mu}}\,\delta\phi^*, \tag{3.56}$$

where we have expanded the evaluation of j^μ to include both fields. In terms of the infinitesimals, the current is

$$j^\mu = i\,\theta\,\left[\phi_{,}^{*\,\mu}\,\phi - \phi_{,}^{\ \mu}\,\phi^*\right], \tag{3.57}$$

and of course, we are free to get rid of the constants out front.

Problem 3.16 For the Lagrangian governing the Schrödinger equation (3.40), find the conserved density and current (the 0 and spatial components of the conserved four-vector) associated with $\psi \rightarrow e^{i\theta}\,\psi$. If you set the overall constant so that the 0 component is $j^0 = c\,\psi^*\,\psi$, what does $\mathbf{j}$ (the spatial components of j^μ) look like?

3.4 Stress Tensor

There are a few details that need pinning down before we're ready to define the stress tensor, the conserved object associated with coordinate-transformation invariance of the action. The first problem is with the action integrand: $\mathcal{L}\,d^4x$. The product of $\mathcal{L}$ and d^4x is a scalar, but neither component is. To see this, think of the volume element d^4x. When we change coordinates, this transforms: $d^4\bar{x} = |\frac{d\bar{x}}{dx}|\,d^4x$. The factor of the (determinant of the) Jacobian of the transformation spoils the scalar form of d^4x (which otherwise looks like a scalar quantity with no dangling indices). As a simple example, when you go from two-dimensional Cartesian coordinates: $d^2x \equiv dxdy$ to polar, you do not just replace $dxdy$ with $dsd\phi$ (dimensionality issues aside) in integrals.

If d^4x is not a scalar (because it does not transform like one), then neither is $\mathcal{L}$. The fix here is to introduce a factor of the metric determinant. For a metric $g_{\mu\nu}$, define $g \equiv \det g_{\mu\nu}$. Then we'll put a factor of $\sqrt{g}$ together with d^4x, to get a scalar. we'll work on a two-dimensional example to see the pattern: for Cartesian coordinates, the metric and its determinant are

$$g_{\mu\nu} \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad g = 1. \tag{3.58}$$

In polar coordinates, we have

$$\bar{g}_{\mu\nu} \doteq \begin{pmatrix} 1 & 0 \\ 0 & s^2 \end{pmatrix}, \qquad \bar{g} = s^2. \tag{3.59}$$

We know, in polar coordinates, that the integration "volume" element is $s\,ds\,d\phi$, and we could write this as $\sqrt{\bar{g}}\,ds\,d\phi = \sqrt{\bar{g}}\,d^2\bar{x}$. Similarly, the "volume" element in Cartesian coordinates is $dx\,dy = \sqrt{g}\,d^2x$, so it is $\sqrt{g}\,d^2x$ that is the scalar – to get the transformed form, all you do is put bars on everything.

Problem 3.17 Show that $\sqrt{g}\,d^3x$ transforms appropriately by evaluating it in both Cartesian and spherical coordinates.

For our four-dimensional space-*time* the metric is $g_{\mu\nu} = \eta_{\mu\nu}$, and we have the minus sign from the temporal component, so the determinant is $\eta = -1$ in Cartesian coordinates. In order to avoid imaginary actions, we use $\sqrt{-\eta}$ in this context. The anatomy of the relativistic field action is, then,

$$S = \int \underbrace{\hat{\mathcal{L}}\sqrt{-\eta}}_{\equiv\mathcal{L}}\, d^4x, \tag{3.60}$$

where $\sqrt{-\eta}\,d^4x$ is a scalar, and $\hat{\mathcal{L}}$ is also a scalar. The combination $\mathcal{L} \equiv \hat{\mathcal{L}}\sqrt{-\eta}$ is known as a "Lagrangian *density*." Up until now, we have worked exclusively in Cartesian coordinates, where $\sqrt{-\eta} = 1$, so the metric determinant hasn't shown up explicitly. In many cases, we'll work in Cartesian coordinates and implicitly set $\sqrt{-\eta} = 1$. The only reason not to do that in this section is that we are probing a coordinate transformation and have to consider the response of the metric to that transformation (i.e., we cannot work exclusively in Cartesian coordinates here, since the coordinates must change).

Going back to (3.49), we have $\mathcal{T}^{\mu\nu} = -2\frac{\partial\mathcal{L}}{\partial\eta_{\mu\nu}}$, but then $\mathcal{T}^{\mu\nu}$ is not a tensor since $\mathcal{L}$ isn't a scalar. Clearly, there is metric dependence both in $\hat{\mathcal{L}}$ and in $\sqrt{-\eta}$. For the latter

$$\frac{\partial\sqrt{-\eta}}{\partial\eta_{\mu\nu}} = \frac{1}{2}\sqrt{-\eta}\,\eta^{\mu\nu}, \tag{3.61}$$

and then we can express $\mathcal{T}^{\mu\nu}$ in terms of the scalar $\hat{\mathcal{L}}$ and the metric determinant piece, $\sqrt{-\eta}$:

$$\mathcal{T}^{\mu\nu} \equiv -2\,\frac{\partial \mathcal{L}}{\partial \eta_{\mu\nu}} = -\sqrt{-\eta}\left[\eta^{\mu\nu}\,\hat{\mathcal{L}} + 2\,\frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}}\right], \tag{3.62}$$

from which we define the stress *tensor*

$$T^{\mu\nu} \equiv \frac{1}{\sqrt{-\eta}}\,\mathcal{T}^{\mu\nu} = -\left[\eta^{\mu\nu}\,\hat{\mathcal{L}} + 2\,\frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}}\right]. \tag{3.63}$$

The advantage of the stress tensor is precisely that it is a tensor. We know how $T^{\mu\nu}$ transforms, and can use it in tensorial equations. The "stress tensor density," $\mathcal{T}^{\mu\nu} = \sqrt{-\eta}\,T^{\mu\nu}$, does not transform as a tensor because $\sqrt{-\eta}$ is not a scalar.

The conservation statement now reads, in Cartesian coordinates,

$$\partial_\mu\, T^{\mu\nu} = 0. \tag{3.64}$$

The stress tensor itself is naturally contravariant, since it comes from the derivative of $\mathcal{L}$ with respect to the covariant $\eta_{\mu\nu}$. The derivative ∂_μ is naturally covariant, as we have seen before in Section 1.10.3. So the conservation statement requires no factors of the metric in its evaluation.

Problem 3.18 There is a nice way to express $\nabla^2\psi$ in any coordinate system:

$$\nabla^2\psi = \frac{1}{\sqrt{g}}\,\partial_j\left[\sqrt{g}\,g^{ji}\,\partial_i\,\psi\right], \tag{3.65}$$

where we are dealing with pure spatial derivatives (i and j go from 1 to 3, covering the spatial coordinates), and $g = \det(g_{ij})$ with g^{ij} the matrix inverse of g_{ij}, the metric. Verify that you get what you expect from this expression in Cartesian, cylindrical, and spherical coordinates.

3.5 Scalar Stress Tensor

We have only one field Lagrangian so far, and we can find the stress tensor for this scalar field Lagrangian. Starting with $\hat{\mathcal{L}} = \frac{1}{2}\,\partial_\alpha\phi\,\eta^{\alpha\beta}\,\partial_\beta\phi$, the derivative of $\hat{\mathcal{L}}$ with respect to the metric is

$$\frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}} = \frac{1}{2}\,\partial_\alpha\phi\,\frac{\partial \eta^{\alpha\beta}}{\partial \eta_{\mu\nu}}\,\partial_\beta\phi. \tag{3.66}$$

The upper and lower forms of the metric are matrix inverses of one another:

$$\eta^{\alpha\sigma}\,\eta_{\sigma\gamma} = \delta^\alpha_\gamma \tag{3.67}$$

and taking the $\eta_{\mu\nu}$ derivative of both sides gives

$$\frac{\partial \eta^{\alpha\sigma}}{\partial \eta_{\mu\nu}}\,\eta_{\sigma\gamma} + \eta^{\alpha\sigma}\,\delta^\mu_\sigma\,\delta^\nu_\gamma = 0. \tag{3.68}$$

Isolating the derivative on the left and contracting with $\eta^{\gamma\beta}$,

$$\frac{\partial \eta^{\alpha\sigma}}{\partial \eta_{\mu\nu}} \delta^{\beta}_{\sigma} = -\eta^{\alpha\sigma} \delta^{\mu}_{\sigma} \delta^{\nu}_{\gamma} \eta^{\gamma\beta}, \tag{3.69}$$

so that

$$\frac{\partial \eta^{\alpha\beta}}{\partial \eta_{\mu\nu}} = -\eta^{\alpha\mu} \eta^{\beta\nu}. \tag{3.70}$$

Going back to the original term of interest:

$$\frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}} = -\frac{1}{2} \partial_\alpha \phi \, \partial_\beta \phi \, \eta^{\alpha\mu} \eta^{\beta\nu}, \tag{3.71}$$

and the stress tensor, from (3.63), is

$$T^{\mu\nu} = \partial_\alpha \phi \, \partial_\beta \phi \, \eta^{\alpha\mu} \eta^{\beta\nu} - \frac{1}{2} \eta^{\mu\nu} \eta^{\alpha\beta} \partial_\alpha \phi \, \partial_\beta \phi. \tag{3.72}$$

What are the individual elements here? Start with the 00 component, working in Cartesian coordinates:

$$\begin{aligned} T^{00} &= \frac{1}{c^2}\left(\frac{\partial \phi}{\partial t}\right)^2 + \frac{1}{2}\left[-\frac{1}{c^2}\left(\frac{\partial \phi}{\partial t}\right)^2 + \nabla\phi \cdot \nabla\phi\right] \\ &= \frac{1}{2}\left[\frac{1}{c^2}\left(\frac{\partial \phi}{\partial t}\right)^2 + \nabla\phi \cdot \nabla\phi\right]. \end{aligned} \tag{3.73}$$

If we take ϕ to be static, then we get back $T^{00} = \frac{1}{2} \nabla\phi \cdot \nabla\phi$, which looks a lot like the energy density of the electrostatic field (written in terms of the electric potential). We call the 00 component of the stress tensor the "energy density." Remember that we have four conservation laws in $\partial_\nu T^{\mu\nu} = 0$, one for each value of μ. If we take $\mu = 0$, we get a conservation statement for T^{00}, and then the T^{0i} components (the spatial ones, for $i = 1$, 2, 3) form the "current density" vector in the conservation law:

$$\partial_\nu T^{0\nu} = 0 = \frac{1}{c}\frac{\partial T^{00}}{\partial t} + \partial_i T^{0i} = 0. \tag{3.74}$$

Evaluating this three-vector,[4]

$$T^{0i} = -\frac{1}{c}\frac{\partial \phi}{\partial t}\frac{\partial \phi}{\partial x_i}, \tag{3.75}$$

and this is a sort of momentum for the field. That momentum is itself conserved by the other three conservation laws (setting $\mu = i \equiv 1, 2, 3$):

$$\frac{1}{c}\frac{\partial T^{0i}}{\partial t} + \partial_j T^{ji} = 0, \tag{3.76}$$

with

$$T^{ji} = \frac{\partial \phi}{\partial x_i}\frac{\partial \phi}{\partial x_j} - \frac{1}{2}\eta^{ij}\left[-\frac{1}{c}^2\left(\frac{\partial \phi}{\partial t}\right)^2 + \nabla\phi \cdot \nabla\phi\right]. \tag{3.77}$$

[4] The derivative with respect to x_i, $\frac{\partial \phi}{\partial x_i}$, is numerically identical to $\frac{\partial \phi}{\partial x^i}$ for Cartesian coordinates.

There is one last quantity of interest we can form from the stress tensor, the trace: $T \equiv T^{\mu\nu}\,\eta_{\mu\nu}$, a scalar. Using the components, we have

$$T = -\partial_\alpha\phi\,\partial_\beta\phi\,\eta^{\alpha\beta} = \frac{1}{c^2}\left(\frac{\partial\phi}{\partial t}\right)^2 - \nabla\phi\cdot\nabla\phi \tag{3.78}$$

(in four dimensions).

Problem 3.19 Under what circumstances is $\phi = \phi_0\, e^{i p_\mu x^\mu}$ (constant p_μ) a solution to the field equations for $\hat{\mathcal{L}} = \frac{1}{2}\,\partial_\alpha\phi\,\eta^{\alpha\beta}\,\partial_\beta\phi$? Evaluate the stress tensor elements for this solution.

Problem 3.20 Show that the stress tensor (3.72) is conserved when ϕ satisfies its field equation (i.e., show that $\partial_\mu T^{\mu\nu} = 0$ when $\Box\phi = 0$, working in Cartesian coordinates).

Problem 3.21 Find the stress tensor for the "massive" Klein–Gordon field with Lagrangian $\hat{\mathcal{L}} = \frac{1}{2}\,\partial_\alpha\phi\,\eta^{\alpha\beta}\,\partial_\beta\phi + \frac{1}{2}\,\mu^2\,\phi^2$. Evaluate the 00 component of the stress tensor, written in terms of temporal and spatial derivatives of ϕ (and ϕ itself). Find the trace of the stress tensor $T \equiv T^{\mu\nu}\,\eta_{\mu\nu}$ in D dimensions (the Minkowski metric, in D dimensions, has a -1 in the 00 spot, with 1s on the diagonal all the way down).

3.6 Electricity and Magnetism

For E&M, the fundamental field is A_μ,[5] and we want to mimic the $\partial_\mu\,\phi\,\eta^{\mu\nu}\,\partial_\nu\phi$ structure of the scalar Lagrangian. There are now two ways to array the indices. Should $\partial_\mu\,A_\nu\,\partial^\mu\,A^\nu$ or $\partial_\mu\,A_\nu\,\partial^\nu\,A^\mu$ appear?[6] The particular combination of interest can be obtained by thinking about natural combinations of "covariant" derivatives from differential geometry.[7]

But we can also reduce the number of relevant combinations by demanding that the action (and hence field equations) governing this *vector* field should be independent of any scalar fields. You could get scalar dependence in a vector field theory through the gradient: $\partial_\nu\phi$ is a vector that depends only on the scalar ϕ. We know how to make scalar actions and find field equations in that setting. What we're after is the new *vector* contribution. Later, when we try to combine scalar and vector fields, we will see that the procedure is to start with a pure scalar action and add a pure vector action, and then we have control over how the two fields couple to one another. If we have implicit scalar contribution in the vector sector, then there is an automatic coupling that we may or may not want.

To see this, take the provisional Lagrangian:

$$\mathcal{L} = \alpha\,\partial_\mu\,A_\nu\,\partial^\mu\,A^\nu + \beta\,\partial_\mu\,A_\nu\,\partial^\nu\,A^\mu \tag{3.79}$$

[5] But remember that it is the upper form that is expressible directly in terms of V/c and $\mathbf{A}$, $A^\mu \doteq \begin{pmatrix} V/c \\ \mathbf{A} \end{pmatrix}$.

[6] There is also $(\partial_\mu\,A^\mu)\,(\partial_\nu\,A^\nu)$, a new term that we omit for convenience, but which could be included.

[7] See, for example, [11].

for arbitrary constants α and β. Now decompose: $A_\mu = \hat{A}_\mu + \phi_{,\mu}$ where $\hat{A}_\mu$ has *no* scalar piece. The Lagrangian can be written in terms of two different fields, ϕ and $\hat{A}_\mu$, with both free and coupled contributions:

$$\begin{aligned}\mathcal{L} = \alpha\, \partial_\mu \hat{A}_\nu\, \partial^\mu \hat{A}^\nu + \beta\, \partial_\mu \hat{A}_\nu\, \partial^\nu \hat{A}^\mu \\ +(\alpha+\beta)\; \phi_{,\mu\nu}\, \phi_{,}^{\;\mu\nu} \\ +\, 2\,(\alpha+\beta)\; \phi_{,\mu\nu}\, \partial^\mu \hat{A}^\nu\end{aligned} \tag{3.80}$$

where the first two lines represent the free field Lagrangians governing $\hat{A}_\mu$ and ϕ separately, and the third line gives the coupling term. Right from the start, then, this Lagrangian refers to both fields, and the portion governing ϕ is already strange in that it involves second derivatives of ϕ. The field equation for the pure scalar piece is $\sim \Box(\Box\phi) = 0$. We can easily satisfy this field equation with a massless Klein–Gordon scalar, so we already know how to get the basic dynamics implied by the scalar portion, and if we keep that (scalar) piece separate, we avoid the enforced cross coupling from the third line of (3.80).

Now we want to see how removing the scalar bit from the action pins down the pure vector Lagrangian unambiguously. We can see that the scalar portion of the field equation from the Lagrangian above would vanish identically for $\alpha = -\beta$. There is, then, a natural combination of derivatives of A_μ to use in building the vector Lagrangian. To see it from the start, using cross-derivative equality; it is clear that $\partial_\mu A_\nu - \partial_\nu A_\mu = 0$ if $A_\nu = \partial_\nu \phi$, so it is the *difference* that we should take as the "derivative" term in the Lagrangian. Let

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{3.81}$$

and we'll take the vector Lagrangian to be

$$\hat{\mathcal{L}} = \frac{1}{4\,\mu_0}\, F_{\alpha\beta}\, F^{\alpha\beta} \equiv \frac{1}{4\,\mu_0}\, F_{\alpha\beta}\, \eta^{\alpha\rho}\, \eta^{\beta\sigma}\, F_{\rho\sigma} \tag{3.82}$$

where the factor out front is just to set the units correctly.

First, we'll see what this Lagrangian has as its field equations. The analog of (3.35) for this vector field is

$$\frac{\partial \hat{\mathcal{L}}}{\partial A_\nu} - \partial_\mu \left(\frac{\partial \hat{\mathcal{L}}}{\partial A_{\nu,\mu}} \right) = 0, \tag{3.83}$$

where we have four field equations, one for each of the values of $\nu = 0 \to 3$. The first term in (3.83) is zero, since $\hat{\mathcal{L}}$ does not depend on the field itself. To evaluate the derivatives of the Lagrangian with respect to the $A_{\nu,\mu}$, we use the definition of $F_{\alpha\beta}$,

$$\frac{\partial F_{\alpha\beta}}{\partial A_{\nu,\mu}} = \frac{\partial}{\partial A_{\nu,\mu}} \left(A_{\beta,\alpha} - A_{\alpha,\beta} \right) = \delta^\nu_\beta\, \delta^\mu_\alpha - \delta^\nu_\alpha\, \delta^\mu_\beta, \tag{3.84}$$

and using this in (3.83), we have the field equation

$$-\frac{1}{\mu_0}\, \partial_\mu \left(F^{\mu\nu} \right) = 0 = -\frac{1}{\mu_0} \left(\Box A^\nu - \partial^\nu \partial_\mu A^\mu \right). \tag{3.85}$$

We chose the combination of derivatives of A_μ appearing in $\mathcal{L}$, namely, the $F_{\mu\nu}\, F^{\mu\nu}$, so that the field equations would make no reference to scalars (that could creep into A_μ

through the gradient: $A_\mu \sim \phi_{,\mu}$). But this means that we can add precisely the gradient of a scalar to A_μ without changing anything. This choice of scalar gradient is the gauge freedom of E&M, and we can use it to set the divergence of A^μ to zero. Take $\bar{A}_\mu = A_\mu + \phi_{,\mu}$ with $\bar{F}_{\mu\nu} = \bar{A}_{\nu,\mu} - \bar{A}_{\mu,\nu}$. Rewriting in terms of the unbarred four-potential, the second derivatives of ϕ cancel each other by cross-derivative equality, so $\bar{F}_{\mu\nu} = F_{\mu\nu}$, and both the action and field equation $-\partial_\mu F^{\mu\nu} = 0$ are unchanged.

But with $\bar{A}_\mu = A_\mu + \phi_{,\mu}$, we can choose ϕ so that $\partial_\mu \bar{A}^\mu = 0$ by taking (using our Green's function, $G(\mathbf{r}, \mathbf{r}', t, t')$)

$$\Box\phi = -\partial_\mu A^\mu \longrightarrow \phi(\mathbf{r}, t) = \int \left(\frac{\partial A^\mu(\mathbf{r}', t')}{\partial x'^\mu} \right) G(\mathbf{r}, \mathbf{r}', t, t')\, dt'\, d\tau'. \tag{3.86}$$

We can then assume that the four-potential has $\partial_\mu A^\mu = 0$ (Lorenz gauge again) from the start, and then the field equation (3.85) becomes $-\frac{1}{\mu_0}\Box A^\nu = 0$.

Problem 3.22 For

$$A^\mu \doteq \begin{pmatrix} V/c \\ A^x \\ A^y \\ A^z \end{pmatrix}, \tag{3.87}$$

write the entries of the "field strength" tensor $F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}$ and $F^{\mu\nu}$ in terms of the electric and magnetic field components (remember that it is ∂_μ that has $\partial_0 = \frac{1}{c}\frac{\partial}{\partial t}$ with $\partial_i = \frac{\partial}{\partial x^i}$). What is the scalar $F^{\mu\nu} F_{\mu\nu}$ in terms of $\mathbf{E}$ and $\mathbf{B}$?

Problem 3.23 A second-rank tensor like $F^{\mu\nu}$ transforms with two factors of the Lorentz transformation: $\bar{F}^{\mu\nu} = L^\mu_{\ \alpha} L^\nu_{\ \beta} F^{\alpha\beta}$. For an infinite line of charge lying along the $\hat{\mathbf{x}}$-axis at rest with constant λ_0 (charge per unit length), write the electric and magnetic fields a distance d from the line (and in the x–z plane, let's say) using Cartesian basis vectors. Construct $F^{\mu\nu}$ and use the transformation properties to find the electric and magnetic fields for an infinite line of charge moving with constant velocity $\mathbf{v} = v$. Are they what you expect? Compute $F^{\mu\nu} F_{\mu\nu}$ and $\bar{F}^{\mu\nu} \bar{F}_{\mu\nu}$. Are they equal (as they must be)?

Problem 3.24 Take the scalar portion of the Lagrangian in (3.80): $\mathcal{L} = \alpha\, \phi_{,\mu\nu}\, \phi_{,}^{\ \mu\nu}$ with constant α, and vary to get the field equation.

Problem 3.25 Our vector field equation, with source in place, reads:

$$\partial^\rho\, \partial_\sigma\, A^\sigma - \Box A^\rho = \mu_0\, J^\rho, \tag{3.88}$$

for

$$A^\mu \doteq \begin{pmatrix} V/c \\ A^x \\ A^y \\ A^z \end{pmatrix}, \qquad J^\mu \doteq \begin{pmatrix} \rho\, c \\ J^x \\ J^y \\ J^z \end{pmatrix}. \tag{3.89}$$

In Lorenz gauge, $\partial_\sigma A^\sigma = 0$, and we recover the familiar $\Box A^\rho = -\mu_0\, J^\rho$. But there are other gauge choices. What do the four field equations (break them up into the 0 component and the spatial vector components) above look like for the gauge choice $\nabla \cdot \mathbf{A} = 0$? This choice is called "Coulomb gauge." Do your field equations look Lorentz covariant?

Problem 3.26 Find the Coulomb-gauge electric potential, V, for a point charge moving with constant velocity.

3.6.1 Stress Tensor

To get the stress tensor for our vector field theory, we again need the two pieces of (3.63). The Lagrangian is already in place, so we just need to evaluate its derivative:

$$\begin{aligned} \frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}} &= \frac{1}{4\,\mu_0} F_{\alpha\beta}\, F_{\rho\sigma} \left(-\eta^{\alpha\mu}\, \eta^{\rho\nu}\, \eta^{\beta\sigma} - \eta^{\alpha\rho}\, \eta^{\beta\mu}\, \eta^{\sigma\nu}\right) \\ &= -\frac{1}{2\,\mu_0} F^{\mu\sigma}\, F^{\nu}_{\ \sigma}. \end{aligned} \tag{3.90}$$

The stress tensor is

$$T^{\mu\nu} = \frac{1}{\mu_0}\left(F^{\mu\sigma}\, F^{\nu}_{\ \sigma} - \frac{1}{4}\, \eta^{\mu\nu}\, F^{\alpha\beta}\, F_{\alpha\beta}\right). \tag{3.91}$$

It is (relatively) easy to verify that $T^{00} = \frac{1}{2}\,\epsilon_0\,(E^2 + c^2\, B^2) \equiv u$, the energy density associated with the fields. But we also get $T^{0i} = \epsilon_0\, c\, (\mathbf{E} \times \mathbf{B})^i \equiv \frac{1}{c}\,\mathbf{S}^i$ (where $\mathbf{S}$ is the Poynting vector from Section 2.5.1), and

$$\partial_\mu T^{0\mu} = \frac{1}{c}\,\frac{\partial T^{00}}{\partial t} + \partial_i\, T^{0i} = 0 \tag{3.92}$$

recovers the relation we already know:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{S}. \tag{3.93}$$

This is energy conservation for the electromagnetic fields.

How about the other three components of the conservation law?

$$\partial_0\, T^{0i} + \partial_j\, T^{ji} = 0, \tag{3.94}$$

which we can write as

$$\frac{1}{c^2}\,\frac{\partial S^i}{\partial t} = -\partial_j\, T^{ji} \tag{3.95}$$

and

$$T^{ij} = \epsilon_0 \left(\frac{1}{2}\,\delta^{ij}\, E^2 - E^i\, E^j\right) + \frac{1}{\mu_0}\left(\frac{1}{2}\,\delta^{ij}\, B^2 - B^i\, B^j\right). \tag{3.96}$$

The statement in (3.95) is field momentum conservation, three equations, one for each spatial component of the field momentum. The "current" here, T^{ij}, is the "Maxwell stress tensor," a momentum flux density.

Problem 3.27 An infinite wire lying along the $\hat{\mathbf{z}}$-axis moves with constant velocity $\mathbf{v} = v\,\hat{\mathbf{z}}$. When moving, it has constant charge-per-unit-length λ. Find the Maxwell stress tensor components T^{ij} (write out the terms in Cartesian coordinates).

Problem 3.28 What is the trace of the electromagnetic stress tensor (in four dimensions)? This result is a defining property of E&M and indicates why it would be difficult to have a scalar theory of gravity (?!).

Problem 3.29 From the stress tensor for E&M, show that $T^{0i} = \frac{1}{c} S^i$ where $i = 1, 2, 3$ (the spatial components) and $\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}$ is the Poynting vector (its components are labeled S^i).

Problem 3.30 An infinite line of charge with constant λ (charge-per-unit-length) lying along the $\hat{\mathbf{z}}$-axis is pulled along the $\hat{\mathbf{z}}$-axis with constant speed v. Find $\mathbf{E}$ and $\mathbf{B}$ and calculate the energy density and Poynting vector. The Poynting vector is non-zero, and yet this configuration is equivalent to a stationary one. Pick a domain for which you can evaluate the left- and right-hand sides of the integrated energy conservation statement:

$$\frac{d}{dt} \int_{\Omega} u \, d\tau = - \oint_{\partial\Omega} \mathbf{S} \cdot d\mathbf{a}, \tag{3.97}$$

to establish that both the left and right are zero (no energy flow).

Problem 3.31 What is the field equation analogous to (3.83) governing a symmetric, second-rank tensor field $h_{\mu\nu}$, starting from a Lagrangian $\mathcal{L}(h_{\mu\nu}, \partial_\alpha h_{\mu\nu})$?

Problem 3.32

a. For $F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}$, show that

$$F_{\mu\nu,\alpha} + F_{\alpha\mu,\nu} + F_{\nu\alpha,\mu} = 0. \tag{3.98}$$

b. Verify that the stress tensor for E&M is conserved, $\partial_\mu T^{\mu\nu} = 0$, when $\partial_\sigma F^{\sigma\rho} = 0$ (i.e., when the field satisfies its vacuum field equation).

3.7 Sources

So far, our field Lagrangians produce only "vacuum" field equations. For the scalar field ϕ, we get $\Box\phi = 0$. For the vector field A_μ, we get back $\Box A_\mu = 0$; where are all the sources? Provided sources, like J^μ (a ρ and $\mathbf{J}$ for E&M), arise from terms that are linear (in the field) in a Lagrangian. As an example, take the scalar Lagrangian, with some given source function s:

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \, \eta^{\mu\nu} \, \partial_\nu \phi + \alpha \, s \, \phi, \tag{3.99}$$

where s is some specified function of position (and time) and α is a constant (to set units). Our field equation (3.35) gives

$$\alpha \, s - \Box\phi = 0 \longrightarrow \Box\phi = \alpha \, s \tag{3.100}$$

and we now have a sourced wave equation. We can add a similar term to the E&M Lagrangian, $\alpha\, J^\mu A_\mu$ (for given four-current J^μ), to get sources there:

$$\mathcal{L} = \frac{1}{4\,\mu_0} F^{\mu\nu} F_{\mu\nu} + \alpha\, J^\mu A_\mu, \tag{3.101}$$

with field equations

$$-\frac{1}{\mu_0} \partial_\mu F^{\mu\nu} + \alpha\, J^\nu = 0 \longrightarrow \partial_\mu F^{\mu\nu} = \mu_0\, \alpha\, J^\nu, \tag{3.102}$$

or, in Lorenz gauge:

$$\Box A^\mu = \alpha\, \mu_0\, J^\mu, \tag{3.103}$$

which suggests $\alpha = -1$.

Notice that in the form (3.102), it is clear that we must have $\partial_\nu J^\nu = 0$, or what we would normally call "charge conservation." If you take the ∂_ν derivative of both sides of (3.102), you get $\partial_\nu \partial_\mu F^{\mu\nu} = 0$ since that's a symmetric–anti-symmetric contraction[8] and then $\partial_\nu J^\nu = 0$ follows on the right.

That conservation statement, $\partial_\nu J^\nu = 0$, should be related to a symmetry (of an action), by Noether's theorem. Indeed it is, and you can guess it from the start. It's the anti-symmetric nature of $F_{\mu\nu}$ that makes the second derivative zero, and that anti-symmetry is a statement of the gauge-invariance of the theory. Going back to the E&M Lagrangian with source in place

$$S[A_\mu] = \int \left[\frac{1}{4\,\mu_0} F_{\alpha\beta} F^{\alpha\beta} - A_\mu J^\mu\right] d^4x \tag{3.104}$$

(working in Cartesian coordinates so that $\sqrt{-\eta} = 1$), take $A_\mu \rightarrow A_\mu + \phi_{,\mu}$ for arbitrary ϕ (just a special type of arbitrary variation for A_μ). We know that $F_{\alpha\beta}$ is unchanged, and so

$$S[A_\mu + \phi_{,\mu}] = \int \left[\frac{1}{4\,\mu_0} F_{\alpha\beta} F^{\alpha\beta} - A_\mu J^\mu\right] d^4x - \int \phi_{,\mu} J^\mu\, d^4x, \tag{3.105}$$

where the first term is $S[A_\mu]$. If we want the action to be unchanged by the choice of ϕ, we must have

$$\int \partial_\mu \phi\, J^\mu\, d^4x = 0 \tag{3.106}$$

for all ϕ. As usual, we could interpret this as the requirement $J^\mu = 0$, but that is overly restrictive. Instead, use integration by parts to write

$$\int \partial_\mu \phi\, J^\mu\, d^4x = -\int \left(\partial_\mu J^\mu\right) \phi\, d^4x = 0 \tag{3.107}$$

for all ϕ, giving $\partial_\mu J^\mu = 0$ as expected. Charge conservation comes from the gauge invariance of the action.

Problem 3.33 Show that $A^\mu \doteq (V/c, \mathbf{A})$ is indeed a four-vector if J^μ is (hint: work from the equation relating the two).

[8] The situation is reminiscent of the divergence of $\nabla \times \mathbf{B} = \mu_0\, \mathbf{J}$ in static E&M, where you get $\nabla \cdot (\nabla \times \mathbf{B}) = 0$ as a matter of differential geometry.

The problem with providing an s or J^μ as a source is that there is no guarantee that those objects are scalar (in the first case) or vector (in the second). We can't allow sources to be provided by some end-user, since they may or may not have the appropriate character. In the case of J^μ for E&M, we further require conservation, $\partial_\mu J^\mu = 0$, and again, unless you are handed a pre-made conservative current density, you have no way to ensure that conservation holds. Instead of providing sources, we must couple fields to *other fields*, where all fields are incorporated into a master Lagrangian. The field equations ensure that all the fields have appropriate character, and we can structure the Lagrangian to enforce conservation laws.

3.7.1 Example: Single Scalar Field

The simplest source for E&M is a scalar field. Take $J^\mu = f\,\phi_{,}{}^{\mu}$ (for a constant f that sets the units and provides the notion of "charge" for the field ϕ). We know that this is a four-vector, and the field equation for ϕ: $\partial_\mu\,\phi_{,}{}^{\mu} = 0$ will give conservation (this pre-supposes that the field equation for ϕ is unchanged by the coupling with E&M, an assumption that is untrue!). Start with

$$S[\phi, A_\mu] = \int \left[\frac{1}{4\,\mu_0} F_{\alpha\beta}\,F^{\alpha\beta} + \frac{1}{2}\,\phi_{,\alpha}\,\phi_{,}{}^{\alpha} - f\,A^\mu\,\phi_{,\mu}\right] d^4x. \tag{3.108}$$

From left to right, the elements are: the vacuum action for A_μ, the vacuum action for ϕ, and the coupling term. Now we get a pair of field equations, one for A_μ:

$$\partial_\mu\,F^{\mu\nu} = -\mu_0 f\,\partial^\nu\phi \tag{3.109}$$

and one for ϕ:

$$\partial_\mu\partial^\mu\,\phi = -f\,\partial_\mu\,A^\mu. \tag{3.110}$$

This second equation is problematic. We are required from (3.109) to have $\partial_\nu\,\partial^\nu\,\phi = 0$, which is true only if $f = 0$ (no coupling) or $\partial_\mu\,A^\mu = 0$. We can require that $\partial_\mu\,A^\mu = 0$, but we have just *lost* the gauge freedom of E&M, because the Lorenz gauge is now imposed as a matter of self-consistency. The full set, including the Lorenz gauge requirement, is

$$\begin{aligned} \Box\phi &= 0 \\ \Box A^\mu &= -\mu_0 f\,\partial^\mu\,\phi \\ \partial_\mu\,A^\mu &= 0. \end{aligned} \tag{3.111}$$

Problem 3.34 Insert $\phi = \phi_0\,e^{i\,p_\mu\,x^\mu}$ and $A^\mu = P^\mu\,e^{i\,q_\alpha\,x^\alpha}$ into the coupled field equations (3.111) and find the constraints on ϕ_0, P^μ, p_μ, and q^μ (all constant scalars/vectors here) that satisfy the set.

3.7.2 Example: A Pair of Scalar Fields

Suppose we want to couple a pair of scalar fields to E&M. We know that a pair of free scalar fields is equivalent to a single complex scalar field, for which the vacuum Lagrangian is (in Cartesian coordinates)

$$\mathcal{L} = \phi^*_{,\mu}\, \eta^{\mu\nu}\, \phi_{,\nu} \tag{3.112}$$

with field equations $\Box\phi = 0 = \Box\phi^*$.

There is a phase symmetry to this Lagrangian: if we take $\bar{\phi} = \phi\, e^{i\theta}$ for constant (real) θ, the $\mathcal{L}$ is unchanged. That symmetry corresponds to a conservation law, a divergenceless j^μ that we can calculate by considering the infinitesimal form of the symmetry, as in Section 3.3.2, where we got (3.57) (omitting the constant out front),

$$j^\mu = i\left[\left(\partial^\mu\, \phi\right)^*\, \phi - \phi^*\left(\partial^\mu \phi\right)\right], \tag{3.113}$$

with $\partial_\mu\, j^\mu = 0$ by construction.

We have a natural current with which to couple the electromagnetic field. Our action now involves the vector E&M portion, the complex scalar piece, and a coupling term (with constant f to set the units):

$$S[A_\mu, \phi] = \int \left[\frac{1}{4\,\mu_0}\, F_{\alpha\beta}\, F^{\alpha\beta} + \phi^*_{,\mu}\, \eta^{\mu\nu}\, \phi_{,\nu} - f\, A_\mu\, j^\mu\right] d^4x, \tag{3.114}$$

with $F_{\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta}$, as always. When we vary with respect to A_μ, we get:

$$\partial_\nu\, F^{\nu\mu} = -\mu_0 f\, j^\mu, \tag{3.115}$$

as expected and desired.

Varying the action with respect to ϕ and ϕ^* gives the field equations

$$\Box\phi = 2\, i f\, A^\mu\, \phi_{,\mu} \quad \text{and} \quad \Box\phi^* = -2\, i f\, A^\mu\, \phi^*_{,\mu} \tag{3.116}$$

in Lorenz gauge. There is now a problem with the current j^μ: it is no longer conserved. When we generated j^μ, we used the vacuum field equations $\Box\phi = 0$. But $\Box\phi$ is no longer equal to zero, and the divergence of the current is now

$$\partial_\mu\, j^\mu = i\left(\phi\, \Box\phi^* - \phi^*\, \Box\phi\right) = 2 f\, A^\mu\, \partial_\mu\left(\phi^*\, \phi\right) = 2 f\, \partial_\mu\left(\phi^*\, \phi\, A^\mu\right), \tag{3.117}$$

where the final equality holds because of Lorenz gauge. This combination is *not* zero, and so this j^μ is unavailable as a source in (3.115).

How might we fix the situation? If we could add a piece to the action that gave a modified (3.115) that looked like:

$$\partial_\nu\, F^{\nu\mu} = -\mu_0 f\, \underbrace{\left(j^\mu - 2 f\left(\phi^*\, \phi\, A^\mu\right)\right)}_{\equiv J^\mu}, \tag{3.118}$$

then the divergence of this new J^μ would be zero. So what we want is a term that looks like $2 f^2\, \phi^*\, \phi\, A^\mu$ under A_μ-variation, and the only scalar that fits the bill is $f^2\, \phi^*\, \phi\, A^\mu\, A_\mu$.

Let's start again with the modified action:

$$S[A_\mu, \phi] = \int \left[\frac{1}{4\,\mu_0}\, F_{\alpha\beta}\, F^{\alpha\beta} + \phi^*_{,\mu}\, \eta^{\mu\nu}\, \phi_{,\nu} - f\, A_\mu\, j^\mu + f^2\, \phi^*\, \phi\, A^\mu\, A_\mu\right] d^4x. \tag{3.119}$$

The field equation for the E&M portion looks like

$$\partial_\nu\, F^{\nu\mu} = -\mu_0 f\left(j^\mu - 2 f\, \phi^*\, \phi\, A^\mu\right) = -\mu_0 f\, J^\mu, \tag{3.120}$$

which is just what we want (by design of course). The scalar field equations again come in complex conjugate pairs, with

$$\Box\phi = 2\,i f\,\phi_{,\mu}\,A^{\mu} + f^{2}\,\phi\,A^{\mu}\,A_{\mu} \text{ and complex conjugate.} \tag{3.121}$$

The field equation for ϕ has changed once again, and now we have to check the divergence of J^{μ} to ensure that the full set is self-consistent. Using the field equation, we have

$$\begin{aligned}
\partial_{\mu} J^{\mu} &= i\left(\phi\,\Box\phi^{*} - \phi^{*}\,\Box\phi\right) - 2f\,\partial_{\mu}\left(\phi^{*}\,\phi\,A^{\mu}\right) \\
&= i\left(-2\,i f\,\phi^{*}_{,\mu}\,A^{\mu} + f^{2}\,\phi^{*}\,A^{\mu}\,A_{\mu}\right)\phi \\
&\quad - i\left(2\,i f\,\phi_{,\mu}\,A^{\mu} + f^{2}\,\phi\,A^{\mu}\,A_{\mu}\right)\phi^{*} - 2f\,\partial_{\mu}\left(\phi^{*}\,\phi\,A^{\mu}\right) \\
&= 0
\end{aligned} \tag{3.122}$$

so that while the field equation for ϕ (and ϕ^{*}) has changed, the result does not spoil the conservation of J^{μ}.

The process is now complete. We have a self-consistent, self-coupled theory that combines E&M with a pair of scalar fields. The action from which the field equations came is something of a mess. Written out in terms of ϕ and its conjugate explicitly, it reads

$$\begin{aligned}
S[A_{\mu},\phi] = \int \Bigg[&\frac{1}{4\,\mu_0}\,F_{\alpha\beta}\,F^{\alpha\beta} + \phi^{*}_{,\mu}\,\eta^{\mu\nu}\,\phi_{,\nu} - i f\,A^{\mu}\left(\phi^{*}_{,\mu}\,\phi - \phi^{*}\,\phi_{,\mu}\right) \\
&+ f^{2}\,\phi^{*}\,\phi\,A^{\mu}\,A_{\mu}\Bigg]\,d^{4}x.
\end{aligned} \tag{3.123}$$

The terms involving ϕ are tantalizingly factorizable; we even have the factors of the coupling constant f to guide the way. In the end, you can write the action succinctly as

$$S[A_{\mu},\phi] = \int \left[\frac{1}{4\,\mu_0}\,F_{\alpha\beta}\,F^{\alpha\beta} + \left[\left(\partial_{\mu} - i f\,A_{\mu}\right)\phi\right]^{*}\,\eta^{\mu\nu}\,\left[\left(\partial_{\nu} - i f\,A_{\nu}\right)\phi\right]\right] d^{4}x, \tag{3.124}$$

or if we define the "covariant derivative": $D_{\mu} \equiv \partial_{\mu} - i f\,A_{\mu}$,

$$S[A_{\mu},\phi] = \int \left[\frac{1}{4\,\mu_0}\,F_{\alpha\beta}\,F^{\alpha\beta} + \left(D_{\mu}\,\phi\right)^{*}\,\eta^{\mu\nu}\,\left(D_{\nu}\,\phi\right)\right] d^{4}x. \tag{3.125}$$

One way of telling this story is to start with the free field Lagrangian for A_{μ} and ϕ, but rewrite the scalar piece of the action using the new definition of derivative while omitting the explicit coupling. You just take $\partial \to D$ in the free scalar portion $\partial_{\mu}\phi^{*}\,\eta^{\mu\nu}\,\partial_{\nu}\phi$. That certainly works, and there are potentially deep reasons for *why* it works (see Problems 3.35 and 3.36). In particular, you are asking for the combined theory to have *more* invariance structure than the scalar field by itself, turning a global (constant) phase choice into a local (depending on position) one related to the E&M gauge choice. That expansion of invariance may be viewed as a fundamental building block for "good" field theories.

It is interesting that this notion of coupling field theories together, where you take a "normal" free field theory and introduce some sort of modified derivative, is precisely how general relativity works. The new derivative there even has the same name.

Problem 3.35 Show that the Lagrangian $\mathcal{L} = \left(D_\mu\,\phi\right)^*\,\eta^{\mu\nu}\,(D_\nu\,\phi)$ (with $f = 1$ here) is invariant under the "local" gauge transformation: $\phi \to \phi\,e^{i\,\psi}$ and $A_\mu \to A_\mu + \psi_{,\mu}$ for ψ an arbitrary function of position and time. The action (3.125) is invariant under a natural, but *larger* set of transformations than either of its constituent free field actions, and even the starting point (3.114).

Problem 3.36 In Section 3.3.2, we considered a simple (single, real) scalar Lagrangian that was unchanged under $\phi \to \phi + \delta\phi$ where $\delta\phi$ is a constant. When coupling to E&M, we could try to generate a new field theory in which the gauge freedom has been expanded from global (constant) to local (where $\delta\phi$ is a function of position). The E&M Lagrangian is already unchanged under $A_\mu \to A_\mu - \psi_{,\mu}/m$ for arbitrary ψ and constant m. Write down the simplest Lagrangian, quadratic in the fields, which is unchanged by

$$\phi \to \phi + \psi, \qquad A_\mu \to A_\mu - \frac{1}{m}\,\psi_{,\mu} \tag{3.126}$$

where ψ is now a function of position.[9]

3.8 Particles and Fields

While field sources are natural and easy to work with in the context of field theories, particle sources are more familiar, and in order to fit them into a field-theoretic framework, we start by rewriting their action in the four-dimensional space-time setting. Let's go back to our notion of a particle action. We had

$$S_\circ = m\,c \int \sqrt{-\frac{dx^\mu}{dt}\,\eta_{\mu\nu}\,\frac{dx^\nu}{dt}}\,dt \tag{3.127}$$

where we have used the reparametrization invariance to choose t as the parameter. This action is not an integral over all of the coordinates, but we can artificially write it as a space-time volume integral by introducing $\delta^3(\mathbf{r} - \mathbf{r}(t))$, a term that gets rid of the spatial integration by putting the particle on its trajectory $\mathbf{r}(t)$. So we could start with the four-volume integral[10]

$$S = \int \left[m\,c \sqrt{-\frac{dx^\mu}{dt}\,\eta_{\mu\nu}\,\frac{dx^\nu}{dt}}\,\delta^3(\mathbf{r} - \mathbf{r}(t)) \right] d^4x, \tag{3.128}$$

and this would allow us to combine particle "Lagrangians" with field Lagrangians. Notice also that the Lagrangian here is a Lagrange "density" – the delta function throws in a factor of $1/\text{volume}$.

Since we now have an appropriate S, we can generate a stress tensor density: $\mathcal{T}^{\mu\nu} = -2\,\frac{\partial\mathcal{L}}{\partial\eta_{\mu\nu}}$ that is conserved. Let's look at the form of this stress tensor (density); taking the integrand of S above as $\mathcal{L}$, we can compute the derivative with respect to the metric:

9 Thanks to Colin Vangel for pointing out this problem to me.

10 In this integral, we have thrown in an additional factor of c relative to (3.127) to get the space-time volume element $d^4x = d\tau\; d(c\,t)$.

$$\mathcal{T}^{\mu\nu} = \frac{m \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}}{\sqrt{1 - \frac{v^2}{c^2}}} \delta^3(\mathbf{r} - \mathbf{r}(t)) \tag{3.129}$$

giving back the usual relativistic momentum (density) as the $\mathcal{T}^{\mu 0}$ component.[11] Now think about what happens if we combine this particle action with the electromagnetic field action, together with coupling:

$$\begin{aligned} S = \int \Bigg[& \frac{1}{4\,\mu_0} F_{\alpha\beta} F^{\alpha\beta} + m\,c \sqrt{-\frac{dx^\mu}{dt} \eta_{\mu\nu} \frac{dx^\nu}{dt}} \, \delta^3(\mathbf{r} - \mathbf{r}(t)) \\ & - q\,A_\alpha \frac{dx^\alpha}{dt} \delta^3(\mathbf{r} - \mathbf{r}(t)) \Bigg] d^4x. \end{aligned} \tag{3.130}$$

We get a combined stress tensor density (setting the metric to Minkowski in Cartesian coordinates)[12]

$$\mathcal{T}^{\mu\nu} = \frac{1}{\mu_0} \left[F^{\mu\rho} F^{\nu}{}_{\rho} - \frac{1}{4} \eta^{\mu\nu} F_{\alpha\beta} F^{\alpha\beta} \right] + \frac{m \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}}{\sqrt{1 - \frac{v^2}{c^2}}} \delta^3(\mathbf{r} - \mathbf{r}(t)), \tag{3.131}$$

and now the integral form of the $\nu = 0$ conservation law reads

$$\frac{d}{dt} \int u\, d\tau + \oint \mathbf{S} \cdot d\mathbf{a} + \frac{d}{dt} \left(\frac{m\,c^2}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = 0, \tag{3.132}$$

where the integral was used to eliminate the delta function. The boundary term for the particle goes away because the delta function is zero out there (away from the particle trajectory). Now we have a full conservation of energy statement, complete with particle energy that must be in place. We can understand the $\frac{dE}{dt}$ term for the particle as the time derivative of the work done on the particle by the fields.

Notice that we also get back both the single-particle source for the electromagnetic field, when we vary the action with respect to A_ν, and recover the Lorentz force acting on the particle in the relativistic form of Newton's second law from the particle–action variation of the two terms with delta functions on them in the action.

What happens to the spatial conservation law: $\frac{1}{c} \frac{\partial \mathcal{T}^{0j}}{\partial t} + \frac{\partial \mathcal{T}^{ij}}{\partial x^i} = 0$? Again, using the integral form and dropping the surface term coming from the particle, we have

$$\frac{d}{dt} \int \frac{1}{c^2} \mathbf{S}\, d\tau + \oint T^{ij}\, da_j + \frac{d}{dt} \left(\frac{m\,\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = 0. \tag{3.133}$$

We can think of the particle term as the momentum in a domain containing the particle. The remaining integrals lend themselves to physical interpretation. The first term looks like the time derivative of another momentum, the *field* momentum, leading us to identify $\frac{1}{c^2}\mathbf{S}$ with

[11] Here, all components of $\mathcal{T}^{\mu 0}$ have dimension of energy (per unit volume), unlike the "usual" energy–momentum four-vector, which has dimension of momentum – they just differ by a factor of c.

[12] There is a factor of $\sqrt{-\eta}$ on the E&M portion of the stress tensor density, but in Cartesian coordinates, $\sqrt{-\eta} = 1$.

a field momentum density. The stress tensor components then look like pressures (force per unit area), and we associate T^{ij} with the ith component of pressure acting on a surface oriented in the jth direction, so that, for example, T^{i1} represents the three components of $\mathbf{P}$ (pressure, the force per unit area) acting on a surface with unit normal pointing in the $\hat{\mathbf{x}}$-direction.

In the absence of particles within our volume, the conservation statement (3.133) is a field-momentum conservation law: the momentum stored in the field within a domain Ω changes because of momentum flux (density, T^{ij}) through the surface of the domain. Looking at the differential expression again, we can see what all those E&M components are really saying – using

$$T^{ij} = -\frac{1}{\mu_0}\left[\frac{1}{c^2}\,E^i\,E^j + B^i\,B^j - \frac{1}{2}\,\delta^{ij}\left(\frac{E^2}{c^2} + B^2\right)\right], \tag{3.134}$$

from (3.96), the divergence is

$$\begin{aligned}\partial_j\,T^{ij} &= -\frac{1}{\mu_0}\left[\frac{1}{c^2}\left(E^j\,\partial_j\,E^i + E^i\,\partial_j\,E^j\right) + \left(B^j\,\partial_j\,B^i + B^i\,\partial_j\,B^j\right) - \frac{1}{2}\partial^i\left(\frac{E^2}{c^2} + B^2\right)\right]\\ &= -\frac{1}{\mu_0}\left[\frac{1}{c^2}\left(E^j\left(\partial_j\,E^i - \partial^i\,E_j\right) + E^i\,\partial_j\,E^j\right) + B^j\left(\partial_j\,B^i - \partial^i\,B_j\right) + B^i\,\partial_j\,B^j\right]\\ &\equiv k^i.\end{aligned} \tag{3.135}$$

In vector form, we have

$$\mathbf{k} = -\frac{1}{\mu_0}\left[-\frac{1}{c^2}\,\mathbf{E}\times(\nabla\times\mathbf{E}) - \mathbf{B}\times(\nabla\times\mathbf{B}) + \frac{1}{c^2}\,\mathbf{E}(\nabla\cdot\mathbf{E}) + \mathbf{B}(\nabla\cdot\mathbf{B})\right]. \tag{3.136}$$

Going back to Maxwell's equations, we can write the vector $\mathbf{k}$ in terms of ρ and $\mathbf{J}$:

$$\begin{aligned}\mathbf{k} &= -\left[\rho\,\mathbf{E} + \epsilon_0\,\mathbf{E}\times\frac{\partial\mathbf{B}}{\partial t} - \mathbf{B}\times\left(\mathbf{J} + \epsilon_0\frac{\partial\mathbf{E}}{\partial t}\right)\right]\\ &= -\left[\rho\,\mathbf{E} + \mathbf{J}\times\mathbf{B} + \frac{1}{c^2}\,\frac{\partial}{\partial t}\left(\frac{1}{\mu_0}\,\mathbf{E}\times\mathbf{B}\right)\right].\end{aligned} \tag{3.137}$$

The first two terms in $\mathbf{k}$ are the (negative of the) force density terms from the Lorentz force law (ρ is charge/volume and $\mathbf{J} = \rho\,\mathbf{v}$ plays the density role in $q\,\mathbf{v}\times\mathbf{B}$). The remaining term in $\mathbf{k}$ is just the time-derivative of the Poynting vector. Putting these together in

$$\int_\Omega\left(\frac{1}{c^2}\,\frac{\partial T^{0i}}{\partial t} + \partial_j\,T^{ji}\right)d\tau + \frac{d}{dt}\left(\frac{m\,v^i}{\sqrt{1-\frac{v^2}{c^2}}}\right) = 0 \tag{3.138}$$

where the domain Ω includes the particle, we have

$$\frac{d}{dt}\left(\frac{m\,v^i}{\sqrt{1-\frac{v^2}{c^2}}}\right) = \int_\Omega\left(\rho\,\mathbf{E} + \mathbf{J}\times\mathbf{B}\right)\,d\tau, \tag{3.139}$$

and the right-hand side becomes the Lorentz force if we shrink Ω down to include just the test particle. So this second conservation law recovers Newton's second law for the particle (along with the correct identification of the Lorentz force as an "ordinary," not Minkowski, force).

3.9 Model Building

If we think about what is possible, given the constraints of special relativity and boundary conditions, there is not as much freedom, in constructing physical theories, as you might expect. For example, if we require that the field equations are:

- Lorentz covariant
- at most second-order partial differential equations (PDEs), and
- linear,

then for scalars, you end up with Klein–Gordon as the most general field equation (in vacuum, no sources),

$$\alpha\,\partial_\mu\,\partial^\mu\phi + \beta\,\phi = 0, \tag{3.140}$$

for constants α and β. This field equation comes from the Lagrangian

$$\mathcal{L} = \frac{1}{2}\,\alpha\,\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}\,\beta\,\phi^2. \tag{3.141}$$

Because we could divide by β in the field equation, the scalar field theory here has only one free parameter.

For a vector field A_μ, there are more options. The field equation that supports our three ingredients is

$$\alpha\,\partial_\mu\,\partial^\mu\,A_\nu + \beta\,\partial_\nu\,\partial_\mu\,A^\mu + \gamma\,A_\nu = 0. \tag{3.142}$$

We can impose additional constraints. In particular, we will require that our vector field theory contains no scalar information. This notion is what leads to the gauge freedom of E&M, and we have seen the procedure before. Take $A_\mu = \phi_{,\mu}$, a scalar gradient,[13] insert in the general field equation:

$$\alpha\,\partial_\mu\,\partial^\mu\,\partial_\nu\,\phi + \beta\,\partial_\nu\,\partial_\mu\,\partial^\mu\,\phi + \gamma\,\partial_\nu\phi = 0 \tag{3.143}$$

and cancel the left-hand side by taking $\beta = -\alpha$ and $\gamma = 0$. That leaves

$$\partial_\mu\,\partial^\mu\,A_\nu - \partial_\nu\,\partial_\mu\,A^\mu = 0 \tag{3.144}$$

as the field equation. You could put in a mass term (take $\gamma \neq 0$) by hand and spoil the scalar independence (and lose the associated gauge freedom, as in Section 3.7.1). If we stick with (3.144), then we have the freedom to introduce an arbitrary scalar component, and we can use that to fix $\partial_\mu\,A^\mu = 0$, so the final field equation is the simple

$$\partial_\mu\,\partial^\mu\,A_\nu = 0. \tag{3.145}$$

[13] This choice is often referred to as a "pure gauge."

Moving on to second-rank tensor fields, we'll focus on symmetric ones for which the most general field equation satisfying our three principles is

$$\alpha\, \partial_\mu\, \partial^\mu h_{\rho\sigma} + \beta\, \partial_\rho \partial^\mu h_{\mu\sigma} + \gamma\, \partial_\sigma\, \partial^\mu h_{\mu\rho} + \tau\, \partial_\rho\, \partial_\sigma\, h^\mu_{\ \mu} = 0. \tag{3.146}$$

Once again, we need additional information to constrain further. If we require that there be no scalar component in the field equation, by setting $h_{\mu\nu} = \phi_{,\mu\nu}$, then we learn that $\alpha + \beta + \gamma + \tau = 0$. Moving up one, we can take $h_{\mu\nu} = K_{\mu,\nu} + K_{\nu,\mu}$, a symmetrized pure vector field, and ask that the field equations make no reference to K_μ. That gives us, once we've used the corresponding gauge freedom,

$$\partial_\mu\, \partial^\mu h_{\rho\sigma} = 0 \tag{3.147}$$

as the vacuum field equation. What exactly this theory means physically requires additional inputs. What are the sources (which must be represented by a divergenceless second-rank tensor), and with what strength do they couple to the field? How do particles respond to the field? These are issues that must be determined experimentally (although once we know what the source is, we can couple with particles to find the predicted interaction).

Problem 3.37 Find the plane-wave solutions for $\Box A_\mu = 0$ in Lorenz gauge; i.e., what constraints must be placed on the "polarization" P^μ and "wave vector" k_ν (both constant) appearing in $A^\mu = P^\mu\, e^{i\, k_\nu\, x^\nu}$ such that the vacuum field equation and gauge condition are satisfied? How about the constraints on the polarization tensor, $Q_{\mu\nu}$, and wave vector, k_σ, in the symmetric $h_{\mu\nu} = Q_{\mu\nu}\, e^{i\, k_\sigma\, x^\sigma}$ satisfying $\Box h_{\mu\nu} = 0$ with $\partial^\mu\, h_{\mu\nu} = 0$?

Problem 3.38 The Klein–Gordon equation (for a free particle) is (with units in place)

$$\frac{1}{c^2}\,\frac{\partial^2 \Psi}{\partial t^2} - \nabla^2\, \Psi + \left(\frac{m\, c}{\hbar}\right)^2\, \Psi = 0. \tag{3.148}$$

a. Send in the ansatz $\Psi = \Psi_0\, e^{i(\omega\, t + \mathbf{k}\cdot\mathbf{r})}$ to find a relationship between ω, k, and m.

b. For a relativistic particle, we have energy–momentum four-vector

$$p^\mu \doteq \begin{pmatrix} \frac{E}{c} \\ \mathbf{p} \end{pmatrix}. \tag{3.149}$$

Expand $p^\mu\, p_\mu = \bar{p}^\mu\, \bar{p}_\mu$ where p^μ is as above, and $\bar{p}^\mu$ is the energy–momentum four-vector evaluated in the particle rest frame (the two expressions must be equal since $p^\mu\, p_\mu$ is a scalar); i.e., write out $p^\mu\, p_\mu = \bar{p}^\mu\, \bar{p}_\mu$ in terms of E, $\mathbf{p}$, m (the mass of the particle), and c.

c. By comparing your expressions from parts (a) and (b), find the relations between E and ω and between $\mathbf{p}$ and $\mathbf{k}$ that allow the plane wave ansatz to be interpreted as a relativistic particle.

Problem 3.39 Find a Lagrangian that leads to (3.147).

Problem 3.40 Find the constraint on the coefficients in (3.146) that removes reference to pure scalar and vector contributions.

Problem 3.41 What is the most general field equation for an anti-symmetric second-rank tensor field $h_{\mu\nu} = -h_{\nu\mu}$, with scalar and vector contributions projected out? This type of field shows up in the Kalb–Ramond theory of electromagnetic interaction for strings.

3.9.1 Dimension

We know that the dimension of space-time changes physics. The Green's function for $\partial_\mu \partial^\mu A_\nu = -\mu_0 J_\nu$, for example, depends on what space-time dimension you are in. But dimension also brings with it additional options for the physical content of a theory. For example, in $D = 3 + 1$, a vector term for the Lagrangian that looks like $\epsilon^{\alpha\beta\mu\nu} A_{\alpha,\beta} A_{\mu,\nu}$ is a total divergence,

$$\partial_\nu \left(\epsilon^{\alpha\beta\mu\nu} A_{\alpha,\beta} A_\mu\right) = \epsilon^{\alpha\beta\mu\nu} A_{\alpha,\beta} A_{\mu,\nu}, \tag{3.150}$$

so will contribute nothing to the field equations governing A_μ as you established in Problem 3.13.

The analogous term, in $D = 2 + 1$ (two spatial, one temporal) dimensions, namely, $\epsilon^{\alpha\beta\mu} A_{\alpha,\beta} A_\mu$, does *not* come from a pure divergence and so is allowed. In this reduced space-time, the most general field equation is now (for a complete development, see [9])

$$\alpha\, \partial_\mu \partial^\mu A_\nu + \beta\, \partial_\nu \partial_\mu A^\mu + \gamma\, A_\nu + \delta\, \epsilon_{\rho\sigma\nu} A^{\rho}{}_{,}{}^{\sigma} = 0 \tag{3.151}$$

for constants α, β, γ, and δ.

Problem 3.42 Find the constraint on α, β, γ, and δ that renders (3.151) free of scalar contribution. Working in Lorenz gauge ($\partial_\mu A^\mu = 0$ as usual), find the static, spherically symmetric solution for the resulting field equation.

Problem 3.43 Find the field equations for

$$\mathcal{L} = \frac{1}{4\,\mu_0} F^{\mu\nu} F_{\mu\nu} + \kappa\, \epsilon^{\mu\nu\alpha} A_{\mu,\nu} A_\alpha \tag{3.152}$$

in $D = 2 + 1$, Cartesian coordinates, with a constant κ. Here, $F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}$ as usual; the field is A_μ.

3.9.2 Allowing Nonlinearity

Of the three building blocks of field theory from the start of Section 3.9, Lorentz covariance is required by special relativity, and second-order field equations are a statement of our observational experience – we "see" the boundaries experimentally, and use that with the field equations to infer the internal (meaning in the space-time between the boundaries) physics. To the extent that this works, we expect second-order PDEs as the governing equations. The third component, linearity, reflects a notion of superposition when it is observed, as in E&M, or simplicity,[14] when used for more speculative or toy theories.

[14] Remember that most interesting couplings end up being nonlinear – as in Section 3.7.2 – so starting with a free field Lagrangian that is linear keeps the free forms as simple as possible.

Note that superposition must be *observed*, as it is for E&M, and then linearity is implied. Of course, experimental observation of superposition is subject to experimental error. Maybe E&M has only approximate superposition but is secretly a nonlinear theory. Any nonlinearity makes the theory much more complicated, since we no longer have the utility of Green's functions in constructing solutions.

There must be some compelling reason to introduce nonlinearity. In the case of gravity, we are forced to consider nonlinear field theories because linear ones do not couple to their own field energy, and one of the defining properties of gravity is its universal coupling, as we shall see in the next chapter. So there is almost no way to imagine a gravity theory that is not nonlinear. But people also allow nonlinearity in other settings, like E&M, in order to cure basic illnesses. As an example, Born–Infeld E&M, with action (depending on a constant p)[15]

$$S = \int \left[-p^2 \sqrt{-\det\left(\eta_{\mu\nu} + \frac{F_{\mu\nu}}{p} \right)} \right] d^4x \tag{3.153}$$

provides point sources that have a finite amount of energy stored in the field of a point particle.[16] The point source solution is available, but since we lack superposition, the dipole solution, for example, is unknown (analytically).

Problem 3.44 Find the field equation and stress tensor density for the action (3.153) (for a first-rank tensor A_μ with $F_{\mu\nu} = A_{\nu 1 \mu} - A_{\mu 1 \nu}$).

Problem 3.45 Introduce a source J^μ in (3.153) and adjust coefficients so that you recover Maxwell's E&M in the appropriate limit for p (which is what?). Find the Born–Infeld electric field associated with an infinite line of charge with constant λ at rest.

15 The original paper is [5]. For a modern discussion of the Born–Infeld action, see [27].
16 But then you have to figure out what that finite amount of energy *is*.

4 Gravity

In this chapter, we will apply some of our field theory ideas to gravity.[1] We'll start with Newtonian gravity and think about what could happen in that familiar setting if we include elements of special relativity. The current relativistic theory of gravity is general relativity, and most of its qualitative predictions exist already at the extended Newtonian level. After developing those predictions, we will think about how to make gravity more like E&M as a field theory, and that introduces the notion of gravitational radiation, the gravitational version of electromagnetic radiation. Finally, we'll see that the full theory of gravity cannot just be E&M with funny signs; there must be a new field governing gravitational interaction, and we'll explore some of the implications of that second-rank, symmetric tensor field.

4.1 Newtonian Gravity

Let's review the original theory of gravity and see where and how it might be updated. Newton's theory of gravity begins with a specified source of mass, $\rho(\mathbf{r})$, the mass density, as a function of position. Then the gravitational field $\mathbf{g}$ is related to the source ρ by

$$\nabla \cdot \mathbf{g} = -4\,\pi\, G\,\rho, \tag{4.1}$$

where G is the gravitational constant ($G \sim 6.67 \times 10^{-11}$ N m^2/kg^2). Particles respond to the gravitational field through the force $\mathbf{F} = m\,\mathbf{g}$ used in Newton's second law:[2]

$$m\,\ddot{\mathbf{r}}(t) = m\,\mathbf{g}, \tag{4.2}$$

for a particle of mass m at location $\mathbf{r}(t)$.

The field $\mathbf{g}$ is a force-per-unit-mass (similar to $\mathbf{E}$, a force-per-unit-charge), and it has $\nabla \times \mathbf{g} = 0$, so it is natural to introduce a potential ϕ with $\mathbf{g} = -\nabla\phi$, just as we do in E&M. Then the field equation and equation of motion read (again, given ρ)

[1] In particular, we'll focus on motivating the need for a modified theory of gravity, ultimately, general relativity. From here, there are a number of good resources, such as [6], [11], and [24].

[2] We'll use the non-relativistic form of Newton's second law for now; you can put in the relativistic momentum if you like.

$$
\begin{aligned}
\nabla^2 \phi &= 4\pi G \rho \\
m\ddot{\mathbf{r}}(t) &= -m\nabla\phi.
\end{aligned}
\tag{4.3}
$$

4.1.1 Solutions

From the field equation, we can immediately obtain (by analogy with E&M if all else fails) the field $\mathbf{g}$ for a variety of symmetric source configurations. Start with the point source solution. If $\rho = m\,\delta^3(\mathbf{r})$ for a point source located at the origin, we have

$$
\phi = -\frac{G m}{r}, \tag{4.4}
$$

similar to the Coulomb potential $V = \frac{q}{4\pi\epsilon_0 r}$, but with $\frac{1}{4\pi\epsilon_0} \rightarrow -G$ (and $q \rightarrow m$, of course).

We can generate, from (4.1), the integral form of "Gauss's law" for gravity. Pick a domain Ω and integrate both sides of (4.1) over that domain, using the divergence theorem, to get

$$
\oint_{\partial\Omega} \mathbf{g}\cdot d\mathbf{a} = -4\pi G \underbrace{\int_{\Omega} \rho(\mathbf{r}')\, d\tau'}_{\equiv M_{\text{enc}}}. \tag{4.5}
$$

As usual, this equation is always true, but not always useful for finding $\mathbf{g}$.

Suppose we have an infinite line of mass with constant mass-per-unit-length λ, lying along the vertical axis. Then symmetry demands that $\mathbf{g} = g(s)\,\hat{\mathbf{s}}$, and using a Gaussian cylinder of height h and radius s centered on the wire, we have

$$
\oint_{\partial\Omega} \mathbf{g}\cdot d\mathbf{a} = g(s)\, 2\pi\, h\, s, \tag{4.6}
$$

with $M_{\text{enc}} = \lambda\, h$, so that

$$
g(s) = -\frac{2G\lambda}{s}, \tag{4.7}
$$

which mimics the electrostatic result $E(s) = \frac{\lambda}{2\pi\epsilon_0 s}$, and the two are again related under the mapping: $\frac{1}{4\pi\epsilon_0} \rightarrow -G$ (and appropriate source modification: charge goes to mass).

Problem 4.1 Find the gravitational field above and below an infinite sheet of mass with constant mass-per-unit-area σ.

Problem 4.2 What is the potential ϕ (with $\mathbf{g} = -\nabla\phi$) for the infinite line of mass and for the infinite sheet?

Problem 4.3 What is the potential inside and outside a sphere of radius R with constant mass density ρ_0? (Set the potential to zero at spatial infinity and use continuity and derivative continuity to set constants.)

4.2 Source Options

4.2.1 Three Types of Mass

Is it clear that the masses appearing on the left and right of (4.2) are the same? In E&M, for example, Newton's second law reads: $m\,\ddot{\mathbf{r}}(t) = q\,\mathbf{E}$, and we don't associate m and q (units aside), so why do we have the same m appearing on both sides of (4.2)? We could imagine that the inertial mass m on the left is different from the "passive" gravitational mass, m_p, on the right:

$$m\,\ddot{\mathbf{r}}(t) = m_p\,\mathbf{g}. \tag{4.8}$$

While experiment suggests $m \approx m_p$, it is impossible (within the confines of Newtonian gravity and dynamics) to rule out the possibility that $m \neq m_p$. Einstein's theory of general relativity *requires* that $m = m_p$, so its correctness suggests that $m = m_p$ exactly, i.e., that the inertial and passive gravitational masses are identical.

There is a third type of mass built into Newtonian gravity: the "active" gravitational mass that generates the gravitational field. In theory, then, there are three different masses that could be present in the equations of Newtonian gravity:

$$\begin{aligned} \nabla \cdot \mathbf{g} &= -4\,\pi\,G\,\rho_a \\ m\,\ddot{\mathbf{r}}(t) &= m_p\,\mathbf{g}, \end{aligned} \tag{4.9}$$

where ρ_a is the active mass density of some source mass. Again, up to experimental error, the active and passive masses are the same, so we won't worry about these distinctions.[3]

4.2.2 Negative Mass

In E&M, there are two types of charge, positive and negative, yet we only (typically) allow one sign for mass. What would happen if you had a negative mass? Take a pair of masses m_ℓ and m_r (left and right) separated a distance x, as shown in Figure 4.1. The force on each mass is

$$\mathbf{F}_\ell = \frac{G\,m_\ell\,m_r}{x^2}\,\hat{\mathbf{x}}, \qquad \mathbf{F}_r = -\frac{G\,m_\ell\,m_r}{x^2}\,\hat{\mathbf{x}}, \tag{4.10}$$

and if we assume all forms of mass are the same (inertial, active, and passive), then the acceleration of each mass is

$$\mathbf{a}_\ell = \frac{G\,m_r}{x^2}\,\hat{\mathbf{x}}, \qquad \mathbf{a}_r = -\frac{G\,m_\ell}{x^2}\,\hat{\mathbf{x}}. \tag{4.11}$$

So far, we have made no assumptions about the signs of the masses. Take $m_r > 0$ and $m_\ell < 0$; what happens? The details of the dynamics are, as always, complicated, but of interest is the direction of the acceleration: both accelerations point to the right (in the $\hat{\mathbf{x}}$-direction) and the masses "chase" each other.

[3] Other people worry about them all the time. If you could find, say, a difference between m and m_p, then general relativity would automatically be unavailable as the theory of gravity.

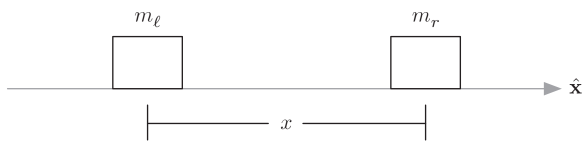

Fig. 4.1 Two masses separated by a distance x.

That implies that both masses are speeding up and traveling in the same direction. What happens to energy and momentum conservation? The total energy of the system is

$$E = \frac{1}{2} m_\ell \, v_\ell^2 + \frac{1}{2} m_r \, v_r^2 - \frac{G \, m_\ell \, m_r}{x}, \tag{4.12}$$

for $\mathbf{v}_\ell$ and $\mathbf{v}_r$, the velocity vectors. The energy can remain constant for a constant separation x while *both* v_ℓ and v_r increase because $m_\ell < 0$. Similarly, the momentum of the initial configuration is zero (assuming the masses start from rest). If the two masses move in the same direction with increasing speed, it would appear that the linear momentum is no longer zero. Again, since $m_\ell < 0$, we have total momentum $\mathbf{p} = m_\ell \, \mathbf{v}_\ell + m_r \, \mathbf{v}_r$, which is zero even though both masses are moving in the same direction.

4.3 Predictions: Non-relativistic

We can go back to Newton's original motivation for introducing the $1/r^2$ force of gravity associated with two massive bodies. That form produces motion obeying Kepler's laws, obtained by Kepler through consideration of observational data. Our first goal will be to identify the motion of a test mass induced by the gravitational field of a spherically symmetric central body of mass M. Then we'll move on to cases in which the "test body" (which feels the effect of the field) is not a point particle.

4.3.1 Elliptical Orbits

For a particle of mass m that travels in the presence of the gravitational field of a spherical central body of mass M, the total energy is

$$H \equiv \frac{1}{2} m \left(\dot{r}^2 + r^2 \, \dot{\phi}^2 \right) + U(r) = E, \tag{4.13}$$

where we are working in the $x - y$ plane in spherical coordinates and $U(r) = -G \, M \, m/r$. Angular momentum is conserved here, from $\dot{p}_\phi = -\frac{\partial H}{\partial \phi} = 0$, so that $J_z \equiv m \, r^2 \, \dot{\phi}$ is a conserved quantity. Using this, we can write the total energy as

$$H = \frac{1}{2} m \left(\dot{r}^2 + \frac{J_z^2}{m^2 \, r^2} \right) + U(r) = E. \tag{4.14}$$

Fig. 4.2 A sketch of the effective potential from (4.15) with $U = -G\,M\,m/r$, as a function of r.

This expression defines the Hamiltonian governing one-dimensional motion in r, with an effective potential energy given by

$$U_{\text{eff}} = \frac{J_z^2}{2\,m\,r^2} + U(r). \tag{4.15}$$

Given the gravitational potential energy $U(r) = -G\,M\,m/r$, we can look at this effective potential and predict the type of motion we can get. The effective potential is large and positive for small r where the $1/r^2$ term dominates, and it is small and negative for large r. A representative sketch is shown in Figure 4.2. Notice that nothing with non-zero angular momentum can actually fall to the center of the central body (assuming a point source, so we don't have to worry about running into the surface of the central body). Newtonian gravity does not predict that orbiting bodies fall to $r = 0$ (unlike general relativity, in which particles with non-zero angular momentum can fall into central bodies).

Because the potential energy here goes like $1/r$, it is natural to define a new coordinate $\rho \equiv 1/r$ to make the potential energy linear in ρ. With this substitution in place, our energy expression becomes

$$H = \frac{1}{2}\,m\left(\frac{\dot{\rho}^2}{\rho^4} + \frac{J_z^2}{m^2}\,\rho^2\right) + U(\rho) = E. \tag{4.16}$$

Finally, we want a geometric description of the trajectory. We don't particularly care how long it takes the particle to move from one place to another; we're more interested in the shape of its motion. To obtain a geometric description, reparametrize in terms of ϕ instead of t. Temporal and angular derivatives are related by

$$\dot{\rho} = \frac{d\rho}{d\phi}\,\dot{\phi} = \rho'\,\frac{J_z}{m}\,\rho^2 \tag{4.17}$$

using $\dot{\phi} = J_z\,\rho^2/m$ and defining $\rho' \equiv \frac{d\rho}{d\phi}$. Now

$$H = \frac{1}{2}\,\frac{J_z^2}{m}\left(\rho'^2 + \rho^2\right) + U(\rho) = E, \tag{4.18}$$

and we can take the ϕ derivative of both sides to get rid of the constant E, then isolate ρ'':

$$\rho'' = -\rho - \frac{m}{J_z^2} U'(\rho) = -\rho + \frac{G M m^2}{J_z^2}, \tag{4.19}$$

for $U'(\rho) \equiv \frac{dU}{d\rho}$ with solution

$$\rho(\phi) = A \cos(\phi) + B \sin(\phi) + \frac{G M m^2}{J_z^2}. \tag{4.20}$$

We don't need to keep both the sine and cosine terms (either one suffices), so take

$$\rho = \frac{G M m^2}{J_z^2} + A \cos\phi. \tag{4.21}$$

Then the radius, as a function of angle, is

$$r = \frac{1}{\rho} = \frac{J_z^2}{G M m^2} \frac{1}{1 + e \cos\phi} \equiv \frac{p}{1 + e \cos\phi} \tag{4.22}$$

where $e \equiv A J_z^2/(G M m^2)$ is a new constant and $p \equiv J_z^2/(G M m^2)$. For $e \neq 0$, this expression for $r(\phi)$ describes an ellipse with "eccentricity" e and "semilatus rectum" p (when $e = 0$, we have a circle of radius p). The closest point to the central body (situated at the origin) occurs when $\phi = 0$ and is called the "perihelion" (technically, that's the closest point in an orbit about the *sun*, as the name suggests), and the furthest point from the central body is at $\phi = \pi$, called the "aphelion." The details of the orbit provide those values, and from them we can extract information about the central body.

4.3.2 Kepler's Laws

Kepler's first law is that the motion of bodies around the sun is described by elliptical orbits, and we have demonstrated this directly with the solution (4.22). Kepler's second law says that equal areas are swept out in equal times, and we can establish this by showing that $\frac{da}{dt}$ = constant for area da swept out in time dt. Working infinitesimally, and referring

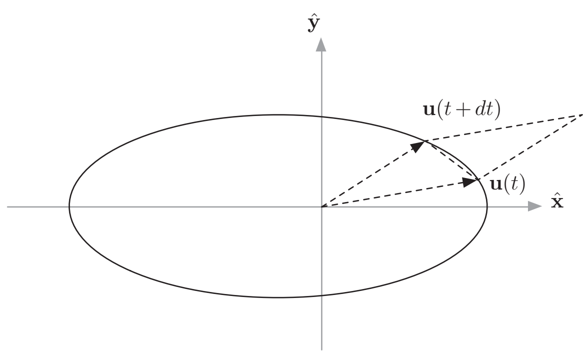

Fig. 4.3 A particle is traveling along an elliptical orbit. At time t it is at $\mathbf{u}(t)$, and at time $t + dt$, it is at $\mathbf{u}(t + dt)$. We want to find the area swept out by the particle, half that of the parallelogram shown.

to Figure 4.3, the area of the triangle formed as the particle moves along the ellipse is given by

$$d\mathbf{a} = \frac{1}{2}\,\mathbf{u}(t) \times \mathbf{u}(t+dt) \tag{4.23}$$

where

$$\mathbf{u}(t) = r(t)\,\left[\cos\phi(t)\,\hat{\mathbf{x}} + \sin\phi(t)\,\hat{\mathbf{y}}\right], \tag{4.24}$$

and we can approximate the location at time $t + dt$ by Taylor expansion:

$$\mathbf{u}(t+dt) = \mathbf{u}(t) + \left[\dot{r}(t)\,\hat{\mathbf{u}}(t) + r(t)\,\dot{\phi}(t)\,\left(-\sin\phi(t)\,\hat{\mathbf{x}} + \cos\phi(t)\,\hat{\mathbf{y}}\right)\right] \tag{4.25}$$

so that

$$d\mathbf{a} = \frac{1}{2}\,\dot{\phi}\,r^2\,\hat{\mathbf{z}}\,dt = \frac{J_z}{2\,m}\,\hat{\mathbf{z}}\,dt \tag{4.26}$$

using $\dot{\phi} = J_z/(m\,r^2)$. Moving dt over, exploiting Leibniz notation, we get

$$\frac{d\mathbf{a}}{dt} = \frac{J_z}{2\,m}\,\hat{\mathbf{z}}, \tag{4.27}$$

a constant. Conclusion: an equal area is swept out by the orbiting body in equal time everywhere along its trajectory.

As for Kepler's third law, that is a statement about the relation between the period of the motion and the elliptical geometry. We want to find the time it takes an orbiting body to go all the way around once. We can get that from the relation $\dot{\phi} = J_z/(m\,r^2)$:

$$\frac{d\phi}{dt} = \frac{J_z}{m\,r^2} = \frac{J_z\,(1 + e\,\cos\phi)^2}{p^2\,m}. \tag{4.28}$$

Then write dt in terms of $d\phi$

$$dt = \frac{p^2\,m\,d\phi}{J_z\,(1 + e\,\cos\phi)^2}. \tag{4.29}$$

Integrating all the way around, from $\phi = 0$ to $2\,\pi$, gives the total period

$$T = \int_0^{2\pi} \frac{p^2\,m\,d\phi}{J_z\,(1 + e\,\cos\phi)^2} = \frac{p^2\,m}{J_z}\,\frac{2\,\pi}{\left(1 - e^2\right)^{3/2}} = \frac{J_z^3}{G^2\,m^3\,M^2}\,\frac{2\,\pi}{\left(1 - e^2\right)^{3/2}}. \tag{4.30}$$

Problem 4.4 For the ellipse with radial coordinate written in the form (4.22), what is the value of the perihelion and aphelion in terms of p and e? The "major" axis of the ellipse is its total horizontal extent (for the orientation we are using in Figure 4.3), and the "minor" axis is the total vertical extent. Write these in terms of p and e.

Problem 4.5 Write Kepler's third law (4.30) in terms of the semi-major (half the major) axis of the ellipse (and the mass of the central body, plus any constants that remain).

4.3.3 Tidal Forces

One of the implications of the equality of all of the inertial, passive, and active masses is the carefully orchestrated stretching and squeezing of extended bodies as they approach a central one. Suppose you were falling toward a spherically symmetric central object (the sun, for example). Different parts of your body would experience acceleration of different magnitudes. Your feet accelerate more than your head (if you're going in feet first), so you would get stretched. You would also get squeezed, since all the pieces of your body are falling in directly toward the center of the sun, and so your left and right sides would have to get closer together.

We can quantify the difference in acceleration from top to bottom and left to right. As is typical, we will model our body as a rectangle of width w and height h (our depth doesn't matter for the calculation, by symmetry). The setup, for a central body of mass M (spherical), is shown in Figure 4.4, with the accelerations of interest marked.

The difference in acceleration between the top and bottom is

$$\mathbf{a}_t - \mathbf{a}_b = -GM\left[\frac{1}{(r+h)^2} - \frac{1}{r^2}\right]\hat{\mathbf{z}} \approx -\frac{2GMh}{r^3}\hat{\mathbf{z}} \tag{4.31}$$

where we have assumed that $h \ll r$ in the approximation above. If the object is to fall rigidly, there must be internal force to oppose this acceleration discrepancy from the top to the bottom. That internal force is provided by your body, up to a point.

For the "squeezing" portion, we'll look at the difference between the right and left accelerations:

$$\begin{aligned}\mathbf{a}_r - \mathbf{a}_\ell &= -\frac{2GM}{\left(r^2 + \left(\frac{w}{2}\right)^2\right)}\sin\theta\,\hat{\mathbf{y}} = -\frac{GMw}{\left(r^2 + \left(\frac{w}{2}\right)^2\right)^{3/2}}\hat{\mathbf{y}} \\ &\approx -\frac{GMw}{r^3}\hat{\mathbf{y}},\end{aligned} \tag{4.32}$$

and once again we'd have to oppose this difference in accelerations in order to avoid a relative squeeze.

4.3.4 Cosmology

Cosmology is the study of the large-scale structure of the universe from a gravitational point of view. That gravity is the dominant force over large scales (spatial and temporal) is clear from its relative weakness. If, for example, there was a pocket of charge lying around in the universe, it would "quickly" (relative to gravity) attract opposite charge and neutralize.

Following a line of reasoning from Newton's time, we assume that the universe is infinite in extent (if the universe were made of a finite set of masses, it would collapse on itself and therefore have a finite lifetime). This immediately leads to a problem. For a universe that is uniformly filled with light-emitting stuff (stars) and is of infinite extent, light should be everywhere. The darkness of the night sky is a statement of Olbers's paradox: if stars are all around emitting light, the night sky should be (infinitely) bright.

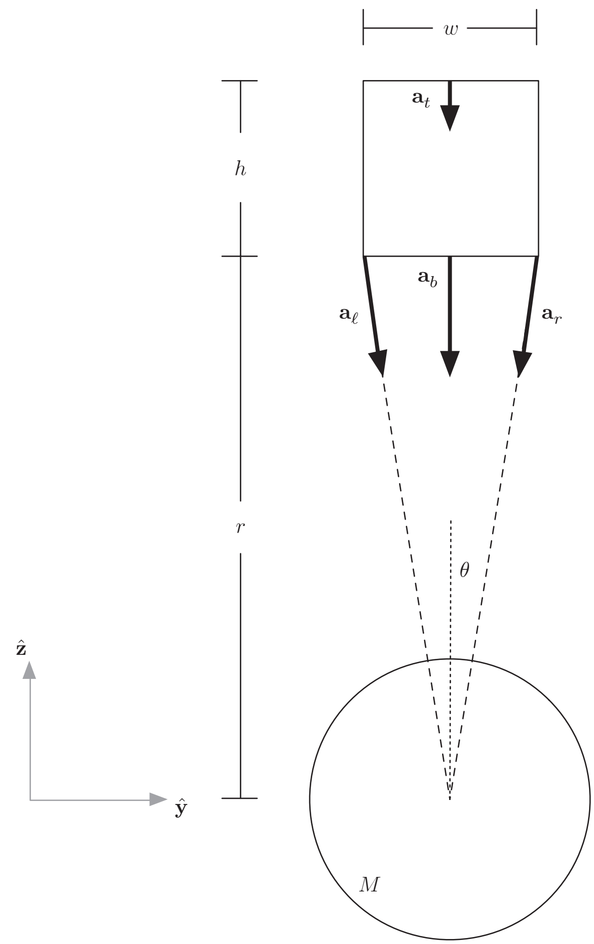

Fig. 4.4 A massive body of width w and height h falls toward a spherical central body of mass M (centered at the origin). We want to find the difference in accelerations from top to bottom and left to right.

As a simple model to establish Olbers's paradox we'll use point dipole sources for the radiation (those will be our single-frequency "stars") and calculate the intensity for an observer at the center of a universe filled with these sources.[4] Take an observer at the origin with a collector plate of area A with normal pointing in the $\hat{\mathbf{z}}$-direction. Put a single source with dipole moment $\mathbf{p}(t) = p_0 \cos(\omega t)\hat{\mathbf{p}}$ making an angle of θ with respect to the z-axis at location $\mathbf{r}'$, then the intensity at the origin, from (2.183) with $a = dw^2$ and $q = p_0/d$, is

$$\mathbf{I} = -\frac{1}{\mu_0 c}\left(\frac{\mu_0 p_0 \omega^2}{4\pi r'}\right)^2 \sin^2\theta\, \hat{\mathbf{r}}'. \tag{4.33}$$

[4] We'll be using the radiation field of the dipole and ignoring the near-field components. If that bothers you, start the integration over all space at a cutoff r_0, already far from the observer.

The power at the collector is $|\mathbf{I} \cdot \mathbf{A}|$, or

$$P = \frac{A \cos\theta'}{\mu_0 c} \left(\frac{\mu_0 p_0 \omega^2}{4\pi r'} \right)^2 \sin^2\theta \tag{4.34}$$

where θ' is the angle between $\mathbf{r}'$ and $\hat{\mathbf{z}}$. We'll average over all dipole angles θ, which gives a factor of $1/2$. The power at the detector, due to this "one" (angle and time-averaged) dipole is

$$P = \frac{A \cos\theta'}{2\mu_0 c} \left(\frac{\mu_0 p_0 \omega^2}{4\pi r'} \right)^2. \tag{4.35}$$

Assuming we have dipoles of equal strength p_0 uniformly distributed throughout the universe with density D_0, the total power collected (integrating over the upper hemisphere and doubling the result) is

$$\begin{aligned} P_{\text{tot}} &= \frac{A D_0}{\mu_0 c} \left(\frac{\mu_0 p_0 \omega^2}{4\pi} \right)^2 \int_0^\infty \int_0^{\pi/2} \int_0^{2\pi} \frac{\cos\theta'}{r'^2} r'^2 \sin\theta' \, d\phi' d\theta' dr' \\ &= \frac{A D_0 \pi}{\mu_0 c} \left(\frac{\mu_0 p_0 \omega^2}{4\pi} \right)^2 \int_0^\infty dr', \end{aligned} \tag{4.36}$$

and the final integral here is infinite. If the universe were filled with light-emitting dipoles, the power at any point would be infinite.

There are ways around this conclusion. Maybe the universe has large pockets of stars, and then mass that does not emit light, so that the night sky is really just the lack of light coming from the "empty" (of starlight) pieces (this seems implausible). Another option is that the "stuff" in the universe is moving, leading to a redshift of starlight out of the visible spectrum. Of course, we would need every source to be redshifted for this idea to work. The observation is that the galaxies in the universe *are* moving radially away from each other,[5] so the redshift of light is a plausible idea. To explain this universal relative motion, the standard picture is points on a balloon that is being inflated. As the balloon inflates, the points get farther away from one another, so that each point (galaxy) sees all the others receding. We'll start with a simple model, in which at time t, the universe has uniform density $\rho(t)$, and generate the function that describes the balloon's "inflation" as a function of time. This motion represents the dynamics of the universe itself.[6]

Newtonian gravity can be used to make predictions about the dynamics of the universe, provided we make some reasonable assumptions. We'll assume that the universe is everywhere "the same," so that there are no preferred points or directions. A preferred point is clearly out. We want mass distributed uniformly without clumping (the universe, then, has a spatially constant mass density ρ); that's what we see (at the large scale) in our vicinity, and there's no reason to imagine other locations are significantly different. No preferred directions means that there's not, for example, an electric field pointing in one direction throughout the universe.

[5] This is established by measuring the shift in known frequencies of light emitted by stars. The shift comes from the relative motion of the star and observer, as in Section 1.5.

[6] There is, of course, random relative motion of galaxies, which is in addition to the universal expansion.

For a sphere of uniform density ρ, we know that a particle of mass m located a distance r from the center of the sphere experiences a force due to all the mass enclosed by the sphere of radius r, while the mass at points $> r$ does not contribute; that's the "shell" theorem familiar from Newtonian gravity (and E&M, common to all $1/r^2$ forces). At time t, the total mass enclosed by a sphere of radius r is $M = \frac{4}{3}\pi r^3 \rho(t)$. A particle of mass m at $\mathbf{r} = r\,\hat{\mathbf{r}}$ has total energy

$$E = \frac{1}{2} m \dot{\mathbf{r}} \cdot \dot{\mathbf{r}} - \frac{G M m}{r} = \frac{1}{2} m \dot{\mathbf{r}} \cdot \dot{\mathbf{r}} - \frac{4}{3} G m \pi r^2 \rho. \tag{4.37}$$

The particle of mass m can go toward or away from the origin, in this scenario, so that its position at time t can be thought of as a function $\mathbf{r}(t) = S(t)\,\mathbf{w}$ where $S(t_0) = 1$ for a particle with initial vector location $\mathbf{w}$ (and note that this assumed form affects all particles at radius r at time t in the same way). So in general, the energy (which is, of course, constant) is given by

$$E = \frac{1}{2} m w^2 \dot{S}^2 - \frac{4}{3} \pi G m w^2 \rho S^2, \tag{4.38}$$

and this is typically rearranged to read

$$\left(\frac{\dot{S}(t)}{S(t)}\right)^2 = \frac{2E}{m w^2 S(t)^2} + \frac{8}{3}\pi G \rho(t) = \frac{2E}{m r(t)^2} + \frac{8}{3}\pi G \rho(t), \tag{4.39}$$

which is known as Friedmann's equation.

We write the equation in terms of $\dot{S}/S$ because of the interpretation of this quantity. Suppose we have two points on a sphere, $\mathbf{r}_1 = S(t)\,\mathbf{w}_1$ and $\mathbf{r}_2 = S(t)\,\mathbf{w}_2$ (where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the locations at time t_0). The vector that points from 2 to 1 is given by $\mathbf{r}_{12} \equiv S(t)\,(\mathbf{w}_1 - \mathbf{w}_2)$. The magnitude of this vector is the distance between the two points at time t. Then $\dot{\mathbf{r}}_{12}$ gives the relative velocity of the two points on the sphere, and we can relate it to the distance vector at time t via

$$\dot{\mathbf{r}}_{12} = \dot{S}(t)\left(\frac{1}{S(t)}\,\mathbf{r}_1 - \frac{1}{S(t)}\,\mathbf{r}_2\right) = \frac{\dot{S}(t)}{S(t)}\,\mathbf{r}_{12}, \tag{4.40}$$

so that $\dot{S}/S$ is the relation between the relative speed of the two points and their distance apart at time t. This combination, $H \equiv \dot{S}/S$, is known as Hubble's constant, and as the time-dependence suggests, it is not necessarily a constant.

A feature of this pure gravitational model is that the universe cannot be static. This was problematic (before people realized that the universe can't be static), yet there is no natural fix. One way to "fudge" the original (4.37) is to introduce an additional potential energy term to counteract the gravitational attraction, but this new term must be quadratic in r – the energy E that is constant in time must go like r^2 so that (4.39) has *no* r-dependence (the left-hand side is r-independent, so the right-hand side must be as well). That means the only thing you could add would be something like $\Lambda r^2 = \Lambda w^2 S(t)^2$ for constant Λ, and this is the so-called cosmological constant. It corresponds to a new force, linear in r, that acts in addition to gravity. Start with

$$E = \frac{1}{2} m w^2 \dot{S}^2 - \frac{4}{3} \pi G m w^2 \rho S^2 + \frac{1}{6} \Lambda w^2 S^2, \tag{4.41}$$

then you get an updated form of the Friedmann equation:

$$\left(\frac{\dot{S}(t)}{S(t)}\right)^2 = \frac{2\,E}{m\,r(t)^2} + \frac{8}{3}\,\pi\,G\,\rho(t) + \frac{1}{3}\,\frac{\Lambda}{m}. \tag{4.42}$$

Problem 4.6 Look up the current value of Hubble's constant. Use it to find the redshift: $z \equiv (f - \bar{f})/\bar{f}$ for emitted frequency f and observed (on earth) frequency $\bar{f}$ assuming the light was emitted from a galaxy a distance $d = 2.5$ million light years away.

4.4 Predictions: Relativistic

There are a variety of interesting predictions that we can make by introducing a "little" special relativity, and most of these are borne out in the full theory of gravity. The following requires a slight suspension of disbelief, but it's worth it for building physical intuition about general relativity.

4.4.1 Event Horizon

When we define the escape speed of an object in a potential energy landscape, we use the minimum launch speed necessary for the object to get to spatial infinity, which means we want the speed of the object to be zero at spatial infinity. For example, if a mass m blasts off from the surface of a spherically symmetric central body of mass M and radius R at speed v_{esc}, the total energy is

$$E = \frac{1}{2}\,m\,v_{\text{esc}}^2 - \frac{G\,M\,m}{R}. \tag{4.43}$$

At spatial infinity, the potential energy term is zero, and we assume the kinetic term is also zero (the mass is at rest out there as in Figure 4.5). So, $E = 0$, and we can solve (4.43) for v_{esc}:

$$v_{\text{esc}} = \sqrt{\frac{2\,G\,M}{R}}. \tag{4.44}$$

Notice that if you shrink down the radius, R, of the central body, the escape speed increases. Then we can turn the escape speed problem around and ask: for what spherical radius R is $v_{\text{esc}} = c$ the escape speed? From (4.44), with $v_{\text{esc}} = c$ (although the logic is flawed; can you spot the hole?), the answer is

$$R = \frac{2\,G\,M}{c^2}. \tag{4.45}$$

For massive spheres of this radius or less, then, even light would not escape the gravitational attraction of the central body.

This radius, first predicted by John Michell in 1784 and independently by Laplace in 1799, represents the "event horizon" of a massive body. If all of the mass of a spherical body is squeezed beneath this radius, light cannot escape from its surface to infinity. This

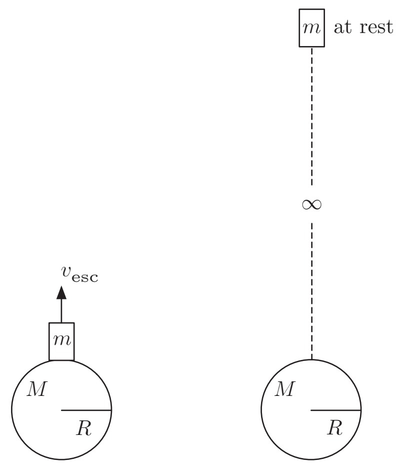

Fig. 4.5 The initial configuration on the left: a mass m leaves the surface of a spherically symmetric body of mass M and radius R with speed v_{esc}. On the right: the mass escapes to spatial infinity where it is at rest.

is an example of a "black hole" and is both qualitatively and quantitatively correct;[7] the result from general relativity is identical (there, this radius is known as the "Schwarzschild radius," and it shows up in a very different way).

Problem 4.7 What is the Schwarzschild radius for the sun? For the earth? Is the mass of the sun packed into a sphere of radius less than or equal to its Schwarzschild radius?

4.4.2 Bending of Light

Light will respond to a gravitational field. This is plausible: light has energy associated with it, and so has some effective "mass," via $m = E/c^2$, that could interact with a gravitational field. Start with the general solution to the gravitational equations of motion, written in $\rho(\phi)$ form as in (4.20). There was nothing explicitly non-luminal about that motion's setup, so we'll use it, assuming motion in the x–y plane as before. The setup is shown in Figure 4.6. Take $M = 0$ for a moment, then we know the motion must be a straight line, and we'll arrange that line to graze the central body right at R (corresponding to the straight, solid line in Figure 4.6), in this case, we can set $A = 0$ and $B = 1/R$ in (4.20) (so that $r = 1/\rho = R$ when $M = 0$ there). If we wrote out the x and y coordinates of this motion, we'd have

$$x = R \cot\phi, \qquad y = R, \tag{4.46}$$

which does indeed describe a line a distance R from the x-axis – the $\cot\phi$ parameter goes from $-\infty$ to ∞ as ϕ goes from π to 0 while y is fixed at R. Using this $M = 0$ description of the line in (4.20), we have

[7] This is unfortunate. Ironically, I would prefer if the result was qualitatively correct but was off by a factor of 2 or 4, as most pseudo-classical results in gravity are.

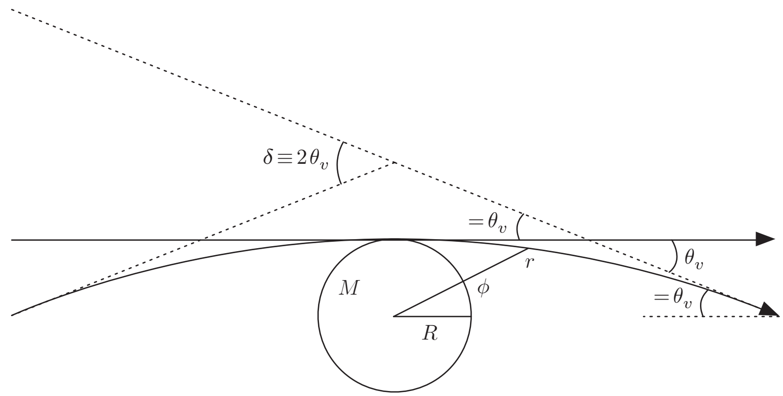

Fig. 4.6 The motion of a particle near a central body. The curved solid line is an approximate trajectory, and the solid straight line is the trajectory if there were no central mass.

$$\rho(\phi) = \frac{1}{R}\sin\phi + \frac{G\,M\,m^2}{J_z^2}. \tag{4.47}$$

Suppose we are at spatial infinity in the x-direction where $\rho = 1/\infty = 0$, and the massive object arrives at our location making an angle θ_v as shown in Figure 4.6; that angle would be defined by

$$0 = \frac{1}{R}\sin(-\theta_v) + \frac{G\,M\,m^2}{J_z^2} \longrightarrow \theta_v = \sin^{-1}\left(\frac{G\,M\,m^2\,R}{J_z^2}\right). \tag{4.48}$$

We can define the "deflection angle," δ, for the motion. This is the difference between the actual starting point of the trajectory and the straight line defined by our viewing angle (its value is $\delta = 2\,\theta_v$, referring to Figure 4.6). The deflection angle is a strange quantity for massive particle motion since we have no reason to assume the particle traveled in a straight line. But for light, we always assume constant velocity vector with speed c, so $2\,\theta_v$ is the difference between the perceived location of the particle (of light) and the actual location.[8]

So far, there has been no reference to the type of particle we have. It is with the definition of the constant J_z that we can actually tailor our solution to refer to light. At the $\phi = \pi/2$ grazing point, all of the motion is in the $\hat{\boldsymbol{\phi}}$-direction, and the light is traveling at speed c, so $J_z = m\,R^2\,\dot{\phi} = m\,R\,c$, and this must be its value at all times by conservation of angular momentum. We can put this into our deflection angle expression,

$$\delta = 2\,\theta_v = 2\,\sin^{-1}\left(\frac{G\,M}{R\,c^2}\right) \approx 2\,\frac{G\,M}{R\,c^2}, \tag{4.49}$$

where the approximation comes from assuming that $G\,M/(R\,c^2) \ll 1$.

The point of all this is that we would say that light moves along trajectories that are not straight lines in the presence of Newtonian gravity. The defining characteristic of light,

[8] If someone asks you to point to the location of a particular star, you point directly at the star – that's the straight-line assumption we normally make about light's path.

here, is that it travels at c. This result, that light is bent by gravity, again holds in general relativity, although the current prediction for the deflection angle is off by a factor of 2 (thank goodness!).

Problem 4.8 Is the approximation $G M/(R c^2) \ll 1$ justified for the sun? What is the deflection angle δ in this case?

Problem 4.9 Find the radius at which light can orbit in a circle around a (spherical) central body of mass M. Why would you never see this happen in this Newtonian setting?[9]

4.4.3 Energy Coupling and Charge

Moving along to some more dramatic modifications that are motivated by special relativity, suppose we go back and reimagine the original field equation,

$$\nabla \cdot \mathbf{g} = -4\pi G \rho, \tag{4.50}$$

for ρ, a mass density. Special relativity provides a relation between mass and energy, so that if you had a source of *energy* density u, you could write the effective mass density of it as $\rho = u/c^2$, and that energy density would show up in the field equation

$$\nabla \cdot \mathbf{g} = -4\pi G \frac{u}{c^2}. \tag{4.51}$$

Consider a point charge with mass M and charge Q. An electric field has energy density $u = \frac{1}{2}\epsilon_0 E^2$, so for a point source's field, the energy density is

$$u = \frac{Q^2}{32\pi^2 \epsilon_0 r^4}. \tag{4.52}$$

We assume that the gravitational field is spherically symmetric, $\mathbf{g}(\mathbf{r}) = g(r)\,\hat{\mathbf{r}}$, and then we have to solve the ordinary differential equation (ODE), from (4.51),

$$\frac{1}{r^2}\left(r^2 g(r)\right)' = -\frac{G Q^2}{8\pi \epsilon_0 c^2 r^4}, \tag{4.53}$$

which gives[10]

$$\mathbf{g} = \frac{G Q^2}{8\pi \epsilon_0 c^2 r^3}\,\hat{\mathbf{r}}. \tag{4.54}$$

Note that this is just the contribution due to the extended energy source; we must also include the massive point-source (with $\nabla \cdot \mathbf{g} = 0$ away from the origin) piece, so the full field is really

[9] In general relativity, these circular orbits exist and are visible.

[10] You may be tempted to integrate (4.51) to get an updated integral form for Gauss's law. Be careful: the energy stored in an electric field is not integrable (for a point particle). We have swept the issue under the rug by focusing on spherically symmetric solutions to the PDE, where we omit the $1/r^2$ term that would come from putting a finite cut-off on the particle radius, along with the $1/r^2$ term that comes from the normal Newtonian point source.

$$\mathbf{g} = \left(-\frac{G\,M}{r^2} + \frac{G\,Q^2}{8\,\pi\,\epsilon_0\,c^2\,r^3}\right)\hat{\mathbf{r}}. \tag{4.55}$$

It is interesting that in theory you could detect the charge (up to sign) of a body by measuring its gravitational force: take a neutral mass m moving with speed v in a circular orbit around a point source at radius R. The charge of the central body is given by

$$Q = \pm\sqrt{8\,\pi\,\epsilon_0\,c^2\,r\left(M + \frac{R\,v^2}{G}\right)}. \tag{4.56}$$

Of course, this presumes we know the mass M of the source; otherwise, we have no way to separate the effective mass contribution $R\,v^2/G$ from M.

Problem 4.10 What is the gravitational potential associated with (4.55)? What would the classical event horizon look like here?

Problem 4.11 Suppose you had a magnetic monopole, giving a magnetic field of the form: $\mathbf{B} = \frac{q_m}{r^2}\,\hat{\mathbf{r}}$. Find the gravitational field that the energy stored in this magnetic field would generate. This provides a magnetic (monopole) analogue to (4.54).

Problem 4.12 Just as the electric field stores energy, the gravitational field does as well. Develop the expression for the energy density $u \sim \mathbf{g}\cdot\mathbf{g}$.

Problem 4.13 How long would it take a test particle of mass m to fall from rest at R to $\frac{1}{2}\,R$ moving under the influence of (4.55)? (Use non-relativistic dynamics.)

4.4.4 Perihelion Precession

We know, from Bertrand's theorem (see, e.g., [17]), that universally closed stable orbits can come only from $1/r$ or r^2 potentials. The potential associated with (4.55) has an extra $1/r^2$ piece. But it is easy to get potential contributions that are of the form $1/r^n$ from less exotic considerations. The multipole moments of a central body start with $1/r$ (monopole), then have $\cos\theta/r^2$ (dipole) etc., so by including contributions from the various multipole moments, we can get different radial (and angular) dependence.

We can consider a small perturbing potential (from whatever mechanism) on top of the dominant central body Newtonian potential. Take a test particle of mass m moving with potential energy described by

$$U(r) = -\frac{G\,M\,m}{r} + V(r) \tag{4.57}$$

where $V(r) \ll \frac{G\,M\,m}{r}$ for all r. Going back to the ϕ-parametrized equation of motion for $\rho \equiv 1/r$ that we developed in (4.19), we have letting $U'(\rho) \equiv \frac{dU}{d\rho}$, $V'(\rho) \equiv \frac{dV}{d\rho}$

$$\rho'' = -\rho - \frac{m}{J_z^2}\,U'(\rho) = -\rho + \frac{G\,M\,m^2}{J_z^2} - \frac{m}{J_z^2}\,V'(\rho). \tag{4.58}$$

For $V = 0$, we get back (4.22), the original elliptical orbits with $p = J_z^2/(G\,M\,m^2)$.

With the perturbing potential in place, the ellipse description changes. Let $\rho = \rho_0 + \epsilon\,\rho_1(\phi)$ for constant ρ_0 (we are perturbing about a circular orbit), then $V(\rho_0 + \epsilon\,\rho_1) \approx V(\rho_0) + \epsilon\,V'(\rho_0)\,\rho_1$. We can insert this expansion into (4.58) and collect in powers of ϵ:

$$\begin{aligned} \epsilon^0 &: 0 = -\rho_0 + \frac{1}{p} - \frac{m}{J_z^2}\,V'(\rho_0) \\ \epsilon^1 &: \rho_1'' = -\rho_1\left(1 + \frac{m}{J_z^2}\,V''(\rho_0)\right). \end{aligned} \tag{4.59}$$

The first equation can, in principle, be solved for the constant ρ_0. The second equation is solved by

$$\rho_1 = A\,\cos\left(\sqrt{1 + \frac{m}{J_z^2}\,V''(\rho_0)}\,\phi\right), \tag{4.60}$$

again keeping just one of the two (cosine and sine) solutions. Then our full solution looks like

$$\rho = \rho_0 + A\,\epsilon\,\cos\left(\sqrt{1 + \frac{m}{J_z^2}\,V''(\rho_0)}\,\phi\right) \tag{4.61}$$

where ρ_0 is just a constant. This gives a radius that is of the form

$$r = \frac{1}{\rho} = \frac{\beta}{1 + \tilde{\epsilon}\,\cos\left(\sqrt{1 + \frac{m}{J_z^2}\,V''(\rho_0)}\,\phi\right)} \tag{4.62}$$

for constants β and $\tilde{\epsilon}$, a new small parameter $\sim \epsilon$.

Our solution would be a closed ellipse if $V''(\rho_0) = 0$. With $V''(\rho_0) \neq 0$, the form of $r(\phi)$ does define an elliptical trajectory of sorts. Look at the point of closest approach, though. At $\phi = 0$, we start off at $\beta/(1 + \tilde{\epsilon})$, but we don't *return* to this location until

$$\sqrt{1 + \frac{m}{J_z^2}\,V''(\rho_0)}\,\phi = 2\pi, \tag{4.63}$$

and that gives

$$\phi = \frac{2\pi}{\sqrt{1 + \frac{m}{J_z^2}\,V''(\rho_0)}} \approx 2\pi\left(1 - \frac{m}{2J_z^2}\,V''(\rho_0)\right) \equiv 2\pi - \psi \tag{4.64}$$

where the approximation comes from assuming that $m\,V''(\rho_0)/J_z^2$ is small. We return, then, to the same closest approach distance, but not at $\phi = 2\pi$; instead, it's a little less than that (for $V''(\rho_0) > 0$), at $2\pi - \psi$ with $\psi = \pi\,m\,V''(\rho_0)/J_z^2$. This motion of the perihelion is known as "precession," and it occurs for any potential with $V''(\rho_0) \neq 0$. At each pass, we maintain an elliptical shape but pick up an additional deficit of ψ. The picture, for the first two orbits, is shown in Figure 4.7.

The story can be complicated further by considering the angular dependence of perturbing potentials, i.e., additional terms of the form $V(\rho, \theta)$, but even this simplest case makes the point. In particular, you could use the measured precession to pin down the charge of a central body as in the previous section. General relativity predicts an

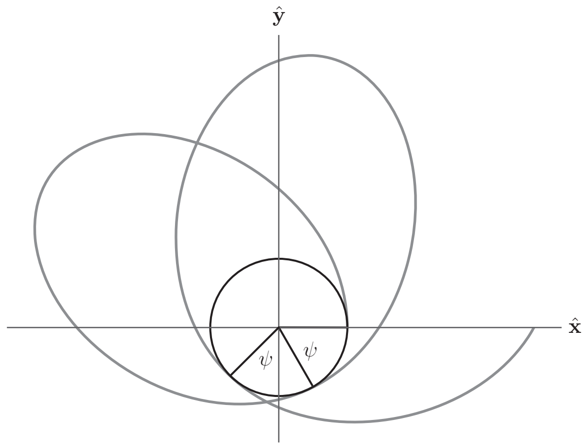

Fig. 4.7 A particle starts on the $\hat{\mathbf{x}}$-axis and moves around in an ellipse, returning to the point of closest approach (drawn as a circle of radius $p/(1-e)$) at an angle $2\pi - \psi$, as shown.

effective $V(\rho) = -G M J_z^2 \rho^3/(m c^2)$, and that can be treated perturbatively to calculate, for example, the perihelion precession of Mercury.

Problem 4.14 Use $V(\rho) = -G M J_z^2 \rho^3/(m c^2)$ in $\psi = \pi m V''(\rho_0)/J_z^2$ to find the advance in the precession of the perihelion of Mercury. Hint: first find the circular orbit radius ρ_0 from the top equation in (4.59), and write it in terms of $G M/c^2$ and p (the semilatus rectum of the orbit). Then use the orbital parameters of Mercury to evaluate ψ.

4.5 Issues

We have looked at some of the physical predictions of Newtonian gravity, and they are not so different qualitatively from the predictions of general relativity, especially once we introduce a little special relativity. That introduction, however, brings up some issues with the form of Newtonian gravity.

4.5.1 Lorentz Covariant Form

Our immediate objection is with the form of the field equation. Working with the potential, the field equation is $\nabla^2\phi = 4\pi G\rho$, and this depends on ∇^2, which is not a Lorentz scalar. The fix is easy, though: we just move to $\Box \equiv \partial_\mu \partial^\mu$, and the static limit is recovered trivially. Similarly, we know how to update Newton's second law so as to work with special relativity. Our field equation and equation of motion for the potential ϕ of Newtonian gravity now read

$$\begin{gathered}\Box\phi = 4\pi G\rho \\ \frac{d}{dt}\left[\frac{m\dot{\mathbf{r}}}{\sqrt{1-\frac{\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}}{c^2}}}\right] = -m\nabla\phi.\end{gathered} \tag{4.65}$$

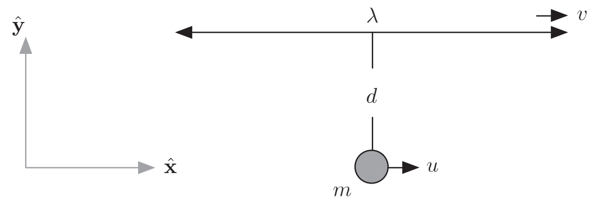

Fig. 4.8 An infinite line of mass moves with constant speed v. It exerts a force on a test mass moving with speed u.

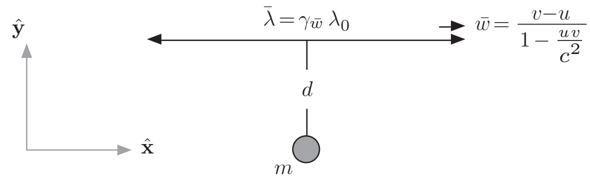

Fig. 4.9 Infinite line of mass viewed in the test particle's rest frame.

The field equation now comes, correctly, from a scalar Lagrangian density that looks like

$$\mathcal{L} = \frac{1}{4\pi G}\frac{1}{2}\,\partial_\mu\phi\,\partial^\mu\phi + \rho\,\phi. \tag{4.66}$$

There is still a relatively serious relativistic problem, though. Think back to the infinite line of charge from E&M described in Section 1.9, and let's tell the story in the context of gravity. In a lab frame, we have an infinite line of mass moving with constant speed v and with constant mass-per-unit-length $\lambda = \gamma_v\,\lambda_0$, where λ_0 is the rest mass density of the line[11] and $\gamma_v \equiv (1 - v^2/c^2)^{-1/2}$. The setup is shown in Figure 4.8. The force on the test particle is

$$\mathbf{F} = \gamma_v\,\frac{2\,G\,\lambda_0\,m}{d}\,\hat{\mathbf{y}}. \tag{4.67}$$

If we analyze the problem in the test particle's rest frame, as in Figure 4.9, where its mass density is $\bar{\lambda} = \gamma_{\bar{w}}\,\lambda_0$, we'd get a force

$$\bar{\mathbf{F}} = \gamma_{\bar{w}}\,\frac{2\,G\,\lambda_0\,m}{d}\,\hat{\mathbf{y}}. \tag{4.68}$$

The problem is that according to the transformation law for forces in special relativity from (1.86), we must have

$$\mathbf{F} = \frac{1}{\gamma_u}\,\bar{\mathbf{F}} \tag{4.69}$$

as the relation between the force in the lab frame and the force in the rest frame. Instead, we have

[11] The mass-per-unit-length has two factors of γ_v associated with it, as we shall see in Problem 4.17 and discuss in Section 4.5.3. One factor comes from the usual length contraction as in E&M, but we get another factor from energy density considerations. For now, we are just copying the electrostatic argument to make a point about vector form. We will later argue that gravity cannot be a vector theory precisely because of the (mass or energy) source transformation.

$$\mathbf{F} = \frac{\gamma_v}{\gamma_{\bar{w}}}\,\bar{\mathbf{F}} \neq \frac{1}{\gamma_u}\,\bar{\mathbf{F}}. \tag{4.70}$$

What's missing from the gravitational picture is, of course, a magnetic counterpart. This argument implies the existence of a gravitational "magnetic" field and force that mimics the usual magnetic one.

Working from the rest frame force, and using the correct transformation (4.69), we must have

$$\mathbf{F} = \frac{\gamma_{\bar{w}}}{\gamma_u}\frac{2\,G\,\lambda_0\,m}{d}\,\hat{\mathbf{y}} = \gamma_v\,\frac{2\,G\,\lambda_0\,m}{d}\left(1 - \frac{u\,v}{c^2}\right)\hat{\mathbf{y}} \tag{4.71}$$

(using (1.96)) in the lab (not (4.67)). We could get this type of force if there was a gravitomagnetic field $\mathbf{h}$ with $\nabla \times \mathbf{h} = -4\,\pi\,G/c^2\,\mathbf{J}$ for mass–current density $\mathbf{J} = \rho\,\mathbf{v}$ and if this gravitomagnetic force acted on moving masses via $\mathbf{F}_h = m\,\mathbf{v} \times \mathbf{h}$.

Problem 4.15 Find $\mathbf{h}$ for the infinite line of mass with mass density λ traveling with constant speed v (λ is the mass density of the moving line of mass).

Problem 4.16 The curl of $\mathbf{h}$ is sourced by moving mass/energy. The divergence of $\mathbf{h}$ is zero, as with the magnetic field: $\nabla \cdot \mathbf{h} = 0$. Introduce a (vector) potential for $\mathbf{h}$ and combine with the potential for Newtonian gravity to form a four-vector with $\sim \phi$ as the zero component, call it A^μ, and write its field equation. Given how this must couple to particles (as $A_\mu\,\dot{x}^\mu$), find the equation of motion of a particle of mass m moving under the influence of A^μ.

Problem 4.17 An infinite massive (electrically neutral) wire lies at rest along the $\hat{\mathbf{x}}$-axis. It has an energy density (one dimensional) of $\bar{u} = m\,c^2/\Delta\bar{x}$ at rest.

a. If the wire moves with constant velocity $\mathbf{v} = v\,\hat{\mathbf{x}}$ through a lab, what is the wire's energy density in the lab?

b. In the rest frame of the wire, its stress tensor is

$$\bar{T}^{\mu\nu} \doteq \begin{pmatrix} \bar{u} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.72}$$

Using the fact that the stress tensor is a second-rank contravariant tensor, find its value in the lab through which the wire is moving; i.e., find $T^{\mu\nu}$. Does the energy density component of the stress tensor in the lab match your result from part (a)?

Once $\mathbf{h}$ is on the scene, we have an augmented Newtonian theory of gravity that looks just like Maxwell's equations with a few sign changes. Again, this result, that gravity looks like E&M, is a prediction of general relativity. It is precisely (up to factors of two) the "weak-field" regime of linearized gravity, structurally identical to Maxwell's equations, and all of the physics you get from those carries over. That means that wrong-signed E&M provides great intuition for the weak-field (far away from sources) regime of gravity – close enough to be different from Newtonian gravity, but far enough away to avoid the full general relativistic description.

There is a fundamental difference between the sources in E&M and gravity, one that spoils the above story slightly (but also provides a provocative hint about the general relativistic fix). When we have a moving line of charge, we have to account for the length contraction of special relativity, so that if a line of charge has constant λ_0 when at rest, it has constant $\lambda = \gamma\,\lambda_0$ when moving. That's not quite right gravitationally, though. For an infinite line of mass with constant λ_0, when the line of mass is moving, there are two effects: length contraction, and the energy (which is really the source here) picks up a factor of γ. We'll come back to this, but file it away for now. It's an indication that this "wrong-signed" E&M approximation can't be the end of the story.

4.5.2 Self-Sourcing

There is a major component of gravity missing both from the Newtonian theory and its naïve relativistic (vector) extension. We saw how the energy stored in the electric field could act as a source for gravity, but what about the energy stored in the gravitational field itself? That should be counted, too. The first pass, which involves taking the energy required to assemble a configuration of mass, gives an energy density of $u = -\frac{1}{8\pi G}\,\mathbf{g}\cdot\mathbf{g}$ or, in terms of the gravitational potential, $u = -\frac{1}{8\pi G}\,\nabla\phi\cdot\nabla\phi$. These negative energy densities are easy to understand in terms of the sign of the work we do in building, say, a spherical shell of charge (we have to push away from the partially built configuration, as it attracts mass), but it's harder to think about as a gravitational source where it is akin to negative mass sources.

Still, if we simply plow ahead, given mass density ρ, we would write the (static) field equation as[12]

$$\nabla^2\phi = 4\pi\,G\,\rho - \frac{1}{2\,c^2}\,\nabla\phi\cdot\nabla\phi. \tag{4.73}$$

Problem 4.18 What is the spherically symmetric vacuum (i.e., $\rho = 0$) solution to (4.73)?

The bigger problem, if you go back to the derivation of the work required to build a configuration of mass (as in Problem 4.12), is that we assume that the field equation itself is just $\nabla^2\phi = 4\pi\,G\,\rho$ in developing the expression for u, and this is no longer the field equation. You can generate an infinite chain and sum it to find a consistent expression for the energy required to build a configuration, but even that falls short since it still does not really count the field energy. We know how to get at that quantity in a field theory: it's the 00 component of the stress tensor for the theory.

Suppose we start with the free relativistic Lagrangian,[13]

$$\mathcal{L} = \frac{1}{2}\,\frac{f(\phi)}{4\pi\,G}\,\partial_\mu\phi\,\eta^{\mu\nu}\,\partial_\nu\phi, \tag{4.74}$$

[12] This expression, and a discussion of its physical interpretation, can be found in [23].
[13] This line of argument comes from [12].

for arbitrary function $f(\phi)$. The stress tensor is given by

$$T^{\mu\nu} = \frac{f(\phi)}{4\pi G}\left[\phi_{,}{}^{\mu}\,\phi_{,}{}^{\nu} - \frac{1}{2}\eta^{\mu\nu}\,\phi_{,}{}^{\alpha}\,\phi_{,\alpha}\right], \tag{4.75}$$

and the 00 component, for *static* fields (with $\dot{\phi} = 0$), is

$$T^{00} = \frac{1}{2}\frac{f(\phi)}{4\pi G}\nabla\phi\cdot\nabla\phi = \mathcal{L} \tag{4.76}$$

so that the Lagrange density *is* the energy density here.

In order to require self-consistent self-coupling, our target will be a vacuum field equation of the form: $\nabla^2\phi = 4\pi G\frac{T^{00}}{c^2} = 4\pi G\frac{\mathcal{L}}{c^2}$. The field equation, from the Lagrangian (4.74) (and again staying with static fields) is

$$f(\phi)\,\nabla^2\phi + \frac{1}{2}f'(\phi)\,\nabla\phi\cdot\nabla\phi = 0 \tag{4.77}$$

and we could write this as

$$\nabla^2\phi = -\frac{1}{2}\frac{f'(\phi)}{f(\phi)}\nabla\phi\cdot\nabla\phi. \tag{4.78}$$

We want the right-hand side to equal $4\pi G\frac{\mathcal{L}}{c^2}$, giving us an ODE for $f(\phi)$:

$$-\frac{1}{2}\frac{f'(\phi)}{f(\phi)}\nabla\phi\cdot\nabla\phi = \frac{1}{2c^2}f(\phi)\,\nabla\phi\cdot\nabla\phi \longrightarrow f'(\phi) = -\frac{1}{c^2}f(\phi) \tag{4.79}$$

with solution $f(\phi) = \frac{c^2}{\phi}$.

The field equation, in vacuum, is then

$$\nabla^2\phi = \frac{1}{2\phi}\nabla\phi\cdot\nabla\phi, \tag{4.80}$$

or if we introduce material sources with mass density ρ (or energetic ones other than gravity), the field equation can be written, from [10],[14] as

$$\nabla^2\left(\sqrt{\phi}\right) = \frac{2\pi G}{c^2}\rho\sqrt{\phi}. \tag{4.81}$$

Problem 4.19 Use the fixed form of $f(\phi)$ in the Lagrange density (4.74) with external sources, ρ, to find the field equation that updates (4.80) to include sources. Show that this is equivalent to the compact (and linear in $\sqrt{\phi}$) form (4.81).

Problem 4.20 Find the spherically symmetric vacuum solution in the self-coupled theory (4.81).

Problem 4.21 A sphere of radius R has constant ρ_0 inside. Find the interior and exterior solutions to (4.81) (assume continuity and derivative continuity for ϕ).

Problem 4.22 Find the energy density, u, for the gravitational field of a point particle in the self-coupled case. What is the total energy stored in the field?

[14] Indeed, this self-coupled scalar theory was introduced by Einstein in 1912 as a precursor to general relativity.

4.5.3 Universal Coupling

We'll conclude by thinking about what the full theory of gravity (as it is currently known) must look like given the deficiencies described in the previous subsection. Our first question is what type of field theory (for the gravitational "potential") we should be investigating. Newtonian gravity starts as a scalar field, then we expanded the gravitational force to include a gravitomagnetic component, based on special relativity. The resulting theory of gravity is a *vector* field theory, like E&M. Finally, we considered self-coupling of gravity back in its scalar form. Each has flaws, but which, if any of these, is the correct starting point?

Let's think about sources first. In E&M, the source is charge and moving charge, combined in a current J^μ that is conserved: $\partial_\mu J^\mu = 0$ (a statement of charge conservation). For gravity, the source is energy and moving energy at the very least. What we want, then, is a structure that combines energy and moving energy and is conserved. The natural object is the stress-energy tensor, $T^{\mu\nu}$, since T^{00} is the energy density and T^{0i} the "moving energy" density. As an added bonus, we know that *every* physical theory has a conserved stress-energy tensor owing to the equivalence of physics in all coordinate systems (the definition and conservation are developed in Section 3.4), making it an ideal candidate for the universal coupling of gravity (every form of mass/energy sources and responds to gravity). So $T^{\mu\nu}$ is always available, and $\partial_\mu T^{\mu\nu} = 0$ is a conservation statement.

Back to the form of gravity: can a scalar be made out of $T^{\mu\nu}$? Sure, the trace of $T^{\mu\nu}$ is a scalar: $T \equiv T^\mu_\mu = T^{\mu\nu}\,\eta_{\mu\nu}$. In a sense, that was the motivation for making a self-consistent scalar theory of gravity. In the previous section, we coupled the field to its energy density, T^{00}. But for static fields, the trace of the stress tensor is directly proportional to T^{00}. The problem with a scalar theory is that, at least in four-dimensional space-time, E&M has $T = 0$, meaning that a scalar theory of gravity would have no way of coupling to E&M. But, as energy carriers, electromagnetic fields must source gravity. You might be tempted to use just the energy density from the electric field as we did in the static case, but that would not lead to field equations that are generally covariant (meaning that they have the same form in all coordinate systems), and there would be preferred frames. If you started with

$$\Box\phi = 4\pi G\,\frac{T^{00}}{c^2}, \tag{4.82}$$

where T^{00} is the energy density from E&M, then a Lorentz boost would change the value of T^{00} while leaving the rest of the equation unchanged. So you'd have a preferred reference frame in which to evaluate T^{00}, and all others would pick up factors of γ^2. That is not in keeping with the first postulate of special relativity; the form of the equation would depend on your inertial frame.

So a pure scalar field is out, how about vector? It's not hard to see that there is no single vector component of $T^{\mu\nu}$, no way to massage the coordinates so as to generate an object that transforms with one factor of the Lorentz boost (for example). The single-indexed $T^{0\nu}$, which appears to be available, does not transform correctly. Consider a boost with matrix entries $L^\mu_{\ \nu}$; the stress tensor transforms as

$$\bar{T}^{\mu\nu} = T^{\alpha\beta} L^{\mu}_{\ \alpha} L^{\nu}_{\ \beta}, \tag{4.83}$$

and then

$$\bar{T}^{0\nu} = T^{\alpha\beta} L^{0}_{\ \alpha} L^{\nu}_{\ \beta} \neq T^{0\beta} L^{\nu}_{\ \beta} \tag{4.84}$$

where the inequality at the end is what you would need (as equality) to make a vector field theory like

$$\Box h^{\nu} = 4\pi G \frac{T^{0\nu}}{c^2} \tag{4.85}$$

generally covariant.

Evidently, the most natural description of a theory of gravity is as a symmetric second-rank tensor, one that couples to the complete stress tensor, something like

$$\Box h^{\mu\nu} = \frac{4\pi G}{c^2} T^{\mu\nu}, \tag{4.86}$$

or if we wanted to consider a dimensionless field ($h^{\mu\nu}$ has dimension of speed squared),

$$\Box\left(\frac{h^{\mu\nu}}{c^2}\right) = \frac{4\pi G}{c^4} T^{\mu\nu}. \tag{4.87}$$

The field $h^{\mu\nu}$ is a symmetric (since $T^{\mu\nu}$ is), second-rank tensor. This will do it, in the end, once we take the field $h^{\mu\nu}/c^2$ and couple it to its own stress tensor, just as we did with the scalar field in Section 4.5.2. With that necessary augmentation, we recover general relativity, including its geometric interpretation.[15]

Problem 4.23 Using the field equation (4.87), argue that you must have $\partial_\mu h^{\mu\nu} = 0$. Take $h_{\mu\nu} = P_{\mu\nu} e^{i k_\alpha x^\alpha}$ and find the constraints placed on the constants k_μ and $P_{\mu\nu}$ such that the vacuum form of (4.87) and gauge condition, $\partial_\mu h^{\mu\nu} = 0$, are satisfied.

The interpretation of $h_{\mu\nu}$ can be motivated by its use in particle actions. In E&M, we argued that a particle Lagrangian, written in proper time parametrization, should couple to the electromagnetic field A_μ via a scalar of the form $\dot{x}^\mu A_\mu$ for $\dot{x}^\mu \equiv \frac{dx^\mu}{d\tau}$. Given the symmetric tensor field $h_{\mu\nu}$, the natural analogue is a scalar term of the form $\dot{x}^\mu h_{\mu\nu} \dot{x}^\nu$. But that term will combine with the "usual" kinetic term. For a relativistic Lagrangian in proper time parametrization, we can start with

$$L = \frac{1}{2} m \dot{x}^\mu \eta_{\mu\nu} \dot{x}^\nu + \alpha \dot{x}^\mu h_{\mu\nu} \dot{x}^\nu, \tag{4.88}$$

where the first term involves the Minkowski metric, and α is a constant that sets the coupling strength. In this form, the story is the normal one: we have a field $h_{\mu\nu}$ that acts on the particle. We could rewrite the Lagrangian as

$$L = \frac{1}{2} m \dot{x}^\mu \left(\eta_{\mu\nu} + \frac{\alpha}{m} h_{\mu\nu}\right) \dot{x}^\nu, \tag{4.89}$$

and in this form, we could consider the particle to be moving freely (no potential) in a background metric with $g_{\mu\nu} \equiv \eta_{\mu\nu} + \alpha/m\, h_{\mu\nu}$. In a sense, any symmetric, second-rank field theory has the potential for this geometric interpretation (anti-symmetric ones do not).

[15] See, e.g., [8].

A Appendix A **Mathematical Methods**

There are some specialized mathematical methods that are useful for exploring the physical content of field theories. Given our interest in point source solutions, Cauchy's integral formula is a powerful tool for evaluating Green's functions, but one that requires some setup for its discussion. The first part of this chapter, then, will provide just enough background to appreciate the proof and role of the integral formula, as well as its application to Green's functions (making contact with our previous derivations). Then we review the "Frobenius" method of series solution for ordinary differential equations (ODEs), important for defining the Bessel functions we find in, for example, the Klein–Gordon Green's function. Finally, there is a section on iterated integral methods.

All of the math methods here are distilled from their formal counterparts so as to focus on the particular physical problems we have encountered. More information can be found in standard mathematical methods books (see [2] and [4], for example).

A.1 Complex Numbers

A.1.1 Definitions

Complex numbers are pairs of real numbers that can be mapped to the x–y plane. We use $z \equiv x + i\,y$ to label the point we would call $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}}$ in the two-dimensional Cartesian plane. Functions of z are similarly defined with two components; $f(z) = u(z) + i\,v(z)$ represents a (sort of) vector function that we could think of as $\mathbf{f}(x, y) = u(x, y)\,\hat{\mathbf{x}} + v(x, y)\,\hat{\mathbf{y}}$. It is typical to associate the coefficient of i with the $\hat{\mathbf{y}}$-axis, and the other coefficient with the $\hat{\mathbf{x}}$-axis (when referring to points in the complex plane).

There are operations we can perform on complex numbers, and some of these have Cartesian analogues. We can multiply a complex number by a real number: for $\alpha \in \mathbb{R}$, $\alpha\,f = \alpha\,u + i\,\alpha\,v$ (like scaling a vector). We can multiply two complex numbers: for $z_1 = x_1 + i\,y_1$ and $z_2 = x_2 + i\,y_2$,

$$z_1\,z_2 = x_1\,x_2 - y_1\,y_2 + i\,(x_2\,y_1 + x_1\,y_2)\,. \tag{A.1}$$

The complex conjugate of a complex number flips the sign of the coefficient of i:

$$z^* = x - i\,y \tag{A.2}$$

(a reflection about the $\hat{\mathbf{x}}$-axis), and then we can define the "inner product" of z_1 and z_2,

$$z_1 \cdot z_2 \equiv z_1^* z_2 = x_1 x_2 + y_1 y_2 + i (x_1 y_2 - x_2 y_1), \tag{A.3}$$

which has the property that the inner product of a complex number with itself is

$$z \cdot z = z^* z = x^2 + y^2, \tag{A.4}$$

a real number, allowing us to associate the inner product with a "length" (squared), just as for two-dimensional vectors.

We can use Cartesian coordinates or polar ones to represent points in the x–y plane, and we can use polar coordinates for complex numbers as well. For $z = x + i y$, we have

$$z = r e^{i\phi} \quad \text{with } r \equiv \sqrt{x^2 + y^2} = \sqrt{z \cdot z} \quad \text{and} \quad \phi \equiv \tan^{-1}(y/x) \tag{A.5}$$

for real r and (real) angle ϕ. In this setting, the inner product of z with itself is just $z \cdot z = r^2$. In the polar representation, rotations can be achieved by complex multiplication. Suppose we have two "unit" complex numbers, $z_1 = e^{i\phi_1}$ and $z_2 = e^{i\phi_2}$, then

$$z_1 z_2 = e^{i(\phi_1 + \phi_2)} \tag{A.6}$$

so that the angle of the product is just the sum of the angle of its constituents (if we hadn't started with normalized complex numbers, there would be a stretching as well).

A.1.2 Derivatives

For a function $f(z) = u(z) + i v(z)$, how should we think about its "derivative" with respect to z? Since z is made up of a real and complex part, this is sort of like asking how to take the derivative of a vector $\mathbf{f}$ with respect to $\mathbf{r} \equiv x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}}$. From a formal point of view:

$$\left.\frac{df(z)}{dz}\right|_{z=z_0} = \lim_{h\to 0} \frac{f(z_0 + h) - f(z_0)}{h}, \tag{A.7}$$

but you have to tell me *how* you want to approach the point z_0: should $h = dx$ or $h = i\,dy$, or some linear combination? There is a special class of functions for which the direction of approach doesn't matter, and those will be our focus. Take $h = d\ell e^{i\phi}$ so that we approach z_0 from an arbitrary angle $\phi \in [0, \pi/2]$. Then the derivative could be written

$$\frac{df(z)}{dz} = \lim_{d\ell\to 0} \frac{du + i\,dv}{d\ell} e^{-i\phi}. \tag{A.8}$$

Suppose we set $\phi = 0$, so that $d\ell = dx$, then we have

$$\frac{df(z)}{dz} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}. \tag{A.9}$$

If we instead take $\phi = \pi/2$ so that $d\ell = dy$,

$$\frac{df(z)}{dz} = -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{A.10}$$

In order for (A.9) to be the same as (A.10), we must have

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \tag{A.11}$$

Under these conditions (and assuming the relevant derivatives *exist*[1]), the derivative of $f(z)$ is defined from every direction and takes on the same value for each. Such functions are called *analytic* functions, and the equations that test for it, in (A.11), are the "Cauchy–Riemann equations."

As a rough rule of thumb, if a function depends on the variable $z = x + iy$ (rather than the individual components x and y in some other combination), it is analytic. So, for example, $f(z) = z^2$ is an analytic function, and we can check it: $f(z) = (x^2 - y^2 + 2\,i\,x\,y)$ has $u \equiv x^2 - y^2$ and $v = 2\,x\,y$, with

$$\frac{\partial u}{\partial x} = 2\,x, \qquad \frac{\partial u}{\partial y} = -2\,y, \qquad \frac{\partial v}{\partial x} = 2\,y, \qquad \frac{\partial v}{\partial y} = 2\,x, \tag{A.12}$$

and clearly (A.11) holds, so the function is analytic. The function $f(z) = z^* = x - iy$ has $\frac{\partial u}{\partial x} = 1$ but $\frac{\partial v}{\partial y} = -1$, so is not analytic.

Problem A.1 Which of the following are analytic functions: $\sin(z)$, $1/z$, $z^* z$?

Problem A.2 Suppose $f_1(z)$ and $f_2(z)$ are both analytic. Is the sum $f_1(z)+f_2(z)$ an analytic function? Is the product?

A.1.3 Integrals

For functions satisfying the conditions (A.11), we have

$$\oint f(z)\, dz = 0 \tag{A.13}$$

around any closed loop. To see it, let's write out the integrand and appeal to our usual vector calculus integral theorem(s):

$$\oint f(z)\, dz = \oint (u + i\,v)\,(dx + i\,dy) = \oint (u\,dx - v\,dy) + i \oint (v\,dx + u\,dy)\,. \tag{A.14}$$

We'll show that both the real and imaginary parts of the integral are zero. Recall Stokes's theorem: for a vector $\mathbf{A} = A^x(x,y)\,\hat{\mathbf{x}} + A^y(x,y)\,\hat{\mathbf{y}}$ and a domain of integration Ω (with boundary $\partial\Omega$),

$$\int_\Omega (\nabla \times \mathbf{A}) \cdot d\mathbf{a} = \oint_{\partial\Omega} \mathbf{A} \cdot d\boldsymbol{\ell}. \tag{A.15}$$

Since we are in the x–y plane, we have $d\mathbf{a} = dxdy\,\hat{\mathbf{z}}$, and the above becomes

$$\int_\Omega \left(\frac{\partial A^y}{\partial x} - \frac{\partial A^x}{\partial y}\right) dxdy = \oint_{\partial\Omega} \left(A^x\, dx + A^y\, dy\right). \tag{A.16}$$

[1] In physics, we tend to assume functions are "nice," so that we are dealing with continuous functions that are differentiable "as many times as necessary." Beware the problem that violates that assumption!

Take the real part of (A.14); let $A^x = u$ and $A^y = -v$, then Stokes's theorem reads

$$-\int_{\Omega}\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right) dxdy = \oint_{\partial\Omega}(u\,dx - v\,dy) \tag{A.17}$$

where the right-hand side is precisely the real part of (A.14) (by construction) and the left-hand side is zero from (A.11). So the real part of the integral in (A.14) is zero.

For the imaginary part, let $A^x = v$ and $A^y = u$ so that Stokes's theorem says

$$\int\left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right) dxdy = \oint(v\,dx + u\,dy), \tag{A.18}$$

and again, the right-hand side is the imaginary part of the integral in (A.14), while the left-hand side is zero from (A.11). We have established that $\oint f(z)\,dz = 0$ around any closed path (the domain Ω is arbitrary and can have any closed path as its boundary $\partial\Omega$). As a consequence of this, we know that for complex numbers a and b,

$$I = \int_a^b f(z)\,dz \tag{A.19}$$

is the same for all paths connecting a to b. Given two arbitrary (non-pathological, whatever that means) paths connecting the points a and b, go from a to b along one path, and then traverse the other path backward from b to a. Those two together form a closed loop, so that

$$0 = \oint f(z)\,dz = \underbrace{\int_a^b f(z)\,dz}_{\text{path I}} + \underbrace{\int_b^a f(z)\,dz}_{\text{path II}} \tag{A.20}$$

and reversing the limits at the cost of a minus sign,

$$0 = \underbrace{\int_a^b f(z)\,dz}_{\text{path I}} - \underbrace{\int_a^b f(z)\,dz}_{\text{path II}} \tag{A.21}$$

so the two integrals, along different paths, are equal. For analytic functions, then, just as the derivative is direction-independent, the integral is path-independent.

Just to see the path-independence in action, take $f(z) = z^2$, which we know is analytic, and we'll connect the point $a = -1$ to $b = 1$ along a rectangular path and a circular path, as depicted in Figure A.1. Evaluating the integral along the rectangular path, we have

$$\begin{aligned} I &= \int_0^h \left(x^2 - y^2 + 2\,i\,x\,y\right)|_{x=-1}\,i\,dy + \int_{-1}^1 \left(x^2 - y^2 + 2\,i\,x\,y\right)|_{y=h}\,dx \\ &\quad + \int_h^0 \left(x^2 - y^2 + 2\,i\,x\,y\right)|_{x=1}\,i\,dy \\ &= i\left(h - \frac{1}{3}h^3 + i\,(-1)h^2\right) + \left(\frac{2}{3} - 2\,h^2 + 0\right) + i\left(-h + \frac{1}{3}h^3 - i\,h^2\right) \\ &= \frac{2}{3}. \end{aligned} \tag{A.22}$$

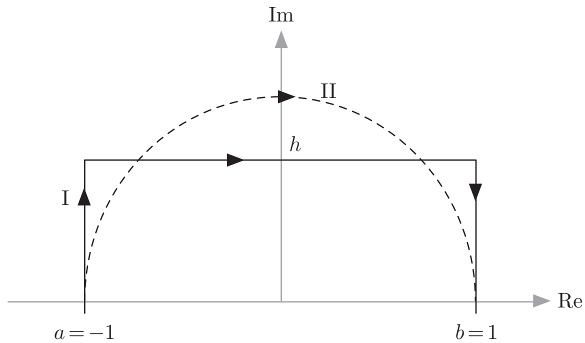

Fig. A.1 Two paths connecting a to b. Path I goes from a up to h, then over and down to b. Path II is a unit semicircle connecting a to b.

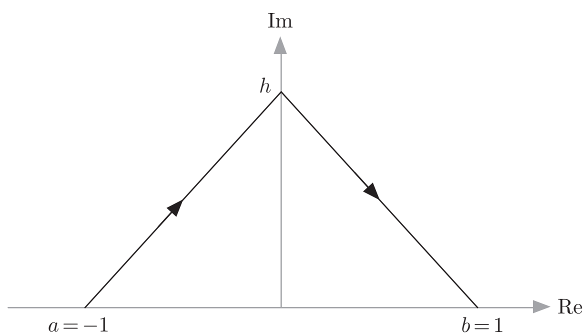

Fig. A.2 Path for evaluating the integral of $f(z) = z^2$ in Problem A.3.

Since we used an arbitrary h, this result is valid for any h (and is h-independent, as it should be).

For the semicircular path, we can use polar coordinates to evaluate the integral. Take $z = e^{i\phi}$ describing points along the unit circle, and $dz = e^{i\phi}\, i\, d\phi$ for $\phi = -\pi \to 0$. The integral is

$$I = \int_{-\pi}^{0} e^{2i\phi}\, e^{i\phi}\, i\, d\phi = \frac{1}{3}\left(e^0 - e^{-3i\pi}\right) = \frac{2}{3}, \tag{A.23}$$

as expected.

Problem A.3 Evaluate the integral of $f(z) = z^2$ from $a = -1$ to $b = 1$ along the path shown in Figure A.2. You already know the answer; this problem is about generating the correct path parametrization.

Problem A.4 Integrate $f(z) = z^*$ along the two paths shown in Figure A.1.

A.1.4 Cauchy Integral Formula

Suppose we have the function $f(z) = 1/z$. Working in polar coordinates, $z = r\,e^{i\phi}$, and we'll integrate around a circle of radius R with $dz = i\,R\,d\phi\,e^{i\phi}$. The integral is

$$\oint \frac{1}{z}\,dz = \int_0^{2\pi} i\,d\phi = 2\pi\,i. \tag{A.24}$$

But this is not zero – why not? The function $f(z)$ satisfies (A.11), as you can verify, so the integral around this simplest possible closed loop must be zero, right? The function $f(z) = 1/z$ is fine everywhere except at $z = 0$ where (A.11) breaks down. Only if the function $f(z)$ satisfies (A.11) over the whole domain of integration (enclosed by the loop around which we integrate) will the integral around a closed loop be zero. If there are points at which (A.11) is not satisfied, the integral will not be zero.

Integration Contours

While we established that $\oint 1/z\,dz = 2\pi\,i$ for a circular integration contour enclosing $z = 0$, we can extend the result to non-circular boundaries. Imagine deforming our circular domain as shown in Figure A.3. We go around the circle, but perform a small excursion $a \to b$, then $b \to b'$ and then $b' \to a'$ before continuing along the circle. The portions of the path $a \to b$ and $b' \to a'$ are anti-parallel and perpendicular to the boundary of the circle, while the path $b \to b'$ is an arc, with no change in radius. The original circle has radius R_1, and the deformed bit is a circular arc of radius R_2. What effect does this deformation have on the integral? Since $1/z = 1/r\,e^{-i\phi}$, if we move along a path perpendicular to the circle (via $dz = dr\,e^{i\phi}$), we just pick up

$$\int_a^b f(z)\,dz = \int_{R_1}^{R_2} \frac{1}{r}\,e^{-i\phi}\,e^{i\phi}\,dr = \log(R_2/R_1) \tag{A.25}$$

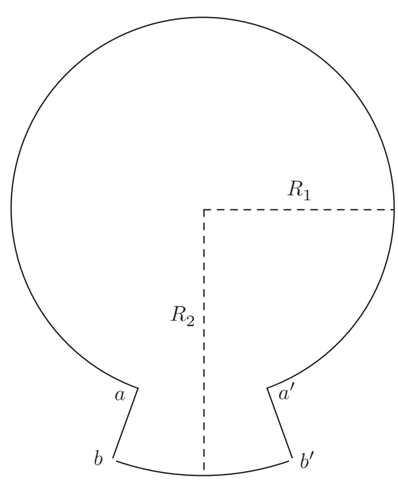

Fig. A.3 A circular domain with a small deformation.

and the integral from $b \to b'$ is the same as the integral from $a \to a'$ (since the angular portion of the integral does not depend on the radius). Finally, the integral from $b' \to a'$ has the same magnitude as the integral from $a \to b$ since they represent the same radial displacement, and since they are traversed in opposite directions, the contributions cancel:

$$\int_a^b f(z)\,dz + \int_{b'}^{a'} f(z)\,dz = 0. \tag{A.26}$$

Conclusion: the deformed path has the same integral value as the original circular path. By piecing together infinitesimal path segments that are purely angular and purely radial, we can deform our circular path to look like almost anything, so we conclude that

$$\oint \frac{1}{z}\,dz = 2\pi i \tag{A.27}$$

for *any* path enclosing $z = 0$ (where we go around the path in a counterclockwise manner), and for a path that does not enclose zero, we have $\oint 1/z\,dz = 0$ (since away from the singularity, $1/z$ satisfies (A.11)).

Integral Formula

For a function $f(z)$ that is everywhere analytic, we will establish that

$$\oint \frac{f(z)}{z}\,dz = \begin{cases} 0 & \text{for paths that do not enclose } z = 0 \\ 2\pi i f(0) & \text{for paths that enclose } z = 0. \end{cases} \tag{A.28}$$

The top equality holds since $f(z)/z$ is analytic everywhere except at $z = 0$.

For the bottom equality, first note that for a very small circle of radius ϵ centered at the origin,

$$\oint \frac{f(z)}{z}\,dz = \lim_{\epsilon\to 0} \oint \frac{f(\epsilon)}{z}\,dz = \lim_{\epsilon\to 0} f(\epsilon) \oint \frac{1}{z}\,dz = \lim_{\epsilon\to 0} 2\pi i f(\epsilon) = 2\pi i f(0), \tag{A.29}$$

where we have gone around counterclockwise (if the circle were traversed in the clockwise direction, we would get a minus sign on the right).

Now consider an arbitrary path I, shown on the left in Figure A.4. We want to evaluate $f(z)/z$ around this arbitrary closed loop, enclosing the origin. Referring to the figure on the

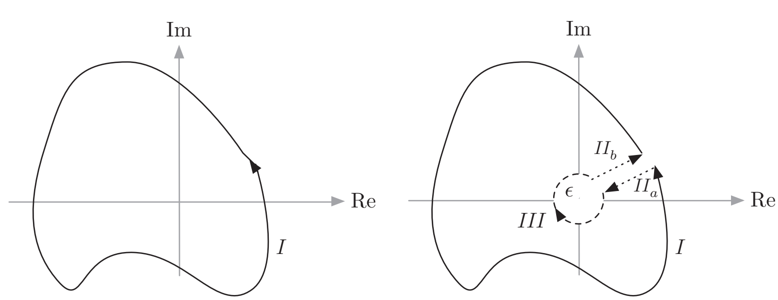

Fig. A.4 Two different paths: the one on the left, I, represents some arbitrary path. The path on the right follows along I for the most part, going in along II_a, circling the origin along III (with radius ϵ), and rejoining I along II_b.

right in Figure A.4, what we'll do is make a closed contour path that follows I along the "outside," and then comes in to the origin along II_a, circles around the origin at radius ϵ, path III, and then goes back out to join up with I along II_b.

The path on the right in Figure A.4 does not enclose the singularity, so that $f(z)/z$ is analytic everywhere, and therefore the integral along this closed path is zero:

$$\int_I \frac{f(z)}{z}\,dz + \int_{II_a} \frac{f(z)}{z}\,dz + \int_{III} \frac{f(z)}{z}\,dz + \int_{II_b} \frac{f(z)}{z}\,dz = 0. \tag{A.30}$$

In the limit as $\epsilon \to 0$, the first integral becomes $\oint_I f(z)/z\,dz$ (which is what we want); the second and fourth terms will cancel and the third term will become $\oint_{III} f(z)/z\,dz$ (which we have already evaluated). The $\epsilon \to 0$ limit of the path integral is still equal to zero, so we have:

$$\begin{aligned} 0 &= \lim_{\epsilon\to 0}\left(\int_I \frac{f(z)}{z}\,dz + \int_{II_a} \frac{f(z)}{z}\,dz + \int_{III} \frac{f(z)}{z}\,dz + \int_{II_b} \frac{f(z)}{z}\,dz\right) \\ &= \oint_I \frac{f(z)}{z}\,dz + \oint_{III} \frac{f(z)}{z}\,dz, \end{aligned} \tag{A.31}$$

or

$$\oint_I \frac{f(z)}{z}\,dz = -\oint_{III} \frac{f(z)}{z}\,dz. \tag{A.32}$$

The closed integral along path III here is $= -2\pi\, i f(0)$ from (A.29) (traversed clockwise). Then for any closed path I, we have

$$\oint_I \frac{f(z)}{z}\,dz = 2\pi\, i f(0) \tag{A.33}$$

and the bottom equality in (A.28) holds.

In the more general case, where we shift the singularity to some new point z_0, we have (for $f(z)$ analytic everywhere, again)

$$\oint \frac{f(z)}{z - z_0}\,dz = \begin{cases} 0 & \text{for paths that do not enclose } z = z_0 \\ 2\pi\, i f(z_0) & \text{for paths that enclose } z_0. \end{cases} \tag{A.34}$$

As a last, technical point, we have assumed that our integration boundaries are oriented so that the path is traversed in the counterclockwise direction (go back to (A.24), where we take $\phi : 0 \to 2\pi$); if we go around a singularity in the *clockwise* direction, we get a minus sign in the right-hand side above. Equation (A.34) (together with the sign convention) is Cauchy's integral formula, and z_0 is called a "pole."

The integral formula (A.28) is very similar to Ampere's law for magnetism. Given a magnetic field $\mathbf{B}$ (the analogue of $f(z)/(z - z_0)$), we have

$$\oint \mathbf{B} \cdot d\boldsymbol{\ell} = \mu_0 I_{\text{enc}}, \tag{A.35}$$

for any closed curve. The right-hand side is the current enclosed by a loop of any shape. If you imagine a line of wire puncturing the x–y plane, then the integral of $\mathbf{B} \cdot d\boldsymbol{\ell}$ around any loop that does not include the puncture point is zero, and if the puncture point is included, you get a constant, just as in the Cauchy integral formula. Away from the wire, the magnetic field is finite, but this is not true at the wire (similar to the idea that $f(z)/z$ be analytic everywhere except at the origin). Even the form of the magnetic field is similar to our starting function $f(z) = 1/z$: for an infinite straight wire carrying constant steady current I, $B = \mu_0 I/(2\pi s)$, so $B \sim 1/s$.

Problem A.5 What is the complex integral of z^n (for integer $n \geq 0$) for a circular path of radius ϵ enclosing the origin? What is the integral of $1/z^n$ with $n \geq 2$ (and n still an integer)?

A.1.5 Utility

There are some physical systems that occur in two dimensions, and for these, the integral formula above is of direct utility. More often, however, the formula is useful for evaluating integrals that occur along the real line. For example, suppose you wanted to evaluate

$$I = \int_{-\infty}^{\infty} \frac{1}{1+x^2}\, dx. \tag{A.36}$$

We can get the anti-derivative by noting that

$$\tan(\underbrace{\arctan(x)}_{\equiv\theta}) = x, \tag{A.37}$$

then taking the x-derivative of both sides gives

$$\frac{1}{\cos^2\theta} \frac{d}{dx} \arctan(x) = 1, \tag{A.38}$$

and since $\sin\theta/\cos\theta = x$, we can square and solve for $\cos^2\theta$ in terms of x: $\cos^2\theta = 1/(1+x^2)$, and conclude

$$\frac{d}{dx} \arctan(x) = \frac{1}{1+x^2}. \tag{A.39}$$

Using this in I gives

$$I = \int_{-\infty}^{\infty} \frac{1}{1+x^2}\, dx = \arctan(x)\big|_{x=-\infty}^{\infty} = \frac{\pi}{2} - \frac{-\pi}{2} = \pi. \tag{A.40}$$

That's easy provided you remember that the derivative of the arctangent is precisely the integrand (or use clever substitutions to reveal the same fact).

Consider instead the complex integral

$$I_C = \oint \frac{1}{1+z^2}\, dz, \tag{A.41}$$

where our closed contour goes from negative infinity to positive infinity along the real axis, then closes with a semi-circle at infinity in the top-half plane as shown in Figure A.5. The integrand is zero for $|z| \to \infty$ along the semi-circular path, so that path contributes nothing. The contour choice makes $I_C = I$, but we can evaluate I_C using the integral formula: write

$$I_C = \oint \frac{\frac{1}{z+i}}{z-i}\, dz \tag{A.42}$$

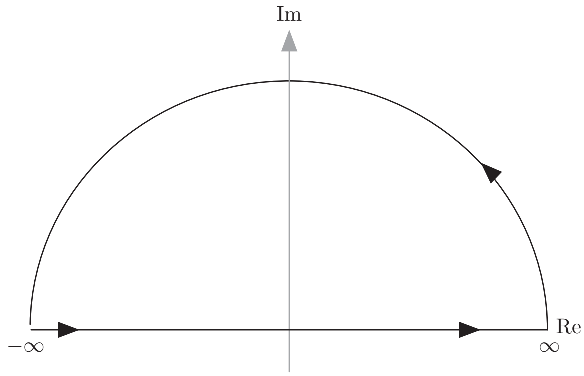

Fig. A.5 A simple loop for evaluating I_C in (A.41) includes integrating along the real axis from $-\infty$ to ∞. The picture is only meant to suggest the correct integration path; formally, we are taking a path that extends from $-R \to R$ on the real axis, and then sending $R \to \infty$.

and note that $f(z) \equiv \frac{1}{z+i}$ is analytic for all points in the upper half-plane (the region enclosed by the contour). Then

$$I_C = 2\pi\, i f(i) = \pi \tag{A.43}$$

since the singularity is at $z_0 = i$. We conclude that $I = I_C = \pi$ (correct), and it is easier to evaluate than the substitution above.

Cauchy's integral formula gives an easy way to evaluate integrals when the singular point is inside the domain enclosed by the integration contour. But what if the singular point is on the boundary of the domain? What should we do with

$$I = \oint \frac{f(z)}{z}\, dz \tag{A.44}$$

for $f(z)$ analytic everywhere in the domain, and a closed contour that goes through $z = 0$? As it turns out, there are a variety of ways to handle this type of situation, and that freedom proves very useful in physical applications. The most important thing is to be clear about which of the following limiting processes you have in mind. We'll review two different limiting procedures. To keep things simple, assume that the analytic function $f(z)$ vanishes for z at infinity in the upper half plane (the other option would be to have $f(z)$ vanish at $z \to \infty$ in the lower half plane). Remember, the goal here is to develop expressions for integrals along the real line, so we want to be able to ignore that "closure" of the domain.

Limit with Excursion

Take the path shown in Figure A.6, the excursion A' (of radius r) encloses the singularity, and the semicircular B section is at infinity. We recover the integral along the real line by taking the limit of the path as $r \to 0$. From Cauchy's integral formula, we know that

$$\int_A \frac{f(z)}{z}\, dz + \int_{A'} \frac{f(z)}{z}\, dz + \int_{A''} \frac{f(z)}{z}\, dz + \int_B \frac{f(z)}{z}\, dz = 2\pi\, i f(0). \tag{A.45}$$

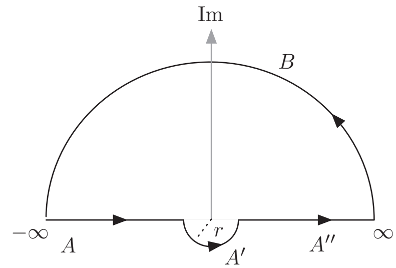

Fig. A.6 A path that encloses the pole at zero. The path is broken into four segments: A, A', A'', and B.

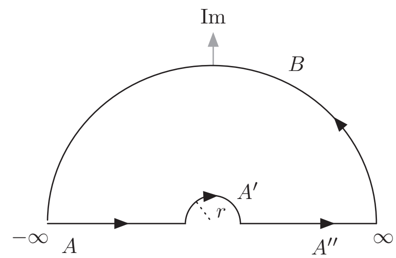

Fig. A.7 The path for evaluating (A.44) for comparison with (A.47).

Let's focus on the A' integral. If we go around this half-circle in the counterclockwise direction as shown, we know that we'll pick up $\pi\, i f(r)$ by the same argument that got us (A.29); then (A.45) reads

$$\int_A \frac{f(z)}{z}\,dz + \pi\, i f(r) + \int_{A''} \frac{f(z)}{z}\,dz + \int_B \frac{f(z)}{z}\,dz = 2\pi\, i f(0). \tag{A.46}$$

If we take the limit as $r \to 0$, then the $\pi\, i f(r)$ term can be combined with the term on the right. The remaining integrals on the left define the simple contour shown in Figure A.5 as $r \to 0$, and for that contour, we have (the B segment vanishes, as usual)

$$\int_{-\infty}^{\infty} \frac{f(x)}{x}\,dx = \pi\, i f(0), \tag{A.47}$$

half the result we would get if the pole was in the domain. When we use this type of prescription to evaluate poles on the boundary, we get back the "Cauchy principle value" for the integral.

Problem A.6 What equation, analogous to (A.47), do you get if you evaluate (A.44) using an excursion that goes above the singularity as in Figure A.7?

Problem A.7 Evaluate

$$\frac{d}{dx}\cos^{-1}(x) \tag{A.48}$$

(That is, what is the derivative of the arccosine?)

Problem A.8 Evaluate the integral

$$\int_{-\infty}^{\infty} \frac{\cos(x)}{1+x^2}\, dx \tag{A.49}$$

using the Cauchy integral formula.

Problem A.9

a. A complex function can be written as a power series in z:

$$f(z) = \sum_{j=-\infty}^{\infty} a_j z^j \tag{A.50}$$

for constants $\{a_j\}_{j=-\infty}^{\infty}$. Evaluate

$$\oint f(z)\, dz \tag{A.51}$$

using the sum in (A.50), where the integration contour is a circle of radius R (traversed counterclockwise).

b. What familiar "function" has the integral representation

$$\frac{1}{2\pi i} \oint z^{m-n-1}\, dz \tag{A.52}$$

for integer m and n, and where the contour is a circle of radius R (traversed counterclockwise)? (Evaluate the integral and see what you get for various m and n.)

Pole Pushing (Contour Limits)

Another way to evaluate closed contour integrals where the pole lies on the boundary is to push the pole up into the upper (or down into the lower) half-plane, evaluate the integral using Cauchy's integral formula, and then take the limit as the pole gets pushed back down, so that we would understand the integral around the path shown in Figure A.5 as

$$\oint \frac{f(z)}{z}\, dz = \lim_{\epsilon \to 0} \oint \frac{f(z)}{z - i\epsilon}\, dz. \tag{A.53}$$

For the expression on the right, the pole is sitting in the upper half-plane, off the real axis, and then we can apply the integral formula

$$\oint \frac{f(z)}{z - i\epsilon}\, dz = 2\pi i f(i\epsilon). \tag{A.54}$$

The limit $\epsilon \to 0$ takes $f(i\epsilon) \to f(0)$,

$$\oint \frac{f(z)}{z}\, dz = \lim_{\epsilon \to 0} \oint \frac{f(z)}{z - i\epsilon}\, dz = \lim_{\epsilon \to 0} 2\pi i f(i\epsilon) = 2\pi i f(0). \tag{A.55}$$

This choice of evaluation gives us back the "usual" result. It can be viewed as a contour deformation: rather than pushing the pole up a distance ϵ, you drop the contour a distance ϵ below the real axis, and then take the limit as $\epsilon \to 0$ to recover the contour in Figure A.5. The two equivalent pictures are shown in Figure A.8. We'll see in the next

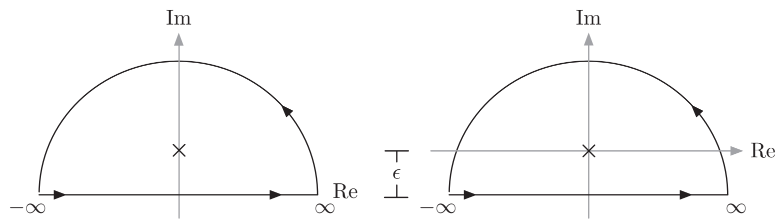

Fig. A.8 Two equivalent views of the limiting process described in this section. On the left, we push the pole up a distance ϵ into the upper half plane, evaluate the integral, and then take the limit as $\epsilon \to 0$. On the right, we leave the pole at 0 on the real axis, but move the contour down a distance ϵ below the real axis, then evaluate the integral and take the limit as $\epsilon \to 0$.

section how these different choices, excursions and pushes, give back different expressions, and how those expressions can be distinguished physically.

A.1.6 Poles and the Residue Theorem

For real functions, we can perform a "Taylor expansion" about zero: given $p(x)$,

$$p(x) = \sum_{j=0}^{\infty} \alpha_j x^j \tag{A.56}$$

where

$$\alpha_j = \frac{1}{j!} p^{(j)}(0) \quad \text{with } p^{(j)}(x) \equiv \frac{d^j}{dx^j} p(x). \tag{A.57}$$

The natural generalization, for complex functions, takes the form

$$f(z) = \sum_{j=-\infty}^{\infty} \alpha_j z^j, \tag{A.58}$$

the "Laurent" series expansion for complex coefficients $\{\alpha_j\}_{j=-\infty}^{\infty}$. Notice that we allow negative powers of z as well as positive ones. From the expansion, it is clear that for any integration contour that encloses the origin,

$$\oint f(z)\, dz = 2\pi i \alpha_{-1}, \tag{A.59}$$

since the integral for any power of z other than z^{-1} is zero (recall Problem A.5, extended to arbitrary contours). The power z^{-1} is special, and when $\alpha_{-1} \neq 0$, we say that the function $f(z)$ has a "simple pole" or "singularity" at $z = 0$. The coefficient α_{-1} is called the "residue" associated with the simple pole. Other negative powers of z are problematic only for evaluating $f(z)$ at the origin, but their integrals vanish. Any z^{-n} for finite (positive) n represents a "singularity." These are "removable"; we can get rid of them by multiplying by z^n. If all the coefficients α_j for $j < 0$ are non-zero, then the singularity at $z = 0$ is an "essential" singularity.

The "residue theorem," indicated by (A.59), says that for an arbitrary closed loop:

$$\oint f(z)\, dz = 2\,\pi\, i \sum_j (\alpha_{-1})_j \;\text{(sum over residues)}, \tag{A.60}$$

so that we have to identify all of the simple poles of a function in the domain of integration, then add their residues (with appropriate sign, positive for counterclockwise, negative for clockwise traversal).

Example

Take the function

$$f(z) = \frac{5}{z-2} + \frac{-\frac{1}{2}}{z-i} - \frac{1}{z^2}. \tag{A.61}$$

This function has simple poles at $z = 2$ and $z = i$, with associated residues $\alpha = 5$ for $z = 2$ and $\alpha = -1/2$ for $z = i$. We'll integrate around (counterclockwise) a closed circle of radius R. Let $z = R\, e^{i\,\phi}$, then the integral is

$$\oint f(z)\, dz = \int_0^{2\,\pi} f(R\, e^{i\,\phi})\, R\, e^{i\,\phi}\, i\, d\phi. \tag{A.62}$$

Suppose we take $R = 1/2$ so that no simple poles are contained inside the contour (the $1/z^2$ piece of $f(z)$ will go away when we integrate), then we expect to get zero from the residue theorem. If we actually perform the integration, then

$$\oint f(z)\, dz = \int_0^{2\,\pi} \left(2\, i\, e^{-i\,\phi} + \frac{5\, i\, e^{i\,\phi}}{2\left(-2 + \frac{e^{i\,\phi}}{2}\right)} - \frac{i\, e^{i\,\phi}}{4\left(-i + \frac{e^{i\,\phi}}{2}\right)} \right) d\phi = 0 \tag{A.63}$$

where each term integrates to zero separately.

If we let $R = 3/2$, enclosing the pole at $z = i$, the residue theorem tells us that the integral should be $2\,\pi\, i\,(-1/2) = -i\,\pi$, which is the value we get by explicit integration. For $R = 3$, enclosing both poles, the residue theorem gives the value

$$\oint f(z)\, dz = 2\,\pi\, i\,(-1/2) + 2\,\pi\, i\,(5) = 9\,\pi\, i, \tag{A.64}$$

and again, this is what you would get doing the integral explicitly. The residue theorem allows us to evaluate integrals without performing the integration, which is always a good idea when possible.

Problem A.10 For the function

$$f(z) = \frac{4\,z - 6\,R}{R^2 - 4\,z^2}, \tag{A.65}$$

evaluate the integral of $f(z)$ for a closed loop that is a circle of radius R centered at the origin (go around counterclockwise).

A.2 Green's Functions

In E&M, we are interested in finding the point-particle solutions to various partial differential equations (PDEs), and those point-source solutions are known as Green's functions (see Section 2.1). The most familiar is the Green's function for the Poisson problem, solving $\nabla^2 G = -\delta^3(\mathbf{r})$. If we multiply this equation by $e^{i\,2\pi\,\mathbf{k}\cdot\mathbf{r}}$ for constant $\mathbf{k}$ and integrate over all space,

$$\int e^{i\,2\pi\,\mathbf{k}\cdot\mathbf{r}}\,\nabla^2 G\,d\tau = -1, \tag{A.66}$$

then integrate by parts to flip the ∇^2 onto the exponential term, where it just produces a factor of $(2\pi\,i\,k)^2$, we get

$$-4\pi^2 k^2 \int e^{i\,2\pi\,\mathbf{k}\cdot\mathbf{r}}\,G\,d\tau = -1. \tag{A.67}$$

The left-hand side has the Fourier transform of G from (2.25) in it, so we get a simple algebraic equation for the Fourier transform of the Green's function

$$\tilde{G}(\mathbf{k}) = \frac{1}{4\pi^2 k^2}, \tag{A.68}$$

and now we want to Fourier transform back (using (2.26)) to find $G(\mathbf{r})$.

For spherical coordinates in k-space, with[2] $\mathbf{r} = r\,\hat{\mathbf{z}}$, the integration reads

$$G(\mathbf{r}) = \int_0^\infty \int_0^\pi \int_0^{2\pi} e^{-i\,2\pi\,k\,r\,\cos\theta}\,\frac{1}{4\pi^2 k^2}\,k^2\,\sin\theta\,d\theta d\phi dk. \tag{A.69}$$

We can do the ϕ and θ integrals easily here:

$$G(\mathbf{r}) = -\frac{i}{4\pi^2 r}\left[\int_0^\infty \frac{e^{i\,2\pi\,k\,r}}{k}\,dk - \int_0^\infty \frac{e^{-i\,2\pi\,k\,r}}{k}\,dk\right] \tag{A.70}$$

and let $k \to -k$ in the second integral, making it $\int_0^{-\infty} \frac{e^{i\,2\pi\,k\,r}}{k}\,dk$, so we can combine the two into one:

$$G(\mathbf{r}) = -\frac{i}{4\pi^2 r}\int_{-\infty}^{\infty} \frac{e^{i\,2\pi\,k\,r}}{k}\,dk. \tag{A.71}$$

Now evaluate the integral using (A.47): the singularity lies at $k = 0$, and $f(z) = e^{i\,2\pi\,z\,r}$, so we pick up a factor of $\pi\,i$ to get

$$G(\mathbf{r}) = \frac{1}{4\pi\,r}, \tag{A.72}$$

the usual (Coulomb) result.

[2] There is no loss of generality here; we know that the inverse Fourier transform of the Green's function must be spherically symmetric, and hence, only depend on r (distance to the origin). The choice $\mathbf{r} = r\,\hat{\mathbf{z}}$ makes the integrals easier to evaluate.

A.2.1 Example: Anti-Helmholtz

We can use the same technique to get the Green's function for the operator $\nabla^2 - \mu^2$. In this case, the Fourier transform solution is

$$\tilde{G} = \frac{1}{4\pi^2 k^2 + \mu^2}, \tag{A.73}$$

and we want to perform the inverse Fourier transform

$$G(\mathbf{r}) = \int \frac{e^{-i 2\pi \mathbf{k}\cdot\mathbf{r}}}{(2\pi k + i\mu)(2\pi k - i\mu)} d^3k. \tag{A.74}$$

Again using spherical coordinates (in **k**-space) with $\mathbf{r} = r\hat{\mathbf{z}}$, and performing the angular integrals, we get

$$G(\mathbf{r}) = -\frac{i}{r}\int_0^\infty \frac{k}{(2\pi k + i\mu)(2\pi k - i\mu)}\left(e^{i2\pi k r} - e^{-i2\pi k r}\right) dk. \tag{A.75}$$

As before, we can combine the complex exponentials to make an integral from k: $-\infty \to \infty$. Substitute $x \equiv 2\pi k$, and we're ready to use our complex integral formula on

$$G(\mathbf{r}) = -\frac{i}{4\pi^2 r}\int_{-\infty}^\infty \frac{x e^{ixr}}{(x + i\mu)(x - i\mu)} dx. \tag{A.76}$$

For the complex contour we'll use the one shown in Figure A.5; take x: $-\infty \to \infty$ and close the contour with an infinite semicircle in the upper half plane where the e^{ixr} will vanish (since x in the complex setting will have positive imaginary value, giving a decaying exponential). Then the singularity at $i\mu$ is enclosed, and our analytic (in the upper half-plane) function is $f(z) = e^{izr} z/(z + i\mu)$, so

$$G(\mathbf{r}) = -\frac{i}{4\pi^2 r}\left(i2\pi \frac{e^{-\mu r} i\mu}{2i\mu}\right) = \frac{e^{-\mu r}}{4\pi r}, \tag{A.77}$$

by the Cauchy integral formula. This reduces correctly, in the case that $\mu = 0$, to $\frac{1}{4\pi r}$.

Problem A.11 Find the Green's function for the anti-Helmoltz case in $D = 3$ using the method from Section 2.1.

A.2.2 Example: Helmholtz

Changing the sign in the previous example leads to a proliferation of interesting options for the inverse Fourier transform integral. The operator is now $\nabla^2 + \mu^2$ so that the Fourier transform of the Green's function is

$$\tilde{G} = \frac{1}{4\pi^2 k^2 - \mu^2}, \tag{A.78}$$

and then performing all the inverse Fourier transform integration/simplifications, we get

$$G(\mathbf{r}) = -\frac{i}{4\pi^2 r}\int_{-\infty}^\infty \frac{x e^{ixr}}{(x + \mu)(x - \mu)} dx. \tag{A.79}$$

Now both poles lie on the real axis, and that means we have to be explicit about how we handle them. We'll try both of the approaches discussed in Section A.1.5.

Limit with Excursion

If we take the path from Figure A.5 and treat the poles as in (A.47), then the complex contour integral that is relevant to the integrand in (A.79) is

$$\int_{-\infty}^{\infty} \frac{x\, e^{i\,x\,r}}{(x+\mu)(x-\mu)}\, dx = \int_{-\infty}^{\infty} \frac{z\, e^{i\,z\,r}}{(z+\mu)(z-\mu)}\, dz. \tag{A.80}$$

Thinking about the poles, for $z = -\mu$, we write the integrand as $f_-(z)/(z+\mu)$ with

$$f_-(z) = \frac{z\, e^{i\,z\,r}}{z-\mu} \tag{A.81}$$

and then $f_-(-\mu) = e^{-i\,\mu\,r}/2$. Similarly, at $z = \mu$, we take the integrand to be $f_+/(z-\mu)$ with

$$f_+(z) = \frac{z\, e^{i\,z\,r}}{z+\mu} \tag{A.82}$$

and we evaluate $f_+(\mu) = e^{i\,\mu\,r}/2$.

For each of these terms, we also multiply by $\pi\, i$ from (A.47). The result is

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{x\, e^{i\,x\,r}}{(x+\mu)(x-\mu)}\, dx &= \int_{-\infty}^{\infty} \frac{z\, e^{i\,z\,r}}{(z+\mu)(z-\mu)}\, dz \\ &= \frac{\pi\, i}{2}\left(e^{-i\,\mu\,r} + e^{i\,\mu\,r}\right).\end{aligned} \tag{A.83}$$

Putting this result into the expression (A.79), we get the Green's function

$$G(\mathbf{r}) = \frac{\cos(\mu\, r)}{4\,\pi\, r}. \tag{A.84}$$

This equation limits correctly in the $\mu \to 0$ case.

Pole Pushing

Referring to the pole pushing of Section A.1.5, we can move poles up and down off the real axis. Since we have two poles, we must choose how we want to handle each pole separately. Let's push the pole at $-\mu$ down into the lower half-plane, and move the pole at μ up into the upper half-plane as shown in Figure A.9. Then we're thinking of the complex integral relevant to (A.79) as a limit,

$$\oint \frac{z\, e^{i\,z\,r}}{(z+\mu)\,(z-\mu)}\, dz = \lim_{\epsilon\to 0} \oint \frac{z\, e^{i\,z\,r}}{(z+\mu+i\,\epsilon)\,(z-\mu-i\,\epsilon)}\, dz. \tag{A.85}$$

The integral, evaluated prior to taking the limit, is

$$\oint \frac{z\, e^{i\,z\,r}}{(z+\mu+i\,\epsilon)\,(z-\mu-i\,\epsilon)}\, dz = i\,2\,\pi\,(\mu+i\,\epsilon)\,\frac{e^{i(\mu+i\,\epsilon)\,r}}{2\,(\mu+i\,\epsilon)}, \tag{A.86}$$

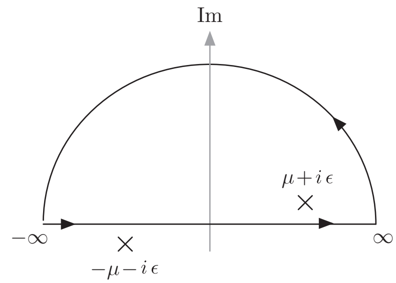

Fig. A.9 Moving the singularities off the real axis, we have chosen to move the left one down a little, and the right one up a little. The contour for integration is also shown.

and the limit gives

$$\oint \frac{z\, e^{i\, z\, r}}{(z+\mu)\,(z-\mu)}\, dz = \pi\, i\, e^{i\,\mu\, r}. \tag{A.87}$$

Using this to evaluate (A.79), we have

$$G_+(\mathbf{r}) = \frac{e^{i\,\mu\, r}}{4\,\pi\, r}. \tag{A.88}$$

If we had gone the other way around, pushing the pole on the left up and the one on the right down, we'd get

$$G_-(\mathbf{r}) = \frac{e^{-i\,\mu\, r}}{4\,\pi\, r}. \tag{A.89}$$

Then our original solution (A.84) is just the average of these.

As solutions to the PDE, $\left(\nabla^2+\mu^2\right) G = -\delta^3$, the situation is completely reasonable – we have two independent solutions, G_+ and G_- and the equation is linear, so linear combinations (appropriately normalized) of the solutions are also solutions (we can also add any function ϕ that solves $\left(\nabla^2+\mu^2\right)\phi = 0$, expanding our options). Our choice of contour basically represents a choice of boundary condition for $G(\mathbf{r})$. But what do we make of the multiple expressions for an integral? Using complex contours to isolate a real integral introduces ambiguities in the details of the path, so we must be careful to completely specify the path and singularity-handling scheme in order to nail down the value of the integral.

Problem A.12 Evaluate the function

$$F(x) = \lim_{\epsilon\to 0}\left(-\frac{1}{i\,2\,\pi}\right)\int_{-\infty}^{\infty}\frac{e^{-i\,x\,q}}{q+i\,\epsilon}\, dq \tag{A.90}$$

for (real) $x<0$ and $x>0$. What familiar "function" is $F(x)$?

A.2.3 Example: D'Alembertian

The potential in E&M satisfies $\Box A^\mu = -\mu_0\, J^\mu$ in Lorenz gauge. The relevant Green's function solves $\Box G = -\delta^4(x)$, just the usual Poisson problem Green's function but in the expanded space-time setting. This time, we'll multiply by $e^{i\,2\,\pi\,k_\mu\,x^\mu}$ and integrate over all space-time to define the Fourier transform of the equation

$$\left(i\,2\,\pi\,k_\mu\right)\left(i\,2\,\pi\,k^\mu\right)\tilde{G} = -1. \tag{A.91}$$

Written in terms of the components of k_μ, the product $k_\mu\,k^\mu$ is $-k_0^2 + k^2$ (where $k^2 \equiv \mathbf{k}\cdot\mathbf{k}$), and we have

$$\tilde{G} = \frac{1}{4\,\pi^2\,(k^2 - k_0^2)}, \tag{A.92}$$

similar to (A.78). To Fourier transform back, we take the usual three-dimensional Fourier transform, and tack on the dk^0 integral:[3]

$$G = \int e^{-i\,2\,\pi\,k_\mu\,x^\mu}\,\tilde{G}\,d^3k\,dk^0. \tag{A.93}$$

But the three-dimensional (spatial) part of the inverse transform is just given by $G_\pm$ from (A.88) and (A.89) (with $\mu = 2\,\pi\,k^0$), so we are left with

$$G = \frac{1}{4\,\pi\,r}\int_{-\infty}^{\infty} e^{i\,2\,\pi\,k^0\,c\,t}\,e^{\pm i\,2\,\pi\,k^0\,r}\,dk^0. \tag{A.94}$$

This final integral can be evaluated using the delta function representation: $\delta(x) = \int_{-\infty}^{\infty} e^{-i\,2\,\pi\,x\,y}\,dy$,

$$G = \frac{1}{4\,\pi\,r}\int_{-\infty}^{\infty} e^{i\,2\,\pi\,k^0(c\,t\pm r)}\,dk^0 = \frac{\delta(c\,t \pm r)}{4\,\pi\,r}, \tag{A.95}$$

matching our PDE-based derivation in Section 2.2.2.[4]

A.2.4 D'Alembertian Redux

We can do the integration from the previous section in a slightly different order that will make the next section a little easier and highlight the role of complex contour integration in Green's function determination. Go back to the algebraic solution for the Fourier transform of G in (A.92): we have

$$\tilde{G} = -\frac{1}{4\,\pi^2\left(k^0 - k\right)\left(k^0 + k\right)} \tag{A.96}$$

where we have switched to the physical variable k^0 (with $k^0 = -k_0$, which makes no difference in (A.92)). The inverse Fourier transform is

$$G = -\frac{1}{4\pi^2}\int\left[\int_{-\infty}^{\infty}\frac{e^{i\,2\,\pi\,k^0\,c\,t}}{\left(k^0 - k\right)\left(k^0 + k\right)}\,dk^0\right]e^{-i\,2\,\pi\,k\,r\,\cos\theta}\,k^2\,\sin\theta\,dk d\theta d\phi \tag{A.97}$$

[3] The zero component of the wave vector has upper form $k^0 = -k_0$ as always.

[4] The Green's function in (A.95) has dimension of $1/\ell^2$, unlike our previous Green's function – it all has to do with whether you want to integrate in time or in $x^0 = c\,t$ when using G to solve the Poisson problem. In the present section, where we are interested in four-dimensional space-time, we use dx^0 to integrate over time, while previously we integrated explicitly with dt.

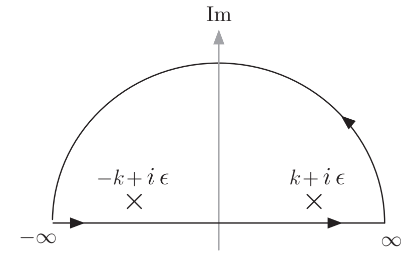

Fig. A.10 Contour and pole placement for the k^0 integration used in (A.97).

where once again we have $k^0 = -k_0$ in the exponand. This time, let's do the k^0 integration first. Since the point source is at the origin at $t' = 0$, we want $t > 0$ to enforce causality and so we'll close the integral in the upper half-plane. The poles are on the real axis, and we'll move both of them up into the upper half-plane so that they each contribute. The contour and pole positions are shown in Figure A.10. Performing that integration (and sending $\epsilon \to 0$ as always) gives

$$G = -\theta(t)\,\frac{i\,2\,\pi}{4\,\pi^2\,2}\int\left[-e^{-i\,2\,\pi\,k\,c\,t} + e^{i\,2\,\pi\,k\,c\,t}\right]e^{-i\,2\,\pi\,k\,r\,\cos\theta}\,k\,\sin\theta\,dk d\theta d\phi, \tag{A.98}$$

where we have thrown in a step function to remind us that the result only holds for $t > 0$.

Performing the angular integrals leaves

$$\begin{aligned}
G &= -\theta(t)\,\frac{1}{4\,\pi\,r}\int_0^\infty\left[-e^{-i\,2\,\pi\,k\,c\,t} + e^{i\,2\,\pi\,k\,c\,t}\right]\left[e^{i\,2\,\pi\,k\,r} - e^{-i\,2\,\pi\,k\,r}\right]dk\\
&= -\theta(t)\,\frac{1}{4\,\pi\,r}\int_0^\infty\left[-e^{i\,2\,\pi\,k\,(r-c\,t)} + e^{-i\,2\,\pi\,k\,(r+c\,t)} + e^{i\,2\,\pi\,k\,(r+c\,t)} - e^{-i\,2\,\pi\,k\,(r-c\,t)}\right]dk\\
&= -\theta(t)\,\frac{1}{4\,\pi\,r}\int_{-\infty}^\infty\left[-e^{i\,2\,\pi\,k\,(r-c\,t)} + e^{i\,2\,\pi\,k\,(r+c\,t)}\right]dk.
\end{aligned} \tag{A.99}$$

We recognize these infinite oscillatory integrals as delta functions, again, and our final expression is

$$G = -\theta(t)\,\frac{1}{4\,\pi\,r}(-\delta(r-c\,t) + \delta(r+c\,t)) = \frac{\delta(r-c\,t)}{4\,\pi\,r}, \tag{A.100}$$

where we used the step function to eliminate $\delta(r+c\,t)$. We're back to (A.95) and have identified the pole placements that define the retarded Green's function.

A.2.5 Example: Klein–Gordon

Let's apply the same technique to the case of $\Box - \mu^2$, and find the Green's function for the Klein–Gordon operator. This will be more involved, but relies on the same basic procedure as the $\mu = 0$ case. The algebraic solution for the Fourier transform of G is, as always (taking $\mu \equiv 2\pi\sigma$ to keep factors of π outside),

$$\tilde{G} = -\frac{1}{4\pi^2\left(\left(k^0\right)^2 - \left(k^2 + \sigma^2\right)\right)}. \tag{A.101}$$

Let $p = \sqrt{k^2 + \sigma^2}$, then the inverse Fourier transform will be

$$G = -\frac{1}{4\pi^2}\int\left[\int_{-\infty}^{\infty}\frac{e^{i2\pi k^0 ct}}{\left(k^0 - p\right)\left(k^0 + p\right)}\,dk^0\right]e^{-i2\pi kr\cos\theta}\,k^2\sin\theta\,dkd\theta d\phi, \tag{A.102}$$

similar to (A.97).

Thinking about the k^0 integration, there are two poles on the real axis, at $k^0 = \pm p$. Let's work out the Green's function contribution due to each separately, and from there, we can add or subtract as desired. Push the pole at $-p$ up into the complex plane, to $-p + i\epsilon$, and the pole at p down into the lower half-plane, to $p - i\epsilon$. That prescription will define what we'll call G_-, and focuses our attention on the role of the negative pole (the positive one has been pushed out of the domain enclosed by our path). Similarly, if we push the pole on the left down and the one on the right up as in Figure A.9, that will define G_+ (note that we will be closing the contour in the upper half-plane, so we are keeping contributions from each of the poles we move up). The integrals proceed as before, working through the k^0 and angular ones, and keeping $t > 0$ (for a source at $t' = 0$, $t > 0$ represents the potentially causal solutions):

$$\begin{aligned} G_- &= \frac{\theta(t)}{4\pi r}\int_0^{\infty} e^{-i2\pi pct}\left[e^{i2\pi kr} - e^{-i2\pi kr}\right]\frac{k}{p}\,dk \\ &= \frac{\theta(t)}{4\pi r}\int_{-\infty}^{\infty} e^{i2\pi(kr - pct)}\frac{k}{p}\,dk, \end{aligned} \tag{A.103}$$

with

$$G_+ = -\frac{\theta(t)}{4\pi r}\int_{-\infty}^{\infty} e^{i2\pi(kr + pct)}\frac{k}{p}\,dk. \tag{A.104}$$

The presence of $p = \sqrt{k^2 + \sigma^2}$ makes it much harder to evaluate the final integrals. They are not just delta functions (although (A.103) and (A.104) clearly reduce correctly in the $\sigma = 0$ limit).

Going back to the definition of p, $p^2 - k^2 = \sigma^2$, we can reparametrize[5] using $p = \sigma\cosh\eta$, $k = \sigma\sinh\eta$, to automatically satisfy the defining relation. The integral can be written in terms of η, since $dk = \sigma\cosh\eta\,d\eta = p\,d\eta$,

$$G_- = \frac{\theta(t)}{4\pi r}\int_{-\infty}^{\infty} e^{i2\pi\sigma(r\sinh\eta - ct\cosh\eta)}\sigma\sinh\eta\,d\eta. \tag{A.105}$$

[5] Many of the manipulations in this section, including this substitution, come from [18], where they are applied to the problem of finding a particular type of Green's function for the Klein–Gordon equation (in the quantum field theory setting). The complex contour techniques can also be found in [3].

We can clean things up a little more by noting that the r-derivative of the integrand pulls down an overall factor of $i\,2\,\pi\,\sigma\,\sinh\eta$, so that

$$G_- = \frac{\theta(t)}{4\,\pi\,r}\,\frac{1}{i\,2\,\pi}\,\frac{d}{dr}\int_{-\infty}^{\infty} e^{i\,2\,\pi\,\sigma\,(r\,\sinh\eta - c\,t\,\cosh\eta)}\,d\eta. \tag{A.106}$$

Suppose we have a field point with time-like separation, so that $-(c\,t)^2 + r^2 < 0$, then we can set $r = \sqrt{(c\,t)^2 - r^2}\,\sinh(\tau)$, $c\,t = \sqrt{(c\,t)^2 - r^2}\,\cosh(\tau)$ (for constant τ), which preserves the norm and has $r < c\,t$. The integral becomes

$$G_- = \frac{\theta(t)}{4\,\pi\,r}\,\frac{1}{i\,2\,\pi}\,\frac{d}{dr}\int_{-\infty}^{\infty} e^{-i\,2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta-\tau)}\,d\eta, \tag{A.107}$$

and a change of variables eliminates the τ (i.e., take $\eta - \tau \to Q$, then take $Q \to \eta$). The pair of Green's functions is

$$\begin{aligned} G_- &= \frac{\theta(t)}{4\,\pi\,r}\,\frac{1}{i\,2\,\pi}\,\frac{d}{dr}\int_{-\infty}^{\infty} e^{-i\,2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta)}\,d\eta \\ G_+ &= -\frac{\theta(t)}{4\,\pi\,r}\,\frac{1}{i\,2\,\pi}\,\frac{d}{dr}\int_{-\infty}^{\infty} e^{i\,2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta)}\,d\eta, \end{aligned} \tag{A.108}$$

and it is clear that $G_+ = (G_-)^*$.

To evaluate the integral in G_-, note the following integral representations of Bessel functions:[6]

$$\begin{aligned} K_0(q) &= \int_0^{\infty} \cos(q\,\sinh\eta)\,d\eta \\ J_0(q) &= \frac{2}{\pi}\int_0^{\infty} \sin(q\,\cosh\eta)\,d\eta \\ Y_0(q) &= -\frac{2}{\pi}\int_0^{\infty} \cos(q\,\cosh\eta)\,d\eta, \end{aligned} \tag{A.109}$$

and we'll use these to perform the η integration

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{-i\,2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta)}\,d\eta \\ &= 2\int_0^{\infty} \cos(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta))\,d\eta \\ &\quad - 2\,i\int_0^{\infty} \sin(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta))\,d\eta \\ &= -\pi\,Y_0(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}) - i\,\pi\,J_0(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}). \end{aligned} \tag{A.110}$$

The combination $J_p(q) + i\,Y_p(q)$ is called a "Hankel function" of the first kind and is written $H_p^{(1)} \equiv J_p(q) + i\,Y_p(q)$. The complex conjugate is called a Hankel function of the second kind: $H_p^{(2)}(q) \equiv J_p(q) - i\,Y_p(q)$, so the integral could again be rewritten as

$$\int_{-\infty}^{\infty} e^{-i\,2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}\,\cosh(\eta)}\,d\eta = -i\,\pi\,H_0^{(2)}(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}). \tag{A.111}$$

[6] These can be found in most mathematical methods books, e.g., [2].

The derivative of the Hankel function is $\frac{d}{dq} H_0^{(2)}(q) = -H_1^{(2)}(q)$, so our final Green's function form is

$$G_- = -\frac{\sigma\,\theta(t)}{4\sqrt{(c\,t)^2 - r^2}}\, H_1^{(2)}(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}). \tag{A.112}$$

Taking the complex conjugate, we obtain the expression for G_+:

$$G_+ = -\frac{\sigma\,\theta(t)}{4\sqrt{(c\,t)^2 - r^2}}\, H_1^{(1)}(2\,\pi\,\sigma\,\sqrt{(c\,t)^2 - r^2}). \tag{A.113}$$

Going back to the previous section, where we found the retarded Green's function for the space-time Poisson problem, our contour included both poles, so it is natural to associate

$$G^{\text{ret}} = G_- + G_+ = -\frac{\mu\,\theta(t)}{4\,\pi\,\sqrt{(c\,t)^2 - r^2}}\, J_1(\mu\,\sqrt{(c\,t)^2 - r^2}) \tag{A.114}$$

with the retarded Green's function. Note that it is real and contains an oscillatory portion, unlike the corresponding solution for $\mu = 0$.

The above expression for G^{ret} requires that $-(c\,t)^2 + r^2 < 0$. The result for space-like separations, where $-(c\,t)^2 + r^2 > 0$, is zero (which makes sense; there is no contribution prior to the arrival of the signal that travels at the maximum propagation speed), so we could combine these two cases by introducing another step function,

$$G^{\text{ret}} = -\frac{\mu\,\theta(t)}{4\,\pi\,\sqrt{(c\,t)^2 - r^2}}\, J_1(\mu\,\sqrt{(c\,t)^2 - r^2})\,\theta((c\,t)^2 - r^2). \tag{A.115}$$

There is one final case of interest: light-like separation between source and field points, for which $r = c\,t$. Clearly, expressions like (A.108) are divergent when $r = c\,t$ (an integral with infinite extent and unit integrand). To extract the form of infinity (which we already vaguely recognize), go back to G_- from (A.103). Suppose we have μ small, so that $k \sim p$, then

$$G_- \approx \frac{\theta(t)}{4\,\pi\,r} \int_{-\infty}^{\infty} e^{i\,2\,\pi\,k(r - c\,t)}\, dk, \tag{A.116}$$

and this is just a delta function,

$$G_- = \frac{\theta(t)}{4\,\pi\,r}\,\delta(r - c\,t), \tag{A.117}$$

the original retarded Green's function for the D'Alembertian. From (A.104), we see that the G_+ solution will involve $\delta(r + c\,t)$, which does not contribute, because of the $\theta(t)$ out front.

Our final (final) expression for the retarded Green's function looks like

$$G^{\text{ret}} = \theta(t)\left[\frac{\delta(r - c\,t)}{4\,\pi\,r} - \frac{\mu}{4\,\pi\,\sqrt{(c\,t)^2 - r^2}}\, J_1(\mu\,\sqrt{(c\,t)^2 - r^2})\,\theta((c\,t)^2 - r^2)\right]. \tag{A.118}$$

Remember that in this form, we are thinking about a source at $\mathbf{r}' = 0$ and time $t' = 0$. For a source at an arbitrary $\mathbf{r}'$ at time t', we would write

$$
\begin{aligned}
G^{\text{ret}} = \theta(t-t') \Bigg[& \frac{\delta(|\mathbf{r}-\mathbf{r}'| - c\,(t-t'))}{4\,\pi\,|\mathbf{r}-\mathbf{r}'|} \\
& - \frac{\mu}{4\,\pi\,\sqrt{(c\,(t-t'))^2 - |\mathbf{r}-\mathbf{r}'|^2}}\,J_1(\mu\,\sqrt{(c\,(t-t'))^2 - |\mathbf{r}-\mathbf{r}'|^2}) \\
& \times\,\theta((c\,(t-t'))^2 - |\mathbf{r}-\mathbf{r}'|^2)\Bigg].
\end{aligned}
\tag{A.119}
$$

Lorentz Covariance

The Bessel function term in (A.119) contains the natural Minkowski separation four-vector length. Let R^μ be the separation four-vector for the source and field point (so that $R^0 = c\,(t-t')$ and $\mathbf{R} = \mathbf{r} - \mathbf{r}'$), then the quantity

$$
(c\,(t-t'))^2 - |\mathbf{r}-\mathbf{r}'|^2 = -R^\mu\,R_\mu, \tag{A.120}
$$

a Lorentz scalar. The second term in brackets in (A.119) depends on $R^\mu\,R_\mu$, so is manifestly scalar. The first term looks problematic, and we should have worried earlier, since that first term is the "usual" retarded Green's function from E&M. Is it a scalar or not?

Go back to its original form (to make the calculation a little easier); we want to evaluate $\frac{\delta(r-c\,t)}{4\,\pi\,r}$. We'd like to write this term to depend explicitly on $r^2 - (c\,t)^2$, which is a Lorentz scalar. The delta function is a natural place to start; think of $\delta(r^2 - (c\,t)^2)$ as a delta function that depends on $f(c\,t) \equiv r^2 - (c\,t)^2$. Then we know that under an r-integration, we would have (from (2.41))

$$
\delta(r^2 - (c\,t)^2) = \frac{\delta(r-c\,t)}{|f'(c\,t=r)|} + \frac{\delta(r+c\,t)}{|f'(c\,t=-r)|} \tag{A.121}
$$

since the two roots of $f(c\,t)$ are $c\,t = \pm r$. But $f'(c\,t) = -2\,(c\,t)$, so that

$$
\delta(r^2 - (c\,t)^2) = \frac{\delta(r-c\,t)}{2\,r} + \frac{\delta(r+c\,t)}{2\,r}. \tag{A.122}
$$

The first term appears in the Green's function, and the second term is canceled by the $\theta(t)$ that sits in front (in (A.118)), so we can write:

$$
\theta(t)\,\frac{\delta(r-c\,t)}{2\,r} = \theta(t)\,\delta(r^2 - (c\,t)^2). \tag{A.123}
$$

Now we can move the source to $\mathbf{r}'$ at t' and write the whole retarded Green's function in terms of $R^\sigma\,R_\sigma$:

$$
G^{\text{ret}} = \theta(t-t')\left[\frac{\delta(R^\sigma\,R_\sigma)}{2\,\pi} - \frac{\mu\,J_1(\mu\,\sqrt{-R^\sigma\,R_\sigma})}{4\,\pi\,\sqrt{-R^\alpha\,R_\alpha}}\,\theta(-R^\beta\,R_\beta)\right]. \tag{A.124}
$$

The step function out front is not clearly Lorentz invariant. Is it possible that some frame could have $t - t' > 0$ while another has $t - t' < 0$? The Lorentz boost for $\Delta t \equiv t - t'$ and

$\Delta r = r - r'$ (working in the one dimension defined by the line connecting the source point to the field point) gives

$$\Delta \bar{t} = \gamma \left(\Delta t - \frac{v}{c^2} \Delta r \right). \tag{A.125}$$

If $\Delta t > 0$ and $\Delta \bar{t} < 0$, then

$$\Delta t - \frac{v}{c^2} \Delta r < 0 \longrightarrow c^2 < v \frac{\Delta r}{\Delta t}. \tag{A.126}$$

For $v < c$, we would need $\Delta r / \Delta t > c$, and neither the delta function nor the step function inside the brackets in (A.124) would contribute at that space-time separation (which would have $-R^\sigma R_\sigma < 0$) in the original frame. So the step function in the Green's function evaluates to zero in both the original and boosted frame, and we conclude that the entire Green's function is Lorentz covariant.

As a final note: the retarded Green's function for $\mu \neq 0$ is much more complicated than its $\mu = 0$ counterpart. Field points depend sensitively on the relative location of the source, and there are many points in the past history of the source that can contribute after initial "contact" is made.

A.3 Series Solutions

The "Frobenius method," or "series solution" method for solving ODEs, is a useful tool and shows up in a variety of different contexts. From the point of view of field theories, ODEs arise as the symmetric or separable solutions to field equations (point source, line source, etc.). We'll review the series method and some of its applications to ODEs of interest to us.

Given an ODE governing a function $f(x)$, the idea behind the series solution method is to expand $f(x)$ in powers of x with unknown coefficients, then to use the differential equation to set up a relation between the unknown coefficients that can ultimately be used to solve for the coefficients. As an example, let's solve the familiar case of

$$\frac{d^2 f(x)}{dx^2} + \omega^2 f(x) = 0. \tag{A.127}$$

Start with

$$f(x) = x^p \sum_{j=0}^{\infty} a_j x^j, \tag{A.128}$$

where we put x^p out front, and we'll determine p as part of the solution process. Putting x^p in front means we can start the sum off at $j = 0$ (suppose, for example, that the solution to a differential equation was $7x^{50}$ – without the x^p in front, we'd have to set the first 49 coefficients to zero to uncover the solution). Inserting the series ansatz into the original differential equation gives

$$x^p \left[\sum_{j=0}^{\infty} a_j (j+p)(j+p-1) x^{j-2} + \omega^2 \sum_{j=0}^{\infty} a_j x^j \right] = 0. \tag{A.129}$$

We'd like to combine powers of x and eliminate each one separately by choosing coefficients appropriately. If you don't cancel each term individually, then you can find some value of x that will make the left-hand side of the above non-zero, which is not good (think of the left-hand side as a simple polynomial – it can have roots, but we require that it be zero everywhere). In order to collect terms, take $k = j - 2$ in the first sum, then

$$\sum_{j=0}^{\infty} a_j\,(j+p)\,(j+p-1)\;x^{j-2} = \sum_{k=-2}^{\infty} a_{k+2}\,(k+p+2)\,(k+p+1)\;x^k, \tag{A.130}$$

and relabel $k \to j$ to put it back in the form of (A.129). Now we can combine terms in the two sums in (A.129), but only at $j \geq 0$; the first two terms from the derivative portion of the ODE stand on their own:

$$a_0\,p\,(p-1)+a_1\,(p+1)\,p\,x^{-1}+\sum_{j=2}^{\infty}\Big[(j+p+2)\,(j+p+1)\;a_{j+2}+\omega^2\,a_j\Big]\;x^j = 0. \tag{A.131}$$

The summand provides a recursive relation between the coefficients $\{a_j\}_{j=2}^{\infty}$ such that each power of x is zero independently:

$$a_{j+2} = -\frac{\omega^2\,a_j}{(j+p+2)\,(j+p+1)}. \tag{A.132}$$

If we take $p = 0$, then both of the first two terms in (A.131) are zero, and the recursion ensures that the infinite sum is also zero, so we have a solution. Let's write out the first few coefficients, starting at arbitrary a_0 (nothing constrains it):

$$\begin{aligned} a_2 &= -\frac{\omega^2\,a_0}{2\times 1} \\ a_4 &= -\frac{\omega^2\,a_2}{4\,\times 3} = \frac{\omega^4\,a_0}{4\,\times 3\,\times 2\,\times 1}. \end{aligned} \tag{A.133}$$

This is an alternating series, and we can solve it from these first few terms

$$a_{2j} = \frac{(-1)^j\,\omega^{2j}}{(2j)!}. \tag{A.134}$$

Putting this into the series ansatz, and taking $a_1 = 0$ to cancel the odd terms (for now) gives

$$f(x) = a_0\sum_{j=0}^{\infty}\frac{(-1)^j}{(2j)!}\,(\omega\,x)^{2j} = a_0\,\cos(\omega\,x), \tag{A.135}$$

where the final equality defines the cosine series.

This time, let's set $a_0 = 0$ and take a_1 arbitrary. The recursion is the same,

$$\begin{aligned} a_3 &= -\frac{\omega^2\,a_1}{3\,\times 2} \\ a_5 &= -\frac{\omega^2\,a_3}{5\,\times 4} = \frac{\omega^4\,a_1}{5\,\times 4\,\times 3\,\times 2\,\times 1}, \end{aligned} \tag{A.136}$$

giving

$$a_{2j+1} = \frac{(-1)^j\,\omega^{2j}}{(2j+1)!}\,a_1, \tag{A.137}$$

and the resulting form for $f(x)$ is

$$f(x) = a_1 \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!}\,\frac{(\omega\,x)^{2j+1}}{\omega} = \frac{a_1}{\omega}\,\sin(\omega\,x). \tag{A.138}$$

We recover the sine solution (notice the funny business with the extra factor of ω, that's just to clean up the expression; it can get absorbed into a_1). The full solution would be a combination of these two,

$$f(x) = a_0\,\cos(\omega\,x) + \frac{a_1}{\omega}\,\sin(\omega\,x). \tag{A.139}$$

Problem A.13 Use the series solution method to solve $f'(x) = \alpha\,f(x)$ for constant α, with $f(0) = f_0$ given.

Problem A.14 What happens if you use $p = 1$ and $p = -1$ to cancel the first and second terms (respectively) in (A.131)?

Problem A.15 Suppose you change the sign of the ODE. Find the pair of solutions for $\frac{d^2f(x)}{dx^2} - \omega^2 f(x) = 0$ using the series method.

A.3.1 Example: Legendre Polynomials

Another equation of interest to us is Legendre's ODE:

$$(1-x^2)\,\frac{d^2f(x)}{dx^2} - 2\,x\,\frac{df(x)}{dx} + n\,(n+1)\,f(x) = 0 \tag{A.140}$$

for integer n. This time, we need both the first and second derivative of $f(x)$. Starting with (A.128), the ODE can be written

$$\begin{aligned} &a_0\,p\,(p-1) + a_1\,p\,(p+1)\,x^{-1} \\ &\quad + \sum_{j=0}^{\infty} \big[a_{j+2}\,(j+p+2)\,(j+p+1) \\ &\quad - a_j\,((j+p)\,(j+p-1) + 2\,(j+p) - n\,(n+1))\big]\,x^j = 0. \end{aligned} \tag{A.141}$$

The recursion relation gives:

$$a_{j+2} = \frac{(j+p)\,(j+p+1) - n\,(n+1)}{(j+p+2)\,(j+p+1)}\,a_j. \tag{A.142}$$

Setting $p = 0$, we again have a pair of series, one even, one odd. Take $a_1 = 0$ and let a_0 be constant; the recursion is

$$a_{j+2} = \frac{j\,(j+1) - n\,(n+1)}{(j+2)\,(j+1)}\,a_j. \tag{A.143}$$

The index j moves in steps of 2, and if n is even, the coefficient a_{n+2} will vanish, and hence all subsequent coefficients will also vanish. The solution, in this case, will be a polynomial of degree n. Similarly, if we had chosen $a_0 = 0$ with a_1 some constant, we would get an odd polynomial. These polynomials of degree n solving (A.140) are called the Legendre polynomials.

Problem A.16 Write out the Legendre polynomials for $n = 1, 2, 3$.

A.3.2 Example: Bessel Functions

We'll do one last case, relevant to our Green's function solutions in this appendix. Bessel's ODE is

$$x^2 \frac{d^2 f(x)}{dx^2} + x \frac{df(x)}{dx} + (x^2 - \nu^2) f(x) = 0. \tag{A.144}$$

In general, ν can be anything (integer or not). For simplicity, we'll take $\nu = 0$ to see how the series solution works in this case. Using the familiar starting point (A.128), and inserting into the Bessel ODE with $\nu = 0$,

$$a_0 p^2 + a_1 (1+p)^2 x + \sum_{j=2}^{\infty} \left[a_j (j+p)^2 + a_{j-2} \right] x^j = 0, \tag{A.145}$$

giving the recursion

$$a_j = -\frac{a_{j-2}}{(j+p)^2}. \tag{A.146}$$

Note the similarity between this recursive update and the one for sine and cosine (A.132) (with $\omega = 1$). Take $p = 0$ and $a_1 = 0$ to define the even function with constant a_0, then the first few terms of the recursion give

$$\begin{aligned} a_2 &= -\frac{a_0}{4} \\ a_4 &= -\frac{a_2}{16} = \frac{a_0}{4 \times 16} \\ a_6 &= -\frac{a_4}{36} = -\frac{a_0}{4 \times 16 \times 36}. \end{aligned} \tag{A.147}$$

From these we can write

$$a_j = \frac{(-1)^j a_0}{\prod_{k=1}^{j} (2k)^2} = \frac{(-1)^j a_0}{4^j (j!)^2}, \tag{A.148}$$

and the solution reads

$$f(x) = a_0 \sum_{j=0}^{\infty} \frac{(-1)^2 \left(\frac{x}{2}\right)^{2j}}{(j!)^2}. \tag{A.149}$$

Compare this with the cosine series; they are similar in many ways. It is interesting that if, for example, the photon had mass (see Problem 3.38 for this association), and so had a Green's function more like the Klein–Gordon case, it is Bessel's ODE and the resulting

Bessel functions that would be the most familiar, while sine and cosine would have less utility (at least for field theories!).

The general series solution here, for arbitrary ν, is

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\ (k+\nu)!} \left(\frac{x}{2}\right)^{\nu+2k}. \tag{A.150}$$

A.4 Iterated Integrals

When presented with an equation like $\nabla^2\phi - \mu^2\,\phi = 0$, there is an interesting way to think about solutions. We already have the Green's function for this operator (from Section A.2.1), and so can form most solutions for non-zero sources via integrals, or infinite sums (if the boundary conditions are different from the point source one). But we could also imagine an iterative approach. Write the equation as a Poisson problem with self-sourcing: $\nabla^2\phi = \mu^2\,\phi$, then we can write down the integral solution for ϕ,

$$\phi(\mathbf{r}) = \bar{\phi}(\mathbf{r}) + \mu^2 \int G(\mathbf{r}, \mathbf{r}')\,\phi(\mathbf{r}')\,d\tau' \tag{A.151}$$

where we pick $\bar{\phi}(\mathbf{r})$ harmonic: $\nabla^2\bar{\phi}(\mathbf{r}) = 0$. In the integrand, $G(\mathbf{r}, \mathbf{r}')$ is the Green's function for ∇^2, that's (2.12), but we'll leave it unevaluated for now. Insert this $\phi(\mathbf{r})$ into the integral in (A.151) to get

$$\begin{aligned}\phi(\mathbf{r}) &= \bar{\phi}(\mathbf{r}) + \mu^2 \int G(\mathbf{r}, \mathbf{r}') \left[\bar{\phi}(\mathbf{r}') + \mu^2 \int G(\mathbf{r}', \mathbf{r}'')\,\phi(\mathbf{r}'')\,d\tau''\right] d\tau' \\ &= \bar{\phi}(\mathbf{r}) + \mu^2 \int G(\mathbf{r}, \mathbf{r}')\,\bar{\phi}(\mathbf{r}')\,d\tau' + \mu^4 \int G(\mathbf{r}, \mathbf{r}') \left(\int (G(\mathbf{r}', \mathbf{r}'')\,\phi(\mathbf{r}'')\,d\tau''\right) d\tau'.\end{aligned} \tag{A.152}$$

We could continue the process, inserting $\phi(\mathbf{r})$ into the double integral, and incurring more integrals, each with an additional factor of μ^2 attached. Each integral involves only the original harmonic function $\bar{\phi}$, while pushing back the dependence on ϕ itself to a higher order in μ. That's all fine provided $|\mu^2| < 1$, so that each term contributes less and less, then we could sensibly write:

$$\begin{aligned}\phi(\mathbf{r}) = \bar{\phi}(\mathbf{r}) &+ \mu^2 \int G(\mathbf{r}, \mathbf{r}')\,\bar{\phi}(\mathbf{r}')\,d\tau' + \mu^4 \int G(\mathbf{r}, \mathbf{r}') \left(\int G(\mathbf{r}', \mathbf{r}'')\,\bar{\phi}(\mathbf{r}'')\,d\tau''\right) d\tau' \\ &+ \mu^6 \int G(\mathbf{r}, \mathbf{r}') \left(\int G(\mathbf{r}', \mathbf{r}'') \left[\int G(\mathbf{r}'', \mathbf{r}''')\,\bar{\phi}(\mathbf{r}''')\,d\tau'''\right] d\tau''\right) d\tau' + \cdots.\end{aligned} \tag{A.153}$$

Another way of approaching the same series is to write $\phi = \phi_0 + \mu^2\,\phi_1 + \mu^4\,\phi_2 + \cdots$. Then, inserting this into $\nabla^2\phi = \mu^2\,\phi$,

$$\nabla^2\left(\phi_0 + \mu^2\,\phi_1 + \mu^4\,\phi_2 + \cdots\right) = \mu^2\left(\phi_0 + \mu^2\,\phi_1 + \mu^4\,\phi_2 + \cdots\right), \tag{A.154}$$

and collecting in powers of μ^2, we have a tower of equations to solve:

$$\begin{aligned}
\nabla^2 \phi_0 &= 0 \\
\nabla^2 \phi_1 &= \phi_0 \\
\nabla^2 \phi_2 &= \phi_1 \\
\nabla^2 \phi_3 &= \phi_2 \\
&\vdots
\end{aligned} \tag{A.155}$$

The top equation is solved by any harmonic function (and that function can be used to establish boundary conditions appropriate to the problem at hand). Let $\phi_0 = \bar{\phi}$ (harmonic, as above). Then the second equation gives ϕ_1 with ϕ_0 as its source, so we would write the formal solution as

$$\phi_1(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}')\, \phi_0(\mathbf{r}')\, d\tau' = \int G(\mathbf{r}, \mathbf{r}')\, \bar{\phi}(\mathbf{r}')\, d\tau', \tag{A.156}$$

and with ϕ_1 in hand, we move on to ϕ_2, which has formal solution

$$\phi_2(\mathbf{r}) = \int G(\mathbf{r}, \mathbf{r}')\, \phi_1(\mathbf{r}')\, d\tau' = \int G(\mathbf{r}, \mathbf{r}') \left(\int G(\mathbf{r}', \mathbf{r}'')\, \bar{\phi}(\mathbf{r}'')\, d\tau'' \right) d\tau', \tag{A.157}$$

again leading to a sequence like (A.153).

A.4.1 Example: Anti-Helmholtz Green's Function

As an example of the process, let's find the Green's function for $\nabla^2 - \mu^2$. We want ϕ to solve

$$\nabla^2 \phi - \mu^2\, \phi = -\delta^3(\mathbf{r}) \tag{A.158}$$

obtained by iterated integral. This time, we'll take ϕ_0 such that

$$\nabla^2 \phi_0 = -\delta^3(\mathbf{r}), \tag{A.159}$$

which is solved by the familiar $\phi_0 = \frac{1}{4\pi r}$. Next, we want to use this as a source for the ϕ_1 equation:

$$\nabla^2 \phi_1 = \phi_0 \longrightarrow \frac{1}{r}\, (r\, \phi_1)'' = \frac{1}{4\pi r}, \tag{A.160}$$

with primes denoting r-derivatives as usual. The solution here is

$$\phi_1 = \frac{1}{4\pi r} \frac{1}{2!} r^2 + \alpha + \frac{\beta}{r} \tag{A.161}$$

for constant α and β that we'll set to zero (our choice, although keeping those constants would allow us to tune the final result for particular types of boundary conditions).

Continuing,

$$\nabla^2 \phi_2 = \phi_1 \longrightarrow \frac{1}{r}\, (r\, \phi_2)'' = \frac{1}{4\pi} \frac{1}{2!}\, r \tag{A.162}$$

with solution $\phi_2 = \frac{1}{4\pi r}\frac{1}{4!}r^4$. The pattern continues, with

$$\phi_j = \frac{1}{4\pi r}\frac{1}{(2j)!}r^{2j}, \tag{A.163}$$

and then the complete solution is

$$\phi = \sum_{j=0}^{\infty} \mu^{2j}\phi_j = \frac{1}{4\pi r}\sum_{j=0}^{\infty}\frac{(\mu r)^{2j}}{(2j)!}. \tag{A.164}$$

The infinite sum is just $\cosh(\mu r)$, so we end up with

$$\phi = \frac{\frac{1}{2}\left(e^{-\mu r} + e^{\mu r}\right)}{4\pi r}, \tag{A.165}$$

which is the real (meaning real exponentials instead of complex ones) version of (2.18). Notice that we ended up with a mixture of growing and decaying exponentials (analogues, here, of advanced and retarded time from the Fourier transform of the D'Alembertian's Green's function). Presumably, we'd pick the decaying solution, and this would come from keeping the coefficients α and β (and their successors) from above. Regardless, we have "a" Green's function for $\nabla^2 - \mu^2$, obtained by iteration.

Appendix B Numerical Methods

There is a class of numerical methods that can be used to solve problems that come from field theories. In this appendix, we will review some of those problems and introduce numerical methods that can be used to solve them. Some of this material is adapted from [13]; [22] and [16] are good for additional reading.

B.1 Root-Finding

Given a function $F(x)$, find some or all of the set $\{\bar{x}_i\}_{i=1}^n$ such that $F(\bar{x}_i) = 0$. This defines the root-finding problem, and we have encountered it in the context of the retarded time condition from Section 2.3. Specifically, if we want to find the electric field of a point charge moving along according to $\mathbf{w}(t)$ as in (2.177), we need to solve the implicit equation:

$$c\,(t - t_r) = |\mathbf{r} - \mathbf{w}(t_r)| \tag{B.1}$$

for t_r. To cast this in the form of the root-finding problem, define

$$F(x) \equiv c\,(t - x) - |\mathbf{r} - \mathbf{w}(x)| \tag{B.2}$$

and then the root of the function $F(x)$ (there will be only one provided the motion described by $\mathbf{w}(t)$ is subluminal) gives the retarded time, from which we can evaluate the separation vector, velocity and acceleration vectors, and the rest of the elements in $\mathbf{E}$ and $\mathbf{B}$.

The procedure we will use is called "bisection." Start with a pair of values, x_ℓ^0 and x_r^0 with $x_\ell^0 < x_r^0$ (hence the ℓ and r designations) and $F(x_\ell^0)\,F(x_r^0) < 0$ (so that a root is in between the points). Then calculate the midpoint between this pair, $x_m^0 \equiv \frac{1}{2}\left(x_\ell^0 + x_r^0\right)$, and evaluate the product $p \equiv F(x_\ell^0)\,F(x_m^0)$. If p is less than zero, the root lies between the left and middle points, so we can move the left and right points over by setting $x_\ell^1 = x_\ell^0$, $x_r^1 = x_m^0$. If p is greater than zero, the root is between the middle and right points, and we update $x_\ell^1 = x_m^0$, $x_r^1 = x_r^0$. The iteration is continued until the absolute value of F at the current midpoint is smaller than some tolerance ϵ,

$$|F(x_m^n)| \le \epsilon \tag{B.3}$$

and then x_m^n is an approximation to the root. The process is shown pictorially in Figure B.1 and is described with pseudocode in Algorithm B.1.

Bisection is simple and requires only that we evaluate the function $F(x)$ over and over. As written in Algorithm B.1, we have not checked that a single root is enclosed by the

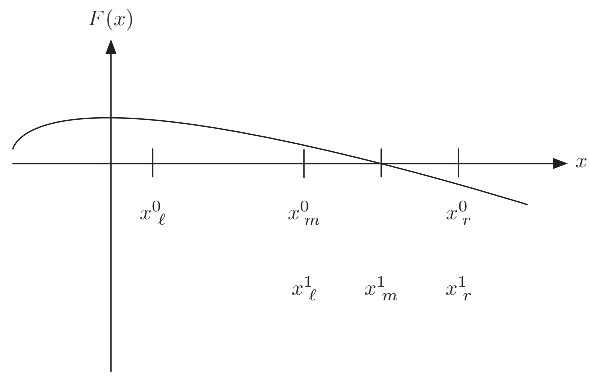

Fig. B.1 Two iterations of the bisection procedure.

Algorithm B.1 Bisection($F, x^0_\ell, x^0_r, \epsilon$)

$x_\ell \leftarrow x^0_\ell$
$x_r \leftarrow x^0_r$
$x_m \leftarrow 1/2\,(x_\ell + x_r)$
while $|F(x_m)| > \epsilon$ **do**
  if $F(x_m)\,F(x_r) < 0$ **then**
    $x_\ell \leftarrow x_m$
  else
    $x_r \leftarrow x_m$
  end if
  $x_m \leftarrow 1/2\,(x_\ell + x_r)$
end while
return x_m

initial bracketing (a requirement for correct functioning). To use this routine, you would first take a look at some representative points; making a plot of $F(x)$ is a good idea. Then you can pick the initial values for x^0_ℓ and x^0_r and be reasonably sure that you have bracketed one, and only one, root.[1]

B.1.1 Example: The Electric Field of a Moving Charge

Given a vector describing the location of a charge q at time t, $\mathbf{w}(t)$, we can use bisection to find the retarded time associated with a field point $\mathbf{r}$ at time t. With t_r in hand, we can evaluate all of the pieces of (2.177). Take $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$ as in Section 2.7, then we can use the routine described in Algorithm B.2 to generate the electric field at an individual

[1] There are functions that have an infinite number of roots between any two x values, and it is not easy to find those roots.

Algorithm B.2 Efield$(x, y, z, t, \mathbf{w}, \epsilon)$

$\mathbf{r} \leftarrow \{x, y, z\}$
$F(\bar{x}) \leftarrow c\,(t - \bar{x}) - |\mathbf{r} - \mathbf{w}(\bar{x})|$
$tr \leftarrow$ Bisection$(F, -\infty, t, \epsilon)$
$\boldsymbol{\mathscr{r}} \leftarrow \mathbf{r} - \mathbf{w}(t_r)$
$\hat{\boldsymbol{\mathscr{r}}} \leftarrow \boldsymbol{\mathscr{r}} / \mathscr{r}$
$\mathbf{u} \leftarrow c\,\hat{\boldsymbol{\mathscr{r}}} - \dot{\mathbf{w}}(t_r)$
$\mathbf{E} \leftarrow \frac{q}{4\pi\epsilon_0} \frac{|\boldsymbol{\mathscr{r}}|}{(\boldsymbol{\mathscr{r}}\cdot\mathbf{u})^3} \left[\left(c^2 - \dot{\mathbf{w}}(t_r)\cdot\dot{\mathbf{w}}(t_r)\right)\mathbf{u} + \boldsymbol{\mathscr{r}} \times (\mathbf{u} \times \ddot{\mathbf{w}}(t_r))\right]$
return E

field point, and by evaluating the field at a variety of points (in a grid, for example), we can generate a map of the electric field.

In order to generate a figure like Figure 2.6, we apply Algorithm B.2 with $d = 0.01$ m, $\omega = 5$ Hz, to a set of points: $x = 0$, $y = -5 \rightarrow 5$, $z = 2\,d \rightarrow 5$ in steps of 0.1, and we run time from 0 to $3\,T$ with $T = 2\,\pi/\omega$ in steps of $T/10$. To make the numbers manageable, we set $c = 1$ and $q/(4\,\pi\,\epsilon_0) = 1$. The bisection present in Algorithm B.2 requires a parameter "$-\infty$" to set the left side of the initial root bracketing, and we take this to be $-\infty$"="$-10,000$. You can calculate, for a given grid of field points, the maximum retarded time, and set $-\infty$ to that number in practice.

Problem B.1 A charged particle moves along the trajectory described by

$$\mathbf{w}(t) = R\left(\cos(\omega\,t)\,\hat{\mathbf{x}} + \sin(\omega\,t)\,\hat{\mathbf{y}} + \omega\,t\,\hat{\mathbf{z}}\right). \tag{B.4}$$

Use the bisection routine (with $\epsilon = 10^{-10}$ as your tolerance) to find the retarded time for the field point $\mathbf{r} = 0$ at $t = 0$ if $R = 10^5$ m, $\omega = 100$ s^{-1}.

Problem B.2 The bisection procedure has a convergence rate that is set by the initial bracketing (the choice of x_ℓ^0 and x_r^0) and the desired tolerance ϵ. Roughly how many steps must you take to achieve $|F(x_m^n)| \le \epsilon$ given x_ℓ^0, x_r^0, and ϵ? (Assume that near the root, F is linear, a good approximation provided ϵ is small.)

Problem B.3 For the electric field portion of (2.177), given $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{x}} + d\,\sin(\omega\,t)\,\hat{\mathbf{y}}$ with $d = 0.01$ m, $\omega = 5$ s^{-1}, and taking $c = 1$, $q/(4\,\pi\,\epsilon_0) = 1$ for simplicity, find $\mathbf{E}$ and make a contour plot of its magnitude at $t = 0$ for points with $x = 0$, and between -5 and 5 in y, and between $2\,d$ and 5 in z.

Problem B.4 Find (and plot the contours of) the magnetic field associated with $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$ using the parameters described in the previous problem.

Problem B.5 Isolate the acceleration portion of the electric field for a charge moving according to $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$ and plot it separately for the parameters given above.

Problem B.6 Write a function that finds the potential V from (2.66) given a particle trajectory $\mathbf{w}(t) = d\,\cos(\omega\,t)\,\hat{\mathbf{z}}$. Using the same parameters as above, make a contour plot of the potential.

B.2 Solving ODEs

Once we can evaluate the electric and magnetic fields, we have to be able to use them in equations of motion (relativistic or not), and those equations of motion themselves require numerical solution in many cases. As a simple example, take the electric and magnetic fields of an infinite line of charge with constant charge-per-unit-length λ moving with constant speed v along the $\hat{\mathbf{z}}$-axis. The static solutions for the fields are

$$\mathbf{E} = \frac{\lambda}{2\pi\,\epsilon_0\,s}\,\hat{\mathbf{s}}, \qquad \mathbf{B} = \frac{\mu_0\,\lambda\,v}{2\pi\,s}\,\hat{\boldsymbol{\phi}}, \tag{B.5}$$

leading to (non-relativistic) equations of motion for a particle of charge q (mass m) that has velocity in the y–z plane initially:

$$\begin{aligned} m\,\ddot{y} &= \frac{\lambda\,q}{2\pi\,\epsilon_0\,y} - \frac{q\,\dot{z}\,\mu_0\,\lambda\,v}{2\pi\,y} \\ m\,\ddot{z} &= \frac{q\,\dot{y}\,\mu_0\,\lambda\,v}{2\pi\,y}. \end{aligned} \tag{B.6}$$

The moving line of charge is one of the simplest configurations of moving charge we can imagine, and yet solving the equations of motion (B.6) is not analytically tractable. The relativistic version is even worse.

Most ODE-solvers prefer to treat equations as first-order vector ODEs.[2] For (B.6), we could accomplish this by assigning the elements of a vector with four entries, **f**, as follows: $f_1 \equiv y, f_2 \equiv z, f_3 \equiv \dot{y}, f_4 \equiv \dot{z}$. Then we can write the time-evolution of **f**, from (B.6), as

$$\frac{d}{dt}\begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} = \begin{pmatrix} f_3 \\ f_4 \\ \frac{\lambda\,q}{2\pi\,m\,\epsilon_0\,f_1} - \frac{q f_4\,\mu_0\,\lambda\,v}{2\pi\,f_1} \\ \frac{q f_3\,\mu_0\,\lambda\,v}{2\pi\,f_1} \end{pmatrix}. \tag{B.7}$$

The more general form is

$$\frac{d}{dt}\,\mathbf{f} = \mathbf{G}(\mathbf{f}) \tag{B.8}$$

where **f** is a vector of whatever length is appropriate to the problem, and **G** is a vector of functions that depend on **f** (and potentially time itself, although we will omit that explicit dependence to avoid clutter). For the current example, **G**(**f**) is the vector appearing on the right-hand side of (B.7). We must also be given the initial vector **f**(0), which will contain all the initial values of position and velocity for our current example.

Problem B.7 Write out the relativistic equations of motion that generalize (B.6).

Problem B.8 Write the third order ODE $\dddot{x} = -\alpha\,x^2 - \beta\,\dot{x} + F(x)$ in vector form (here, α and β are constants, and $F(x)$ is some provided function).

[2] Vector here does not come with any of the usual coordinate-transformation interpretation. In this section, the "vector" is, finally, just a collection of objects.

Problem B.9 Predict the qualitative form of the motion of a charge moving under the influence of the electric and magnetic fields given in (B.5) assuming the charge starts from rest a distance s_0 from the line of charge.

B.2.1 Discretization and Approach

We want a numerical method that will allow us to generate an approximate solution for $\mathbf{f}(t)$ in (B.8), and one way to go about doing that is to introduce a temporal "grid." Imagine chopping time up into small chunks of size Δt, then we could refer to the jth chunk as $t_j \equiv j\,\Delta t$. We'll develop a numerical method for approximating the values of $\mathbf{f}$ at these special time points, and we can use the notation $\mathbf{f}_j \equiv \mathbf{f}(t_j)$, the true solution evaluated at the discrete time.

The first approach to approximating the set $\{\mathbf{f}_j\}_{j=0}^N$ (assuming we are interested in the solution up to finite time t_N) is to note that

$$\mathbf{f}_{j+1} \approx \mathbf{f}_j + \Delta t \left.\frac{d\mathbf{f}(t)}{dt}\right|_{t=t_j} = \mathbf{f}_j + \Delta t\,\mathbf{G}(\mathbf{f}_j) \tag{B.9}$$

from Taylor expansion, and where we used the differential equation that $\mathbf{f}$ solves in writing the equality.

That observation suggests that the sequence defined by $\bar{\mathbf{f}}_0 = \mathbf{f}_0$ (the provided initial condition) with update

$$\bar{\mathbf{f}}_{j+1} = \bar{\mathbf{f}}_j + \Delta t\,\mathbf{G}(\bar{\mathbf{f}}_j) \tag{B.10}$$

defines a set $\{\bar{\mathbf{f}}_j\}_{j=0}^N$ that approximates the desired set $\{\mathbf{f}_j\}_{j=0}^N$. The barred approximates do match the true solution up to a point, but at each step an error of order Δt^2 is made, and those accumulate in an unfortunate fashion. The method described by (B.10) is called "Euler's method." It is not "stable" (referring to that accumulation of error) and cannot be reliably used.

Example: Stability

To see how a method can be unstable, let's take Euler's method applied to the simple problem of $\frac{df(t)}{dt} = i\,\alpha\,f(t)$. The solution to this problem is $f(t) = f_0\,e^{i\,\alpha\,t}$. The Euler's method update gives

$$\bar{f}_{j+1} = \bar{f}_j + \Delta t\,G(\bar{f}_j) = \bar{f}_j\,(1 + i\,\alpha\,\Delta t)\,, \tag{B.11}$$

which can be solved:

$$\bar{f}_{j+1} = (1 + i\,\alpha\,\Delta t)^j\,f_0. \tag{B.12}$$

The magnitude of the solution, at time level $j+1$, is

$$|\bar{f}_{j+1}| = \sqrt{\left(1 + \alpha^2\,\Delta t^2\right)^j}\,|f_0|, \tag{B.13}$$

and the magnitude grows as time goes on. That growth is a general feature of unstable methods and renders them useless over long time periods.

To make progress, we can continue the Taylor expansion from (B.9),

$$\begin{aligned}\mathbf{f}_{j+1} &\approx \mathbf{f}_j + \Delta t \left.\frac{d\mathbf{f}(t)}{dt}\right|_{t=t_j} + \frac{1}{2}\Delta t^2 \left.\frac{d^2\mathbf{f}(t)}{dt^2}\right|_{t=t_j} \\ &= \mathbf{f}_j + \Delta t\,\mathbf{G}(\mathbf{f}_j) + \frac{1}{2}\Delta t^2 \left.\frac{d\mathbf{G}}{dt}\right|_{t=t_j}\end{aligned} \tag{B.14}$$

where we kept an additional term from the Taylor expansion, and used the ODE again. The time-derivative of $\mathbf{G}$ can be written in terms of its dependence on $\mathbf{f}$,

$$\left.\frac{d\mathbf{G}}{dt}\right|_{t=t_j} = \frac{\partial \mathbf{G}}{\partial \mathbf{f}} \cdot \frac{d\mathbf{f}}{dt} = \frac{\partial \mathbf{G}}{\partial \mathbf{f}} \cdot \mathbf{G}, \tag{B.15}$$

using the ODE once more. There is a matrix of derivatives here; the shorthand $\frac{\partial \mathbf{G}}{\partial \mathbf{f}}$ refers to a matrix with i–j entry $\frac{\partial G_i}{\partial f_j}$. The expansion in (B.14) can be expressed in terms of the derivative with respect to $\mathbf{f}$,

$$\mathbf{f}_{j+1} \approx \mathbf{f}_j + \Delta t\left(\mathbf{G}(\mathbf{f}_j) + \frac{1}{2}\Delta t\,\frac{\partial \mathbf{G}}{\partial \mathbf{f}} \cdot \mathbf{G}(\mathbf{f}_j)\right). \tag{B.16}$$

It would be nice to avoid constructing and calculating the matrix $\frac{\partial \mathbf{G}}{\partial \mathbf{f}}$ explicitly. There is a clever trick to avoid that construction, and it comes from a reimagining of the term in parentheses in (B.16). That term appears in the Taylor expansion of

$$\mathbf{G}\left(\mathbf{f}_j + \frac{1}{2}\Delta t\,\mathbf{G}(\mathbf{f}_j)\right) \approx \mathbf{G}(\mathbf{f}_j) + \frac{1}{2}\Delta t\,\frac{\partial \mathbf{G}}{\partial \mathbf{f}} \cdot \mathbf{G}(\mathbf{f}_j) \tag{B.17}$$

so that we could, finally, write (B.16) as

$$\mathbf{f}_{j+1} \approx \mathbf{f}_j + \Delta t\,\mathbf{G}\left(\mathbf{f}_j + \frac{1}{2}\Delta t\,\mathbf{G}(\mathbf{f}_j)\right) \tag{B.18}$$

without significantly changing the accuracy of the approximation.[3] This observation can be used to define the "midpoint method" or "midpoint Runge–Kutta," where the approximates are defined by $\bar{\mathbf{f}}_0 = \mathbf{f}_0$ and the update

$$\begin{aligned}\mathbf{k}_1 &= \Delta t\,\mathbf{G}(\bar{\mathbf{f}}_j) \\ \mathbf{k}_2 &= \Delta t\,\mathbf{G}(\bar{\mathbf{f}}_j + (1/2)\,\mathbf{k}_1) \\ \bar{\mathbf{f}}_{j+1} &= \bar{\mathbf{f}}_j + \mathbf{k}_2\end{aligned} \tag{B.19}$$

where the intermediates $\mathbf{k}_1$ and $\mathbf{k}_2$ are conventional. The set $\{\bar{\mathbf{f}}_j\}_{j=0}^N$ approximates the projection of the true solution $\{\mathbf{f}_j\}_{j=0}^N$, and they do a better job at approximation as $\Delta t \to 0$.

Problem B.10 Given explicit temporal dependence in $\mathbf{G}$, so that we are solving $\frac{d\mathbf{f}}{dt} = \mathbf{G}(t, \mathbf{f})$, update (B.19) to include the time-dependence.

[3] Here we have used a sort of "inverse" Taylor expansion to rewrite a messy term as a shifted evaluation of a function.

Algorithm B.3 RK4($\mathbf{G}, \mathbf{f}_0, \Delta t, N$)

$\bar{\mathbf{f}}_0 \leftarrow \mathbf{f}_0$
for $j = 0 \to N-1$ **do**
  $\mathbf{k}_1 \leftarrow \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j)$
  $\mathbf{k}_2 \leftarrow \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + (1/2)\, \mathbf{k}_1)$
  $\mathbf{k}_3 \leftarrow \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + (1/2)\, \mathbf{k}_2)$
  $\mathbf{k}_4 \leftarrow \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + \mathbf{k}_3)$
  $\bar{\mathbf{f}}_{j+1} \leftarrow \bar{\mathbf{f}}_j + \frac{1}{3}\left(\frac{1}{2}\,\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \frac{1}{2}\,\mathbf{k}_4\right)$
end for
return $\{\bar{\mathbf{f}}_k\}_{k=0}^N$

B.2.2 Fourth-Order Runge–Kutta

One can continue the process of improving the accuracy of the Taylor series expansion of $\mathbf{f}_{j+1}$, rewriting the expansion using the ODE and expressing derivatives of $\mathbf{G}$ in terms of shifted evaluations as above. An end result is "fourth-order Runge–Kutta," and it is the standard for ODE solving. It is defined by the update, analogous to (B.19),

$$\begin{aligned}
\mathbf{k}_1 &= \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j) \\
\mathbf{k}_2 &= \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + (1/2)\, \mathbf{k}_1) \\
\mathbf{k}_3 &= \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + (1/2)\, \mathbf{k}_2) \\
\mathbf{k}_4 &= \Delta t\, \mathbf{G}(\bar{\mathbf{f}}_j + \mathbf{k}_3) \\
\bar{\mathbf{f}}_{j+1} &= \bar{\mathbf{f}}_j + \frac{1}{3}\left(\frac{1}{2}\,\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \frac{1}{2}\,\mathbf{k}_4\right)
\end{aligned} \tag{B.20}$$

where we start the iteration off with $\bar{\mathbf{f}}_0 = \mathbf{f}_0$ as usual. This method will work for most initial value ODE problems and is used for a variety of motion calculations. The pseudo-code is shown in Algorithm B.3.

B.2.3 Dimensionless Equations

When working on problems numerically, it is useful to make the equations of interest dimensionless.[4] Instead of solving an equation of motion like (B.6) for a bunch of different values of q and m, for example, we notice that only q/m will matter, and we can go further to make dimensionless ratios. This is both structurally revealing (for the equation) and also safe for numerical work, where you want to keep numbers close to 1 (avoiding both zero and infinity, which are difficult to represent on a computer).

To demonstrate the process, we'll focus on (B.6). Let $y = \ell_0\, Y$ for ℓ_0 with dimension of length and Y dimensionless, and similarly, take $z = \ell_0\, Z$, $t = t_0\, T$ for t_0 with dimension of time, T dimensionless. Then the equations of motion read

[4] This has utility outside of numerical work, too.

$$\begin{aligned}\frac{d^2Y}{dT^2} &= \frac{\lambda\, q\, t_0^2}{2\,\pi\,\epsilon_0\, m\,\ell_0^2}\,\frac{1}{Y} - \frac{q\,\mu_0\,\lambda\, v\, t_0}{2\,\pi\,\ell_0\, m}\,\frac{dZ}{dT}\,\frac{1}{Y}\\ \frac{d^2Z}{dT^2} &= \frac{q\,\mu_0\,\lambda\, v\, t_0}{2\,\pi\,\ell_0\, m}\,\frac{dY}{dT}\,\frac{1}{Y}.\end{aligned} \tag{B.21}$$

Now we can choose to set t_0 such that the dimensionless combination

$$\frac{q\,\mu_0\,\lambda\, v\, t_0}{2\,\pi\,\ell_0\, m} = 1 \longrightarrow t_0 = \frac{2\,\pi\,\ell_0\, m}{q\,\mu_0\,\lambda\, v}. \tag{B.22}$$

Define the other dimensionless constant, $\frac{\lambda\, q\, t_0^2}{2\,\pi\,\epsilon_0\, m\,\ell_0^2} \equiv \alpha$, and we now have the simplified dimensionless pair of equations:

$$\begin{aligned}\frac{d^2Y}{dT^2} &= \frac{1}{Y}\left(\alpha - \frac{dZ}{dT}\right)\\ \frac{d^2Z}{dT^2} &= \frac{dY}{dT}\,\frac{1}{Y}.\end{aligned} \tag{B.23}$$

To set ℓ_0, let's introduce some initial conditions – suppose the particle starts a distance $y(0) = a$ from the line of charge (take $z(0) = 0$) and has no initial velocity. To make the initial condition dimensionless, we take $\ell_0 = a$ (that was unconstrained by the above), then $Y(0) = 1$ is the starting condition. We still have our parameter α in (B.23), but this is the *only* parameter left in the problem.

In order to use the Runge–Kutta method, we need to identify the vectors **f** and **G** from (B.23). In vector form, the equations will look like (B.7) but with the new dimensionless variables. Let $f_1 = Y$, $f_2 = Z$, $f_3 = \frac{dY}{dT}$, and $f_4 = \frac{dZ}{dT}$, then the content of (B.23) can be written as

$$\frac{d}{dT}\begin{pmatrix} f_1\\ f_2\\ f_3\\ f_4\end{pmatrix} = \begin{pmatrix} f_3\\ f_4\\ \frac{1}{f_1}\,(\alpha - f_4)\\ \frac{1}{f_1}\, f_3\end{pmatrix}. \tag{B.24}$$

The initial values are given by the vector

$$\mathbf{f}(0) \doteq \begin{pmatrix} 1\\ 0\\ 0\\ 0\end{pmatrix}. \tag{B.25}$$

Using Algorithm B.3, with $\alpha = 1$, the motion shown in Figure B.2 is produced.

Problem B.11 Express the speed of light, c, in terms of the dimensionless variables defined in this section.

Problem B.12 Run RK4 to solve (B.24) with initial values (B.25) to generate Figure B.2. Does the particle travel faster than the speed of light?

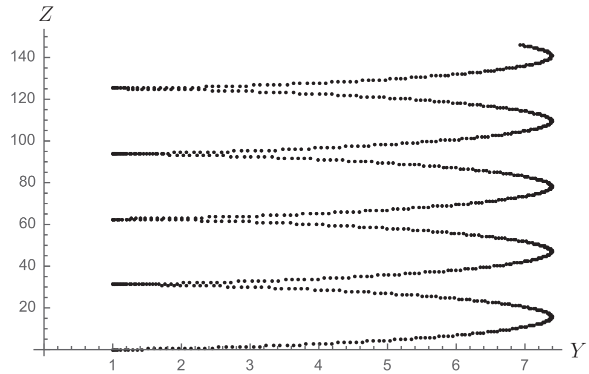

Fig. B.2 The trajectory of a charged particle moving under the influence of the electric and magnetic fields of a moving, infinite, charged wire.

B.3 Solving PDEs

So far, we have focused on numerical methods that are relevant when we already know the fields of interest, and we want to find out how particles respond to those. Next, we'll turn to problems where we are given a source distribution and want to find the associated fields. There are two main approaches. First, we have integral solutions to partial differential equations (PDEs) as in (2.1), and those come from Green's functions. In most cases, we cannot evaluate the integrals in terms of simple functions, so a numerical solution approximating the integration is appropriate. The Green's function approach comes with specific boundary values, but in situations involving different types of boundary conditions (where finding the Green's function itself is difficult), a direct discretization of the PDE itself can be used to generate a matrix equation that can be solved; this is the second technique we will consider.

B.3.1 Integration

Start with the simple one-dimensional integration problem: Given a function $F(x)$, find the value of

$$I = \int_a^b F(x)\, dx. \tag{B.26}$$

We'll introduce a grid in x as we did in the previous section for t. Let $x_j = a + j\,\Delta x$ with $x_n = b$ for given n (the number of points in the grid) so that j goes from zero to n and $\Delta x = (b-a)/n$, and take $F_j \equiv F(x_j)$. The first approximation to I comes from replacing the integral sign with a summation and dx with Δx:

$$I \approx \sum_{j=0}^{n-1} F_j\, \Delta x \equiv I_b. \tag{B.27}$$

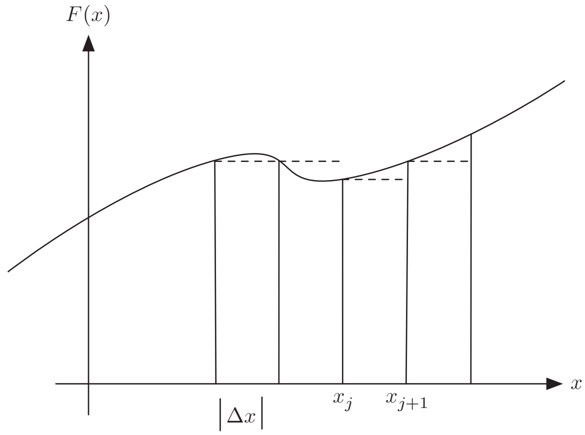

Fig. B.3 A piece of the integral approximation we make in using (B.27). The dashed lines represent the piecewise function that (B.27) integrates exactly.

You can think of this approximation as follows: assuming the function $F(x)$ is constant over the interval from x_j to x_{j+1} with value $F_j \equiv F(x_j)$, the exact integral over the interval is

$$\int_{x_j}^{x_{j+1}} F_j\, dx = F_j \left(x_{j+1} - x_j\right) = F_j\, \Delta x, \tag{B.28}$$

so we are integrating a constant function exactly over each interval, then adding those up. The approximation comes from the fact that the function $F(x)$ does not take on the constant value F_j over the interval. The idea is shown geometrically in Figure B.3.

We can refine the integral approximation (without changing Δx) by using better approximations to $F(x)$ on the interior of each interval. For example, suppose we make a linear approximation to $F(x)$ between x_j and x_{j+1} that matches $F(x)$ at x_j and x_{j+1}, then integrate that exactly. Take

$$F(x) \approx F_j + (x - x_j)\, \frac{F_{j+1} - F_j}{x_{j+1} - x_j} \tag{B.29}$$

for $x = x_j$ to x_{j+1}, then the exact integral of this linear function is

$$\begin{aligned} \int_{x_j}^{x_{j+1}} \left(F_j + (x - x_j)\, \frac{F_{j+1} - F_j}{x_{j+1} - x_j}\right) dx &= \frac{1}{2}\left(F_j + F_{j+1}\right)\left(x_{j+1} - x_j\right) \\ &= \frac{1}{2}\left(F_j + F_{j+1}\right)\, \Delta x, \end{aligned} \tag{B.30}$$

and if we add these up over the entire domain, we get a new approximation to I in (B.26):

$$I \approx \sum_{j=0}^{n-1} \frac{1}{2}\left(F_j + F_{j+1}\right)\, \Delta x \equiv I_t. \tag{B.31}$$

This approximation for I is shown pictorially in Figure B.4 and is known as the "trapezoidal approximation." Notice, in Figure B.4, that the dashed lines indicating the piecewise linear function that we use to approximate $F(x)$ over each interval are a much better

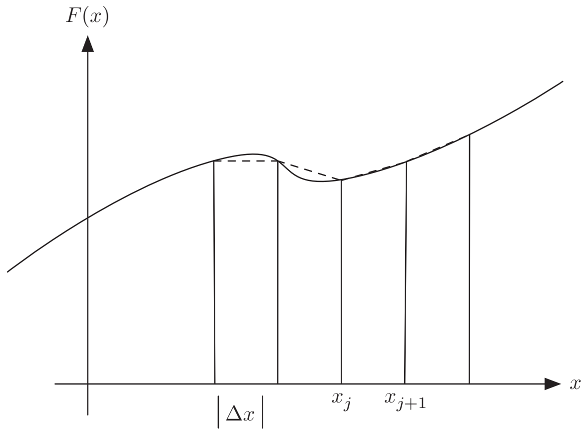

Fig. B.4 An exact integration of a piecewise linear approximation to $F(x)$, giving (B.31).

approximation to the function (and lie "on top of it" visually in places) than the piecewise constant approximation shown in Figure B.3.

Problem B.13 Write (B.31) in terms of (B.27); i.e., relate I_t to I_b. You should need to evaluate two extra points, so that $I_t = I_b + X + Y$ where X and Y involve single evaluations of $F(x)$.

Continuing, we can replace our piecewise linear approximation with a quadratic one. This time, we need to include more points in our approximation to $F(x)$. To do linear interpolation, we only needed the values of $F(x)$ at x_j and x_{j+1}, but now to set all the coefficients in a quadratic interpolation, we must use the values of $F(x)$ at x_j, x_{j+1}, and x_{j+2}. Over the interval x_j to x_{j+2}, then, we'll approximate $F(x)$ by

$$\begin{aligned} F(x) \approx \frac{1}{2\,\Delta x^2} \Big[&(x - x_{j+1})\,(x - x_{j+2})\,F_j - 2\,(x - x_j)\,(x - x_{j+2})\,F_{j+1} \\ &+ (x - x_j)\,(x - x_{j+1})\,F_{j+2} \Big], \end{aligned} \tag{B.32}$$

where the quadratic approximation matches $F(x)$ at the grid points x_j, x_{j+1}, and x_{j+2}. The exact integral of the quadratic approximation function, over the interval of interest, is

$$\int_{x_j}^{x_{j+2}} F(x)\,dx = \frac{1}{3}\,\Delta x \left(F_j + 4\,F_{j+1} + F_{j+2}\right), \tag{B.33}$$

giving us the "Simpson's rule" approximation to I,

$$I \approx \sum_{j=0,2,4,\ldots}^{n-2} \frac{1}{3}\,\Delta x \left(F_j + 4\,F_{j+1} + F_{j+2}\right) \equiv I_s. \tag{B.34}$$

The implementation of the Simpson's rule approximation is shown in Algorithm B.4. There, you provide the function F, the limits of integration a to b, and the number of grid points n (that *must* be a factor of two, as is clear from the sum in (B.34)).

Algorithm B.4 Simpson(F, a, b, n)

Check that n is a factor of 2, exit if not.
$\Delta x \leftarrow (b-a)/n$
$I_s \leftarrow 0$
for $j = 0 \rightarrow n-2$ in steps of 2 **do**
  $I_s \leftarrow I_s + \frac{1}{3}\,\Delta x\,\left(F_j + 4\,F_{j+1} + F_{j+2}\right)$
end for
return I_s

Algorithm B.5 Simpson2D$(F, a_x, b_x, n_x, a_y, b_y, n_y)$

$I_y(x) =$Simpson$(F(x,y), a_y, b_y, n_y)$
return Simpson$(I_y(x), a_x, b_x, n_x)$

B.3.2 Higher Dimensional Integration

Integrals like (2.1) are three-dimensional, but we can use our one-dimensional integration routine to solve them. Think of the two-dimensional integral

$$I = \int_{a_x}^{b_x} \left(\int_{a_y}^{b_y} F(x,y)\,dy \right) dx. \tag{B.35}$$

For each value of x, we have a one-dimensional integral in y to carry out,

$$I_y(x) = \int_{a_y}^{b_y} F(x,y)\,dy. \tag{B.36}$$

The approach, for Cartesian coordinates, is shown in Algorithm B.5. Inside the Simpson2D function, we must make the one-dimensional function $I_y(x)$ that carries out the y-integral for a given value of x (so that in its return statement, $F(x,y)$ has fixed x; it is really a function only of y).

Problem B.14 Integrate (B.32) from $x = x_j \rightarrow x_{j+2}$ and show that you get the right-hand side of (B.33).

Problem B.15 Write out the pseudocode analogous to Algorithm B.5 for three-dimensional integration. Assume Cartesian coordinates and add the z component running from a_z to b_z with n_z steps.

Problem B.16 Implement Simpson's method (Algorithm B.4) and use it to approximate the integral

$$I = \int_0^2 \frac{\sin(x)}{x}\,dx. \tag{B.37}$$

The correct value, up to 10 digits, is 1.6054129768. How many grid points do you need to match these 10 digits?

Problem B.17 For a circle of radius R carrying constant charge-per-unit-length λ, write out the integral (in cylindrical coordinates) solution for the electrostatic potential at $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$. Write a function that uses Simpson's method to approximate this one-dimensional integral at a given point for $\lambda = 1$ C/m and $R = 1$ m.

Example

Take a disk of radius R with constant charge-per-unit-area σ smeared over it. The electric field at $\mathbf{r}$ is given by the formal integral

$$\mathbf{E}(x, y, z) = \frac{1}{4\,\pi\,\epsilon_0} \int \frac{\sigma\,\delta(z')\,\boldsymbol{\mathscr{r}}}{\mathscr{r}^3}\,d\tau', \tag{B.38}$$

for $\boldsymbol{\mathscr{r}} \equiv \mathbf{r} - \mathbf{r}'$ as usual. We can write the integrand explicitly, using cylindrical coordinates for the source:

$$\mathbf{E}(x, y, z) = \frac{\sigma}{4\,\pi\,\epsilon_0} \int_0^R \int_0^{2\pi} \frac{\left(x - r'\,\cos\phi'\right)\,\hat{\mathbf{x}} + \left(y - r'\,\sin\phi'\right)\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}}{\left((x - r'\,\cos\phi')^2 + (y - r'\,\sin\phi')^2 + z^2\right)^{3/2}}\,r'\,d\phi'\,dr'. \tag{B.39}$$

The solution, then, is a set of three two-dimensional integrals. A plot of the field lines above the disk is shown in Figure B.5. In order to make this plot, we have to evaluate $\mathbf{E}$ at a number of different points (different values for x, y, and z).

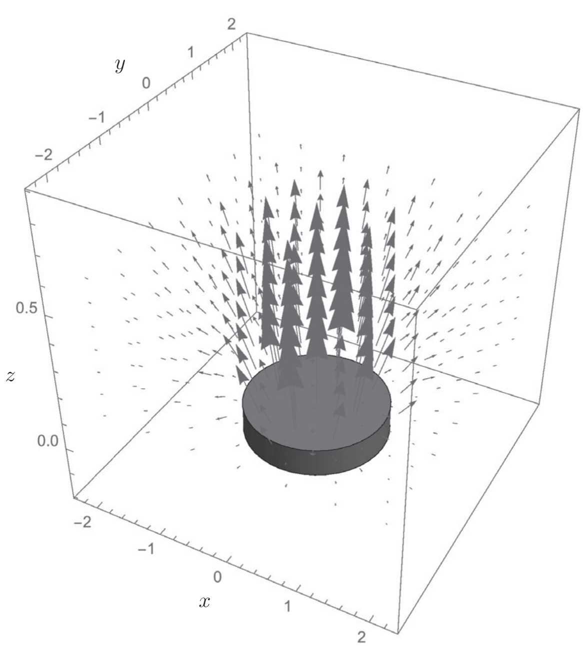

Fig. B.5 Some representative electric field vectors for a finite disk with constant charge-per-unit-area σ calculated from (B.39) using Algorithm B.5.

Problem B.18 Try making the plot analogous to Figure B.5 using $\sigma(r) = \sigma_0 \, r/R$ where $\sigma_0/(4\pi\epsilon_0) = 1/\text{N m}^2\,\text{C}$ numerically, to make the expressions simpler. Take $R = 1$ m, and sample four points between -2 and 2 in each of the x- and y-directions, with four points from 0.1 to 1 in the z-direction to make the plot.

B.3.3 Finite Difference

The integral solutions above are appropriate as long as the Green's function you are using is correctly adapted to your boundary values. In many cases, boundary values are given for which there is no obvious Green's function (or for which the Green's function cannot be represented analytically). There are a variety of techniques for handling general boundary value problems, and finite difference is one of them. The idea is to replace the derivatives appearing in Poisson's equation with approximations that come from a Taylor series expansion, which will turn the PDE problem into a linear algebra one.

One-Dimensional Form

As an example of the procedure, take the one-dimensional case. We want $F(x)$ solving

$$\frac{d^2F(x)}{dx^2} = -s(x), \qquad F(0) = F_0, \qquad F(X) = F_X, \tag{B.40}$$

for given source $s(x)$ and boundary values provided at $x = 0$ and $x = X$. Introduce a grid in x: $x_j = j\,\Delta x$, for $j = 0 \to n+1$ and $x_{n+1} = X$, with $F_j \equiv F(x_j)$ as usual. The second derivative of $F(x)$ evaluated at x_j can be approximated by noting that

$$F_{j\pm1} = F_j \pm \Delta x \left.\frac{dF}{dx}\right|_{x=x_j} + \frac{1}{2}\Delta x^2 \left.\frac{d^2F}{dx^2}\right|_{x=x_j} \pm \frac{1}{6}\Delta x^3 \left.\frac{d^3F}{dx^3}\right|_{x=x_j} + \cdots \tag{B.41}$$

and then we can isolate the second derivative using a linear combination

$$\left.\frac{d^2F}{dx^2}\right|_{x=x_j} \approx \frac{F_{j+1} - 2F_j + F_{j-1}}{\Delta x^2} \tag{B.42}$$

where the error we make in using this approximation is of order[5] Δx^2.

Putting the approximation (B.42) into (B.40), and writing $s_j \equiv s(x_j)$, we get a set of algebraic equations governing the approximations to $F(x)$ on the grid (we'll use bars to indicate that this is now a set of approximate numerical values, obtained by replacing $\approx$ with $=$ in (B.42)):

$$\frac{\bar{F}_{j+1} - 2\bar{F}_j + \bar{F}_{j-1}}{\Delta x^2} = -s_j, \tag{B.43}$$

for $j = 1 \to n$, and we need to be careful with the cases $j = 1$ and n, since those will involve the boundary values F_0 and F_X. The full set of equations, including those special cases, is

[5] Meaning, roughly, that the error is bounded by some constant times Δx^2. The constant that sits out front depends on the fourth derivative of $F(x)$ evaluated at x_j, in the present case.

$$
\begin{aligned}
\frac{\bar{F}_2 - 2\bar{F}_1}{\Delta x^2} &= -s_1 - \frac{F_0}{\Delta x^2} \\
\frac{\bar{F}_{j+1} - 2\bar{F}_j + \bar{F}_{j-1}}{\Delta x^2} &= -s_j \quad \text{for } j = 2 \to n-1 \\
\frac{-2\bar{F}_n + \bar{F}_{n-1}}{\Delta x^2} &= -s_n - \frac{F_X}{\Delta x^2},
\end{aligned}
\tag{B.44}
$$

where we have moved the F_0 and F_X terms over to the right-hand side since those are known values. We want to solve this set of equations for $\{\bar{F}_j\}_{j=1}^n$.

We can write (B.44) in matrix-vector form. Define the vector $\bar{\mathbf{F}} \in \mathbb{R}^n$ with entries that are the unknown values:

$$
\bar{\mathbf{F}} \doteq \begin{pmatrix} \bar{F}_1 \\ \bar{F}_2 \\ \vdots \\ \bar{F}_n \end{pmatrix},
\tag{B.45}
$$

and the tridiagonal matrix that acts on this to produce the left-hand side of (B.44) is

$$
\mathbb{D} \doteq \frac{1}{\Delta x^2} \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots \\ \cdots & 0 & 0 & 1 & -2 \end{pmatrix}.
\tag{B.46}
$$

Finally, define the slightly modified source "vector"

$$
\mathbf{s} \doteq \begin{pmatrix} -s_1 - \frac{F_0}{\Delta x^2} \\ -s_2 \\ \vdots \\ -s_{n-1} \\ -s_n - \frac{F_X}{\Delta x^2} \end{pmatrix},
\tag{B.47}
$$

and we can write (B.44) as

$$
\mathbb{D}\,\bar{\mathbf{F}} = \mathbf{s},
\tag{B.48}
$$

which has solution obtained by inverting $\mathbb{D}$. Formally, $\bar{\mathbf{F}} = \mathbb{D}^{-1}\,\mathbf{s}$ is what we want. How we obtain that matrix inverse depends on the particular problem (there are direct matrix inversion routines implemented in almost any programming language, and a suite of approximate inverses that can also be used).

Problem B.19 Using the finite difference approach from above, solve the damped, driven harmonic oscillator problem

$$
\frac{d^2F(x)}{dx^2} + \alpha\,\frac{dF(x)}{dx} + \beta\,F(x) = -\sin(2\,\pi\,x),
\tag{B.49}
$$

with $F(0) = 0$ and $F(1) = 0$. Use a grid with $N = 1{,}000$, $\Delta x = 1/(N+1)$, and take $\alpha = 1$, $\beta = 2$. Make a plot of your solution.

Two Dimensions

The two-dimensional version of the problem is identical in spirit to the above, but the details are a little more involved: there are more boundary values, and it is harder to set up the relevant matrices and vectors. We'll stay with Cartesian coordinates and work with the Poisson problem on the finite domain bounded by a unit square (you can always rescale your problem so as to make the boundary have unit length):

$$\begin{aligned}
\nabla^2 F(x,y) &= -s(x,y) \\
F(0,y) &= L(y) \\
F(1,y) &= R(y) \\
F(x,0) &= B(x) \\
F(x,1) &= T(x)
\end{aligned} \tag{B.50}$$

where the source and all the boundary functions must be provided.

Our grid is now two-dimensional, with $x_j = j\,\Delta x$ and $y_k = k\,\Delta y$, and we'll take n_x and n_y grid points in the x- and y-directions ($j = 0$ and $n_x + 1$ are boundaries in x, $k = 0$, $n_y + 1$ are boundaries in y). Define $F_{jk} \equiv F(x_j, y_k)$, the value of $F(x,y)$ at x_j, y_k, then the Laplacian, evaluated at x_j and y_k, can be approximated by

$$\begin{aligned}
\nabla^2 F|_{x_j,y_k} &= \left.\frac{\partial^2 F}{\partial x^2}\right|_{x_j,y_k} + \left.\frac{\partial^2 F}{\partial y^2}\right|_{x_j,y_k} \\
&\approx \frac{F_{(j+1)k} - 2\,F_{jk} + F_{(j-1)k}}{\Delta x^2} + \frac{F_{j(k+1)} - 2\,F_{jk} + F_{j(k-1)}}{\Delta y^2}
\end{aligned} \tag{B.51}$$

and the *method* comes from using this finite difference approximation exactly in (B.50):

$$\frac{\bar{F}_{(j+1)k} - 2\,\bar{F}_{jk} + \bar{F}_{(j-1)k}}{\Delta x^2} + \frac{\bar{F}_{j(k+1)} - 2\,\bar{F}_{jk} + \bar{F}_{j(k-1)}}{\Delta y^2} = -s_{jk}, \tag{B.52}$$

ignoring boundaries for now.

How do we formulate (B.52) as a matrix-vector equation? The unknown values are now $\{\bar{F}_{jk}\}_{j=1,k=1}^{j=n_x,k=n_y}$, a "matrix" of unknown values on the grid. We need a way to map points on the grid to vector elements. There are $n_x\,n_y$ total points in the discretized grid, so we'll make a vector that has this length. Define the function

$$g(j,k) = (k-1)\,n_x + j. \tag{B.53}$$

This function assigns a vector location to the grid location (j,k). Then we can build a vector of unknown values $\bar{\mathbf{F}} \in \mathbb{R}^{n_x\,n_y}$ that has entries indexed by $g(j,k)$. For example, the value of $\bar{F}_{11}$ (at the $j = 1$, $k = 1$ grid location) becomes entry $g(1,1) = 1$ in $\bar{\mathbf{F}}$. The assignment scheme taking grid locations to vector entries using $g(j,k)$ is shown in Figure B.6 where $n_x = 3$ and $n_y = 4$.

In terms of the elements of the vector $\bar{\mathbf{F}}$, then, we could write (B.52) as

$$\frac{\bar{F}_{g(j+1,k)} - 2\,\bar{F}_{g(j,k)} + \bar{F}_{g(j-1,k)}}{\Delta x^2} + \frac{\bar{F}_{g(j,k+1)} - 2\,\bar{F}_{g(j,k)} + \bar{F}_{g(j,k-1)}}{\Delta y^2} = -s_{g(j,k)} \tag{B.54}$$

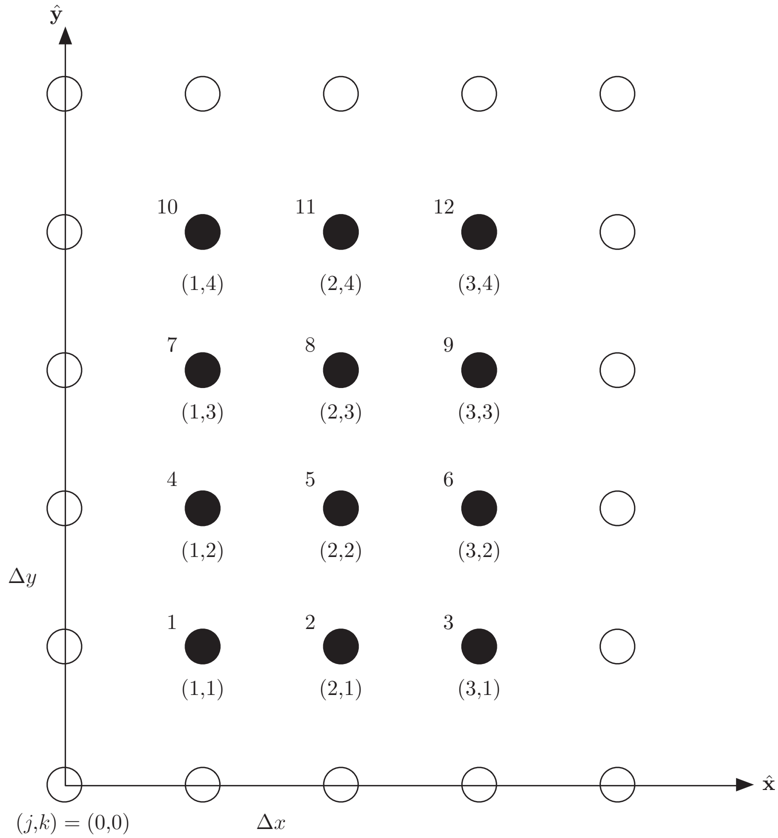

Fig. B.6 An example of a grid with $n_x = 3$, $n_y = 4$. Open circles represent boundary points. Underneath each grid location is its (j, k) pair, and above and to the left is the corresponding element of the vector embedding set up by $g(j, k)$ from (B.53).

where the elements of the vector $\mathbf{s} \in \mathbb{R}^{n_x n_y}$ appear on the right, ordered using the same embedding.[6] In this form, it is clear how to set the values of the matrix $\mathbb{D} \in \mathbb{R}^{n_x n_y \times n_x n_y}$ representing the discretized Laplacian: in the $g(j, k)$ row, we'll set values in the columns $g(j, k)$, $g(j \pm 1, k)$, and $g(j, k \pm 1)$. The pseudocode that generates $\mathbb{D}$ is shown in Algorithm B.6.

Making the right-hand side, which contains both the source and the boundary values, can be done with some care. Think of the set of points with constant k (a "row" in our grid from Figure B.6). At $j = 1$, there will be a term in (B.54) that is a boundary term, $\bar{F}_{g(j-1,k)}$, whose value is given by $L(y)$ from (B.50), evaluated at y_k. Specifically, $\bar{F}_{g(0,k)} = L(y_k)$. We'll move this known value over to the right-hand side, so that for $j = 1$ (B.54) reads

$$\frac{\bar{F}_{g(2,k)} - 2\,\bar{F}_{g(1,k)}}{\Delta x^2} + \frac{\bar{F}_{g(1,k+1)} - 2\,\bar{F}_{g(1,k)} + \bar{F}_{g(1,k-1)}}{\Delta y^2} = -s_{g(1,k)} - \frac{L(y_k)}{\Delta x^2}. \tag{B.55}$$

[6] We'll redefine this vector slightly to include the known boundary values in a moment.

Algorithm B.6 MakeDmat($\Delta x, n_x, \Delta y, n_y$)

```
for j = 1 → n_x do
  for k = 1 → n_y do
    if j ≠ n_x then
      D_{g(j,k),g(j+1,k)} ← 1/Δx^2
    end if
    if j ≠ 1 then
      D_{g(j,k),g(j−1,k)} ← 1/Δx^2
    end if
    D_{g(j,k),g(j,k)} ← −2/Δx^2 − 2/Δy^2
    if k ≠ n_y then
      D_{g(j,k),g(j,k+1)} ← 1/Δy^2
    end if
    if k ≠ 1 then
      D_{g(j,k),g(j,k−1)} ← 1/Δy^2
    end if
  end for
end for
return 𝔻
```

Similarly, at $j = n_x$, we have $\bar{F}_{g(n_x+1,k)} = R(y_k)$, and (B.54) is

$$\frac{-2\,\bar{F}_{g(n_x,k)} + \bar{F}_{g(n_x-1,k)}}{\Delta x^2} + \frac{\bar{F}_{g(n_x,k+1)} - 2\,\bar{F}_{g(n_x,k)} + \bar{F}_{g(n_x,k-1)}}{\Delta y^2} = -s_{g(n_x,k)} - \frac{R(y_k)}{\Delta x^2}. \tag{B.56}$$

The same story holds for the top and bottom boundaries, obtained at constant j when $k = 1$ (bottom) and $k = n_y$ (top). We can combine all of these elements of the right-hand side of our problem into one function that takes the source function, all four boundary functions, and constructs the right-hand side of (B.54) for all values of j and k. That function is shown in Algorithm B.7; it returns the vector **s** that consists of both source and boundary information.

Using the pair of functions in Algorithm B.6 and Algorithm B.7, we can again generate a matrix-vector equation representing the discretized form of (B.50), $\mathbb{D}\,\bar{\mathbf{F}} = \mathbf{s}$, and solve by matrix inversion to get $\bar{\mathbf{F}}$. The solution takes some unpacking to get in a form appropriate for displaying (remember that the solution is a vector of grid values – we need to "invert" $g(j,k)$ to get back the grid location given a vector index, but that's not too bad).

As an example, we'll work in the E&M setting, where $\nabla^2 V = -\rho/\epsilon_0$ is our special case of $\nabla^2 F = -s$. Take a unit square and put sinusoidal boundary conditions on two sides: $T = \sin(2\,\pi\,x)$ and $L = \sin(4\,\pi\,y)$ with the other sides grounded, $B = R = 0$. Inside the box, we'll set $\rho = 0$, no charges. The numerical solution for the potential, obtained by using finite differences, is shown in Figure B.7.

We can also ground all four sides and put equal and opposite lines of charge on the interior of the box. This time $B = T = L = R = 0$ and $\rho(x, y)$ will be such that there is a

Algorithm B.7 PoissonRHS($\Delta x, n_x, \Delta y, n_y, s, L, R, T, B$)

```
for j = 1 → n_x do
  for k = 1 → n_y do
    s_g(j,k) ← −s(x_j, y_k)
  end for
end for
for k = 1 → n_y do
  s_g(1,k) ← s_g(1,k) − L(y_k)/Δx^2
  s_g(n_x,k) ← s_g(n_x,k) − R(y_k)/Δx^2
end for
for j = 1 → n_x do
  s_g(j,1) ← s_g(j,1) − B(x_j)/Δy^2
  s_g(j,n_y) ← s_g(j,n_y) − T(x_j)/Δy^2
end for
return s
```

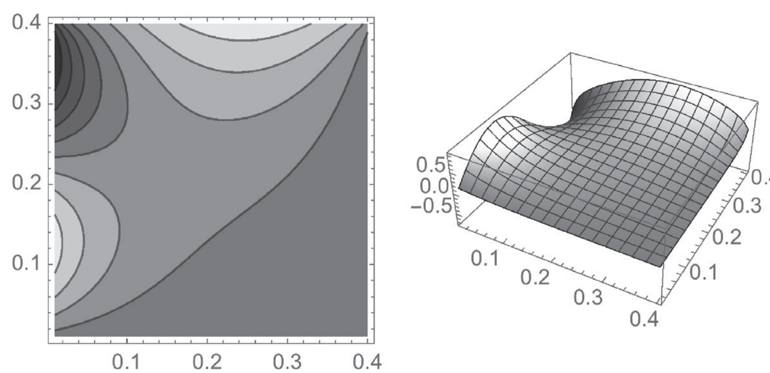

Fig. B.7 The potential associated with an empty (of charge) box that is grounded on two sides and has oscillatory boundaries on the left and top. The contour plot of the potential is on the left and the three-dimensional height plot is on the right.

line of positive source and a parallel line of negative source (a cross section of a parallel plate capacitor). The solution for the potential in that case is shown in Figure B.8.

Finally, we can combine the two problems. Take the boundary functions from the example in Figure B.7 with the source from Figure B.8; the potential is shown in Figure B.9. This would be a difficult case to handle by direct integration of a Green's function.

Problem B.20

a. On a grid with $n_x = n_y = 49$ for a unit square (so that $\Delta x = 1/(n_x + 1)$, $\Delta y = 1/(n_y + 1)$), make a source function that is zero everywhere except at x_{25}, y_{25}, where the source should have value $s = 1$. Ground all four walls and solve for the potential. It should look like a monopole that is "squared off" near the boundary.

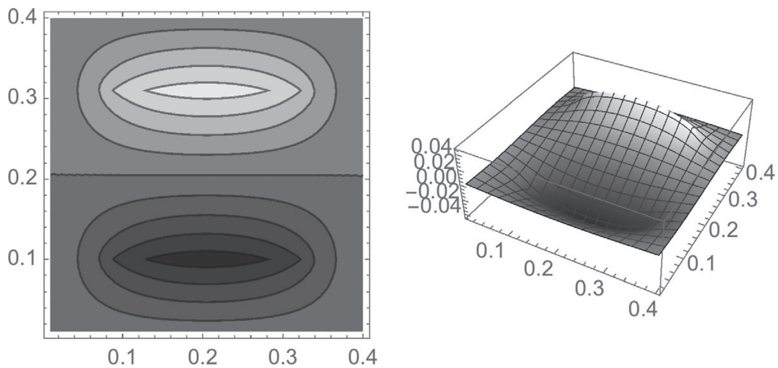

Fig. B.8 The potential for a “parallel-line" capacitor, with a line of positive charge on top, negative charge on bottom. The contour plot is on the left, with heights on the right.

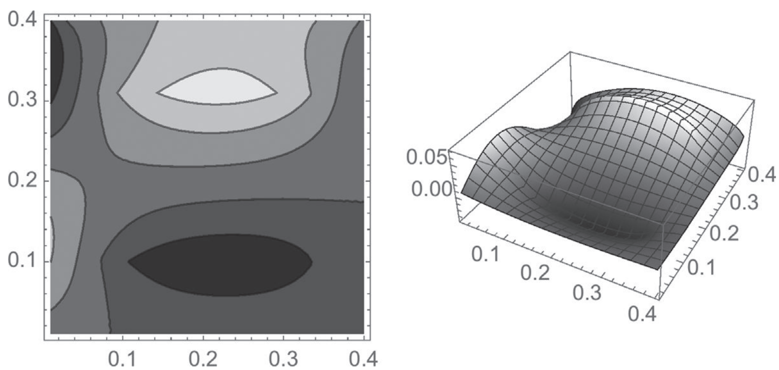

Fig. B.9 The potential for a parallel plate capacitor with oscillatory boundary conditions.

b. We can also get dipole behavior. Take $n_x = 49$ and $n_y = 49$ (again for a unit square) and let $s = 0$ except at x_{25}, y_{25} where it is -1 and at x_{25}, y_{26} where it is 1. Using grounded boundaries, solve for the potential and plot it.

Problem B.21 Take $n_x = n_y = 50$ for a unit square and set the source $s = x$; ground the walls and find the potential.

Appendix C E&M from an Action

In this appendix, we start from the action for E&M and use it to develop the field equations, gauge structure, integral form of the solution, etc. In short, all of the physics of E&M come from this action, and we explicitly demonstrate this as an example of the power of the action as a shorthand device.

C.1 The E&M Action

We start with a scalar field Lagrangian defined by the action

$$S[A_\nu] = \int \left[\frac{1}{4\,\mu_0} \left(A_{\beta,\alpha} - A_{\alpha,\beta}\right) \left(A_{\sigma,\rho} - A_{\rho,\sigma}\right) \eta^{\beta\sigma}\, \eta^{\alpha\rho} - A_\alpha\, J^\alpha \right] \sqrt{-\eta}\, d^4x \tag{C.1}$$

for field A_ν and source J^ν (we are assuming this is a four-vector). The field equations for the scalar $\hat{\mathcal{L}}$ (the integrand with $\sqrt{-\eta} = 1$) are

$$-\partial_\mu \left(\frac{\partial \hat{\mathcal{L}}}{\partial A_{\nu,\mu}} \right) + \frac{\partial \hat{\mathcal{L}}}{\partial A_\nu} = 0, \tag{C.2}$$

which read, for this Lagrangian,

$$-\partial_\mu \left(\frac{1}{\mu_0} \left(\partial^\mu A^\nu - \partial^\nu A^\mu \right) \right) - J^\nu = 0, \tag{C.3}$$

giving sourced field equation

$$\partial^\mu\, \partial_\mu A^\nu - \partial^\nu\, \partial^\mu A_\mu = -\mu_0\, J^\nu. \tag{C.4}$$

C.1.1 Gauge Freedom

The starting (vacuum, $J^\nu = 0$) Lagrangian is insensitive to the introduction of a pure scalar field via $A_\nu = \phi_{,\nu}$. To see this, and figure out what it implies about the source function J^μ, let $F_{\alpha\beta} \equiv A_{\beta,\alpha} - A_{\alpha,\beta}$, so that the action can be written succinctly:

$$S[A_\nu] = \int \left[\frac{1}{4\,\mu_0} F_{\alpha\beta}\, F^{\alpha\beta} - A_\alpha\, J^\alpha \right] \sqrt{-\eta}\, d^4x. \tag{C.5}$$

It is clear that if we take $A_\nu \rightarrow A_\nu + \delta A_\nu$ with $\delta A_\nu \equiv \phi_{,\nu}$, the combination $F_{\alpha\beta}$ does not change:

$$F_{\alpha\beta}[A_\nu + \delta A_\nu] = A_{\beta,\alpha} - A_{\alpha,\beta} + \phi_{,\beta\alpha} - \phi_{,\alpha\beta} = A_{\beta,\alpha} - A_{\alpha,\beta} = F_{\alpha\beta}[A_\nu] \tag{C.6}$$

by cross-derivative equality ($\partial_\alpha \partial_\beta = \partial_\beta \partial_\alpha$). The choice $\delta A_\mu = \phi_{,\mu}$ is just one of the infinite family of variations with respect to which the action must be a minimum. Evaluating (in Cartesian coordinates) the action at this perturbed field gives

$$S[A_\nu + \phi_{,\nu}] = \underbrace{\int \left[\frac{1}{4\,\mu_0} F_{\alpha\beta} F^{\alpha\beta} - A_\alpha J^\alpha \right] d^4x}_{=S[A_\nu]} - \underbrace{\int \phi_{,\mu} J^\mu \, d^4x}_{\equiv \delta S}. \tag{C.7}$$

If we demand that $S[A_\nu + \phi_{,\nu}] = S[A_\nu]$ for all ϕ, then we must have $\delta S = 0$:

$$\delta S = \int \phi_{,\mu} J^\mu \, d^4x = -\int \phi \, \partial_\mu J^\mu \, d^4x = 0 \quad \text{for all } \phi \longrightarrow \partial_\mu J^\mu = 0. \tag{C.8}$$

We have a conservation law for our four-vector source J^μ, and going back to the field equation, written in terms of $F_{\mu\nu}$:

$$\partial_\mu F^{\mu\nu} = -\mu_0 J^\nu. \tag{C.9}$$

It is clear that the field equation itself is insensitive to the choice of ϕ. We can pick any ϕ we like to add to A_μ and the field equation is unchanged. Choose ϕ such that $\partial^\mu \left(A_\mu + \phi_{,\mu} \right) = 0$, and define our new A_μ to be $A_\mu + \phi_{,\mu}$; we have $\partial^\mu A_\mu = 0$ (and if it doesn't, introduce ϕ and solve). This is the Lorenz gauge choice.

C.2 Field Equations and Particle Interactions

In Lorenz gauge, the field equation for A_ν from (C.4) reads

$$\partial_\mu \partial^\mu A^\nu = -\mu_0 J^\nu. \tag{C.10}$$

Given a source J^ν we can, in theory, solve for A^ν, but what would we do with such a solution? Suppose we have a particle, with relativistic action

$$S = m\,c \int \sqrt{-\frac{dx^\mu}{dt} \eta_{\mu\nu} \frac{dx^\nu}{dt}} \, dt, \tag{C.11}$$

what could we insert for interaction with the field? The simplest scalar term that is itself reparametrization invariant (and hence will "go" with our temporal parametrization for the first term) is

$$S_A = \alpha \int A_\mu \frac{dx^\mu}{dt} \, dt \tag{C.12}$$

where α sets the strength of the interaction. Putting this together with S gives

$$S = m\,c \int \sqrt{-\frac{dx^\mu}{dt} \eta_{\mu\nu} \frac{dx^\nu}{dt}} \, dt + \alpha \int A_\mu \frac{dx^\mu}{dt} \, dt. \tag{C.13}$$

When we vary the associated (particle) Lagrangian, we get, splitting up the spatial and temporal parts,

$$\frac{d}{dt}\left(\frac{m\,\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}}\right) = \alpha\left(\nabla A^0\,c + \frac{\partial \mathbf{A}}{\partial t}\right) - \alpha\left(\mathbf{v}\times(\nabla\times\mathbf{A})\right). \tag{C.14}$$

We'll call the coupling constant $\alpha \equiv -q$ (just a name – our choice!) and define the spatial vectors:

$$\begin{aligned} \mathbf{E} &= -\nabla A^0\,c - \frac{\partial \mathbf{A}}{\partial t} \\ \mathbf{B} &= \nabla\times\mathbf{A}. \end{aligned} \tag{C.15}$$

The equation of motion for the particle is now

$$\frac{d}{dt}\left(\frac{m\,\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}}\right) = q\,\mathbf{E} + q\,\mathbf{v}\times\mathbf{B}. \tag{C.16}$$

In the low-speed limit ($v \ll c$), we recover the pre-relativistic form of Newton's second law, $m\,\mathbf{a} = q\,(\mathbf{E} + \mathbf{v}\times\mathbf{B})$.

C.3 Static Solutions to the Field Equations

Returning to the field equation for A_ν: $\Box A^\nu = -\mu_0\,J^\nu$,[1] we can consider solutions for sources that do not depend on time. If the sources are time-independent, then it is reasonable to expect the fields to be time-independent – in this limited setting, we get (for the field equation(s), gauge condition, and charge conservation, respectively):

$$\begin{aligned} \nabla^2 V &= -\mu_0\,J^0 \\ \nabla^2 \mathbf{A} &= -\mu_0\,\mathbf{J} \\ \nabla\cdot\mathbf{A} &= 0 \\ \nabla\cdot\mathbf{J} &= 0 \end{aligned} \tag{C.17}$$

for $V \equiv A^0\,c$. We'll call $J^0 \equiv \rho\,c$ for (charge) density ρ, and then we can also write these in terms of our spatial $\mathbf{E}$ and $\mathbf{B}$ vectors. Since $\mathbf{E} = -\nabla V$ (for these time-independent fields) we get $\nabla\cdot\mathbf{E} = \rho/\epsilon_0$. Separately taking the curl of $\mathbf{B} = \nabla\times\mathbf{A}$,

$$\nabla\times\mathbf{B} = \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A} = \mu_0\,\mathbf{J} \tag{C.18}$$

using $\nabla\cdot\mathbf{A} = 0$. We also learn, as a geometric feature, that $\nabla\cdot\mathbf{B} = 0$ and $\nabla\times\mathbf{E} = 0$ (owing to their pure curl/gradient form).

[1] We'll use the constants μ_0 and ϵ_0 with $\mu_0\,\epsilon_0 = c^{-2}$ as in E&M, but here they are appearing as arbitrary couplings.

From this static set, we can extract all sorts of solutions – if we integrate the equation for **E** over a domain Ω, and apply the divergence theorem,

$$\oint_\Omega \mathbf{E} \cdot d\mathbf{a} = \frac{1}{\epsilon_0} \int \rho \, d\tau, \tag{C.19}$$

and this equation can be used to *find* the field **E** in cases of highly symmetric source distribution. For example, if there is a point source, we expect **E** to be spherically symmetric: $\mathbf{E} = E(r)\,\hat{\mathbf{r}}$. Choose the integration domain to be a sphere with the point source at its center: $\rho = q\,\delta^3(\mathbf{r})$ (q sets the units), then

$$E(r)\,4\,\pi\,r^2 = \frac{1}{\epsilon_0}\,q \longrightarrow \mathbf{E} = \frac{q}{4\,\pi\,\epsilon_0\,r^2}\,\hat{\mathbf{r}}. \tag{C.20}$$

From **E** we can also recover V for the point source

$$V = \frac{q}{4\,\pi\,\epsilon_0\,r} + \alpha \tag{C.21}$$

where α is just a constant of integration (it sets the value of the potential at spatial infinity – take it to be zero). What we really have, for V here, is the Green's function for the operator ∇^2 (the Green's function *is* the point source solution). That means that we can solve the general case, $\nabla^2 V = -\rho/\epsilon_0$, at least formally:

$$V(\mathbf{r}) = \int \rho(\mathbf{r}')\,G(\mathbf{r},\mathbf{r}')\,d\tau' = \int \frac{\rho(\mathbf{r}')}{4\,\pi\,\epsilon_0\,|\mathbf{r}-\mathbf{r}'|}\,d\tau'. \tag{C.22}$$

We have used "superposition," the notion that the sum of V for point sources gives a solution (to $\nabla^2 V = -\rho/\epsilon_0$) for the sum of the sources – that is built in to the linear nature of the field equation (sums of solutions are themselves solutions).

Using (C.19), we can find **E** for other configurations, too – suppose we have an infinite line of "charge" lying along the $\hat{\mathbf{z}}$-axis: $\rho = \lambda\,\delta(x)\,\delta(y)$. This time, the symmetry of the source leads to the ansatz (in cylindrical coordinates) $\mathbf{E} = E(s)\,\hat{\mathbf{s}}$, and we'll take a cylindrical volume of radius s and height ℓ as the domain in (C.19):

$$E(s)\,2\,\pi\,s\,\ell = \frac{1}{\epsilon_0}\,\lambda\,\ell \longrightarrow \mathbf{E} = \frac{\lambda}{2\,\pi\,\epsilon_0\,s}\,\hat{\mathbf{s}}. \tag{C.23}$$

Again, we can solve $\mathbf{E} = -\nabla V$ for V here, to get

$$V = -\frac{\lambda}{2\,\pi\,\epsilon_0}\,\log(s/s_0) \tag{C.24}$$

where s_0 is a constant that sets the zero of V.

Moving on to the **A** (static) field equation, $\nabla^2\mathbf{A} = -\mu_0\,\mathbf{J}$, our first question is, what are the sources **J**? Remember that we required J^μ to be a four-vector, and we have named the zero component $J^0 = \rho\,c$, so we can connect the vector components of **J** to J^0 by Lorentz transformation. Suppose we have a point source $\rho = q\,\delta^3(\bar{\mathbf{r}})$ sitting at the origin of a coordinate system $\bar{L}$ that is moving through the "lab" (coordinates given in L) with constant velocity $\mathbf{v} = v\,\hat{x}$. In the lab, the values of J^μ are

$$\begin{pmatrix} J^0 \\ J^x \\ J^y \\ J^z \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\,\beta & 0 & 0 \\ \gamma\,\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q\,\delta(\bar{x})\,\delta(\bar{y})\,\delta(\bar{z})\,c \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{C.25}$$

where we have set the vector component of $\bar{J}^\mu$ to zero – a point source has no associated vector with which to make a $\bar{\mathbf{J}}$. The delta functions require some thought – since the coordinates in the directions perpendicular to $\hat{\mathbf{x}}$ are unchanged, we have $\delta(\bar{y}) = \delta(y)$ and $\delta(\bar{z}) = \delta(z)$, but $\bar{x} = \gamma\,(x - v\,t)$ and

$$\delta(\bar{x}) = \delta(\gamma\,(x - v\,t)) = \frac{1}{\gamma}\,\delta(x - v\,t) \tag{C.26}$$

where we drop the absolute value signs on γ since $\gamma > 0$. The 0-component of J^μ is $J^0 = \gamma\,q\,c\,\delta(y)\,\delta(z)\,\delta(x - v\,t)/\gamma$, so we just recover $\rho = q\,c\,\delta(x - v\,t)\,\delta(y)\,\delta(z)$. For the vector components, we have $J^x = q\,v\,\delta(x - v\,t)\,\delta(y)\,\delta(z)$. The conclusion here is that $\mathbf{J} = \rho\,\mathbf{v}$ for a moving point source, and we can extrapolate from a single moving point source to a continuum description for ρ. Here, we are again exploiting the linearity of the field equation to sum the sources (then we know we just sum up the associated fields).

Taking the field equation $\nabla \times \mathbf{B} = \mu_0\,\mathbf{J}$ and integrating over a surface S, using Stokes's theorem, we have

$$\oint_{\partial S} \mathbf{B} \cdot d\boldsymbol{\ell} = \int_S \mathbf{J} \cdot d\mathbf{a}, \tag{C.27}$$

and again, we can use this to find $\mathbf{B}$ for cases that are simplified by their source geometry. As an example, take an infinite line of source, with constant λ as above, and move it with constant speed v along the $\hat{\mathbf{z}}$-axis. Then we have $\rho = \lambda\,\delta(x)\,\delta(y)$ and $\mathbf{J} = \rho\,\mathbf{v} = \lambda\,\delta(x)\,\delta(y)\,v\,\hat{\mathbf{z}}$. Given the "curl"-y relation to the source, and its azimuthal symmetry, we can guess $\mathbf{B} = B(s)\,\hat{\boldsymbol{\phi}}$ for the field. Using this assumed form, and the moving source in (C.27) applied to a circular area of radius s, we get

$$B(s)\,2\,\pi\,s = \mu_0\,\lambda v \longrightarrow \mathbf{B} = \frac{\mu_0\,\lambda\,v}{2\,\pi\,s}\,\hat{\boldsymbol{\phi}}. \tag{C.28}$$

C.3.1 Particle Motion

We have $\mathbf{E}$ and $\mathbf{B}$ for an infinite line of point sources moving with constant speed v – how would a particle respond to these? We know the relativistic form of Newton's second law from (C.16) – take a point "charge"[2] that sits in the y–z plane, initially a distance s_0 from the line (and lying on the $\hat{\mathbf{y}}$-axis), and starting from rest. In this case, there will be no

[2] I'm trying to be careful to highlight that we have not used *any* E&M input in this discussion – the only place(s) that we can make contact to this familiar theory are through units and suggestive variable names.

force component that moves the point charge out of the plane, so we can focus on the pure two-dimensional motion. Newton's second law reads

$$\begin{aligned}\frac{dp^y}{dt} &= \left[\frac{q\,\lambda\,\mu_0\,c^2}{2\,\pi\,y} - \frac{q\,\lambda\,v\,\dot{z}\,\mu_0}{2\,\pi\,y}\right]\\ \frac{dp^z}{dt} &= \left[\frac{q\,\lambda\,v\,\dot{y}\,\mu_0}{2\,\pi\,y}\right].\end{aligned} \tag{C.29}$$

The relativistic momenta are related to the velocities via $\mathbf{p} = m\,\mathbf{v}/\sqrt{1 - v^2/c^2}$, but we can invert that to get

$$\mathbf{v} = \frac{\mathbf{p}}{m}\frac{1}{\sqrt{1 + \left(\frac{p}{m\,c}\right)^2}}, \tag{C.30}$$

so that the equations of motion are

$$\begin{aligned}\frac{dp^y}{dt} &= \frac{q\,\lambda\,\mu_0\,c^2}{2\,\pi}\frac{1}{y}\left[1 - \frac{v}{c^2}\frac{p^z}{m}\frac{1}{\sqrt{1 + \left(\frac{p}{m\,c}\right)^2}}\right]\\ \frac{dp^z}{dt} &= \frac{q\,\lambda\,v\,\mu_0}{2\,\pi\,y}\frac{p^y}{m}\frac{1}{\sqrt{1 + \left(\frac{p}{m\,c}\right)^2}}.\end{aligned} \tag{C.31}$$

To remove the dimensions, take $y = \ell\,Y$, $z = \ell\,Z$, where ℓ has dimension of length and Y, Z are dimensionless; then take $t = \frac{\ell}{c}\,T$ for dimensionless T; and, finally, set $p^y = m\,c\,P^y$, $p^z = m\,c\,P^z$ (dimensionless P^y and P^z), $v = c\,V$. The equations of motion are now

$$\begin{aligned}\frac{dP^y}{dT} &= \frac{q\,\lambda\,\mu_0}{2\,\pi\,m}\frac{1}{Y}\left[1 - \frac{V\,P^z}{\sqrt{1 + P^2}}\right]\\ \frac{dP^z}{dT} &= \frac{q\,\lambda\,\mu_0}{2\,\pi\,m}\frac{V}{Y}\frac{P^y}{\sqrt{1 + P^2}}.\end{aligned} \tag{C.32}$$

Call the dimensionless constant out front $\alpha \equiv \frac{q\,\lambda\,\mu_0}{2\,\pi\,m}$. Solving this system numerically with $\alpha = 1$, $Y(0) = 1$ (i.e., $\ell = s_0$) and no initial momentum, we get the trajectory shown in Figure C.1 (using dimensionless time-step $\Delta T = 0.001$, $V = 0.9$, and running out to $T = 400$). The particle velocities, V^y and V^z (reconstructed from the momenta), are shown in Figure C.2.

The trajectory of the particle is familiar and physically reasonable – starting from rest, the "electric" field of the line of charge pushes the particle away (in the $\hat{\mathbf{y}}$-direction), leading to a "magnetic" force that points along the $\hat{\mathbf{z}}$-axis. The two forces compete, with the magnetic one tending to cause the particle to circle back toward the wire, while the electric one always pushes away from the wire (with increasing strength as the particle approaches the line of charge). Notice, in the velocity plots, that the y-component goes through zero just as the z-component is a maximum – that zero is the turnaround point in the trajectory, where the particle starts heading back toward the line. Because the equations of motion are relativistic, the speed of the particle never exceeds c (which is 1 in this dimensionless setting).

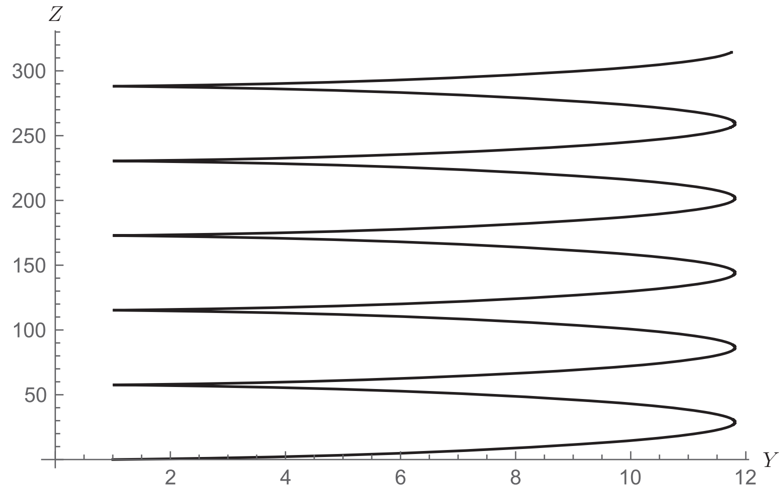

Fig. C.1 The trajectory of a particle moving according to (C.32).

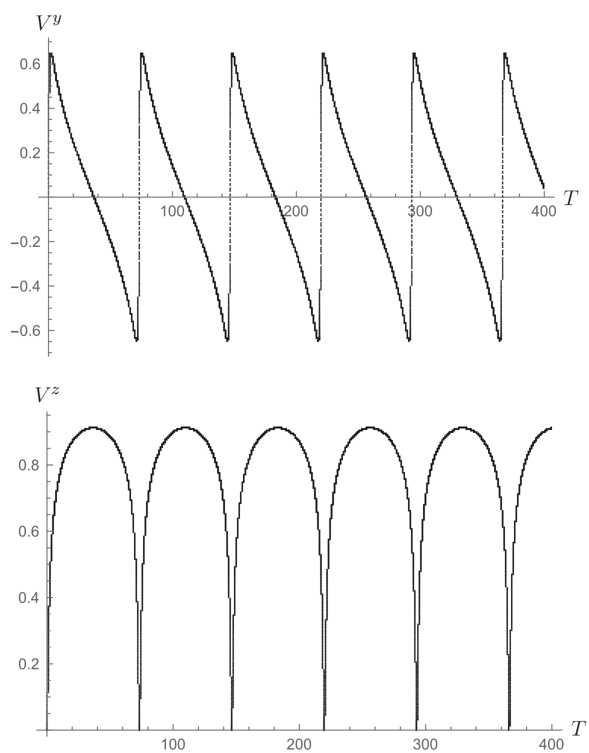

Fig. C.2 The *y* and *z* (dimensionless) components of velocity for a particle moving according to (C.32).

C.4 Point Source Solution

Going back to the full field equations, $\Box A^\mu = -\mu_0 J^\mu$, we'd like an integral solution for A^μ. To find it, we'll solve for the Green's function of $\Box$ – we want $G(\mathbf{r}, \mathbf{r}', t, t')$ satisfying

$$\left[-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right] G(\mathbf{r}, \mathbf{r}', t, t') = -\delta^3(\mathbf{r} - \mathbf{r}')\,\delta(t - t'). \tag{C.33}$$

We'll put the "point particle"[3] at the origin $\mathbf{r}' = 0$ at time $t' = 0$ and generalize later. To set convention, define the temporal Fourier transform of G to be

$$\tilde{G}(\mathbf{r}, f) = \int_{-\infty}^{\infty} G(\mathbf{r}, t)\, e^{i 2\pi f t}\, dt. \tag{C.34}$$

To use the Fourier transform, take (C.33), multiply it by $e^{i2\pi f t}$, and integrate in time (from $-\infty \to \infty$), using integration by parts on the term with temporal derivatives. We then have the spatial problem

$$\left[\nabla^2 + \left(\frac{2\pi f}{c}\right)^2\right] \tilde{G} = -\delta^3(\mathbf{r}), \tag{C.35}$$

and this we can solve: take $\tilde{G}(\mathbf{r}, f) = \tilde{G}(r, f)$ to focus on the spherically symmetric solution, then for $\mathbf{r} \neq 0$,

$$\frac{1}{r}\left(r\tilde{G}\right)'' = -\left(\frac{2\pi f}{c}\right)^2 \tilde{G} \longrightarrow r\tilde{G} = A\, e^{i\left(\frac{2\pi f}{c}\right) r} + B\, e^{-i\left(\frac{2\pi f}{c}\right) r} \tag{C.36}$$

for constants A and B. If $f = 0$, we should recover the Green's function for ∇^2, which tells us to set $A = \frac{a}{4\pi}$ with $B = \frac{1-a}{4\pi}$ for a new constant a. Now performing the inverse Fourier transform, we get back

$$G(\mathbf{r}, t) = \frac{1}{4\pi r}\left(a\,\delta(t - r/c) + (1 - a)\,\delta(t + r/c)\right). \tag{C.37}$$

The second delta function here suggests that there is a contribution to G whose influence occurs *before* the point source has "flashed" (at $t' = 0$) – we'll choose $a = 1$ to exclude this acausal piece of the Green's function. When we move the source to $\mathbf{r}'$ at t', we get the final Green's function:

$$G(\mathbf{r}, \mathbf{r}', t, t') = \frac{\delta(t' - (t - |\mathbf{r} - \mathbf{r}'|/c))}{4\pi\,|\mathbf{r} - \mathbf{r}'|}. \tag{C.38}$$

The general solution to the original field equation is

$$\begin{aligned} A^\mu(\mathbf{r}, t) &= \mu_0 \int \frac{\delta(t' - (t - |\mathbf{r} - \mathbf{r}'|/c))\, J^\mu(\mathbf{r}', t')}{4\pi\,|\mathbf{r} - \mathbf{r}'|}\, d\tau'\, dt' \\ &= \mu_0 \int \frac{J^\mu(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{4\pi\,|\mathbf{r} - \mathbf{r}'|}\, d\tau', \end{aligned} \tag{C.39}$$

using the delta function to perform the t' integration.

[3] The source here is not really a point particle at all – it instantaneously appears at $\mathbf{r}'$ at t' and then, just as quickly, disappears, violating our notion of conservation.

C.5 Stress Tensor

The stress tensor of this field theory can be obtained from its scalar Lagrangian (omitting the source portion),

$$\begin{aligned}\hat{\mathcal{L}} &= \frac{1}{4\,\mu_0}\,F_{\alpha\rho}\,F_{\beta\sigma}\,\eta^{\alpha\beta}\,\eta^{\rho\sigma}\\ F_{\alpha\rho} &\equiv A_{\rho,\alpha} - A_{\alpha,\rho} = -F_{\rho\alpha}.\end{aligned} \tag{C.40}$$

We need to evaluate the derivative of $\hat{\mathcal{L}}$ with respect to $\eta_{\mu\nu}$ to construct the stress tensor. Note that

$$\frac{\partial \eta^{\alpha\beta}}{\partial \eta_{\mu\nu}} = -\eta^{\alpha\mu}\,\eta^{\beta\nu} \tag{C.41}$$

and, using this identity, together with the product rule, gives

$$\frac{\partial \hat{\mathcal{L}}}{\partial \eta_{\mu\nu}} = -\frac{1}{2\,\mu_0}\,F^{\mu\sigma}\,F^{\nu}{}_{\sigma}. \tag{C.42}$$

From the coordinate-invariance of the action comes the stress tensor definition

$$T^{\mu\nu} = -\left(2\,\frac{\partial\hat{\mathcal{L}}}{\partial\eta_{\mu\nu}} + \eta^{\mu\nu}\,\hat{\mathcal{L}}\right) = \frac{1}{\mu_0}\left(F^{\mu\sigma}\,F^{\nu}{}_{\sigma} - \frac{1}{4}\,\eta^{\mu\nu}\,F_{\alpha\beta}\,F^{\alpha\beta}\right), \tag{C.43}$$

with its set of four conservation laws: $\partial_\mu\,T^{\mu\nu} = 0$. It is easiest to express the stress tensor components using the auxiliary **E** and **B** we identified in Section C.2; in terms of those,

$$F^{\mu\nu} \doteq \begin{pmatrix} 0 & E^x/c & E^y/c & E^z/c \\ -E^x/c & 0 & B^z & -B^y \\ -E^y/c & -B^z & 0 & B^x \\ -E^z/c & B^y & -B^x & 0 \end{pmatrix}. \tag{C.44}$$

Then we can write the components as (using i, j ranging from 1 to 3):

$$\begin{aligned} T^{00} &= \frac{1}{2\,\mu_0}\left(\frac{E^2}{c^2} + B^2\right)\\ T^{0j} &= \frac{1}{\mu_0\,c}\,(\mathbf{E}\times\mathbf{B})^j\\ T^{ij} &= -\frac{1}{\mu_0}\left[\frac{1}{c^2}\,E^i\,E^j + B^i\,B^j - \frac{1}{2}\,\eta^{ij}\left(\frac{E^2}{c^2} + B^2\right)\right].\end{aligned} \tag{C.45}$$

Conservation of energy, for the field, reads (letting $T^{00} \equiv u$, the energy density, and $T^{0j} \equiv \frac{1}{c}\,S^j$)

$$\partial_\mu\,T^{\mu 0} = 0 \longrightarrow \frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S} = 0. \tag{C.46}$$

C.5.1 Vacuum Solutions

Let's return to the full field equations, this time with zero source: $\Box A^\mu = 0$, the wave equation for all four components of A^μ. The simplest, and in some ways most natural, solution to this equation is

$$A^\mu = P^\mu \, e^{i k_\alpha x^\alpha} \tag{C.47}$$

for constant P^μ, k_α – a single Fourier mode. We know that we can add these together (by superposition) to make anything (because of the completeness of the exponentials), so this is a basic "unit" for vacuum. In order for the field equation to be satisfied ($\Box A^\mu = 0$) in Lorenz gauge ($\partial_\mu A^\mu = 0$), we must have

$$k^\alpha k_\alpha = 0, \qquad k^\alpha P_\alpha = 0. \tag{C.48}$$

We can perform additional gauge fixing to eliminate the P^0 component of the field – take $\bar{A}_\mu = A_\mu + \phi_{,\mu}$ with $\partial^\mu \phi_{,\mu} = 0$ (so as not to spoil the already-Lorenz gauge choice), then let $\phi = Q\, e^{i k_\alpha x^\alpha}$, ensuring that $\Box\phi = 0$. The new field is

$$\bar{A}_\mu = \left(P_\mu + Q\, k_\mu\right) \, e^{i k_\alpha x^\alpha}. \tag{C.49}$$

We know that $k_\mu k^\mu = 0$, and then $k_0 \neq 0$; if $k_0 = 0$, then $k_\mu k^\mu = \mathbf{k}\cdot\mathbf{k} = 0$ (i.e., the spatial portion of the vector has zero length) means $k^j = 0$, too. So we can pick $Q = -P_0/k_0$ to eliminate the 0 component of $\bar{A}_\mu$. Finally, we have

$$\begin{aligned} A^0 &= 0 \\ \mathbf{A} &= \mathbf{P}\, e^{i k_\alpha x^\alpha} = \mathbf{P}\, e^{i\,(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r})}. \end{aligned} \tag{C.50}$$

How about the $\mathbf{E}$ and $\mathbf{B}$ fields? We have $\mathbf{E} = -c\,\nabla A^0 - \frac{\partial \mathbf{A}}{\partial t}$, $\mathbf{B} = \nabla\times\mathbf{A}$:

$$\begin{aligned} \mathbf{E} &= -\mathbf{P}\,(i\, k_0\, c)\; e^{i\,(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r})} \\ \mathbf{B} &= i\,\mathbf{k}\times\mathbf{P}\, e^{i\,(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r})}. \end{aligned} \tag{C.51}$$

The $\mathbf{E}$ field points in the $\mathbf{P}$-direction,[4] while $\mathbf{B}$ is perpendicular to both $\mathbf{P}$ and $\mathbf{k}$. We can calculate the energy density and $\mathbf{S}$ (energy flux vector) for these plane waves; we'll use the real part of the fields in (C.51) to perform the calculation, so we'll take

$$\begin{aligned} \mathbf{E} &= \mathbf{P}\,(k_0\, c)\; \sin(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r}) \\ \mathbf{B} &= -\mathbf{k}\times\mathbf{P}\; \sin(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r})\,. \end{aligned} \tag{C.52}$$

The energy density is

$$\begin{aligned} u &= \frac{1}{2\,\mu_0}\left(\frac{E^2}{c^2} + B^2\right) = \frac{1}{2\,\mu_0}\left(P^2\, k_0^2 + \left(k^2\, P^2 - (\mathbf{k}\cdot\mathbf{P})^2\right)\right)\, \sin^2(k_0\, c\, t + \mathbf{k}\cdot\mathbf{r}) \\ &= \frac{k^2\, P^2}{\mu_0}\; \sin^2(k\, c\, t + \mathbf{k}\cdot\mathbf{r}) \end{aligned} \tag{C.53}$$

[4] Notice that in this case, we have done enough gauge fixing to make the polarization four-vector in A^μ physically relevant – it is the vector direction of $\mathbf{E}$, the field that shows up in Newton's second law.

since $P^\mu k_\mu = 0$ with $P^0 = 0$ gives $\mathbf{P} \cdot \mathbf{k} = 0$, and $k^\mu k_\mu = 0$ gives $k_0 = k$. The vector that tells us how energy is transported out of a region is

$$\mathbf{S} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} = \frac{(k^0 c)}{\mu_0} P^2 \sin^2(k c t + \mathbf{k} \cdot \mathbf{r})\ \mathbf{k}, \tag{C.54}$$

and it points (appropriately) along the direction of travel for the plane wave.

C.6 Radiation

Finally, we will explore the *source* of the vacuum solutions from the previous section – somewhere, there is a source that produces the initial A^μ, which then propagates in the vacuum of space, where $J^\mu = 0$. To connect the vacuum solutions to their sources, we will use the integral solution (C.39) with some physically motivated approximations. Suppose we are given $J^0 = \rho c$ and $\mathbf{J}$ as the sources. Then the 0 component of the field equation has solution

$$\begin{aligned} V &= \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}', t')\, \delta(t' - (t - |\mathbf{r} - \mathbf{r}'|/c))}{|\mathbf{r} - \mathbf{r}'|}\, d\tau'\, dt' \\ &= \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|}\, d\tau', \end{aligned} \tag{C.55}$$

and we can write the distance between the source at $\mathbf{r}'$ and the field point at $\mathbf{r}$ as

$$|\mathbf{r} - \mathbf{r}'| = r \left[1 - 2\frac{\hat{\mathbf{r}} \cdot \mathbf{r}'}{r} + \left(\frac{r'}{r}\right)^2\right]^{1/2}. \tag{C.56}$$

If we assume that the motion of the source is such that $\mathbf{r}'$ remains in a bounded domain with some typical length ℓ, then we can focus on field locations for which $r \gg \ell$ or, inside the integral, $r \gg r'$. Under this assumption, we will have

$$\begin{aligned} \frac{1}{|\mathbf{r} - \mathbf{r}'|} &= \frac{1}{r}\left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{r}'}{r}\right) + O((r'/r)^2) \\ \rho(\mathbf{r}', t - |\mathbf{r} - \mathbf{r}'|/c) &\approx \rho(\mathbf{r}', t - r/c + \hat{\mathbf{r}} \cdot \mathbf{r}'/c) = \rho(\mathbf{r}', t - r/c) + \dot{\rho}(\mathbf{r}', t - r/c)\frac{\hat{\mathbf{r}} \cdot \mathbf{r}'}{c} \end{aligned} \tag{C.57}$$

where again we are using the notion that $r/c \gg r'/c$ in the charge density expansion (and assuming that the higher-order temporal derivatives of ρ do not get large). Using these together, we get

$$\begin{aligned} V(\mathbf{r}, t) &\approx \frac{1}{4\pi\epsilon_0 r} \int \left[\left(\rho(\mathbf{r}', t - r/c) + \dot{\rho}(\mathbf{r}', t - r/c)\frac{\hat{\mathbf{r}} \cdot \mathbf{r}'}{c}\right)\left(1 + \frac{\hat{\mathbf{r}} \cdot \mathbf{r}'}{r}\right)\right] d\tau' \\ &\approx \frac{1}{4\pi\epsilon_0 r}\left[\int \rho(\mathbf{r}', t - r/c)\, d\tau' + \frac{\hat{\mathbf{r}}}{r} \cdot \int \mathbf{r}'\, \rho(\mathbf{r}', t - r/c)\, d\tau' \right. \\ &\quad \left. + \frac{\hat{\mathbf{r}}}{c} \cdot \int \mathbf{r}'\, \dot{\rho}(\mathbf{r}', t - r/c)\, d\tau'\right] \end{aligned} \tag{C.58}$$

where the first term is just the total "charge," the second term is the dipole moment (evaluated at $t - r/c$) of the charge distribution, and the third term, which goes like $1/r$, is the derivative of the dipole moment. We can write

$$V(\mathbf{r}, t) = \frac{1}{4\pi\,\epsilon_0\, r}\left[Q + \frac{\hat{\mathbf{r}}\cdot\mathbf{p}(t - r/c)}{r} + \frac{1}{c}\,\hat{\mathbf{r}}\cdot\dot{\mathbf{p}}(t - r/c)\right]. \tag{C.59}$$

Similarly, the magnetic vector potential for this configuration becomes

$$\mathbf{A}(\mathbf{r}, t) \approx \frac{\mu_0}{4\pi\, r}\int \mathbf{J}(\mathbf{r}', t - r/c)\, d\tau', \tag{C.60}$$

with $\int \mathbf{J}\, d\tau = \dot{\mathbf{p}}$, so

$$\mathbf{A}(\mathbf{r}, t) \approx \frac{1}{4\pi\,\epsilon_0\, r\, c^2}\,\dot{\mathbf{p}}(t - r/c). \tag{C.61}$$

The magnetic field follows from $\mathbf{B} = \nabla \times \mathbf{A}$:

$$\mathbf{B} = -\frac{\mu_0}{4\pi\, r}\left(\frac{\hat{\mathbf{r}}\times\dot{\mathbf{p}}(t - r/c)}{r} + \frac{\hat{\mathbf{r}}\times\ddot{\mathbf{p}}(t - r/c)}{c}\right). \tag{C.62}$$

We will ultimately be interested in that portion of $\mathbf{E}$ and $\mathbf{B}$ that go like $1/r$; those are the ones that lead to energy loss from the physical system, so we'll focus by writing

$$\mathbf{B} = -\frac{\mu_0}{4\pi\, r\, c}\,\hat{\mathbf{r}}\times\ddot{\mathbf{p}}(t - r/c). \tag{C.63}$$

The electric field comes from $\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}$; evaluating it, and dropping any $1/r^2$ terms, yields

$$\begin{aligned}\mathbf{E} &= \frac{1}{4\pi\,\epsilon_0\, r}\left(\frac{1}{c^2}\left(\hat{\mathbf{r}}\cdot\ddot{\mathbf{p}}(t - r/c)\right)\hat{\mathbf{r}} - \frac{1}{c^2}\,\ddot{\mathbf{p}}(t - r/c)\right)\\ &= \frac{\mu_0}{4\pi\, r}\left(\left(\hat{\mathbf{r}}\cdot\ddot{\mathbf{p}}(t - r/c)\right)\hat{\mathbf{r}} - \ddot{\mathbf{p}}(t - r/c)\right).\end{aligned} \tag{C.64}$$

The pair (C.63) and (C.64) give rise to an energy flux vector (the familiar Poynting vector) that goes like $1/r^2$ and is perpendicular to both $\mathbf{E}$ and $\mathbf{B}$. These fields look, locally, like the plane wave solutions discussed above. The energy carried by these fields is lost to the source, flowing radially away from the source and remaining constant on spheres of arbitrary radius; the energy eventually leaves the "universe."

References

[1] J. M. Aguirregabiria, A. Hernández, and M. Rivas, "δ-function converging sequences," *Am. J. Phys.*, **70**, 180–185, 2002.

[2] G. Arfken, H. Weber, *Mathematical Methods for Physicists*, 5th ed., Harcourt, 2001.

[3] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles*, Dover, 2010.

[4] M. Boas, *Mathematical Methods in the Physical Sciences*, 3rd ed., Wiley, 2005.

[5] M. Born and L. Infeld. "Foundations of the new field theory," *Proceedings of the Royal Society of London A*, **144**, 425–451, 1934.

[6] T. Cheng, *A College Course on Relativity and Cosmology*, Oxford University Press, 2015.

[7] D. Clark, J. Franklin, and N. Mann, "Relativistic linear restoring force," *European Journal of Physics*, **33**, 1041–1051, 2012.

[8] S. Deser, "Self-interaction and gauge invariance," *General Relativity and Gravitation*, **1**, 9–18, 1970. Available at gr-qc/0411023.

[9] S. Deser, R. Jackiw, and S. Templeton, "Topologically massive gauge theories," *Annals of Physics*, **140**, 372–411, 1982.

[10] A. Einstein, "Zur Theorie des statschen Gravitationsfeldes," *Ann. Phys. (Leipzig)*, **38**, 443–458, 1912. (English translation in The Collected Papers of Albert Einstein, trans. Anna Beck, Princeton University Press, 1996, vol. 4, pp. 107–120).

[11] J. Franklin, *Advanced Mechanics and General Relativity*, Cambridge University Press, 2010.

[12] J. Franklin, "Self-consistent, self-coupled scalar gravity," *American Journal of Physics*, **83**, 332–337, 2015.

[13] J. Franklin, *Computational Methods for Physics*, Cambridge University Press, 2013.

[14] J. Franklin and D. J. Griffiths, "The fields of a charged particle in hyperbolic motion," *American Journal of Physics*, **82**, 755–763, 2014.

[15] Thomas A. Garrity, *Electricity and Magnetism for Mathematicians*, Cambridge University Press, 2015.

[16] N. Giordano, *Computational Physics*, Prentice Hall, 1997.

[17] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics*, 3rd ed., Pearson, 2001.

[18] W. Greiner and J. Reinhardt, *Quantum Electrodynamics*, 2nd ed., Springer, 1994.

[19] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed., Pearson, 2013.

[20] J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, 1998.

[21] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Butterworth-Heinemann, 1980.

[22] T. Pang, *An Introduction to Computational Physics*, Cambridge University Press, 2006.

[23] P. C. Peters, "Where is the energy stored in a gravitational field?," *Am. J. Phys.*, **49**(6), 564–569, 1981.

[24] B. Schutz, *A First Course in General Relativity*, 2nd ed., Cambridge University Press, 2009.

[25] Walter Thirring and E. M. Harrell, *Classical Mathematical Physics: Dynamical Systems and Field Theories*, 3rd ed., Springer, 2003.

[26] A. Zangwill, *Modern Electrodynamics*, Cambridge University Press, 2012.

[27] B. Zwiebach, *A First Course in String Theory*, 2nd ed., Cambridge University Press, 2009.

Index