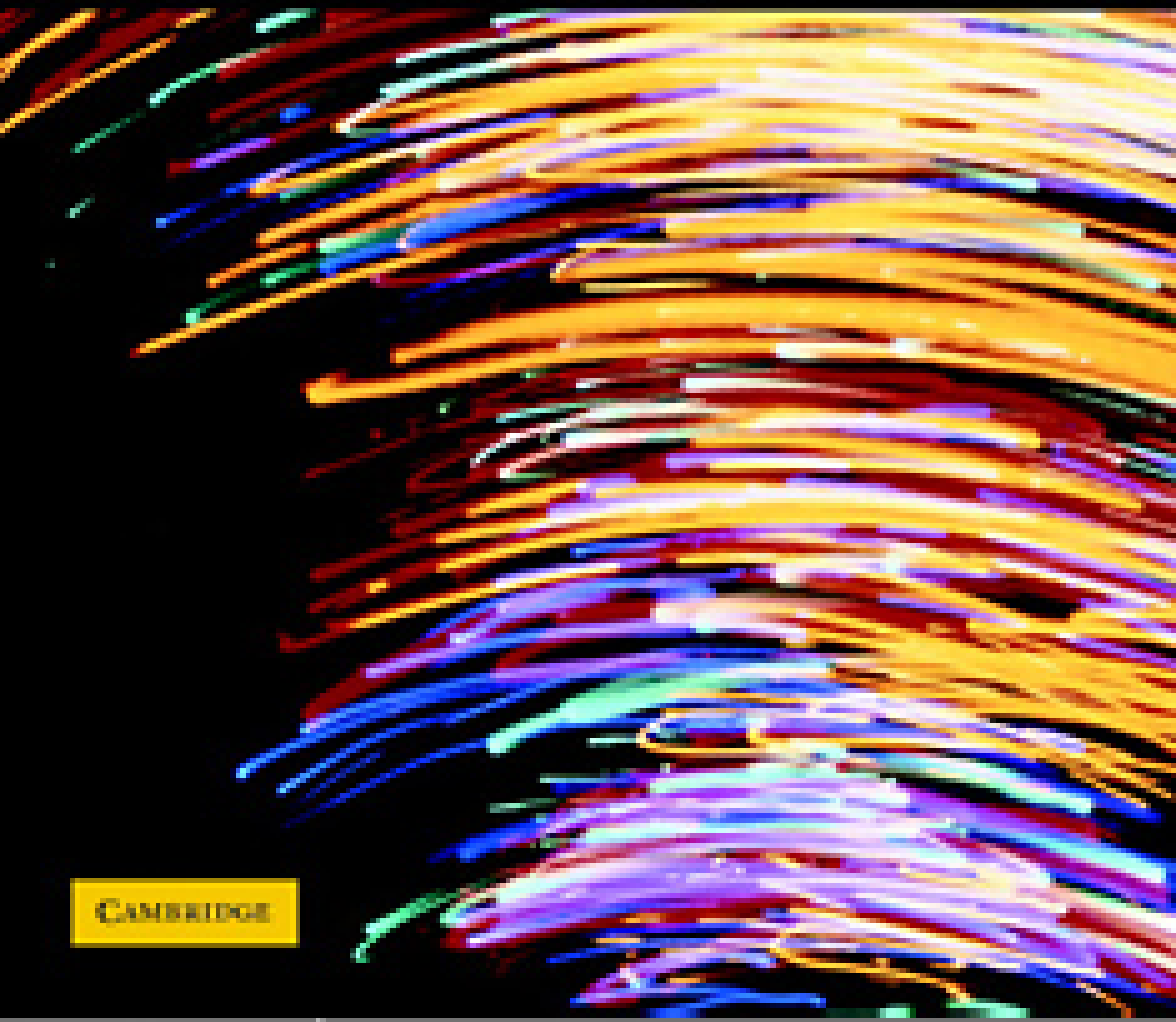
Patricia M. Schwarz
and John H. Schwarz

Special Relativity

From Einstein to Strings



CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521812603

This page intentionally left blank

SPECIAL RELATIVITY
from Einstein to Strings

The traditional undergraduate physics treatment of special relativity is too cursory to warrant a textbook. The graduate treatment of special relativity is deeper, but often fragmented between different courses such as general relativity and quantum field theory. For this reason physics students need one book that ties it all together. With this in mind, this book is written as a textbook for the self-learner whose physics background includes a minimum of one year of university physics with calculus. More advanced mathematical topics, such as group theory, are explained as they arise. The readership is expected to include high school and college physics educators seeking to improve and update their own understanding of special relativity in order that they may teach it better, science and engineering undergraduates who want to extend their cursory knowledge of relativity to greater depth, and physics graduate students looking for a simple unified treatment of material that usually appears in the graduate physics curriculum in a somewhat disconnected fashion.

The main difference between this book and existing books on special relativity is that it extends the topic list beyond the standard basic topics of spacetime geometry and physics, to include the more current and more advanced (but still accessible) topics of relativistic classical fields, causality, relativistic quantum mechanics, basic supersymmetry, and an introduction to the relativistic string. Another difference is that in most cases the dimension of space is allowed to be arbitrary.

A companion CD-ROM contains Flash animations of key examples and problems discussed in the book. Understanding relativity requires that the student be able to visualize relative motion from different points of view, making animated diagrams preferable to static diagrams where relative motion has to be decoded from complicated symbols labeling each observer.

PATRICIA SCHWARZ received a Ph.D. in theoretical physics from the California Institute of Technology. Her research specialty is spacetime geometry in general relativity and string theory. She is also an expert in multimedia and online education technology. Her award-winning multimedia-rich web site at http://superstringtheory.com is popular with the public and is used widely as an educational resource across the world.

JOHN SCHWARZ, the Harold Brown Professor of Theoretical Physics at the California Institute of Technology, is one of the founders of superstring theory. He co-authored a two-volume monograph *Superstring Theory* with Michael Green and Edward Witten in 1987. He is a MacArthur Fellow and a member of the National Academy of Sciences.

SPECIAL RELATIVITY

From Einstein to Strings

PATRICIA M. SCHWARZ AND JOHN H. SCHWARZ
Pasadena, California

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521812603

First published in print format 2004

ISBN-13 978-0-511-18903-6 eBook (Adobe Reader)
ISBN-10 0-511-18903-6 eBook (Adobe Reader)

ISBN-13 978-0-521-81260-3 hardback
ISBN-10 0-521-81260-7 hardback

To spacetime and everyone who has ever tried to understand it

Contents

Preface *page* xi

Part I Fundamentals 1

1 From Pythagoras to spacetime geometry 3
- 1.1 Pythagoras and the measurement of space 4
- 1.2 The differential version, in D dimensions 9
- 1.3 Rotations preserve the Euclidean metric 10
- 1.4 Infinitesimal rotations 12
- 1.5 Could a line element include time? 14
- 1.6 The Lorentz transformation 17

2 Light surprises everyone 21
- 2.1 Conflicting ideas about space and light 22
- 2.2 Maxwell's transverse undulations 25
- 2.3 Galilean relativity and the ether 28
- 2.4 The Michelson–Morley experiment 31
- 2.5 Einstein ponders electromagnetism and relativity 36
- 2.6 Einstein's two postulates 37
- 2.7 From light waves to spacetime geometry 46

3 Elements of spacetime geometry 55
- 3.1 Space and spacetime 55
- 3.2 Vectors on a manifold 67
- 3.3 Vectors in spacetime 75
- 3.4 Tensors and forms 83
- 3.5 The Principle of Relativity as a geometric principle 89

4 Mechanics in spacetime 95
- 4.1 Equations of motion in spacetime 95

4.2 Momentum and energy in spacetime 101
4.3 Energy and momentum conservation in spacetime 105
4.4 Relativistic kinematics 109
4.5 Fission, fusion, and $E = Mc^2$ 119
4.6 Rigid body mechanics 123

5 Spacetime physics of fields 127
5.1 What is a field? 128
5.2 Differential calculus in spacetime 132
5.3 Integral calculus in spacetime 146
5.4 Continuous systems in spacetime 156
5.5 Electromagnetism 169
5.6 What about the gravitational field? 189

6 Causality and relativity 197
6.1 What is time? 197
6.2 Causality and spacetime 205

Part II Advanced Topics 219

7 When quantum mechanics and relativity collide 221
7.1 Yet another surprise about light 222
7.2 The Schrödinger equation is not covariant 225
7.3 Some new ideas from the Klein–Gordon equation 231
7.4 The Dirac equation and the origin of spin 233
7.5 Relativity demands a new approach 242
7.6 Feynman diagrams and virtual particles 251

8 Group theory and relativity 260
8.1 What is a group? 260
8.2 Finite and infinite groups 267
8.3 Rotations form a group 270
8.4 Lorentz transformations form a group 277
8.5 The Poincaré group 282

9 Supersymmetry and superspace 287
9.1 Bosons and fermions 289
9.2 Superspace 293
9.3 Supersymmetry transformations 297
9.4 $\mathcal{N} = 1$ supersymmetry in four dimensions 302
9.5 Massless representations 306

10 Looking onward 312
10.1 Relativity and gravity 312
10.2 The standard model of elementary particle physics 320
10.3 Supersymmetry 323
10.4 The relativistic string 331
10.5 Superstrings 338
10.6 Recent developments in superstring theory 340
10.7 Problems and prospects 344

Appendix 1 Where do equations of motion come from? 349

Appendix 2 Basic group theory 359

Appendix 3 Lie groups and Lie algebras 362

Appendix 4 The structure of super Lie algebras 365

References 367
Index 369

Preface

Towards the end of the nineteenth century many physicists believed that all the fundamental laws that describe the physical Universe were known, and that all that remained to complete the understanding was an elaboration of details. The mind-boggling error of this viewpoint was laid bare within a few short years. Max Planck introduced the quantum in 1899 and Albert Einstein's breakthrough work on special relativity appeared in 1905. The ensuing relativity and quantum revolutions each led to surprising and unexpected concepts and phenomena that have profoundly altered our view of physical reality. The science and the history associated with each of these revolutions has been told many times before. But they are worth coming back to again and again with the added benefit of historical perspective. After all, they have changed the world scientifically, technologically, and philosophically. Perhaps due to the lesson from a century ago, very few people today are so foolish as to speak of an "end of science". In fact, revolutionary advances in theoretical physics are currently in progress, and we seem to be a long way from achieving a settled and final picture of physical reality.

As the title indicates, this book is about the special theory of relativity. This theory overthrew the classical view of space and time as distinct and absolute entities that provide the backdrop on which physical reality is superimposed. In special relativity space and time must be viewed together (as spacetime) to make sense of the constancy of the speed of light and the structure of Maxwell's electromagnetic theory. The basic consequences of special relativity can be described by simple algebraic formulas, but a deeper understanding requires a geometric description. This becomes absolutely crucial for the extension to include gravity.

This book is divided into two parts – entitled "Fundamentals" and "Advanced Topics." The first part gives a detailed explanation of special relativity. It starts with simple mathematics and intuitive explanations and gradually builds up more advanced mathematical tools and concepts. Ultimately, it becomes possible to recast Maxwell's electromagnetic theory in terms of two simple equations

($dF = 0$ and $d * F = *j$) that incorporate relativistic geometry in a simple and beautiful way. Each chapter in Part I of the book starts with a "hands-on exercise." These are intended to help the reader develop spatial awareness. They are not supposed to be scientific experiments, rather they are exercises to limber up the mind.

The second part of the book includes advanced topics that illustrate how relativity has impacted subsequent developments in theoretical physics up to and including modern work on superstring theory. Relativity and quantum mechanics each raised a host of new issues. Their merger led to many more. This is discussed in Chapter 7. One aspect of the structure of spacetime implied by special relativity is its symmetry. To describe this properly requires a branch of algebra called group theory. This is explored in Chapter 8. Chapter 9 raises the question of whether the symmetry of spacetime can be extended in a nontrivial way, and it describes the unique answer, which is supersymmetry. The last chapter gives a brief overview of modern theoretical physics starting with the well-established theories: general relativity and the standard model of elementary particles. It then discusses more speculative current research topics, especially supersymmetry and string theory, and concludes with a list of unsolved problems. These are topics that one would not ordinarily find in a book about special relativity. We hope the reader will enjoy finding them in a form that is more detailed than a popular book, but less technical than a textbook for a graduate-level course.

Part I

Fundamentals

1

From Pythagoras to spacetime geometry

Hands-on exercise:[1] measuring the lengths of lines

Physics is about describing the physical world. In physics courses we get used to doing this using mathematics, and sometimes it can seem as if the mathematics is the physics. But our goal is to learn about the physical world, and so sometimes we have to just put the math aside and let the physical world be our teacher. It is in this spirit that we begin this chapter with a hands-on exercise that requires measuring the physical world with your hands. To complete this exercise you will need the following supplies:

- Three cloth or paper measuring tapes, preferably from computer printouts of the file *measures.html* included on the CD that comes with this book.
- Some Scotch tape.
- One large spherical object such as a large melon, a beach ball or a globe, with a diameter roughly between 15 and 20 cm.
- One flat table or desk.
- A pencil and some graph paper.

If you have printed out the page with the measuring tapes on them from the CD, cut them out with the edges of the paper aligned with the measuring edges of the printed tapes. Tape measures A and B should be taped together at a right angle to one another with the measuring edges facing one another. We will call this taped-together object the Side Measurer. The Side Measurer will be used to measure the lengths of the two sides of a right triangle, while tape measure C, which we will call the Hypotenuse Measurer, will be used to measure the length of the hypotenuse, in the common set of units inscribed on the three measures.

[1] Each chapter in Part I of the book starts with a "hands-on exercise." These are intended to help the reader develop spatial awareness. They are not supposed to be scientific experiments, rather they are exercises to limber up the mind. The reader is free to skip them, of course.

Go to your desk or table and tape the corner of the Side Measurer onto some convenient location on its surface. Now use the Hypotenuse Measurer to measure the distances between the locations on the Side Measurer marked by the numbers 1, 2, 3, 4, 5, 6. In other words, measure the distances from 1 to 1, 2 to 2 and so on. Make a table on your graph paper to record your measurements. Plot the results on the graph paper with the side lengths on the x axis and the hypotenuse lengths on the y axis.

Next untape the Side Measurer from the desk or table. Grab your large spherical object (henceforth referred to as the LSO) and tape the corner of the Side Measurer onto some convenient location on its surface, taking care to preserve the right angle where tape measures A and B are taped together. Now use the Hypotenuse Measurer to measure the same set of distances that you measured previously when the Side Measurer was taped to the table or desk. Write them down in a table as you did above, and then plot the data on the plot you made above.

Now on the same plot, draw the line $y = \sqrt{2}\,x$. Write down any impressions you have or conclusions you arrive at by looking at these data, and save them for later.

1.1 Pythagoras and the measurement of space

What does the previous hands-on exercise have to do with special relativity? Special relativity is a theory of spacetime geometry. Before we try to understand the geometry of spacetime, let's go back over what we've already learned about the geometry of space. In the exercise above we were exploring the applicability of the Pythagorean theorem on two different surfaces. The Pythagorean theorem states that:

Given a right triangle, the sum of the squares bounding the two legs of the triangle is equal to the square bounding the hypotenuse of the triangle.

In Pythagoras' time, there was only geometry – algebra was still 1300 years in the future. Pythagoras wasn't talking about the squares of the lengths of the sides as numbers. He proved his theorem by cutting up the squares on the legs and showing that the pieces could be reassembled into the square on the hypotenuse, so that the two squares truly were equal. But now that we have algebra, we can say that if the lengths of the two sides of a right triangle are denoted by A and B, then the length C of the hypotenuse of the right triangle is given by solving the equation

$$A^2 + B^2 = C^2 \tag{1.1}$$

for the value of C.

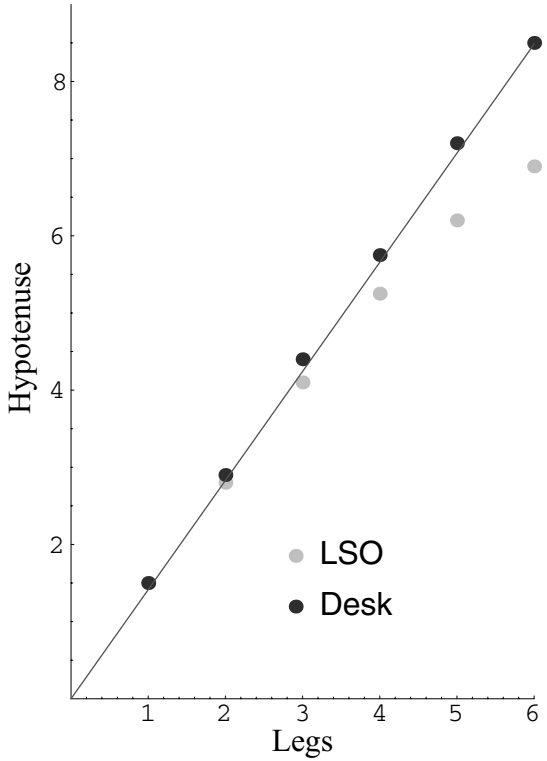


Fig. 1.1. Data from hands-on exercise.

As you should be able to see in the hands-on exercise, this formula works quite reliably when we're measuring right triangles on a desk or table but begins to fail when we measure right triangles on the LSO. One set of data from this exercise is plotted in Figure 1.1.

Now let's use math to explore this issue further. Consider a right isosceles triangle on a two-dimensional sphere of radius R with azimuthal angle θ and polar angle ϕ. A triangle in flat space is determined by three straight lines. The closest analog to a straight line on a sphere is a great circle. Let's make the legs of our right triangle extend from the north pole of the sphere along the great circles determined by $\phi = 0$ and $\phi = \pi/2$, beginning at $\theta = 0$ and terminating at $\theta = \theta_0$. The arc of a great circle of radius R subtending an angle θ_0 has arc length $R\theta_0$, so we can say that $A = B = R\theta_0$. The arc length of the great circle serving as the hypotenuse is given by

$$C = R\cos^{-1}(\cos^2\theta_0). \tag{1.2}$$

We leave the derivation of this result as an exercise for the reader.

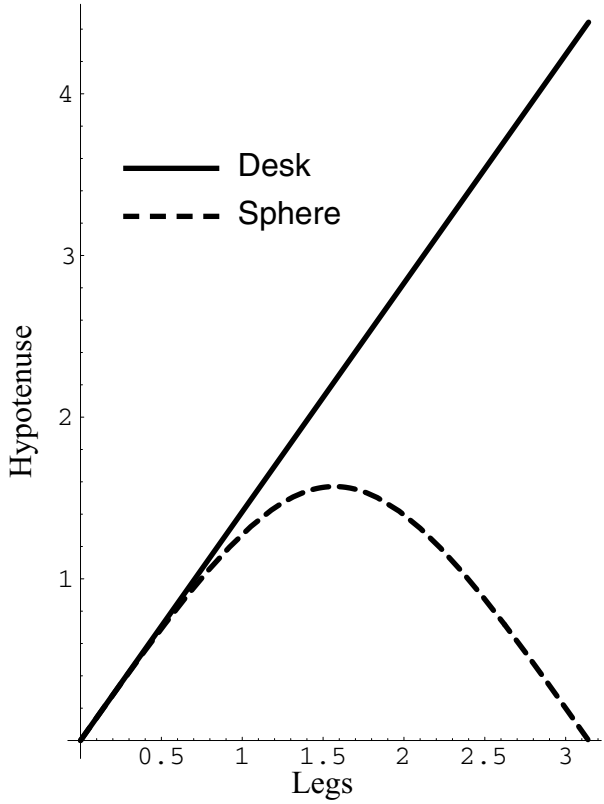


Fig. 1.2. Mathematical solution plotted for $R = 1$.

According to the Pythagorean formula, the hypotenuse should have a length

$$C = \sqrt{2}\, R\theta_0. \tag{1.3}$$

For small values of θ, where $C/R \ll 1$, the Pythagorean rule works fairly well, although not exactly, on the sphere, but eventually the formula fails. We can see how badly it fails in Figure 1.2. This is because the sphere is curved, and the Pythagorean formula only works on flat surfaces. The formula works approximately on the sphere when the distance being measured is small compared to the radius of curvature of the sphere. The mathematical way of saying this is that the sphere is *locally flat*.

But what does this have to do with Einstein's Special Theory of Relativity? In this book we're going to develop the concept of spacetime, but we're only going to study flat spacetime, because that's what special relativity is all about. Everything you will learn in this book will apply to flat spacetime in the same way that the Pythagorean formula applies to flat space. In the real world we experience the force

of gravity, and gravity can only be consistently described in terms of a spacetime that is curved, not flat. But a curved spacetime can be approximated as being flat when the force of gravity is small, or, equivalently, when distance scales being measured are small compared to the radius of curvature of spacetime. So we can learn a lot about the Universe just by studying special relativity and flat spacetime, even though in the strictest sense there is no such thing as a completely flat space or spacetime – these geometries exist as mathematical idealizations, not in the material gravitating world.

Even though the Pythagorean formula is only approximately true, it is true enough at the distance scales accessible to Newtonian physics that all of classical physics depends on it. The mathematical and philosophical revolution that made this possible was the marriage of algebra and geometry in the Cartesian coordinate grid. In 1619 a young philosopher named René Descartes dreamt that an "Angel of Truth" came to him from God with the very Pythagorean message that mathematics was all that was needed to unlock all of the secrets of nature. One outcome of this insight was the description of space in terms of algebraic coordinates on an infinite rectangular grid. If space has two dimensions, the distances between any two points in this grid can be calculated by applying the Pythagorean rule, with the distance L_{12} between the two points P_1 and P_2 given by the length of the hypotenuse of the right triangle whose two legs are the differences Δx and Δy between the x and y coordinates of the two points as projected on the two orthogonal axes of the grid

$$L_{12}^2 = \Delta x^2 + \Delta y^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2. \tag{1.4}$$

The world we know seems to have three space dimensions, but this is no problem because a rectangular coordinate system can be defined just as easily in any number of dimensions. In D space dimensions we can describe each point P at which an object could be located or an event could take place by a position vector $\vec{r}$ representing a collection of D coordinates $(x^1, x^2, \ldots, x^D)$ in a D-dimensional rectangular grid. The distance r_{12} between two points $\vec{r}_1$ and $\vec{r}_2$ is given by the Pythagorean formula generalized to D dimensions

$$|r_{12}|^2 = \sum_{i=1}^{D} (x_1^i - x_2^i)^2. \tag{1.5}$$

Any position vector $\vec{r}$ in this D-dimensional space can be written in terms of the D coordinate components x^i in a basis of D orthonormal vectors $\hat{e}_i$ as

$$\vec{r} = \sum_{i=1}^{D} x^i \hat{e}_i. \tag{1.6}$$

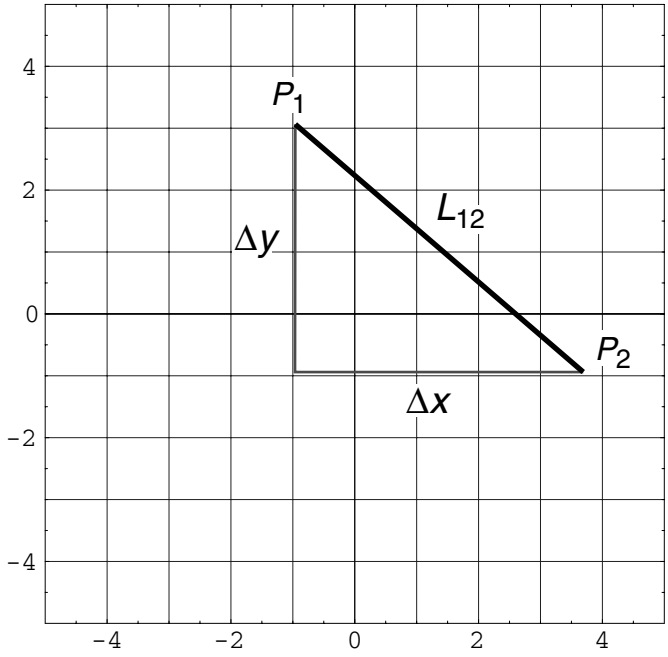


Fig. 1.3. Space as a flat rectangular coordinate grid.

A set of orthonormal basis vectors has the inner product

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij}, \tag{1.7}$$

where δ_{ij} is the Kronecker delta symbol given by the relation

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \tag{1.8}$$

Any other vector $\vec{V}$ in this space can then be represented in this orthonormal basis as

$$\vec{V} = \sum_{i=1}^{D} V^i \hat{e}_i, \quad V^i = \vec{V} \cdot \hat{e}_i, \tag{1.9}$$

where the set of D numbers V^i are said to be the components of the vector $\vec{V}$ in this specific basis. Note that the same vector can be represented in more than one basis. This is an extremely important thing to remember, and we will return to it again and again in this book, in greater and greater detail, because this is one of the basic mathematical ideas behind the principle of relativity, both special and general.

1.2 The differential version, in D dimensions

The Pythagorean formula gives us the length of a straight line between two points in a D-dimensional flat space. In order to calculate the length of a line that isn't straight, we can approximate the line as being made up of an infinite number of tiny straight lines with infinitesimal length dl, each of which satisfies an infinitesimal version of the Pythagorean rule. In two space dimensions we write this as

$$dl^2 = dx^2 + dy^2 \tag{1.10}$$

and in D dimensions it becomes

$$dl^2 = \sum_{i=1}^{D} (dx^i)^2. \tag{1.11}$$

If we use the Kronecker delta function δ_{ij} as defined in Eq. (1.8), and adopt the convention of summing over repeated indices, this expression can be rewritten as

$$dl^2 = \delta_{ij}\, dx^i dx^j. \tag{1.12}$$

In differential geometry this object is called the Euclidean metric in rectangular coordinates. Euclidean space is another name for flat space. A metric is another name for an infinitesimal line element. Using the methods of differential geometry, the curvature of a given space can be calculated from the first and second derivatives of metric functions $g_{ij}(x)$, which replace the constant δ_{ij} if the space has nonzero curvature. If we were to calculate the curvature of the D-dimensional space whose metric is Eq. (1.12), we would find that it is exactly zero, because all of the derivatives of the components of δ_{ij} are zero. But that's a subject for another book.

Now that we have an infinitesimal line element, we can integrate it to find the lengths of lines that are not straight but curved, using the differential version of the Pythagorean formula, also known as the Euclidean metric in rectangular coordinates. If we have some curve C between points P_1 and P_2 then the length ΔL of the curve is given by

$$\Delta L = \int_{P_1}^{P_2} dl. \tag{1.13}$$

A curve in D-dimensional Euclidean space can be described as a subspace of the D-dimensional space where the D coordinates x^i are given by single-valued functions of some parameter t, in which case the length of a curve from $P_1 = x(t_1)$ to $P_2 = x(t_2)$ can be written

$$\Delta L = \int_{t_1}^{t_2} \sqrt{\delta_{ij}\, \dot{x}^i \dot{x}^j}\, dt, \quad \dot{x}^i = \frac{dx^i}{dt}. \tag{1.14}$$

For example, we can calculate the circumference of a circle of radius R in two-dimensional Euclidean space described by $\{x^1 = R\cos t,\ x^2 = R\sin t\}$. In this case,

$$\delta_{ij}\dot{x}^i\dot{x}^j = R^2(\sin^2 t + \cos^2 t) = R^2 \tag{1.15}$$

and

$$\Delta L = \int_0^{2\pi} \sqrt{\delta_{ij}\,\dot{x}^i\dot{x}^j}\,dt = \int_0^{2\pi} R\,dt = 2\pi R. \tag{1.16}$$

Since it's guaranteed that $\delta_{ij}\dot{x}^i\dot{x}^j \geq 0$, formula (1.14) for the length of a curve is always positive as long as the curve itself is well-behaved. This won't continue to be the case when we graduate from space to spacetime.

1.3 Rotations preserve the Euclidean metric

At first glance, a description of space as a rectangular coordinate grid seems like turning the Universe into a giant prison ruled by straight lines that point in fixed directions and tell us how we have to describe everything around us in their terms. We know that, in the real world, we possess free will and can turn ourselves around and look at any object from a different angle, a different point of view. We see that the object looks different from that point of view, but we know it is not a different object, but just the same object seen from a different angle.

Luckily for us, the same wisdom emerges from the mathematics of Euclidean space. We don't have to stick with one rigid coordinate system – we can turn the whole coordinate grid around to see the same object from a different angle. This can be done in any number of dimensions, but for the sake of brevity we will stick with $D = 2$ with the traditional choice $x^1 = x$, $x^2 = y$.

A general linear transformation from coordinates (x, y) to coordinates $(\tilde{x}, \tilde{y})$ can be written in matrix form as

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} + \begin{pmatrix}b_1\\ b_2\end{pmatrix}. \tag{1.17}$$

The constants b_1 and b_2 represent a shift in the origin of the coordinate system. Taking the differential of this expression automatically gets rid of b_1 and b_2, and this reflects the freedom with which we can set the origin of the coordinate system anywhere in the space without changing the metric. This freedom is called translation invariance, and ultimately, as we shall show in a later chapter, it leads to conservation of momentum for objects moving in this space.

If we require that the metric remain unchanged under the rest of the transformation, so that

$$d\tilde{x}^2 + d\tilde{y}^2 = dx^2 + dy^2, \tag{1.18}$$

then it must be true that

$$a_{11}^2 + a_{12}^2 = a_{21}^2 + a_{22}^2 = 1, \qquad a_{11}a_{12} = -a_{21}a_{22}. \tag{1.19}$$

We have three equations for four variables, so instead of a unique solution, we get a continuous one-parameter family of solutions that can be written in terms of an angular parameter θ as

$$a_{11} = a_{22} = \cos\theta, \qquad a_{12} = -a_{21} = \sin\theta. \tag{1.20}$$

This describes a rotation by an angle θ. We also get a second family of solutions

$$a_{11} = -a_{22} = \cos\theta, \qquad a_{12} = a_{21} = \sin\theta, \tag{1.21}$$

which represents a reflection about an axis characterized by θ. Transformations like this will be discussed in more detail in Chapter 8.

Expanding out the matrix multiplication for the solution in (1.20), our coordinate transformation becomes

$$\begin{aligned} \tilde{x} &= x\cos\theta + y\sin\theta \\ \tilde{y} &= -x\sin\theta + y\cos\theta. \end{aligned} \tag{1.22}$$

This transformation is a rotation of the rectangular coordinate system by an angle θ about the origin at $x = y = 0$. To see this, look at the $\tilde{x}$ and $\tilde{y}$ axes, represented by the line $\tilde{y} = 0$ and $\tilde{x} = 0$, respectively. In the (x, y) coordinate system, they become

$$\begin{aligned} \tilde{y} = 0 \;&\rightarrow\; y = x\tan\theta \\ \tilde{x} = 0 \;&\rightarrow\; x = -y\tan\theta, \end{aligned} \tag{1.23}$$

and if we plot this we can see that the $\tilde{x}$ and $\tilde{y}$ axes are both rotated counterclockwise by an angle θ compared to the x and y axes.

As the angle from which we view an object changes, we don't expect the object itself to change. Suppose we have a position vector $\vec{r}$ pointing to some object in two-dimensional Euclidean space. In the (x, y) coordinate system, $\vec{r}$ can be written

$$\vec{r} = x\hat{e}_x + y\hat{e}_y. \tag{1.24}$$

If we require that the position vector itself remain unchanged as we rotate the coordinate system in which the vector is described, so that

$$\vec{r} = x\hat{e}_x + y\hat{e}_y = \tilde{x}\hat{e}_{\tilde{x}} + \tilde{y}\hat{e}_{\tilde{y}}, \tag{1.25}$$

then the transformation of the coordinate components of the vector must cancel the transformation of the unit basis vectors. Therefore the transformation rule for the orthonormal basis vectors must be

$$\begin{aligned}\hat{e}_{\tilde{x}} &= \hat{e}_x \cos\theta + \hat{e}_y \sin\theta \\ \hat{e}_{\tilde{y}} &= -\hat{e}_x \sin\theta + \hat{e}_y \cos\theta.\end{aligned} \tag{1.26}$$

The rotation transformation can be written in matrix form as

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.27}$$

The inverse transformation is a rotation in the opposite direction

$$R(\theta)^{-1} = R(-\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{1.28}$$

The matrix $R(\theta)$ satisfies the conditions $\det R = 1$ and $R^T I\, R = I$, where I is the identity matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.29}$$

The first condition classifies R as a special, as opposed to a general, linear transformation. The second condition classifies R as an orthogonal matrix. The full name of the group of linear transformations represented by $R(\theta)$ is the special orthogonal group in two space dimensions, or $SO(2)$[2] for short. Transformations by $R(\theta)$ take place around a circle, which can be thought of as a one-dimensional sphere, known as S^1 for short.

A transformation matrix can have the properties of being special and orthogonal in any number of dimensions, so rotational invariance is easily generalized to D space dimensions, although the matrices get more complicated as D increases. As one would expect, since we have $SO(2)$ for $D = 2$, we have $SO(D)$ for arbitrary D. A rotation in D space dimensions takes place on a $(D-1)$-dimensional sphere, called S^{D-1} for short. We will examine rotational invariance in D space dimensions in greater detail in a later chapter. For now, everything we want to accomplish in this chapter can be achieved using the simplest case of $D = 2$.

1.4 Infinitesimal rotations

So far this seems like elementary stuff. Why are we looking at rotations in space, when our goal in this book is to learn about spacetime? We will see shortly that the

[2] The group becomes known as $O(2)$ if we include reflections in addition to rotations. For a reflection, the determinant is -1 instead of $+1$.

mathematics of rotational invariance of Euclidean space has a very close analog in the relativistic invariance of spacetime. To build the case for this, and build our first glimpse of spacetime, we need to study infinitesimal rotations, that is, rotations for which θ is close to zero.

A very small rotation with $\theta \sim 0$ can be written:

$$\begin{aligned} \tilde{x} &\simeq x + \theta y + O(\theta^2) \\ \tilde{y} &\simeq y - \theta x + O(\theta^2) \end{aligned} \tag{1.30}$$

In this infinitesimal limit, the rotation matrix $R(\theta)$ can be written in terms of a matrix $\mathbf{r}$

$$R(\theta) \simeq I + \theta \mathbf{r} + \cdots, \qquad \mathbf{r} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.31}$$

which we will call the generator of the rotation transformation. Notice that $\mathbf{r}$ is an antisymmetric matrix, that is $\mathbf{r}^T = -\mathbf{r}$, where the matrix transpose is defined by $(M^T)_{ij} = M_{ji}$.

Written in this manner, an infinitesimal rotation looks like the first two terms in the expansion of an exponential

$$e^{\alpha x} \sim 1 + \alpha x + \cdots \tag{1.32}$$

This is no coincidence. A non-infinitesimal rotation $R(\theta)$ can be obtained from the exponential of the generator matrix $\mathbf{r}$

$$R(\theta) = \exp(\theta \mathbf{r}) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{1.33}$$

as you will be asked to prove in an exercise at the end of this chapter.

Now suppose we want to consider unifying space and time into a two-dimensional spacetime, hopefully something as simple and symmetric as Euclidean space. What would be the analog of rotational invariance in that case? The most obvious difference between space and time is that we can't turn ourselves around to face backwards in time like we can in space. So a rotation transformation presents a problem. Is there a transformation that is like a rotation but without the periodicity that conflicts with the knowledge that we can't rotate our personal coordinate frames to face backwards in time?

An antisymmetric generator matrix will always lead to a rotation when we take the exponential to get the full transformation. Suppose the generator matrix is not antisymmetric but instead symmetric, so that the matrix is equal to its own transpose? Such a generator matrix yields a new transformation that looks almost like a rotation, but is not. Let's call the new generator matrix ℓ and the new

transformation parameter ξ. The new infinitesimal transformation is

$$L(\xi) \simeq I + \xi\ell + \cdots, \qquad \ell = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \tag{1.34}$$

Because of the sign difference in the generator ℓ, the exponential of the generator gives the unbounded functions $\cosh\xi$ and $\sinh\xi$ instead of the periodic functions $\cos\theta$ and $\sin\theta$

$$L(\xi) = \exp(\xi\ell) = \begin{pmatrix} \cosh\xi & -\sinh\xi \\ -\sinh\xi & \cosh\xi \end{pmatrix}. \tag{1.35}$$

The action of $L(\xi)$ on a set of rectangular coordinates axes is not to rotate them but to skew them like scissors. The θ in $R(\theta)$ lives on the interval $(0, 2\pi)$, but the parameter ξ in $L(\xi)$ can vary between $(-\infty, \infty)$. In the limit $\xi \to \pm\infty$, the transformation degenerates and the axes collapse together into a single line, as you will be asked to show in an exercise.

As with the rotation transformation $R(\theta)$, the inverse transformation of $L(\xi)$ is a transformation in the opposite direction

$$L(\xi)^{-1} = L(-\xi) = \begin{pmatrix} \cosh\xi & \sinh\xi \\ \sinh\xi & \cosh\xi \end{pmatrix}. \tag{1.36}$$

This new transformation satisfies the special condition $\det L = 1$. However, the orthogonality condition $R^T I R = I$ is amended to

$$L^T \eta L = \eta, \qquad \eta = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.37}$$

Because of the minus sign, this type of transformation is called a special orthogonal transformation in (1, 1) dimensions, or $SO(1, 1)$ for short. It's going to turn out that this (1, 1) refers to one space and one time dimension. In order to develop that idea further, it's time to bring time into the discussion.

1.5 Could a line element include time?

The journey towards the description of physical space by an infinite Cartesian coordinate grid was a rough one, because enormous spiritual and moral significance was given to the organization of physical space by European culture in the Middle Ages. At one time, even the assertion that empty space existed was considered heresy. By contrast, the question of time was not as controversial. It seemed obvious from common experience that the passage of time was something absolute that was universally experienced by all observers and objects simultaneously in the same way. This didn't conflict with the story of the Creation told in the Bible, so

the concept of absolute time did not present a challenge to devout Christians such as Isaac Newton, who held it to be an obvious and unquestionable truth.

In this old picture of space and time, space and time are inherently separate, and both absolute. An object at a location marked in absolute space by the position vector $\vec{r}$ moves along a path in absolute space that can be written as a function of time such that

$$\vec{r} = \vec{r}(t) = x^i(t)\hat{e}_i, \tag{1.38}$$

where the implied sum over the repeated index i runs over all of the dimensions in the space. For now, let's restrict space to one dimension, so we're only dealing with one function of time $x(t)$.

In Newtonian physics in one space dimension, in the absence of any forces, the motion of an object is determined by the solution to the differential equation

$$\frac{d^2x(t)}{dt^2} = 0. \tag{1.39}$$

This equation is invariant under the transformation

$$\begin{aligned} \tilde{t} &= t \\ \tilde{x} &= x - vt, \end{aligned} \tag{1.40}$$

where v is the velocity of an observer in the $\tilde{x}$ coordinate system relative to an observer in the x coordinate system.

This invariance principle was first proposed by Galileo based on general physical and philosophical arguments, long before Newton's equation existed. Galileo argued that when comparing a moving ship with dry land, the natural laws governing the motion of an object on the ship should not depend on the motion of the ship relative to the dry land, as long as that motion was smooth motion at a constant velocity.

This sounds suspiciously like what we learned about Euclidean space and rotations – an object should not change as we rotate the coordinate system used to describe it. But now we're not talking about rotating space into space, we're talking about space and time, and instead of a rotation, which involves a dimensionless angle, we have motion, which involves the dimensionful quantity of velocity.

Let's consider the possibility that Galilean invariance could be an infinitesimal version of some full spacetime invariance principle, a version valid only for velocities close to zero. Suppose we use as our candidate spacetime transformation the $SO(1, 1)$ transformation developed in the previous section.

At first it would appear that Galilean invariance is not consistent with a small rotation by $L(\xi)$. However, the units by which we measure space and time are not the same. Time is measured in units of time such as seconds or years, and space

is measured in units of length such as feet or meters. If we want to compare a Galilean transformation to a rotation by $L(\xi)$, we should scale the coordinate t by some dimensionful constant c with units of length/time, so that $\tau = ct$ has the dimension of length

$$\tau = ct, \qquad [\tau] = L \to [c] = L/T. \tag{1.41}$$

If we took the transformation $L(\xi)$ literally as a kind of rotation in two-dimensional spacetime with time coordinate τ and space coordinate x, then this rotation would change coordinate components (τ, x) to $(\tilde{\tau}, \tilde{x})$ by

$$\begin{aligned} \tilde{\tau} &= \tau \cosh\xi - x \sinh\xi \\ \tilde{x} &= -\tau \sinh\xi + x \cosh\xi. \end{aligned} \tag{1.42}$$

The Galilean transformation contains a parameter v with units of velocity. Using v and c, we can define a dimensionless parameter $\beta \equiv v/c$. If we assume that $\xi \simeq \beta$ then an infinitesimal rotation by $L(\xi)$ for small ξ looks like

$$\begin{aligned} \tilde{\tau} &\simeq \tau - \beta x \\ \tilde{x} &\simeq x - \beta\, \tau. \end{aligned} \tag{1.43}$$

So far, this looks nothing like a Galilean transformation. However, maybe one of the terms is smaller than the others and can be neglected. If we rewrite the above equations back in terms of t and c then we get:

$$\tilde{t} \;\simeq\; t - \frac{vx}{c^2} \;\simeq\; t, \qquad c \to \infty \tag{1.44}$$

$$\tilde{x} \;\simeq\; x - vt. \tag{1.45}$$

The infinitesimal limit $\beta \to 0$ corresponds to the limit $c \to \infty$. In this limit the second term in the first equation can be safely neglected because it's much smaller than the others. So our postulated spacetime rotation $L(\xi)$ for very small values of $\xi \sim \beta$ does appear to be consistent with a Galilean transformation.

So here comes the big question: What line element is left invariant by the spacetime rotation $L(\xi)$? What is the spacetime analog of the differential version of the Pythagorean theorem? One can show, as you will in an exercise, that the line element

$$ds^2 = -d\tau^2 + dx^2 \tag{1.46}$$

is invariant under the coordinate transformation given by $L(\xi)$ in (1.42) such that

$$-d\tau^2 + dx^2 = -d\tilde{\tau}^2 + d\tilde{x}^2. \tag{1.47}$$

The metric (1.46) gives us at last the analog of the Pythagorean rule for spacetime. If curvature comes from derivatives of the metric, then this strange metric with a minus sign must be flat. This metric for flat spacetime is called the Minkowski metric, and flat spacetime is also known as Minkowski spacetime. Note that unlike the Euclidean metric, the Minkowski metric is not positive-definite. This is extremely important and we will explore the implications of this fact in greater detail in later chapters.

What happens for higher dimensions? A flat spacetime with D space dimensions and one time dimension has a line element

$$ds^2 = -d\tau^2 + dl^2, \qquad dl^2 = \delta_{ij} dx^i dx^j. \tag{1.48}$$

The Latin indices (i, j) refer to directions in space, and by convention take the values $(1, 2, \ldots, D)$. When dealing with spacetime, it has become the convention to use a set of Greek indices (μ, ν) and appoint the 0th direction as being the time direction so that $dx^0 = d\tau$, in which case the Minkowski metric can be written

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \tag{1.49}$$

where

$$\begin{aligned} \eta_{00} &= -1, \\ \eta_{0i} = \eta_{i0} &= 0, \\ \eta_{ij} &= \delta_{ij}. \end{aligned} \tag{1.50}$$

In two spacetime dimensions the metric is invariant under an $SO(1, 1)$ transformation of the coordinates. As one might suspect, in $d = D + 1$ spacetime dimensions, the transformation is called $SO(1, D)$ or $SO(D, 1)$. In either case, this indicates that it pertains to D space dimensions and one time dimension. (We never consider more than one time dimension!)

1.6 The Lorentz transformation

The transformation $L(\xi)$ is known as the Lorentz transformation, the invariance principle is called Lorentz invariance and the transformation group $SO(1, D)$ is known as – no surprise here – the Lorentz group, which we'll study in greater detail in a later chapter. But we don't yet know the relationship between the Lorentz transformation parameter ξ and the velocity parameter $\beta \equiv v/c$ that gives the relative velocity between the two coordinate systems in question.

Suppose we have an observer who is at rest in the (τ, x) coordinate frame in a flat spacetime with metric (1.46). The $(\tilde{\tau}, \tilde{y})$ coordinate frame is moving at velocity β relative to the (τ, x) frame. (Note: Even though β is a dimensionless parameter

proportional to the velocity, for the sake of brevity we shall refer to it as the velocity.) Therefore an observer measuring space and time in the $(\tilde{\tau}, \tilde{y})$ coordinate system sees the observer in the (τ, x) frame not as being at rest, but moving with velocity $-\beta$ so that

$$\frac{d\tilde{x}}{d\tilde{\tau}} = -\beta. \tag{1.51}$$

Since the observer in question is at rest in the (τ, x) coordinate frame, for her/him $dx = 0$ and therefore

$$ds^2 = -d\tau^2. \tag{1.52}$$

But according to the $(\tilde{\tau}, \tilde{y})$ coordinate system,

$$ds^2 = -d\tilde{\tau}^2 + d\tilde{x}^2 = -d\tilde{\tau}^2 + \beta^2 d\tilde{\tau}^2. \tag{1.53}$$

Because the two coordinate systems (τ, x) and $(\tilde{\tau}, \tilde{y})$ differ only by a Lorentz transformation, and because we're in flat spacetime, where the metric is Lorentz-invariant, it must be true that

$$d\tilde{\tau} = \frac{d\tau}{\sqrt{1-\beta^2}}. \tag{1.54}$$

Taking the differential of the Lorentz transformation relating the two frames in Eq. (1.42) gives

$$\begin{aligned} d\tilde{\tau} &= d\tau \cosh\xi - dx \sinh\xi \\ d\tilde{x} &= dx \cosh\xi - d\tau \sinh\xi, \end{aligned} \tag{1.55}$$

so the Lorentz transformation parameter ξ is related to the velocity β by

$$\cosh\xi = \gamma, \qquad \sinh\xi = \gamma\beta, \qquad \gamma \equiv \frac{1}{\sqrt{1-\beta^2}}. \tag{1.56}$$

The Lorentz transformation $L(\xi)$, rewritten in terms of γ and β, becomes

$$L(\beta) = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}. \tag{1.57}$$

Something very strange and interesting has happened. At first it seemed from the infinitesimal transformation (1.43) that we would end up with $\xi = \beta$, and the velocity β would live in the interval $(-\infty, \infty)$. However, the Lorentz transformation is only real and finite for

$$-1 < \beta < 1, \tag{1.58}$$

which means that

$$-c < v < c. \tag{1.59}$$

So the velocity c, at first introduced only to create a dimensionally balanced coordinate transformation, ends up being the maximum allowed velocity in the spacetime.

It's probably no secret that this maximum velocity c imposed by the geometry of flat spacetime is the speed of light. But that association is something physical that can't be proven by geometry alone. To show that c is the speed of light, we need to appeal to the physics of light, which is the subject of the next chapter.

Exercises

1.1 Verify that the rotation $R(\theta)$ leaves the Euclidean line element dl^2 invariant.

1.2 Let's look at a two-sphere of radius R whose center is at the origin of three-dimensional flat Euclidean space, with coordinates related by

$$(x, y, z) = (R\sin\theta\cos\phi, R\sin\theta\sin\phi, R\cos\theta). \tag{E1.1}$$

Consider the three great circles passing through the north pole $\vec{x}_0 = (0, 0, R)$ and the points $\vec{x}_1 = (R\sin\theta_0, 0, R\cos\theta_0)$ and $\vec{x}_2 = (0, R\sin\theta_0, R\cos\theta_0)$ on the sphere. These three circles define the right triangle on the sphere described at the beginning of the chapter. Recall that the Euclidean dot product of two vectors $\vec{x}_i \cdot \vec{x}_j = |x_i||x_j|\cos\theta_{ij}$, where θ_{ij} is the angle between the two vectors in the two-dimensional plane they determine.

(a) Use this result to verify that the arc length of the hypotenuse of this right triangle is given by Eq. (1.2).

(b) Expand Eq. (1.2) for small θ_0 to check whether the Pythagorean rule is obeyed in that limit.

1.3 Using the transformation rule for basis vectors, compute the components of the vector $\vec{V}$ in the $(\tilde{x}, \tilde{y})$ coordinate system of Eq. (1.17).

1.4 On a sheet of graph paper, represent the (x, y) coordinate system as a two-dimensional rectangular grid, with y on the vertical axis and x on the horizontal axis. Using the transformation $R(\theta)$ to relate the two coordinate systems $(\tilde{x}, \tilde{y})$ and (x, y), plot the two lines $\tilde{x} = 0$ and $\tilde{y} = 0$ in the (x, y) coordinate system for the values $\theta = \pi/4, \pi/2, 3\pi/4$ and π. Then draw another coordinate grid with $(\tilde{x}, \tilde{y})$ on the axes, and plot the two lines $x = 0$ and $y = 0$ in the $(\tilde{x}, \tilde{y})$ coordinate system for the same values of θ.

1.5 On a sheet of graph paper, represent the (τ, x) coordinate system as a two-dimensional rectangular grid, as you did above for the (x, y) coordinate

system, but with τ replacing y on the vertical axis, and x on the horizontal axis. Using the transformation $L(\xi)$ to relate the two coordinate systems (τ, x) and $(\tilde{\tau}, \tilde{x})$, plot the two lines $\tilde{\tau} = 0$ and $\tilde{x} = 0$ in the (τ, x) coordinate system for $\xi = 1/3, 1/2, 1$ and 2. Then draw another coordinate grid with $(\tilde{\tau}, \tilde{x})$ on the axes, and plot the two lines $\tau = 0$ and $x = 0$ in the $(\tilde{\tau}, \tilde{x})$ coordinate system for the same values of ξ. Using (1.56), calculate β and γ for $\xi = 1/3, 1/2, 1$ and 2.

1.6 Given the 2×2 antisymmetric matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{E1.2}$$

compute the first four terms in the Taylor expansion of the exponential $e^{\theta A}$ around $\theta = 0$ and derive a general formula for the elements of $e^{\theta A}$ as an infinite sum of powers of θ.

1.7 Given the 2×2 symmetric matrix

$$S = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \tag{E1.3}$$

compute the first four terms in the Taylor expansion of the exponential $e^{\xi S}$ around $\xi = 0$ and derive a general formula for the elements of $e^{\xi S}$ as an infinite sum of powers of ξ.

1.8 Multiply two rotation matrices $R(\theta_1)$ and $R(\theta_2)$. Is the result a third rotation matrix $R(\theta_3)$? If so, what is the angle θ_3 of the resulting transformation in terms of θ_1 and θ_2?

1.9 Multiply two Lorentz transformation matrices $L(\xi_1)$ and $L(\xi_2)$. Is the result another Lorentz transformation matrix $L(\xi_3)$? If so, what is the transformation parameter ξ_3 of the resulting matrix in terms of ξ_1 and ξ_2?

2

Light surprises everyone

Hands-on exercise: wave and particle properties

The purpose of this exercise is for you to observe some basic wave and particle properties. To complete this exercise you will need the following:

- Tub of water, or access to a quiet pond, lake or swimming pool.
- Things to float on the surface of the water.
- Pen or pencil and some drawing paper.
- Small projectile such as a stone.

Disturb the middle of the tub just until you are able to make a visible wave on the surface. Watch how the wave propagates. Wait until the surface of the water returns to being flat and make another wave. Keep doing this as many times as necessary to be able to draw what you see on the paper and answer the following questions:

- Does the wave have a definite location at any one moment in time?
- Does the wave have a definite direction as it propagates?
- Approximately how far does the wave travel in 1 s?
- Describe the motion of the water in which the wave moves.

Throw your small projectile in the air at various angles, letting it drop down (not in the water). Keep doing this as many times as necessary to be able to draw what you see and answer the following questions:

- Does the object have a definite location at any one moment in time?
- Does the object have a definite direction as it propagates?
- Approximately how far does the object travel in 1 s?
- Describe the motion of the air in which the object moves.
- Is there any difference between the vertical and horizontal motion of the object?

2.1 Conflicting ideas about space and light

What we call known physics today, what we learn in school or teach ourselves from books, at one time was the unknown. It was what people did not understand, and sought to understand through exhausting and often frustrating intellectual labor. In the process of moving from the unknown to the known, a lot of wrong ideas can come up that seem very right at the time. The story of the classical understanding of space and light from Aristotle to Einstein is a story in which almost everyone involved was both right and wrong at the same time.

Aristotle started with an idea that seems right enough – *Nature abhors a vacuum* – and used it to argue that empty space could not exist, period. Every last tiny space in the Universe was filled with a universal substance, which later came to be called the *ether*. According to Aristotle, the space taken up by a material object was the surface area, not the volume, of the object. Using Aristotle's logic, the amount of space taken up by a round ball of radius R would be $4\pi R^2$ rather than $4\pi R^3/3$. Aristotle was such a powerful figure in Western culture that it took until the fifteenth century for his argument against spatial volume to be refuted. But even so, Aristotle's argument that empty space could not exist formed the root of the wrong understandings of both space and light that troubled classical physics until Einstein showed up with a brilliant idea that put the controversies to rest, at least until quantum theory showed up.

Even though Descartes' work on analytic geometry laid the mathematical foundation for the Newtonian description of physical space as an empty and absolute backdrop for the actions of matter, Descartes shared Aristotle's abhorrence of the void. Descartes believed that a type of material substance called the *plenum* must fill the entire universe, down to every nook and cranny, and that vortices swirling in this fluid were what moved the planets in their orbits. Descartes believed that light was an instantaneous disturbance in the plenum between the observer and the observed. He believed so strongly that light propagates instantaneously that he swore that if this were ever proved false, he would confess to knowing absolutely nothing.

Newton learned geometry by reading Descartes, but he inserted into physics his own belief that space was both empty and absolute. To Newton, a devout Christian, to question the absoluteness of space was to question the absoluteness of God – not merely an intellectual error, but an actual sin against God. In Newton's Universal Law of Gravitation, the force between two gravitating objects varies inversely as the square of the distance between them, and the distance between gravitating objects is treated as an empty space devoid of any intervening substance. Newton's powerful and concise theory was a huge success in explaining all of known astronomy at the time, but Newton's many critics rightly complained that the Universal

Law of Gravitation provides no mechanical means for transmitting the gravitational force between bodies, other than the literal hand of God – a conclusion that Newton was not unsatisfied with himself.

Field theory had not been invented yet. This was still the age of mechanics. Things happened because one thing pushed or pulled on another. Pushing or pulling is not something that anyone envisioned could be done across empty space. If there is no material substance filling the space between the planets, then how would one planet sense the introduction or removal of any other planet?

The lack of any causal mechanism of force transmission in Newton's theory of gravity led many of his contemporaries to label his theory as nonsense. One such person was Dutch physicist Christian Huygens, who, like Newton, learned his vocation by reading Descartes. Huygens, however, was appointed by fate to be the undoing of his own master. It was Huygens who made the first numerical estimate of the speed of light and proved Descartes wrong about the very thing Descartes was the most certain he was right.

In 1667 Galileo had tried to measure the speed of light using lanterns and mountain tops but he never had a chance, because light travels too fast for the time interval in question to have been measured by any existing timekeeping device. The speed of light is 3×10^8 m/s. At such an enormous speed, light only needs 8 min. to cross the 1.5×10^{11} m from the Sun to the Earth. If we can only measure time to within a few seconds, then in order to measure the speed of light, we have to observe light propagation over the distance scale of the solar system. And this is how the first successful measurement of the speed of light was made in 1676.

Danish astronomer Ole Roemer, who spent 10 years making careful observations of the orbital periods of Jupiter's moon Io, was quite surprised when the period he observed seemed to fluctuate with the distance between Jupiter and Earth, with the period being longer when Jupiter and Earth were moving farther apart. In 1676 he announced that this discrepancy could only come from the time it took light to travel from Io to the Earth. Two years later Huygens provided a numerical estimate for this speed of 144 000 miles/s. Huygens had proved that his hero Descartes was wrong about light. Luckily for Descartes, he didn't live long enough to have to fulfill his promise to confess to knowing absolutely nothing.

Huygens took the finite speed of light as evidence for his wave theory of light. In his treatise on optics *Le traité de la lumière*, he put forward his model of wave propagation that physics students now learn as Huygens' Principle: given a particular wave front, each point on that wave front acts as the source point for a spherical secondary wave that advances the wave front in time. Huygens' Principle is illustrated in Figure 2.1.

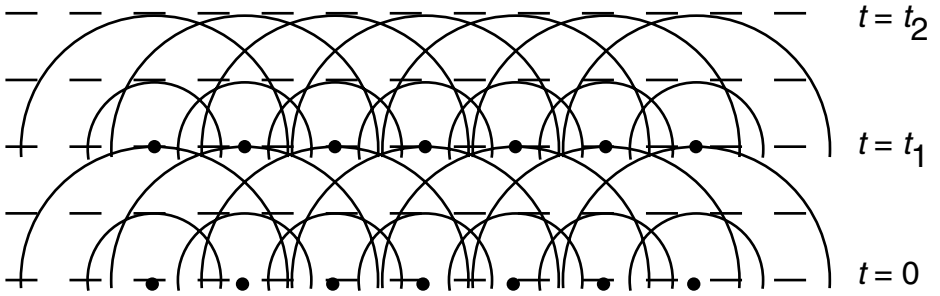


Fig. 2.1. According to Huygens' Principle, each point on a wave front acts as a source point for a spherical secondary wave that determines the wave front at some later time. The dashed lines represent a wave front advancing in time t.

Huygens' optics of wave fronts made little impact in his own time, for he was living in the age of Isaac Newton, Superstar. Consistent with his belief in absolute empty space, Newton envisioned light as a swarm of particles he called "corpuscles" moving through empty space, each corpuscle moving at a different speed depending on the color of the light it represented. Newton had a corpuscular explanation for refraction that was wrong, but the error was not experimentally measurable at the time. Snell's Law of Refraction[1] states that if a light ray is incident on the interface between two transparent media, the angle θ_1 made by the incident ray with the normal to the plane of the interface and the angle θ_2 made by the refracted ray are related through the formula

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}, \tag{2.1}$$

where v_1 and v_2 are the speeds of light in the two transparent media. Snell's law is illustrated in Figure 2.2.

Newton wrongly argued that the ray angle would be smaller in the medium where the speed of light was the largest. The debate over Huygens' wave theory and Newton's particle theory could have been settled over refraction alone, except for one problem: the best known value for the speed of light in air was still off by 25 percent, so they had no hope of being able to measure the difference between the speed of light in air and its speed in water. When this was done in 1850 by Foucault, Newton's theory was conclusively ruled out and Huygens was vindicated.

Despite the fact that he was wrong, Newton made enormous contributions to optics. In 1669 he built the first reflecting telescope, which used a curved mirror instead of a lens, and revolutionized astronomy. Newton ground the mirror himself. When Newton published his book *Opticks* in 1704, it created a scientific and

[1] Also known as Descartes' Law.

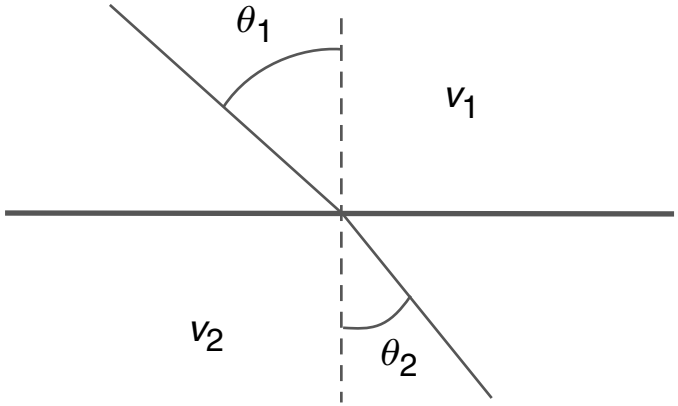


Fig. 2.2. Snell's Law of Refraction describes the bending of a ray of light when it travels from a medium in which the speed of light is v_1 into a medium in which the speed is $v_2 \neq v_1$.

a popular sensation all over England and Europe. Voltaire published a popularization of Newton's work, and discussions of Newtonian optics were all the rage at amateur science and philosophy clubs throughout the educated upper and middle class.

No matter how popular it became and how it revolutionized astronomy, Newtonian optics could not explain the phenomenon of diffraction, where light appears to bend around the edges of objects. Diffraction was first observed in 1665 in Italy by Father Francesco Grimaldi, but both Newton and Huygens regarded it as irrelevant to the wave vs. particle debate. Diffraction could not be ignored forever, however, and eventually wave optics had to be brought back into the picture.

In 1746 in his book *Nova theoria lucis et colorum* (*New Theory of Light and Colour*), Swiss mathematician Leonhard Euler advanced the notion that light consists of wavelike vibrations in the ether. Euler argued that light propagates in the ether, just as sound propagates in the air. Unfortunately, Euler's theory was ultimately no better than Newton's in explaining diffraction. A very bright eye doctor named Thomas Young refined Euler's wave theory to make it consistent with Huygens' Principle. Young and Augustin Fresnel finally proved through their understanding and demonstrations of interference and diffraction that light definitely had wavelike properties that could not be explained by tiny corpuscles flying through space.

2.2 Maxwell's transverse undulations

A real understanding of the nature of light required an understanding of electromagnetism, and that took 100 years to happen, if we start counting from Newton.