

Linear Algebra for Physicists

Flip Tanedo

flip.tanedo@ucr.edu 

Department of Physics & Astronomy, University of California, Riverside, CA 92521

Abstract

Lecture notes for Physics 17, a course on linear algebra in preparation for upper-division undergraduate physics coursework at UC Riverside.

Contents

1 Logistics	4
1.1 Two powerful questions	4
1.2 Exercises and Examples	4
1.3 Obvious-ness	4
2 Motivation	5
3 Basics	6
3.1 Pre-conceptions	6
3.2 Real Three-Vectors	6
3.3 Vectors and Numbers	7
3.4 Notation: Indices	8
3.5 Arithmetic and linear combinations	9
3.6 Rescaling: multiplication by a number	10
3.7 Linear Combination and Span	11
3.8 Basis vectors: an illustrative example	12
3.9 Basis vectors, formally	14
3.10 The meaning of column vectors	17
3.11 Operations that are not (yet) allowed	18
3.12 Euclidean three-space	19
3.13 Length in Euclidean three-space	21
3.14 Angles in Euclidean three-space	22
4 Index Notation and Summation Convention	23
5 Matrices and Linear Transformations	27
5.1 Jargon	27
5.2 Linear transformations	28
5.3 Linear Transformations in One Dimension	29
5.4 Linear Transformations in Two and Three Dimensions	29
5.5 The trivial transformation: identity matrix	30
5.6 Composition of linear transformations	31
5.7 Linearity and Basis Vectors	33
5.8 Linear transformations as maps acting on basis vectors	36
5.9 Inverse transformations	37

5.10	Matrices that aren't invertible: Projections	38
5.11	Funny things you can do to matrices	40
5.12	Matrices that aren't square: maps between vector spaces	42
6	Rotations	44
6.1	Rotations in Two-Dimensions	44
6.2	Transformation of Vectors	46
6.3	Transformation of numbers	46
6.4	Transformation of Row Vectors	47
6.5	Transformation of Matrices	48
6.6	Transformation of Tensors	49
6.7	Rotations in Three-Dimensions	51
6.8	Invariance of Trace and Determinant	52
6.9	Exponentiation of Generators	53
6.10	Active versus Passive	53
7	Tensors as (Multi-)Linear Maps: Tensors	55
7.1	Products of Vector Spaces	55
7.2	Linear Maps to Numbers: Row Vectors	56
7.3	Tensors as Multi-Linear Maps	56
7.4	Duality	59
7.5	A basis for the dual vector space	59
7.6	A basis for multi-linear maps	60
8	Ket Notation	61
8.1	Basis Bras and Kets	62
8.2	Action of bras on kets	63
8.3	Matrices and Tensors	63
8.4	Ket Tricks	63
9	Metric Spaces: Dot Products	64
9.1	Example Polar Coordinates	66
9.2	Length	72
9.3	Orthogonality	73
9.4	Orthonormal Basis	74
9.5	The tensor structure of the metric	75
9.6	Machine to create row vectors	75
9.7	Example: Projection Matrices	77
9.8	Inverse Metric	80
9.9	Isometries: A Fancy Name For Rotations	81
9.10	Discrete isometries	83
9.11	The meaning of indices	85
9.12	Transpose, Adjoint	86
9.13	Complex metric spaces	89
9.14	Gram–Schmidt	91

10 Special Relativity: Minkowski Space	92
10.1 Conventions	92
10.2 Principles of special relativity	94
10.3 Minkowski Space and its Isometries	95
10.4 Example: Muon Decay	96
10.5 Example: What does someone else measure?	98
10.6 Example: the electromagnetic field strength	100
10.7 Example: simultaneity	101
10.8 Example: discrete isometries	101
11 Central Dogmas	102
11.1 Vectors are not columns	102
11.2 Everything is a vector space, everything is a linear map	103
11.3 What the inner product brings	103
11.4 Basis Dependence	104
11.5 A multitude of conventions	105
11.6 The pantheon of nice matrices	106
12 Eigensystem	109
12.1 Self Adjoint (Hermitian) Matrices	109
12.2 Finding Eigenvalues: Characteristic Equation	109
12.3 Finding Eigenvectors	111
12.4 Rotating to the Eigenbasis	114
12.5 Intuition for a linear transformation in the eigenbasis	116
12.6 Degenerate Eigenvalues	117
12.7 Inverting Nice Matrices	124
13 Quantum Mechanics	126
14 Function Space	127
14.1 Functions as Vectors	127
14.2 Histogram Space	129
14.3 Derivatives as Linear Transformations	130
14.4 Derivatives in other function space bases	137
14.5 Locality	137
14.6 Differential Operators	138
14.7 Inner Product	139
14.8 Adjoint	141
14.9 Cracks in the scaffolding as we go to the continuum	144
14.10 Why histogram space is dumb	147
14.11 Legendre Polynomials: salvaging the polynomial basis	148
14.12 Completeness in Function Space	149
14.13 Orthonormality in Function Space	151
14.14 Fourier Series	152
14.15 The Variational Principle	153
A Things we did not get to	156

1 Logistics

1.1 Two powerful questions

At any time in this course, you should feel comfortable asking either of the following questions:

1. Is it obvious that...?
2. Why is this significant?

The first question is the way to ask for on-the-spot clarification—I will either appreciate that I did not properly explain a subtle point, *or* I will explain the intuition¹ for why something should be obvious. The second question is a way to remind me that I may have *lost sight of the forest for the trees*: I want this course to *mathematically connect big ideas in physics*. Asking this question is a reminder that making those connections justifies the hard mathematical work we will put into the course.

1.2 Exercises and Examples

I have tried to insert exercises and examples in these notes. There are still far too few for sound pedagogy. If you really, really want to learn something, you *have* to do exercises. Think of the examples as exercises with solutions—though they are not always written this way. Mull over the exercises: ask yourself why the problems are posed the way they are, challenge the statements to find the domain of validity, think of how one may extend those exercises to other applications. The exercises are a far better gauge of your learning than whether or not you have read a section of the notes. If you are confused reading the text in section 10, it is often the case that you should have been doing the exercises since section 5.

1.3 Obvious-ness

Finally, I want to comment on the word *obvious*. I write this often. It is somewhat dangerous because it can come off as being arrogant: *this is so obvious to me, if you do not understand you must be deficient*. This is never the reason why I use that word. Instead, the word *obvious* serves a very practical purpose. The goal of this class is not just to be able to “do stuff” (e.g. diagonalize a symmetric matrix), but to also build that intuition that comes from a deeper understanding how the mathematics works. In this sense, every time I write the word *obvious* it is a flag: I am saying something that—with the proper perspective—should be self-evident. If it is not self-evident, then you should stop to interrogate why it is not self-evident. Most likely there is something where a change in perspective may (1) make it obvious, and (2) in so doing deepen your understanding of the subject. So when you see the word ‘obvious,’ I want you to do a quick check to confirm whether or not the statement is indeed obvious. If it is not, then welcome the opportunity to learn.²

¹Your sense of mathematical and physical intuition is incredibly valuable. This is one of the key traits that makes a physics training unique.

²There is, of course, the possibility that what I have written is *not* obvious. For example, if I have made a typo... in which case, please let me know.

2 Motivation

Here are three incredibly significant equations in physics:

$$\mathbf{F} = m\mathbf{a} \quad R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G_N}{c^4}T_{\mu\nu} \quad \hat{H}|\Psi\rangle = E|\Psi\rangle . \quad (2.1)$$

These are Newton's force law, Einstein's field equations, and the Schrödinger equation. They govern classical physics, general relativity, and quantum theory.

Each equation looks rather unique: they seem to each be speaking their own mathematical language. Newton's law is written with boldfaced vectorial quantities $\mathbf{F} = (F_x, F_y, F_z)^T$ that should look very familiar to any physics undergraduate. Einstein's equation has these funny μ and ν indices on every term—have you seen these before? Do they look intimidating? If you ever want to make your equations look “technical” and “physicsy,” you should dress them up with indices. The Schrödinger equation has no indices, but instead has these funny angle-brackety things... and that \hat{H} looks suspicious. Where did H get a hat, and what is the content of this equation other than $\hat{H} = E$?

Each of these equations turns out to be a “vectorial” equation. Each one is actually shorthand for a number of equations. Newton's equation is shorthand for three equations, one for each component. Einstein's equation is shorthand for 16 equations, one for each combination of the indices μ and ν that run over four values³. The Schrödinger equation is shorthand for an *infinite* number of equations, one for each allowed energy of a quantum system.

The mathematical formalism that unifies these different ideas (and notations) of ‘vector’ is called linear algebra. It may sound humble: after all, “linear” systems are *easy*, aren't they? Did we not just spend years of our lives learning fancy things like *calculus* and *differential equations* to deal with functions that are more complicated than *lines*? In some sense, yes: linear algebra is about lines and planes in different numbers of dimensions.⁴ However, linear algebra turns out to be far more richer than what you may be used to from high school.

In this course we will see how the three equations in (2.1) are connected by the mathematics of linear algebra. We will dive into the different notation and shamelessly pass between \mathbf{v} , v^i , and $|v\rangle$ to describe the same abstract vector. We will connect to the mathematical description of *symmetry* and see how it is an underlying theme in our descriptions of nature. And we will do all of this in a way that will make the instructors of the linear-algebra-for-mathematicians course and linear-algebra-for-engineers course vomit a little in disgust. Consider that one of privileges of being a physicist.

³The four values are the three directions of space and one direction of time.

⁴On the other hand: a good chunk of the calculus that we do is also implicitly linear. Physicists often Taylor expand and keep only the $\mathcal{O}(\varepsilon)$ term. Integration boils down to summing trapezoids whose angley-bits are given by the first derivative of a function... the linear component.

3 Basics

3.1 Pre-conceptions

If this were a mathematics course, then we would start by very carefully defining words like *vector* and *matrix*. As a physics student, you already have a working definition of these words. It is probably something like this:

A vector has a magnitude and a direction. We write a vector as an array of three numbers arranged in a column. A matrix is an array of nine numbers arranged in a 3×3 block. There is a rule for how to apply (multiply) the matrix to the vector to produce a new vector.

The problem is that you already know too much to learn linear algebra as a mathematics student. You have already seen the tip of the iceberg and so have preconceptions about what vectors are and how they work. You may remember from freshman mechanics that forces are vectors. So are momenta and velocities. You may also recall the idea of a force field—like the electric field—which is actually a whole bunch of vectors: one for each point in space. Examples of matrices are a little more subtle: you may recall that you can represent rotations as matrices. Speaking of rotations, there was another thing that showed up called the moment of inertia *tensor*. It looked like a matrix, but we never called it the “moment of inertia matrix.” What the heck is a tensor, anyway?

And so, you see that starting this course like a mathematics course could cause trouble. The mathematics professor would start by defining a vector. That definition will say nothing about magnitudes or directions, and will not even say anything about arrays of numbers. That definition will clash with the hard-earned intuition that you built from your physics education thus far. It will be perplexing, and may make you feel rather unhappy. What do these mathematicians know, anyway? Or maybe its the physics that is wrong, or have we just completely misunderstood everything and we are just now noticing that we are hopelessly lost? We begin to spiral into a black hole of confusion.

Fortunately, *this is not a mathematics course*.

As a consequence, we will not give a rigorous definition of a vector. We start with a familiar definition of vectors and lay out which qualities are general, and which properties are specific. Then we will come to appreciate the approximation that “*everything is a vector*.” So let us start with something comfortably familiar, even though it constitutes only the simplest example of a vector.

3.2 Real Three-Vectors

Let us write \mathbf{v} to be a vector. This is a standard convention for writing a vector. In this course we will use a few different notations for vectors according to convenience. Notation is neither physics nor mathematics, it is simply a shorthand for a physical or mathematical idea.

In fact, let us focus on a particular type of vector: **real three-vectors**. These are the familiar vectors that we can write as a column of three numbers that effectively represent the

coordinates in three-dimensional space:

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (3.1)$$

where x , y , and z are real numbers. These numbers are called the **components** of the vector \mathbf{v} .

Exercise 3.1. *There is something very perverse about this “vector.” The variable names x , y , and z imply that \mathbf{v} is something that physicists like to call a “position vector.” If you say this to a mathematician they will vomit. By the end of this course, you should appreciate why the notion of a position vector makes no sense. Hint: You may have some intuition for this already: a velocity vector tells you about the instantaneous motion of a particle relative to its present position. Try to write the analogous statement for a “position vector.^a”*

^aI am not a mathematician, but you see that even I have to write “position vector” in condescending quotation marks. In lecture I use even more condescending air quotes.

This three-dimensional space is called [three-dimensional] **real space** and we write it as \mathbb{R}^3 . This is because a vector is an element of three-dimensional real space specified by *three* real numbers.

Three-dimensional real space is an example of a **vector space**, which is just a stupidly formal way of saying that it is where vectors live. Vectors are *elements* of a vector space. A vector space is the set of all possible allowed vectors of a given type. For \mathbb{R}^3 , the vector space is composed of all possible triplets of real numbers.

Example 3.1. *It should be no surprise that we can imagine real two-dimensional space, \mathbb{R}^2 . This is a vector space where each vector may be written as two real numbers. You can also imagine writing real four-dimensional space, \mathbb{R}^4 , or complex two dimensional space, \mathbb{C}^2 .*

From the above example, you should have some intuition for what the **dimension** of a vector space means: the dimension counts how many numbers you need to specify a vector. For real vector spaces, \mathbb{R}^d , the dimension is the number d . We will always assume that d is a positive integer.⁵

3.3 Vectors and Numbers

We should be clear that there are now two different kinds of objects: *vectors* and *numbers*. We will have all sorts of notation for vectors, but let us write them with a boldfaced Roman

⁵The notion of a non-integer-dimensional space does show up occasionally. These do not even have to be particularly exotic: you can look up the dimension of a fractal.

	Vectors	Numbers
Addition (commutative, associative)	✓	✓
Additive null element	0	0
Additive inverse element	$\mathbf{v} + (-\mathbf{v}) = 0$	$a + (-a) = 0$
Multiplication of two of these objects	✗	✓
Multiplication by a number (distributive)	✓	✓ (same as above)
Collection of all allowed objects (space)	vector space	field (“numbers”)
Example of a space	\mathbb{R}^3	\mathbb{R}

Table 1: What you can do with vectors compared to numbers. The glaring difference is that we cannot multiply two vectors. We will need to invent additional mathematical structure to define vector multiplication.

letter for now, e.g. \mathbf{v} . We typically write numbers as lowercase italicized Roman letters like a or sometimes Greek letters like α . These two types of objects are similar, except vectors do not have a built-in definition for multiplication, see Table 1.

You already know everything there is to know about numbers.⁶ Most relevant is that you can multiply numbers with each other (including division, the inverse of multiplication) and you can add them together (including subtraction). For the first part of this course, we will focus on real numbers, \mathbb{R} . Later we will also allow for complex numbers, \mathbb{C} .

Like numbers, vectors can be added and subtracted. In fact, vector arithmetic turns out to be very similar to ‘number arithmetic.’ However, unlike numbers, there is no obvious definition for vector multiplication. This leads to the idea of *defining* functions for various kinds of vector multiplication. Linear algebra is the study of a particular class of these functions. The dot product, for example, which takes two vectors and returns a number, is something we have to “make up” and attach to a vector space.

3.4 Notation: Indices

One theme in this course is that we will repeatedly refine our notation to suit our needs. Let us introduce an *index* notation where we write the components of vectors \mathbf{v} and \mathbf{w} as follows:

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix}. \quad (3.2)$$

We see that a boldfaced Roman letter, u , corresponds to a vector. The *components* of the vector are u^1 , u^2 , u^3 . The “ x -component” of \mathbf{u} is called u^1 : we use the same letter as the

⁶Formally, what I mean by ‘number’ is what mathematicians call a **field**. This simply means some objects where one can add, subtract, multiply, and divide as you would expect. This term is a little tedious for us because physicists usually mean something else when they say ‘field.’ Usually we mean something like the electric field or the field associated with the Higgs boson.

vector, but italicized rather than boldfaced. The upper index is *not* some kind of power, it simply means “the first component.”

Example 3.2. If you see \mathbf{s} , this is understood to be a vector that has multiple components. If it is a three-vector, it has three components. If you see s^2 , then this means that this is the second component of the vector \mathbf{s} . The component of a vector is a number.

You may worry that this notation introduces ambiguity. If we see q^2 , is this the square of some number q , or is it the second component of some vector \mathbf{q} ? The answer depends on context. You should avoid choosing variable names where there is ever the potential for ambiguity. If you have a vector that you call \mathbf{q} , then do not use the letter q for anything else.

You know from \mathbb{R}^3 that you can add together any two vectors \mathbf{v} and \mathbf{w} . Let us call this sum \mathbf{u} so that $\mathbf{u} \equiv \mathbf{v} + \mathbf{w}$. Then we can succinctly write the components of \mathbf{u} in one line:

$$u^i = v^i + w^i . \quad (3.3)$$

The variable i is called an **index**. What values does the index take? In this example, it is clear that (3.3) holds for $i = 1, 2, 3$. That is, i takes values from 1 to the dimension of the space. The typical convention is that we do not have to state the range of index values because it should be understood from the space itself.

With that in mind, it should be clear that if \mathbf{q} is the difference of two vectors, then the components of \mathbf{q} may be succinctly written:

$$\mathbf{q} = \mathbf{v} - \mathbf{w} \quad \Leftrightarrow \quad q^i = v^i - w^i . \quad (3.4)$$

In fact, as physicists we typically use the two statements above interchangeably. If you know the components of a vector, then you know the vector.

3.5 Arithmetic and linear combinations

All vector spaces allow addition and subtraction. This is defined component-wise. The sum of \mathbf{v} and \mathbf{w} is

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v^1 + w^1 \\ v^2 + w^2 \\ v^3 + w^3 \end{pmatrix} . \quad (3.5)$$

What this means is that the *sum* of two vectors is also a vector. That means that if \mathbf{v} and \mathbf{w} are vectors in \mathbb{R}^3 , then $(\mathbf{v} + \mathbf{w})$ is a vector in \mathbb{R}^3 . The components of the vector $(\mathbf{v} + \mathbf{w})$ are simply the sum of the components of \mathbf{v} and \mathbf{w} . A few formal properties that generalize to all vector spaces:

- Vector addition is associative. This means that in the sum $\mathbf{v} + \mathbf{w} + \mathbf{u}$, it does not matter if you add $(\mathbf{v} + \mathbf{w})$ first and then add \mathbf{u} , or if you take \mathbf{v} and then add it to $(\mathbf{w} + \mathbf{u})$. This is the kind of ‘obvious’ property that we tend to take for granted.

- Vector addition is commutative. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$. This is also kind of obvious. But recall that matrix multiplication is not commutative.
- There is a zero vector, $\mathbf{0}$, that does leaves any other vector unchanged under addition. $\mathbf{v} + \mathbf{0} = \mathbf{v}$. This should be totally obvious. The components of $\mathbf{0}$ are obviously all zero.
- There is an additive inverse (negative vectors). If \mathbf{v} is a vector, then $-\mathbf{v}$ is a vector and satisfies $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

Example 3.3. *The first property implies that once you have identified one vector in a vector space, \mathbf{v} , then you can immediately have an infinite number of vectors. This is because $2\mathbf{v} = \mathbf{v} + \mathbf{v}$ must also be a vector. Then $3\mathbf{v} = 2\mathbf{v} + \mathbf{v}$ must also be a vector. And so forth.*

We get another type of operation “for free” with a vector space. This is called scalar multiplication or *rescaling*.

3.6 Rescaling: multiplication by a number

Another operation that exists in a vector space is rescaling: we multiply a vector by a number. Let α be a number. If you want to nitpick, let us restrict α to be a real number. If we have a vector \mathbf{v} with components v^i , then $\alpha\mathbf{v}$ is also a vector.⁷ The components of $\alpha\mathbf{v}$ are

$$(\alpha v)^i = \alpha v^i, \quad (3.6)$$

by which we mean

$$(\alpha\mathbf{v}) = \begin{pmatrix} \alpha v^1 \\ \alpha v^2 \\ \alpha v^3 \end{pmatrix}. \quad (3.7)$$

The parenthesis on the left-hand side is sloppy notation to mean “the vector that is the vector \mathbf{v} rescaled by the number α .” Another way of saying this is that there is a vector $\mathbf{w} \equiv \alpha\mathbf{v}$ whose components are $w^i = \alpha v^i$.

Example 3.4. *Let us do one explicit example with numbers. Suppose the vectors \mathbf{v} and \mathbf{w} have components*

$$\mathbf{v} = \begin{pmatrix} 4.2 \\ -2.6 \\ 7.0 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} 5.3 \\ 2.1 \\ -2.5 \end{pmatrix}. \quad (3.8)$$

⁷“Also a vector” means that it is also an element of the vector space; so $(\alpha\mathbf{v})$ is an element of \mathbb{R}^3 if \mathbf{v} is an element of \mathbb{R}^3 .

I can rescale each vector by different numbers: $\alpha = 10$, $\beta = 2$. We can consider the vector that comes from adding these rescaled vectors:

$$\mathbf{u} \equiv \alpha\mathbf{v} + \beta\mathbf{w} . \quad (3.9)$$

The second component of \mathbf{u} is $u^2 = -26 + 4.2 = -21.8$.

At this point it is useful to define some jargon. A **scalar** is a number. This is in contrast to vectors (and matrices and tensors) which we can think of as arrays of numbers. In fact, every time you see the word scalar, you should just think “number.” Another name for ‘rescaling a vector by a number’ is *scalar multiplication*.

3.7 Linear Combination and Span

Based on our rules for vector space arithmetic, we know that if \mathbf{v} and \mathbf{w} are two vectors in our vector space and if α and β are any two numbers, then

$$\alpha\mathbf{v} + \beta\mathbf{w} \quad (3.10)$$

is also a vector in our vector space. We call any such sum—for any values of α and β —a **linear combination** of the vectors \mathbf{v} and \mathbf{w} . You can of course generalize to the linear combination of more than two vectors, say

$$\alpha\mathbf{v} + \beta\mathbf{w} + \gamma\mathbf{u} . \quad (3.11)$$

Given some number of vectors— \mathbf{v} and \mathbf{w} —you can ask what are all of the possible vectors that you can form from the linear combination of those vectors? This is a vector space.⁸ We say that this vector space is **spanned** by the vectors \mathbf{v} and \mathbf{w} . We call this vector space $\text{Span}(\mathbf{v}, \mathbf{w})$. You can extend this to even more vectors, $\text{Span}(\mathbf{v}, \mathbf{w}, \mathbf{u}, \dots)$.

Exercise 3.2. Show that the vector space spanned by \mathbf{v} and $\alpha\mathbf{v}$ is the same as the vector space spanned by \mathbf{v} .

Exercise 3.3. If \mathbb{R}^3 is the space of vectors with three real components, argue that the span of any four vectors is at most \mathbb{R}^3 but possibly a subset of \mathbb{R}^3 . Give an example where the span of four vectors is \mathbb{R}^2 .

⁸You may want to convince yourself that this satisfies the requirements of vector space arithmetic.

3.8 Basis vectors: an illustrative example

Let us push this idea further. It is useful to start with an example. For simplicity, let us focus on the two-dimensional plane, \mathbb{R}^2 . A vector in \mathbb{R}^2 looks like this:

$$\mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} . \quad (3.12)$$

Any such vector may be written as the linear combination of the following two vectors:

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \quad (3.13)$$

Indeed, it should be obvious that

$$\mathbf{v} = \alpha \hat{\mathbf{e}}_1 + \beta \hat{\mathbf{e}}_2 \quad \text{with} \quad \alpha = v^1 \quad \beta = v^2 . \quad (3.14)$$

In other words, these ‘special’ vectors $\hat{\mathbf{e}}_{1,2}$ satisfy:

1. Any vector in \mathbb{R}^2 may be written as a linear combinations of $\hat{\mathbf{e}}_{1,2}$. We showed this because \mathbf{v} in the above discussion could be any vector in \mathbb{R}^2 . Thus $\text{Span}(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2) = \mathbb{R}^2$.
2. The coefficients in the linear combination are precisely the components of the vector \mathbf{v} . Soon we will see that this observation is actually backwards: it is the choice that $\hat{\mathbf{e}}_1$ are special that defines the components of a vector.

It should be obvious that any pair of vectors that “aren’t pointing in the same direction” can span the entire space \mathbb{R}^2 .

Exercise 3.4. What does “aren’t pointing in the same direction” mean in this context? Use \mathbf{v} and $\alpha\mathbf{v}$ in your answer.

We could try a different pair of vectors and consider its linear combinations:

$$\mathbf{f}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \mathbf{f}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \quad (3.15)$$

Then the vector \mathbf{v} may be written as $\mathbf{v} = \alpha \mathbf{f}_1 + \beta \mathbf{f}_2$. To find α and β , we can simply plug in the components of \mathbf{v} and $\mathbf{f}_{1,2}$ so that:

$$v^1 = \alpha - \beta \quad v^2 = \alpha . \quad (3.16)$$

In other words,

$$\mathbf{v} = (v^2)\mathbf{f}_1 + (v^2 - v^1)\mathbf{f}_2 . \quad (3.17)$$

These coefficients $\alpha = v^2$ and $\beta = v^2 - v^1$ may be written in shorthand. Let’s suggestively write

$$\mathbf{v} = \begin{pmatrix} v^2 \\ v^2 - v^1 \end{pmatrix}_{\mathbf{f}} , \quad (3.18)$$

where we use the subscript \mathbf{f} to mean “coefficients with respect to $\mathbf{f}_{1,2}$. This looks just like a two-component vector, doesn’t it?

Exercise 3.5. Let $\mathbf{v} \in \mathbb{R}^2$ be the following vector in two-dimensional real space:

$$\mathbf{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} . \quad (3.19)$$

Here are two vectors that span \mathbb{R}^2 :

$$\mathbf{g}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \mathbf{g}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} . \quad (3.20)$$

What are the coefficients α and β so that $\mathbf{v} = \alpha\mathbf{g}_1 + \beta\mathbf{g}_2$? Answer: $\alpha = 2$ and $\beta = 1$.

What we're getting at is the following. Define a set of vectors that span a space. We will call this set of vectors a **basis** of that space—we'll give a slightly more formal definition below. Any vector in the space can be written as a linear combination of basis vectors. For example, if $\mathbf{b}_{1,2}$ are a basis of \mathbb{R}^2 , then for any vector $\mathbf{v} \in \mathbb{R}^2$ we may write

$$\mathbf{v} = \alpha\mathbf{b}_1 + \beta\mathbf{b}_2 . \quad (3.21)$$

Then we have encoded all of the data of vector \mathbf{v} into the coefficients (α, β) . In fact, let me be more economical with my symbols and change notation a bit and write $(\alpha^1, \alpha^2) \equiv (\alpha, \beta)$ so that

$$\mathbf{v} = \alpha^1\mathbf{b}_1 + \alpha^2\mathbf{b}_2 . \quad (3.22)$$

Then I can write the information encoded in \mathbf{v} as a column, which I will write with a subscript b to distinguish it from the “actual” vector components, (3.12):⁹

$$\mathbf{v} = \begin{pmatrix} \alpha^1 \\ \alpha^2 \end{pmatrix}_b . \quad (3.23)$$

The last two equations mean exactly the same thing. Now here's something cute: we can treat the two-component array¹⁰ on the right-hand side of (3.23) as if it were a vector. We can do vector arithmetic on it. If we have two vectors with “ \mathbf{b} basis components”

$$\mathbf{v} = \begin{pmatrix} \alpha^1 \\ \alpha^2 \end{pmatrix}_b \quad \mathbf{w} = \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix}_b , \quad (3.24)$$

Then we could take linear combinations of the two with respect to two numbers a and b :

$$a\mathbf{v} + b\mathbf{w} = (\alpha^1 + \beta^1)\mathbf{b}_1 + (\alpha^2 + \beta^2)\mathbf{b}_2 = \begin{pmatrix} \alpha^1 + \beta^1 \\ \alpha^2 + \beta^2 \end{pmatrix}_b . \quad (3.25)$$

If \mathbf{v} and \mathbf{w} span \mathbb{R}^2 , then any vector in the vector space may be written as a linear combination of the form (3.25). This means we may use the “ \mathbf{b} basis components” as

⁹We will soon see that there is nothing holy about (3.12).

¹⁰I'm trying not to call it a vector.

equivalent ways of encoding a vector as the natural description (3.12). But wait a moment: what is so “natural” about (3.12)?

If we reverse the argument for the \mathbf{b} basis, then we see that the “natural” components of the vector \mathbf{v} in (3.12) are simply the coefficients of the linear combinations of the basis vectors $\hat{\mathbf{e}}_{1,2}$ in (3.14). What made the basis vectors $\hat{\mathbf{e}}_{1,2}$ so special, anyway? Nothing at all.

What we’ve come to is that the *components* of a vector depend on the basis that we choose. In \mathbb{R}^3 we usually use the basis $\hat{\mathbf{e}}_{1,2,3}$ where the vectors point respectively along the \hat{x} , \hat{y} , and \hat{z} directions. It is kind of an “obvious” basis, though it completely depends on the \hat{x} , \hat{y} , and \hat{z} directions having some intrinsic meaning. They often do not: we could set up our coordinate system however we wanted. In fact, *nowhere* in our definition of a vector space did we even assume that a coordinate system exists!

Indeed, it’s the other way around: a choice of basis vectors *defines* a ‘coordinate system’ rather than the other way around.¹¹ All of this begs for a re-definition.

3.9 Basis vectors, formally

A **basis** is a *minimal set* of vectors that span a space. Here ‘minimal’ means that if you remove any vector from the basis, then there are vectors in the space that cannot be written as a linear combination of the remaining vectors.

Example 3.5. Consider the following three vectors:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \quad (3.26)$$

These three vectors are not a basis for a subspace because there are vectors that are linear combinations of \mathbf{v} , \mathbf{w} , and \mathbf{u} that can be equivalently written as a linear combination of just \mathbf{v} and \mathbf{u} , for example. To see this, consider the vector

$$\mathbf{t} = 4\mathbf{v} + 2\mathbf{w} + 3\mathbf{u} = \begin{pmatrix} 7 \\ -1 \\ 0 \end{pmatrix}. \quad (3.27)$$

This may equivalently be written as

$$\mathbf{t} = 7\mathbf{v} - \mathbf{w}. \quad (3.28)$$

Indeed, there are an infinite number of ways to write \mathbf{t} . Because $\mathbf{v} - \mathbf{w} + \mathbf{u} = 0$, you can add any multiple of this linear combination to \mathbf{t} to leave \mathbf{t} unchanged.

¹¹Though it really is dangerous to think about a vector space as having coordinates. We will see why when we talk about vector bundles and manifolds.

The **dimension** of a vector space is the number of vectors in the basis. In the example above, the vector space spanned by linear combinations of \mathbf{v} , \mathbf{w} , and \mathbf{u} has dimension two. This is because you only need two vectors write any vector in the space as a linear combination. If you drew all of the vectors in this subspace as arrows with their base at the origin, then the arrow heads with all live on the xy -plane.

Here are some *obvious* statements¹²:

1. The zero vector cannot be part of any basis.
2. The dimension of a vector space does not depend on the choice of basis.
3. If you have a proposed set of basis vectors but there is a *non-trivial* linear combination of those vectors that sums to zero, then the set of vectors is not a basis. Here non-trivial means “the coefficients are not all equal to zero.” This should be evident from Example 3.5.
4. If any two vectors in a proposed basis are proportional to one another, then this set of vectors is not a basis.
5. The number of components v^i to describe a vector is the dimension of the vector space.
6. In the expansion of a vector as a linear combination of basis vectors, the coefficients are unique to the vector. That is: if $\mathbf{v} = \sum_i \alpha^i \mathbf{b}_i$ for a basis $\mathbf{b}_{1,2,3}$, then the set of numbers $(\alpha^1, \alpha^2, \alpha^3)$ uniquely defines \mathbf{v} . There is no other combination of coefficients in a linear combination that sum to \mathbf{v} .
7. When describing a vector, the *coefficients* of the linear combination of basis vectors and the *components* of a column vector with respect to that basis are identical. This is by definition.

The last point is poignant. You may have believed that a vector *is* the column of numbers. We want to move away from this so that we may generalize our definition. A vector is a linear combination of basis vectors, where we are remaining agnostic about what the basis vectors are. Let me say this again: *the column of numbers is not the vector, it is simply a representation of a linear combination of basis vectors. All the “vector-ness” is encoded in the basis vectors.*

Example 3.6. *Color space is a vector space that highlights this idea of a more abstract basis vector. In color theory, all colors are linear combinations of red, green, and blue. This should sound really weird because in physics these colors are simply wavelengths of light: what is special about them? Nothing in nature. What is special is that our eyes have three types of color receptor cells. Each type is sensitive to a certain window of the visible spectrum. We call these human eye responses the colors red, green, and blue. When we add colors, what we really mean is we’re adding “responses” to a particular spectrum of light. When we add colors, we are not adding electromagnetic waves: we are adding neurological responses. For each type of color-sensitive cell, one ‘blip’ of neural response is a basis vector for our color response. The sensation of a particular color is*

¹²This means: if they are not immediately apparent, stop and think about it to make sure you understand.

a linear combination of this basis. An actual human being is not sensitive to the whole vector space: for example, we cannot add negative colors to our sensory response. This is a fascinating subject and a surprising application of linear algebra.^a

^aThere are some great YouTube videos on this. Here are a few: <https://www.youtube.com/watch?v=xAoljeRJ3lU>, <https://www.youtube.com/watch?v=AS10HMW873s>, <https://www.youtube.com/watch?v=99v96TL-tuY>.

What is less obvious is that at this point there is *no preferred basis*. Any minimal set of vectors that span a vector space is a perfectly good vector. Suppose $\mathbf{b}_{1,2,3}$ is one such set for \mathbb{R}^3 . We can write the vector with respect to the coefficients of the linear combination of $\mathbf{b}_{1,2,3}$ basis vectors that reproduces it. If we had another basis, $\mathbf{b}'_{1,2,3}$, we could write the same vector with respect to a different linear combination of the $\mathbf{b}'_{1,2,3}$ basis. The components of these linear combinations, say α^i and α'^i , will be different because the basis elements are different. However, they represent the same vector.

In your heart, you should feel anxious. You *like* the “obvious” basis

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \hat{\mathbf{e}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.29)$$

We can even call this the Cartesian basis. It seems so natural, we even gave these basis vectors little hats to remind us how much we like them! Stop and think about what you like about this basis. I guarantee you that all of those nice features invoke mathematical machinery that are *not* included in a vector space. You may like that the Cartesian basis is orthogonal and each basis vector has unit length. To this I reply: *how do you measure angle or length? These are not concepts that our vector space is equipped with.* You are, of course, correct that there is a way to *define* angle and length—but that is something additional that we have to impose on the space. We will get to this shortly.

Example 3.7. Another surprising example of a vector space is the space of polynomials of finite degree. This means functions of the form

$$f(x) = a^0 + a^1x + a^2x^2 + \cdots a^Nx^N. \quad (3.30)$$

Finite degree means that N is some integer that is not infinity.^a To be clear, our notation has become a bit ambiguous: here x is a variable and x^n means x to the n^{th} power. The coefficients a^i , on the other hand, are numbers and i is an index. We can pick the following basis:

$$\hat{\mathbf{e}}_0 = x^0 = 1 \quad \hat{\mathbf{e}}_1 = x^1 = x \quad \hat{\mathbf{e}}_2 = x^2 \quad \cdots \quad \hat{\mathbf{e}}_N = x^N \quad (3.31)$$

These ‘basis vectors’ are actually functions that are simple monomial powers of x . It should be obvious (there’s that phrase again) that linear combinations of these basis

vectors/functions can give any function $f(x)$ of degree up to N . It should also be obvious that the dimension of this space is $(N + 1)$; don't forget to count the $\hat{\mathbf{e}}_0$ vector.

For example, consider the polynomial $f(x) = 3 + x^2$. The linear combination of basis vectors that gives this has $a^0 = 3$, $a^2 = 1$, and all other coefficients zero:

$$f(x) = 3\hat{\mathbf{e}}_0 + \hat{\mathbf{e}}_2 . \quad (3.32)$$

We could represent this vector/function as a column:

$$f(x) = \mathbf{f} = \begin{pmatrix} 3 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.33)$$

where \mathbf{f} is an $(N + 1)$ -component column of the numbers a^i .

^aThis may seem like a silly point, but one of the key ‘aha’ ideas in this course will be that we can do linear algebra on the space of general functions where we allow $N \rightarrow \infty$.

Exercise 3.6. Consider a vector/function \mathbf{f} with components f^i in the polynomial space in Example 3.7. Now consider the vector/function $\mathbf{f}' \equiv df/dx$. Write out an expression for the i^{th} component of \mathbf{f}' . Hint: for example, the $i = 1$ component is $2f^2$.

3.10 The meaning of column vectors

The previous subsection on bases¹³ is so important that we should really emphasize the mathematical edifice that we have reverse-engineered¹⁴

1. A vector is technically *not* the column of numbers that we usually say it is. That column of numbers is simply a way of writing the *components* of a vector.
2. The components of a vector are simply the coefficients of the basis vectors in the linear combination of basis vectors that sum to the vector. That is: v^i is defined by $\mathbf{v} = v^i \hat{\mathbf{e}}_i$ where the $\hat{\mathbf{e}}_i$ are a set of basis vectors that we all agree upon.
3. I never had to say what the basis vectors *are*. They can be anything where linear combinations of those things are still the same type of thing. In this way, we can treat

¹³The plural of ‘basis’ is ‘bases,’ pronounced *bay-sees*, just like the plural of ‘axis’ is ‘axes’ pronounced *axe-sees*.

¹⁴Do you appreciate why I say ‘reverse engineered’ here? In mathematics classes, one would start with some postulates for what an abstract vector is and then your usual 3-component column vectors pop out as one silly example. We have started those 3-component column vectors and used their properties to motivate a general definition of what vectors are.

the basis vectors abstractly.

You may be used to vectors being forces, momenta, velocities, electric fields, and so forth. We want to be able to use the same machinery of linear algebra on more general objects: particles with quantum spin, functions, the perception of colors by the human eye, and so forth.

Exercise 3.7. *What happens when we do not agree on a basis? Suppose you set up a basis. Stand up. Suppose you are facing north. Your feet are at the origin. If you spread your arms out, your right hand points in the direction of your first basis element (x -axis, pointing east), $\hat{\mathbf{e}}_1$. Your nose points in the direction of your second basis element (y -axis, north), $\hat{\mathbf{e}}_2$. Your head points in the direction of your third basis element (z -axis, skyward), $\hat{\mathbf{e}}_3$.*

However, your friend Riley approaches you from the northeast so Riley is facing southwest. Riley decides to set up their own basis, analogous to you. Their first basis element $\hat{\mathbf{r}}_1$ points in the northwest direction, their second basis element $\hat{\mathbf{e}}_2$ points in the southwest direction, and their third basis element $\hat{\mathbf{e}}_3$ also points skyward.

For simplicity, assume that the length of your basis vectors are all the same—even though we haven't defined what length means. Suppose you 'measure' a vector with components $v^1 = 2$, $v^2 = -1$, and $v^3 = 1.5$. This is a vector pointing southeast and upward. What components does Riley measure with respect to their basis?

Exercise 3.8. *One of my favorite examples of a vector space is the space of Fibonacci sequences. Fibonacci sequences are infinite lists of numbers a_i that satisfy $a_{i+2} = a_i + a_{i+1}$. Once you specify the first two numbers a_0 and a_1 , you can iteratively generate every other number in the sequence. Each sequence is a vector in the space of possible Fibonacci sequences. Show that this is true by confirming that a linear combination of Fibonacci sequences with each i^{th} term added, e.g. $(a + b)^i = a_i + b_i$ is also a Fibonacci sequence. Give an example of a basis for the Fibonacci sequences. What is the dimension of the Fibonacci sequence space? Answer: the dimension is two, even though each element is an infinitely long list of numbers.*

3.11 Operations that are not (yet) allowed

In these definitions, we make a big deal about how the sum of two vectors *is also a vector*. Or how the rescaling of a vector by a number *is also a vector*. This is in contrast to operations that are either not allowed or that do not produce vectors. An example of an operation that is not allowed is adding together vectors from two different vector spaces. The following

proposed sum of a vector in \mathbb{R}^3 with a vector in \mathbb{R}^2 does not make sense:

$$\begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} + \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} = ? \quad (3.34)$$

If you find yourself adding vectors from two different vector spaces, then you have made a mistake.

Another operation that requires care is rescaling a real vector by a complex number. If \mathbf{v} is a vector in \mathbb{R}^3 and we try to multiply it by a complex number, $\alpha = 2 + 3i$, then the resulting “vector” is not a vector in \mathbb{R}^3 :

$$(\alpha\mathbf{v})^i = (2 + 3i)v^i \notin \mathbb{R}, \quad (3.35)$$

that is: the components of $\alpha\mathbf{v}$ are not real numbers, and so this cannot be an element of a vector space that is *defined* to have real components. Later on we will generalize to the case of *complex vector spaces*, but we will treat that with some care.¹⁵

Thus far, we have introduced the *nouns* of this course: vectors. We have identified a few *verbs* that let us do things with these vectors:

1. Addition takes two vectors in a vector space and returns a vector in the same vector space.
2. Rescaling takes a vector and a number and returns a vector in the same vector space.

We can rewrite this in the language of *mappings* (or *functions*) as follows. Let V be a vector space, say $V = \mathbb{R}^3$. Let us write \mathbb{R} mean [real] numbers. Then the above statements tell us that addition and rescaling can be thought of as maps:

1. Vector addition: $V \times V \rightarrow V$
2. Rescaling: $V \times \mathbb{R} \rightarrow V$.

Do not be intimidated by the \times symbol here. This “mapping” notation means nothing more and nothing less than the statements above.

We now know everything there is to know about the vector space \mathbb{R}^3 . We want to learn more about functions (maps) that involve this vector space. How can we combine vectors and numbers to produce other vectors and numbers? What about more complicated objects like matrices and tensors?

3.12 Euclidean three-space

You may object: *wait! I know there are more things you can do with three-vectors!* You remember that there are two types of vector multiplication that we use in physics. The **dot product** and the **cross product**.

¹⁵If you want to be fancy, you can replace ‘number’ with the mathematical notion of a field. Both the real numbers and the complex numbers are examples of fields. In my mind a field is just a class of number, though mathematicians have fancier definitions.

In \mathbb{R}^3 , the **dot product** is a map $V \times V \rightarrow \mathbb{R}$. That means it takes two vectors and returns a number. The particular number that it returns is typically *defined* to be

$$\mathbf{v} \cdot \mathbf{w} = \sum_i v^i w^i = v^1 w^1 + v^2 w^2 + v^3 w^3 . \quad (3.36)$$

The dot product generalizes in linear algebra. It is often called an **inner product** or a **metric** and has a few different notations that we will meet. What is important is that this dot/inner product is an *additional* mathematical function that we attach to a vector space.

Three-dimensional real space combined with the dot product/inner product/metric (3.36) is called Euclidean three-space. In general, a vector space combined with a ‘dot product’ is called a **metric space**. The word metric should invoke some etymological notion of measurement of distance. Indeed, the dot product is a tool that tells us how ‘close’ two vectors are to one another—though it is not yet obvious how.

Example 3.8. Let $\mathbf{r} = (x, y, z)$ be a “position vector” of a point relative to the origin.^a Then the distance of the point from the origin is

$$d = \sqrt{\mathbf{r} \cdot \mathbf{r}} = \sqrt{x^2 + y^2 + z^2} . \quad (3.37)$$

This gives a notion of how the dot product is related to measuring distances, but it turns out to be a bit of a red herring! The real sense in which the dot product measures the ‘closeness’ of two vectors is the sense in which it defines an angle between those vectors. (See below.)

^aIt is dangerous to use the phrase “position vector,” see Exercise 3.1.

The **cross product** is a different story. You may remember the cross product from such hits as¹⁶ angular momentum, $\mathbf{r} \times \mathbf{p}$. It looks like a map that takes two vectors and spits out another vector, $V \times V \rightarrow V$. Indeed, this is the case in Euclidean three-space. However, it had some funny properties compared to the dot product. For example, there was something weird with the order of the two vectors: $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$. It is also a bit funny that the direction of the output vector is completely different¹⁷ from the directions of the input vectors. It will turn out that this product does not generalize as simply as the dot product, though there is a generalization called the **wedge product** which is outside the scope of this course.¹⁸

¹⁶<https://tvtropes.org/pmwiki/pmwiki.php/Main/YouMightRememberMeFrom>

¹⁷The technical meaning of ‘completely different’ is *orthogonal*, which we define below with the help of the metric.

¹⁸That is not to say that the wedge product is not relevant in physics. The wedge product features prominently in a mathematical field called **differential geometry**, which is in turn the framework for general relativity. The wedge product is related to defining volumes and integration measures.

Exercise 3.9. Define the generalization of the Euclidean three-space metric to Euclidean space in d dimensions. (Easy.)

Exercise 3.10. Try to define a generalization of the cross product in two-dimensional Euclidean space. Reflect on why this is much less natural than the generalization of the dot product.

3.13 Length in Euclidean three-space

Euclidean three-space is real space combined with the Euclidean dot product, (3.36). The [Euclidean] **magnitude** (length) of a three vector \mathbf{v} as $|\mathbf{v}|$ in Euclidean three-space. The magnitude is defined to be

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} . \quad (3.38)$$

This definition generalizes to Euclidean d -dimensional space with the appropriate generalization of the dot product.

Example 3.9. Consider the vector

$$\mathbf{v} = \begin{pmatrix} 3 \\ -4 \\ 0 \end{pmatrix} \quad (3.39)$$

in Euclidean three-space. The magnitude of \mathbf{v} is $|\mathbf{v}| = 5$.

Some references prefer to use the double bar notation, $||\mathbf{v}||$ for the length of a vector. This is to distinguish it from the absolute value of a number, $|-3| = 3$. We will be even more perverse: sometimes we will write v to mean the magnitude of \mathbf{v} when there is no ambiguity.

Example 3.10. Consider the vector

$$\mathbf{v} = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix} . \quad (3.40)$$

Then the magnitude of \mathbf{v} is $|\mathbf{v}| = \sqrt{14}$. We could also write this as $v = \sqrt{14}$, but we should be careful when we write things like v^2 which could either mean the second component of \mathbf{v} —which is 3—or the square of the magnitude—which is 14.

We see that the dot product (metric) allows us to define length. Because the length of a vector is a number, we can divide the vector \mathbf{v} by its length $|\mathbf{v}|$ to obtain a **unit vector**:

$$\hat{\mathbf{v}} = \frac{1}{|\mathbf{v}|} \mathbf{v} . \quad (3.41)$$

The right-hand side is simply scalar multiplication by $|\mathbf{v}|^{-1}$. Unit vectors are useful for identifying directions.

Example 3.11. In grade school one may have learned that a vector is an arrow that has a magnitude and a direction. Unit vectors encode the ‘direction’ of a vector.

Example 3.12. Let \mathbf{v} be defined as in (3.41). The unit vector associated with \mathbf{v} is

$$\hat{v} = \begin{pmatrix} 3/5 \\ -4/5 \\ 0 \end{pmatrix} . \quad (3.42)$$

3.14 Angles in Euclidean three-space

Let \mathbf{v} and \mathbf{w} be two vectors in Euclidean three-space. The [Euclidean] angle between these two vectors, θ , is

$$\cos \theta \equiv \hat{\mathbf{v}} \cdot \hat{\mathbf{w}} = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}| |\mathbf{w}|} . \quad (3.43)$$

Exercise 3.11. Confirm that this matches the definition of the angle between two vectors that you learned in your youth.

The above definition of the angle between two vectors is general for any metric space—that is, a vector space equipped with a dot product.

Example 3.13. *The angle between two vectors defines the sense in which two vectors are close to one another. This is the sense in which the dot product (metric) lets you measure the “distance” between two vectors. Note that this is completely different from the notion of distance between two points in space, (3.37).*

Example 3.14. *When I was taking a class like this, the professor kept bringing up something called conformal transformations. At the time, I did not appreciate how significant conformal transformations are in physics. A conformal transformation is a map from a metric space to itself that preserves angles. That means that the angle between two vectors before the transformation is the same as the angle of the two vectors that come out of the transformation, with the reminder that the metric may also transform. When conformal transformations finally came up again in my physics life, it came up in the context of theories that are unchanged under a rescaling of all the vectors. I remember thinking: what does this have to do with preserving angles? We can see from (3.14) that rescaling the lengths of vectors does not change the angle between them. Conformal transformations are really very neat. In the early days of the aerospace industry, mathematics based on conformal transformations helped engineers design efficient airplane wings before there were reliable computers to simulate airflow. I encourage you to look this up, the technique is called 2D conformal mapping.*

4 Index Notation and Summation Convention

There's something that physicists do that tend to drive mathematicians crazy: we write a generic *component of a vector* and refer to it as if it were the vector itself. It is a fairly harmless peccadillo:¹⁹ if I say “the vector v^i ,” then it is not hard to guess that I mean “the vector \mathbf{v} which has components that I label v^i .”

The reason why we have this culture is that this index notation ends up being so damn convenient. In addition to vectors, we will have other objects that have indices: dual vectors, matrices, and tensors. When we write everything in with indices, we can “see” properties of these objects that are not obvious without the indices. Specifically, we can see *how an object transforms under symmetries*. In this course, we will focus on *rotations* of vectors and their generalizations.

There's a second reason why indices are convenient: they allow us to use **Einstein summation convention**. This is a notational shortcut that introduces upper and lower indices to convey sums. Consider, for example, the “matrix multiplication” of a row vector \mathbf{w} on a column vector \mathbf{v} . Nevermind the formal definition of “row vector” as opposed to

¹⁹There are times when you can get into trouble if you drink your own Kool Aid, so to speak. The reason is that the *component v^i* is simply a number, whereas \mathbf{v} is a vector. Some manipulations are only allowed for numbers and not vectors, and you should be clear that you mean ‘the component v^i ’ if you are treating it like a number, and not ‘the vector whose components are v^i .’ See Example 4.3.

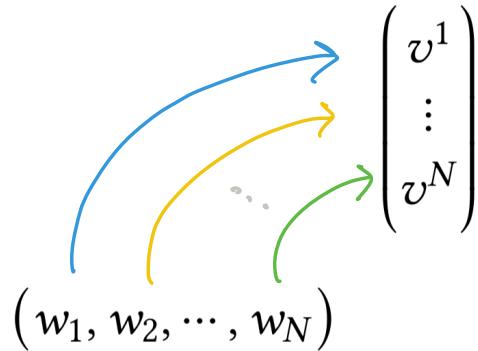


Figure 1: The ‘matrix multiplication rule’ for acting with a row vector on a column vector.

“column vector.” Let’s just write it out in components where it is obvious for \mathbb{R}^3

$$\underline{\mathbf{w}} = (w_1 \ w_2 \ w_3) \quad \mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \quad \underline{\mathbf{w}\mathbf{v}} = w_1 v^1 + w_2 v^2 + w_3 v^3 . \quad (4.1)$$

The final expression is familiar, right? It follows the usual rules of matrix multiplication for a “matrix” that happens to be one row and three columns; we review this rule in Fig. 1. We notice that we chose to write the components of $\underline{\mathbf{w}}$ with lower indices—this is the convention. Row vectors (which have many names) have indices written as subscripts while column vectors have indices written as superscripts. There is no mathematics here, just a choice of notation. The result of the multiplication is simply a number, which we can write as a sum:

$$\underline{\mathbf{w}\mathbf{v}} = \sum_{i=1}^3 w_i v^i \equiv w_i v^i . \quad (4.2)$$

On the right-hand side we have *defined* the summation convention: *whenever there is exactly one upper index and exactly one lower index with the same letter, we should understand that there is a sum over that index over all of its allowed values.* We call pairs of repeated indices where one is upper and one is lower **contracted indices**.

The value $w_i v^i$ is simply a number. It is not a vector. It does not have any “vectorial” (tensorial) structure. It is not an element of the vector space \mathbb{R}^3 . It does not transform under rotations. It is *just a number*. In other words, $w_i v^i$ behaves like an object with *no indices*. Contracted indices “cancel each other out.”

This is significant because we will see that indices tell us how objects transform. Evidently, column vectors and row vectors transform differently since one has an upper index and one has a lower index. Further, when we contract the two indices, we end up with something with no indices: a number that does not transform at all. This may seem like notational overkill—trust me, it is worth building this notation now. We will use it over and over.

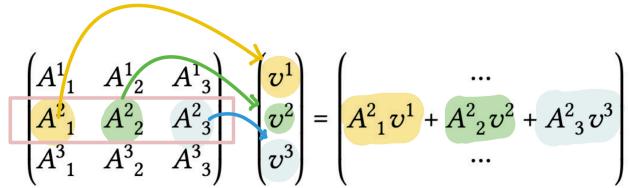


Figure 2: The ‘matrix multiplication’ rule for $A\mathbf{v} = \mathbf{v}'$. We show that the second element of \mathbf{v}' is a sum of terms, where each term is a multiplication of the j^{th} column of the 2nd row of A by the j^{th} row of \mathbf{v} .

Example 4.1. Matrices M have the following index structure: M_j^i . There is a first index and a second index—the order matters. The first index is upper, and the second index is lower. Matrix multiplication boils down to a contraction of indices:

$$(M\mathbf{v})^i = M_j^i v^j . \quad (4.3)$$

Let us read this equation carefully. First, $M\mathbf{v}$ is a vector. The i^{th} component of this vector is $(M\mathbf{v})^i$. How is this related to the components of M and \mathbf{v} ? The right-hand side tells us that we simply take the sum:

$$M_j^i v^j = M_1^i v^1 + M_2^i v^2 + M_3^i v^3 . \quad (4.4)$$

Example 4.2. From the above example, you can then excuse the glib statement: “the vector $M_j^i v^j$.” As we explained above, $M_j^i v^j$ is not a vector, but a component of a vector. However, the point is that even though there are three indices, two of them are contracted so the object effectively only has one upper index. This is the index structure of a vector. This matches the usual matrix multiplication rule shown in Fig. 2.

Exercise 4.1. Consider the following vector, row vector, and matrix:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \underline{\mathbf{w}} = (4 \ 5 \ 6) \quad M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} . \quad (4.5)$$

These have index structure v^i , w_i , and M_j^i respectively. Note that the first index of a matrix is the row and the second is the column, thus $M_2^1 = 2$ while $M_1^2 = 4$. Calculate the following: $(wM)_2$, $(M\mathbf{v})^1$, $(MM)^1_2$. Here MM is understood to be the square of the matrix M , $(M^2)_j^i = M_k^i M^k_j$.

Example 4.3. It should be clear that

$$w_i M_j^i = w_1 M_j^1 + w_2 M_j^2 + w_3 M_j^3 = M_j^1 w_1 + M_j^2 w_2 + M_j^3 w_3 = M_j^i w_i . \quad (4.6)$$

After all, each of the components w_i and M_j^i are simply numbers. However: even though $w_i M_j^i = M_j^i w_i$, it is completely incorrect to say $\underline{\mathbf{w}} M = M \underline{\mathbf{w}}$. This is because $\underline{\mathbf{w}}$ and M are tensorial (vector-y) objects. The order of their ‘multiplication’ matters. You can see this from the matrix notation.

$$\underline{\mathbf{w}} M = (4 \ 5 \ 6) \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad M \underline{\mathbf{w}} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} (4 \ 5 \ 6) . \quad (4.7)$$

The first multiplication gives a row vector, as you expect since $(wM)_j$ has one lower index. The second multiplication does not even make sense. What we see is that expressions like $w_i M_j^i = M_j^i w_i$ are valid as long as you are only talking about the components. The glib “physicist slang” of replacing a component by its vector/matrix/tensor can get you into trouble if you have moved components around in a way that is only allowed for numbers, but not vectory-things.

Since the language is now becoming cumbersome, let us define the word **tensorial** to mean an object with indices. This will replace the phrase “vectory” in our notes.

One neat thing about this is that our convention for contracting indices makes it clear that $(Mv)^i$ is a component of a vector: it has one upper index. Similarly, you may recall that the multiplication of matrices M and N proceeds as follows:

$$(MN)^i_j = M^i_k N^k_j . \quad (4.8)$$

Exercise 4.2. Confirm that (4.8) holds for 2×2 matrices.

On the right-hand side of (4.8), we have one pair of contracted indices (k), one upper index i , and one lower index j . We thus deduce that this object is a matrix: it has one upper and one lower index. Indeed, the product of two matrices is also a matrix. Our indices and contraction rules tell us what kinds of objects we can produce by contracting indices between them.

Example 4.4. You may also contract indices within an object. For example, because a matrix has one upper and one lower index, you may contract them together. This is called the trace, $\text{Tr } M = M^i_i$.

5 Matrices and Linear Transformations

Vectors are the ‘nouns’ in linear algebra. The word ‘linear’ refers to the the *verbs*. That is: we would like to act on vectors.

5.1 Jargon

Let us introduce some jargon here. Rather than formal definitions, we give practical “physicist’s” definitions.²⁰ You can look up proper definitions in your favorite mathematics textbook. The following words are closely related: function, map, transformation. I often use them interchangeably, though this is rather sloppy.

A **function** is a mathematical machine that takes some inputs and produces some output. The inputs can be numbers, vectors, or more sophisticated objects. The outputs may also be numbers, vectors, or more sophisticated objects. The outputs do not have to be the same type of object as the inputs—in general they are not. A function that takes one input and returns one output is called a **map**. A map that takes one type of object and returns the same type of object is called a **transformation**. Mathy-folks like to draw diagrams like Fig. 3.

Example 5.1. *The dot product is a **function** that takes in two vectors and outputs a number.*

Example 5.2. *The magnitude is a **map** that takes a vector and returns a number.*

Example 5.3. *A function $f(x, y) = z$ is not a map because it takes two inputs.*

Example 5.4. *The national weather service website can tell you the temperature in your location as a function of the time of day. In this sense, it is a map from time to temperature.*

²⁰If we do something less formally, we say that we are physicists, not mathematicians. If we choose to be more highbrow, then we say that we are physicists, not engineers.

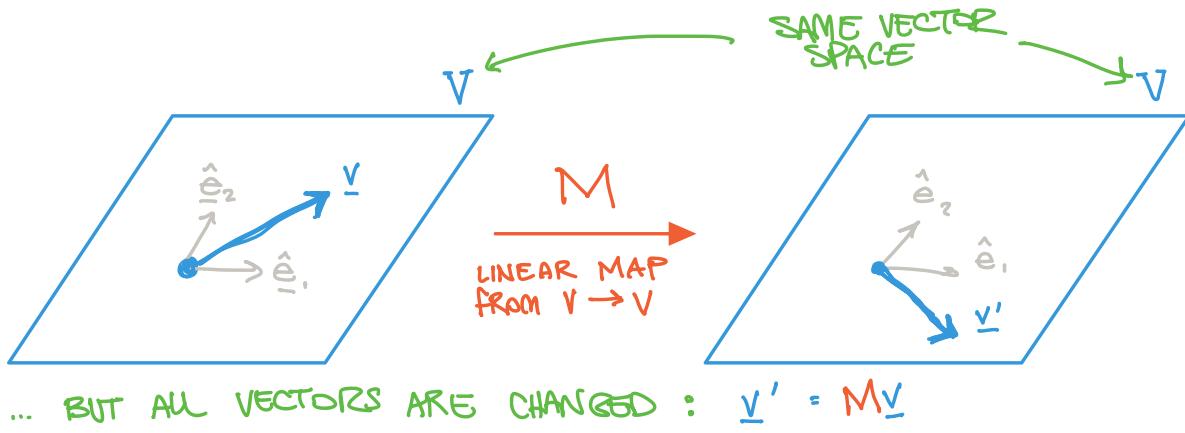


Figure 3: Example of a linear transformation. The transformation M (for matrix) takes a vector \mathbf{v} and turns it into a different vector that we call \mathbf{v}' . This new vector is related to \mathbf{v} by the matrix $\mathbf{v}' = M\mathbf{v}$. *Every* vector is transformed under the map M . This is an *active transformation* where all vectors transform, but the basis ($\hat{\mathbf{e}}_i$) stays fixed.

Example 5.5. *The map that takes a vector and returns its unit vector is a transformation: it is a function that takes a vector and returns a vector.*

Example 5.6. *The function that takes the a time in Pacific standard and converts it to East Coast standard is a transformation. It transforms one time (2:00pm) to another time (5:00pm).*

5.2 Linear transformations

A function f is **linear** if it satisfies:

$$f(\alpha x) = \alpha f(x) \quad f(x + y) = f(x) + f(y) . \quad (5.1)$$

Here α is a number, and x and y are some objects (perhaps vectors) on which the function has been defined to act. We may combine these two properties to write the condition of linearity in equation:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \quad (5.2)$$

Exercise 5.1. Show that equations (5.1) and (5.2) imply one another. Going from (5.1) to (5.2) is slightly more tricky than the other way around.

5.3 Linear Transformations in One Dimension

The definition of ‘linear’ has *nothing to do with lines*. It is true that $y = mx + b$ is the equation for a line in the (x, y) plane. This is not the same as $f(x) = mx + b$ being linear. In fact, $f(x) = mx + b$ is *not* linear as a transformation of numbers, $\mathbb{R} \rightarrow \mathbb{R}$

Exercise 5.2. Show that $f(x) = mx + b$ does not satisfy the definition of linearity, (5.2). It is sufficient, for example, to consider $f(2x) \stackrel{?}{=} f(x) + f(x)$. Another cute way to see this is to consider linear combinations with zero, say $f(x + 0)$.

Exercise 5.3. Show that $f(x) = mx$ is linear.

The problem in $f(x) = mx + b$ is the shift by a constant, b . I will comment that we can make up mathematics where ‘shift by a constant’ is allowed. This, however, is no longer *linear algebra*. You can look up affine spaces²¹ if you’re curious. I only mention it here because the word ‘affine’ may show up later in your mathematical physics studies—I want you to remember that it has something to do with constant shifts.

5.4 Linear Transformations in Two and Three Dimensions

In Section 4 we introduced the idea that a matrix, M has a peculiar index notation M^i_j . This turned out to be convenient because we could use Einstein summation convention to conveniently write out what happens when we apply a matrix M to a vector \mathbf{v} . You end up with a vector $M\mathbf{v}$ with components

$$(M\mathbf{v})^i = M^i_j v^j . \quad (5.3)$$

We recognized that the j indices are *contracted*. From an index point of view, these contracted indices “cancel out” so the resulting expression only has one free index.²² This is good, the left-hand side of the above equation has one free index, so the right-hand side should also have exactly one free index.²³

²¹https://en.wikipedia.org/wiki/Affine_space

²²I say *free* index to contrast it from a contracted (repeated) index that is summed over. We can put any allowed value into a free index and the equation is true.

²³This should remind you of dimensional analysis. If you have different dimensions on two sides of an equation, you have done something wrong.

5.5 The trivial transformation: identity matrix

The identity matrix, $\mathbb{1}$ leaves vectors unchanged: $\mathbb{1}\mathbf{v} = \mathbf{v}$. You already know its components in \mathbb{R}^3 :

$$\mathbb{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.4)$$

Sometimes we will indicate the dimension of the space by writing $\mathbb{1}_3$ or $\mathbb{1}_{3\times 3}$ when we want to be extra clear. There is a useful way of writing the components of the identity matrix. Define the **Kronecker** δ :

$$\delta_j^i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \quad (5.5)$$

Then we say that the components of $\mathbb{1}$ are δ_j^i , no matter what dimension of space we are in.

But wait, you object, *the indices are all messed up!* You are right. The indices on δ_j^i look wrong because we made a big deal that matrices have a first index and a second index that we must distinguish between. The first index is written upper, and the second index is lower. So should we not have written δ_j^i , with the j clearly to the right of the i ?

Yes, this is technically correct. The reason why we made a big deal about this and wrote M_j^i instead of M_i^j is that we will eventually want to consider alternative configurations of indices. For example, T_{ij} or S_i^j or even weirder things like $R^\mu_{\nu\rho\sigma}$. It turns out that for the identity matrix, there is no ambiguity because

$$\delta_j^i = \delta^i_j = \delta_i^j. \quad (5.6)$$

So for convenience, we will just write δ_j^i . Observe, however, that we *never* write δ_{ij} . There is no problem defining δ_{ij} in the same way, but it turns out that this combination rarely shows up.²⁴

It is useful to think about the Kronecker δ as a machine that converts indices. Consider what it does to a vector:

$$(\mathbb{1}\mathbf{v})^i = \delta_j^i v^j = \delta_1^i v^1 + \delta_2^i v^2 + \delta_3^i v^3. \quad (5.7)$$

On the right-hand side, there is exactly one non-zero term because δ_j^i is zero for all terms except when $j = i$. That means

$$\delta_j^i v^j = v^i. \quad (5.8)$$

The right-hand side is, of course, what we expected since $\mathbb{1}\mathbf{v} = \mathbf{v}$. So we can train ourselves to “mechanically” interpret the δ_j^i object as one that replaces indices.

²⁴Instead, there is a different mathematical structure called the metric, g_{ij} that has two lower indices. Sometimes the metric components may be equal to the Kronecker δ , so we may say $g_{ij} = \delta_{ij}$. But in those cases we need to be clear that we are using the metric that happens to be the identity, not that we are multiplying by the identity. Why the big deal? One example is that the metric is where gravity lives—so it would be a rather big omission if we mistook it for “multiplying by one.”

Exercise 5.4. Show that this works for lower indices as well. Consider the multiplication of two matrices, $M\mathbb{1}$. Show that

$$M^i_j \delta_k^j = M^i_k , \quad (5.9)$$

so that the δ_k^j simply replaces the j index on M^i_j and turns it into a k .

Exercise 5.5. Even though we have not introduced any objects with more complicated index structure, suppose you have the following object:

$$R^\mu_{\nu\rho\sigma} . \quad (5.10)$$

The Greek letters are indices (like i and j), we have switched notation because this is the convention for spacetime. An object with this index structure shows up in general relativity, it is called the Riemann tensor. Without knowing any general relativity, write out the value of

$$\delta_\alpha^\mu \delta_\beta^\nu \delta_\gamma^\rho \delta_\tau^\sigma R^\mu_{\nu\rho\sigma} . \quad (5.11)$$

This is some kind of ‘matrix multiplication’ by ‘identity matrices’ hitting a $N \times N \times N \times N$ array of numbers, where $N = 4$ is the dimension of spacetime.

5.6 Composition of linear transformations

Suppose you have two linear transformations from $V \rightarrow V$, call them $f(\mathbf{x})$ and $g(x)$. These are represented by matrices M and N , respectively. Then the **composition** of these two linear transformations is

$$f \circ g(\mathbf{x}) \equiv f(g(\mathbf{x})) \Leftrightarrow f \circ g(\mathbf{x}) = MN\mathbf{x} \quad (5.12)$$

On the right-hand side we have written the matrix that represents $f \circ g$ as the product of the matrices for f and g : MN . Matrix multiplication follows the usual rules. We can write this using summation convention very nicely. The matrices have components and index structure M^i_j and N^i_j respectively. Each has an upper index and a lower index. The product MN has components

$$(MN)^i_j = M^i_k N^k_j . \quad (5.13)$$

Example 5.7. One thing that we notice from (5.13) is that the index structure makes it clear that the order of matrix multiplication matters.

$$(MN)^i_j = M^i_k N^k_j \neq N^i_k M^k_j = (NM)^i_j . \quad (5.14)$$

Thus the order in which two linear transformations are composed matters. On the other hand, the order in which components (which are just numbers) are multiplied does not matter. For example:

$$(MN)^i_j = M^i_k N^k_j = N^k_j M^i_k . \quad (5.15)$$

The sums on the right-hand side both represent the same series of terms. Neither of those sums are the same as

$$(NM)^i_j = N^i_k M^k_j = M^k_j N^i_k . \quad (5.16)$$

This is the kind of manipulation that is pretty obvious at its core, but the first time you see this manipulation done in textbook it can be pretty disorienting.

The above example makes the following point clear: The order in which components are multiplied is *not* the same as the order in which the linear transformations are applied. That is, the order of the components are multiplied not the same as the order in which the matrices are multiplied.

Example 5.8. If you want to recover the “matrix multiplication” MN from the “component contraction” $M^i_k N^k_j = N^k_j M^i_k$, you should start by writing out the terms so that neighboring indices are contracted. The expression $M^i_k N^k_j$ has neighboring k indices contracted, while the equivalent expression $N^k_j M^i_k$ does not. Thus we take $M^i_k N^k_j$ and then we can “promote” this back to a matrix multiplication, MN . A caveat: physicists often prefer to work with component contractions instead of matrix multiplications. In fact, with more complicated tensor contractions, it is not always meaningful or possible to recover a “matrix multiplication” form.

Exercise 5.6. A **diagonal matrix** is one where the only non-zero components are along the diagonal. We will sometimes write diagonal matrices with a hat, \hat{M} , to indicate that they are diagonal. Sometimes we abbreviate the definition of a diagonal matrix by

only specifying its diagonal components:

$$\hat{M} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \text{diag}(2, 7, 4) . \quad (5.17)$$

Show that the product of diagonal matrices is also diagonal whose components are the products of the diagonal components. You can use (5.13) armed with the knowledge that $(\hat{M})_j^i = 0$ if $i \neq j$. Then you must show that $(\hat{M}\hat{N})_j^i = 0$ for $i \neq j$. Finally, check the values of the diagonal elements to show that

$$\hat{M}\hat{N} = \text{diag} (M^1{}_1 N^1{}_1, M^2{}_2 N^2{}_2, M^3{}_3 N^3{}_3) . \quad (5.18)$$

5.7 Linearity and Basis Vectors

A couple of useful exercises Here's a pair of exercises that are worth trying out:

Exercise 5.7. Calculate the following linear transformations by a matrix M on two different vectors, \mathbf{v} and \mathbf{w} :

$$M\mathbf{v} = \begin{pmatrix} 3 & 2 & 1 \\ -1 & 4 & -2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = ? \quad M\mathbf{w} = \begin{pmatrix} 3 & 2 & 1 \\ -1 & 4 & -2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} = ? \quad (5.19)$$

Exercise 5.8. That was tedious, wasn't it? What if I asked you to calculate three more...

$$M \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = ? \quad M \begin{pmatrix} 0 \\ 1 \\ 7 \end{pmatrix} = ? \quad M \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = ? \quad (5.20)$$

Did you do it? Go ahead, do these calculations. Use *Mathematica* if you must. Did you do it the hard way, or did you notice that there's an easy way?

General two-dimensional discussion The easy observation is that each of the vectors being acted on in (5.20) are **linear combinations** of \mathbf{v} and \mathbf{w} . What you should notice is that

$$\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = \mathbf{v} + \mathbf{w} \quad M \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = M(\mathbf{v} + \mathbf{w}) = M\mathbf{v} + M\mathbf{w} . \quad (5.21)$$

In other words: if $\mathbf{u} = \mathbf{v} + \mathbf{w}$ and we already know the action of a matrix M on \mathbf{v} and \mathbf{w} —that is, we know $M\mathbf{v}$ and $M\mathbf{w}$ —then we’re basically done. No further matrix multiplication is needed! We can simply use the linearity of M to take a linear combination of $M\mathbf{v}$ and $M\mathbf{w}$. You should check that this is true for the other matrix multiplications in (5.20).

This is a big deal. We know that \mathbf{v} and \mathbf{w} are the **basis vectors** for a subspace $V_2 = \text{Span}(\mathbf{v}, \mathbf{w})$. We have established that if we know the action of M on the basis vectors of V_2 , then we know the action of M on any vector in V_2 . This is because any vector in V_2 has the form

$$\mathbf{u} = a\mathbf{v} + b\mathbf{w} \quad (5.22)$$

for two numbers a and b . Then the action of M on any such vector is

$$M\mathbf{u} = a(M\mathbf{v}) + b(M\mathbf{w}) . \quad (5.23)$$

We can pose this in the following way. Suppose we have a two-dimensional vector space V_2 with some basis \mathbf{b}_1 and \mathbf{b}_2 . Now suppose we have a linear transformation M that goes from $V_2 \rightarrow V_2$. This is a matrix.

1. We know that the matrix is encoded by $2 \times 2 = 4$ numbers, each of the four elements, M^i_j . If you know these four elements, then you can matrix multiply any vector in V_2 by simply using the matrix multiplication rules. All you need to specify is the vector \mathbf{u} , which is encoded by its two components u^i . Then you get two numbers as output, $(Mu)^i = M^i_j u^j$.
2. Alternatively, we suppose that we do not know (or we deliberately forget) the four elements M^i_j . Instead, suppose we know the components of $M\mathbf{v}$ and $M\mathbf{w}$. These are each two component vectors, so this is a total of four numbers. Then we may specify \mathbf{u} not by its components u^i , by its coefficients a and b . Then we may use (5.23) to specify the two components of $M\mathbf{u}$.

In each picture, you specify the four numbers that encode the transformation and then use them to calculate the action of M on a generic vector \mathbf{u} .

Example 5.9. *The two approaches above are equivalent when the basis vectors are*

$$\mathbf{b}_1 = \hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathbf{b}_2 = \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \quad (5.24)$$

This should be clear since

$$M\hat{\mathbf{e}}_1 = \begin{pmatrix} M^1_1 \\ M^2_1 \end{pmatrix} \qquad M\hat{\mathbf{e}}_2 = \begin{pmatrix} M^1_2 \\ M^2_2 \end{pmatrix} . \quad (5.25)$$

The example above tells us that the ‘components of a matrix’ have a specific meaning: they encode the action of M on a particularly nice basis, $\hat{\mathbf{e}}_{1,2}$. This is significant for the following reason: while a matrix may encode a linear transformation, we should not settle into thinking that the 2×2 of numbers somehow *is* the matrix.

Example 5.10. M^1_1 is simply the coefficient of $\hat{\mathbf{e}}_1$ in the expansion of $M\hat{\mathbf{e}}_1$. M^1_2 is simply the coefficient of $\hat{\mathbf{e}}_2$ in the expansion of $M\hat{\mathbf{e}}_1$.

Another way of saying this is the following:

Key Idea 5.1. *The i^{th} column of a matrix are the components of $M\hat{\mathbf{e}}_i$. This assumes that $\hat{\mathbf{e}}_i$ is “canonical” basis which has the i^{th} element set to one and the rest of the elements are zero.*

General discussion The general version of this discussion is as follows. We write \mathbb{R}^N to be N -dimensional real space. A linear transformation from $\mathbb{R}^N \rightarrow \mathbb{R}^N$ is a matrix. Suppose we have a nice basis. In fact, the ‘obvious’ basis is

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \quad \dots \quad \hat{\mathbf{e}}_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (5.26)$$

Then we know that any vector in \mathbb{R}^N may be written as a linear combination of these basis vectors. In fact, this is *precisely* what we mean by the components of a vector: it is the coefficient of each $\hat{\mathbf{e}}_i$ basis vector in the linear combination that represents the vector:

$$\mathbf{v} = v^i \hat{\mathbf{e}}_i \quad \mathbf{v} \text{ “=} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^N \end{pmatrix}. \quad (5.27)$$

The first equality is *true*. It defines the meaning of v^i no matter what the $\hat{\mathbf{e}}_i$ are, whether they are columns or more abstract objects. The equation on the right has air quotes because it is not really an equation. What it means is \mathbf{v} is represented by the following column of numbers. The column of numbers simply mean the coefficients of the basis vectors—no matter what the basis vectors are.

In physics, we treat lots of different things as vectors. Momenta, electric fields, wavefunctions, and so forth. When we say that the electric field at a point is a vector, we do not mean that there is actually an invisible arrow with some length and direction there. Nor do we mean that the electric field at that point is a column of three numbers. What we mean is that these three numbers represent the components of the electric field with respect to a basis. The basis are usually the electric field in the x , y , and z directions.

If we know the action of M on the basis vectors, what that really means is that we know the components of $M\hat{\mathbf{e}}_i$ for each i . That is to say,

$$M\hat{\mathbf{e}}_i = (M\hat{\mathbf{e}}_i)^k \hat{\mathbf{e}}_k \quad M^k_i \equiv (M\hat{\mathbf{e}}_i)^k. \quad (5.28)$$

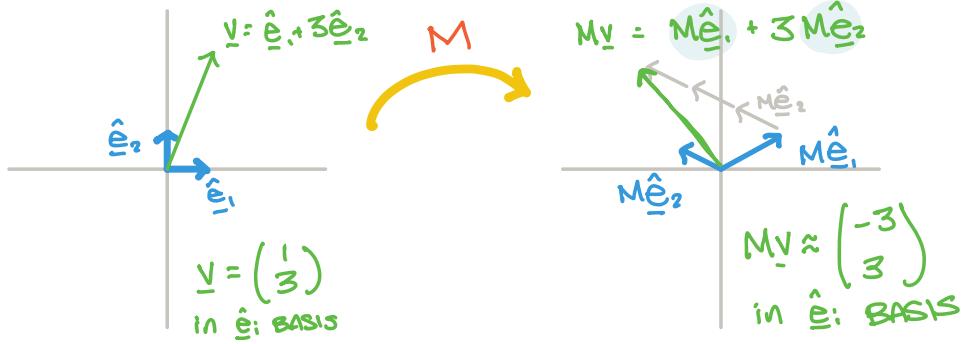


Figure 4: M is a linear transformation from $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. If we know the action of M on the basis vectors $\hat{\mathbf{e}}_i$ of \mathbb{R}^2 , then we know the action of M on any vector \mathbf{v} .

Be very careful with this. What we mean is the N -dimensional generalization of (5.25), so refer back to that if the notation is cumbersome. What is important here is that we now have a definition of the components M_j^i that is fully general.

5.8 Linear transformations as maps acting on basis vectors

The relation between the components of linear transformations and basis vectors is elucidated in Fig. 4. A vector is a linear combination of the basis vectors, so the action of a linear transformation M on a vector \mathbf{v} reduces to taking linear combinations of the action of M on the basis vectors $\hat{\mathbf{e}}_i$:

$$\mathbf{v} = v^i \hat{\mathbf{e}}_i M \mathbf{v} = v^i M \hat{\mathbf{e}}_i . \quad (5.29)$$

There are several observations:

- We could define a basis $\mathbf{f}_i \equiv M \hat{\mathbf{e}}_i$. Then the components of \mathbf{v} in the $\hat{\mathbf{e}}_i$ basis are the same as the components of $M\mathbf{v}$ in the \mathbf{f}_i basis.
- The components of $M\mathbf{v}$ in the $\hat{\mathbf{e}}_i$ basis are *not* the v^i , but rather the linear combinations $M_j^i v^j$. We know this from our matrix multiplication rules, but we can also see it from the linear combination $v^i M \hat{\mathbf{e}}_i$ and the observation that the matrix components M_j^i are simply the components of $M \hat{\mathbf{e}}_i$, as we saw in (5.25).
- What is nice about (5.29) is that we have separated the *components* of a vector v^i from the “vector-y” part, $M \hat{\mathbf{e}}_i$. What we mean by this is that the *components* are just numbers. They are coefficients of a linear combination and are what we write as the components of a column of numbers. However, vectors are not always “columns of numbers.” We want to abstract what a vector is and thus we need to also abstract what a linear transformation is. All of that abstraction is encoded in the basis vectors and the definition of the transformation acting on the basis vectors.

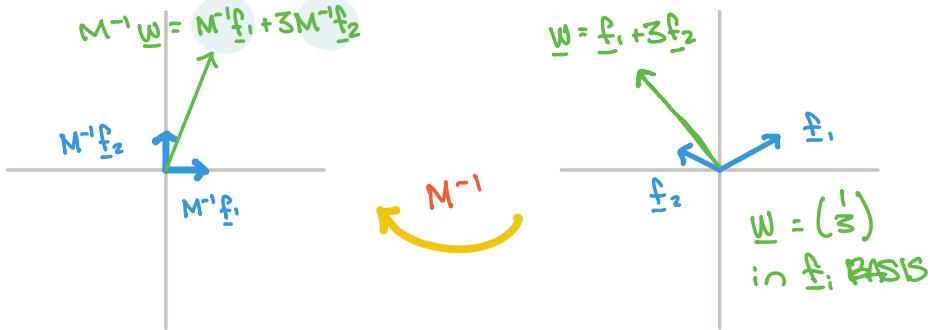


Figure 5: Let M is a linear transformation from $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, as we drew in Fig. 4. Now assume that you have a vector \underline{w} that is the linear combination of some basis vectors \underline{f}_i , see the right-hand side of the sketch. If the \underline{f}_i can be obtained as the transformation by M of some other vectors—the basis vectors we called \underline{e}_i in Fig. 4—then the inverse transformation M^{-1} acts as a map “back” to the original basis.

5.9 Inverse transformations

Let M be a linear transformation from $V \rightarrow V$. This is just to say that M is a matrix that acts on vectors in V . For “nice” matrices, M , there is a unique **inverse** matrix (inverse transformation) M^{-1} with components $(M^{-1})_j^i$. The defining property of the matrix inverse is $M^{-1}M = \mathbb{1}$. That means that $M^{-1}(M\underline{v}) = \underline{v}$ for any vector \underline{v} . The matrix inverse simply undoes the transformation M . This is shown in Fig. 5.

Given a basis, matrices are described by their components, M_j^i . How do you find the components of the inverse matrix, $(M^{-1})_j^i$? The most straightforward way is the hard way. Let us consider the 2×2 case. We know that the defining relationship is

$$M(M^{-1}) = (M^{-1})M = \mathbb{1} \quad M_k^i(M^{-1})_j^k = (M^{-1})_k^i M_j^k = \delta_j^i. \quad (5.30)$$

On the right we have four equations for each combination of $i, j \in \{1, 2\}$. We also have four unknown components, $(M^{-1})_j^i$. This is a system of linear equations that one can usually solve. Sometimes the system will not be solvable—in that case the matrix M is not invertible.

Exercise 5.9. Convince yourself that (5.30) holds for any vector space of any dimension, N . You end up with N^2 equations for N^2 unknown components M_j^i . Does this change if all of the components are allowed to be complex numbers?

The product of two matrices M and N is a matrix, MN . The inverse of the product of matrices is the product of their inverses *in the reverse order*

$$(MN)^{-1} = N^{-1}M^{-1}. \quad (5.31)$$

You can that this is because you have to undo the transformation that was last applied:

$$(MN)^{-1}MN = N^{-1}M^{-1}MN = N^{-1}\mathbb{1}N = \mathbb{1}. \quad (5.32)$$

Example 5.11. The inverse of a diagonal matrix is easy:

$$M = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \quad M^{-1} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \quad (5.33)$$

We have used the earlier observation in Exercise 5.6 that the product of diagonal matrices \hat{M} and \hat{N} is also a diagonal matrix whose elements are the products of the corresponding elements in \hat{M} and \hat{N} .

Physical states are often described by vectors. The dynamics of a theory of physics are then described by linear transformations of vectors. Usually this is some description of the change in the vector over some infinitesimal time. This change is a response to some source of the dynamics. You end up with formulas that always look like

$$M\mathbf{v} = \mathbf{s} \quad \Leftrightarrow (\text{dynamics})(\text{state}) = (\text{source}). \quad (5.34)$$

The source is something like a pebble falling into a pond. The dynamics are a matrix that describes how ripples propagate. The state is the configuration of the water in the pond. The problem that you have to solve in physics is usually framed as follows: If you are given some source \mathbf{s} , what is the value of the state \mathbf{v} ? This requires taking the inverse of the dynamics M .

Thus far we have seen two things: first, finding the components of an inverse matrix is generally very annoying. Secondly, there is at least one case—when the matrix is diagonal—where the annoying task is significantly simplified. It turns out that there are more clever ways to invert a matrix that often apply for the types of matrices that show up in physics. These tricks boil down to converting matrices into their diagonal form.

5.10 Matrices that aren't invertible: Projections

Solving for the components of a matrix inverse boils down to N^2 equations for N^2 unknowns. This does not mean that those equations have a solution. A simple example is the case of a diagonal matrix with one component zero. We illustrate this in Fig. 6 Thinking of the matrix M as a map, we see the the problem is one of uniqueness. A linear map should take each vector \mathbf{v} into a single, well-defined vector $M\mathbf{v}$. M is suspicious because it sends multiple different vectors to the *same* vector, say $M\mathbf{v} = M\mathbf{w}$. That's fine. Each of \mathbf{v} and \mathbf{w} are sent to a vector. The problem shows up when we try to take the inverse. If we write $\mathbf{z} = M\mathbf{v} = M\mathbf{w}$, we know that

$$(M^{-1})\mathbf{z} = (M^{-1})M\mathbf{v} = \mathbf{v} \quad (M^{-1})\mathbf{z} = (M^{-1})M\mathbf{w} = \mathbf{w}. \quad (5.35)$$

However, we assumed that $\mathbf{v} \neq \mathbf{w}$. This means makes the above line inconsistent. The inverse transformation is not defined. Bummer.

Fortunately, most of the *dynamics* in physics *is* invertible. That does not mean that we can ignore non-invertible matrices, though. Non-invertible matrices can be understood as **projections** onto a subspace. You can see this in the example in Fig. 6. The map M is essentially only using the information about the second component of its input vector and

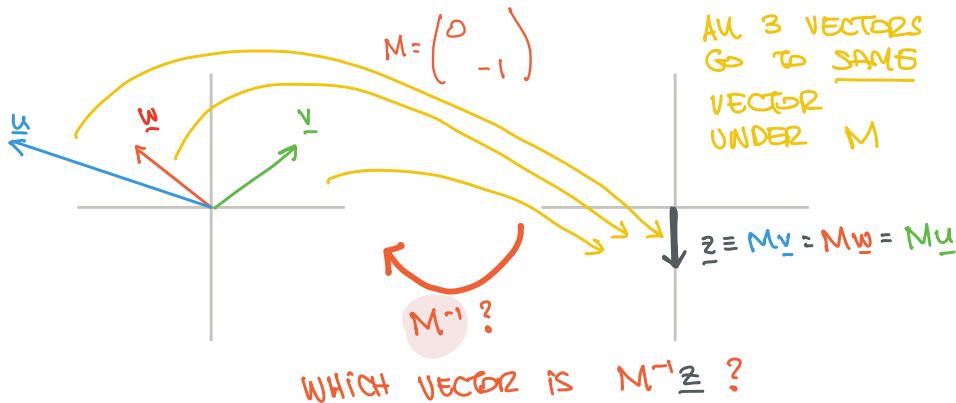


Figure 6: M is a non-invertible matrix. We notice that there's no problem applying M to vectors. The odd thing is that there are multiple vectors— \mathbf{v} , \mathbf{w} , and \mathbf{u} in the picture—that all turn into the same vector, $M\mathbf{v} = M\mathbf{w} = M\mathbf{u} \equiv \mathbf{z}$. The problem shows up when we try to “undo” the transformation M by applying M^{-1} to \mathbf{z} . $M^{-1}\mathbf{z}$ should be a vector. Instead, we have at least three different valid options: \mathbf{v} , \mathbf{w} , and \mathbf{u} . This means that M^{-1} is not a well defined linear map: it can send one vector \mathbf{z} into many possible vectors.

mapping it to just the second component of the output vector. That means that M is really just a map to a one-dimensional subspace. In this way, the two dimensional vector space \mathbb{R}^2 is **projected** onto a one-dimensional subspace of \mathbb{R}^2 . In that example, the one dimensional space is defined by vectors of the form

$$\mathbf{v} = \begin{pmatrix} 0 \\ v^2 \end{pmatrix} . \quad (5.36)$$

Projections are really useful in physics. Sometimes you want to take a big vector space and only focus on a subspace that satisfy certain conditions. For example, it turns out that in nature you can have two types of massless matter particles: left-handed and right-handed according to their quantum spin in the direction of their motion. One may want to project onto only the space of left-handed particles because that is easier to deal with than the entire space. Because projections necessarily “throw out” information, they are non-invertible.²⁵

Example 5.12. One of the most curious puzzles in fundamental physics over the last half century is the black hole information paradox.^a One way of stating the paradox is the question of whether or not “falling into a black hole” is a projection that throws out information.

^a<https://www.quantamagazine.org/the-most-famous-paradox-in-physics-nears-its-end-20201029/>

²⁵I think this was more intuitive to past generations of physicists. When you throw out a piece of paper, it ends up in a recycling plant and it gets torn up and is lost forever. Since the advent of computers, when you “throw out” a file, you can usually control/command-z and undo to bring it back.

5.11 Funny things you can do to matrices

Transpose

The **transpose** of a matrix is what happens when you ‘flip all the elements along the diagonal.’ Given a matrix M , the transpose of M is M^T with components

$$(M^T)_{ij}^i = M_j^i . \quad (5.37)$$

Example 5.13. The following two matrices are transposes of each other.

$$M = \begin{pmatrix} 9 & -3 & 2 \\ 2 & 5 & 7 \\ 1 & 0 & 3 \end{pmatrix} \qquad M^T = \begin{pmatrix} 9 & 2 & 1 \\ -3 & 5 & 0 \\ 2 & 7 & 3 \end{pmatrix} . \quad (5.38)$$

This unusual operation will turn out to be rather useful for us. I leave the following cryptic statement: the transpose of a matrix is *not* its inverse, but it does have a *dual* relationship to the original matrix. Once we define a metric, we will be able to give a more rigorous definition for transpose. For now, let use take (5.37) as the definition of something we can ‘do’ to matrices.

Trace

The **trace** of a matrix M is simply the sum of its diagonal components:

$$\text{Tr } M = M_i^i = M_1^1 + M_2^2 + \dots . \quad (5.39)$$

It is not obviously meaningful from the perspective of M being a linear transformation. However, it ends up being useful as a mechanical procedure that one can do to matrices. The reason for this is something we will see soon: the trace of a matrix is unchanged under rotations.

Determinants

Here’s another seemingly strange operation that you can do on a matrix. The determinant of a 2×2 matrix is defined to be

$$\det M \equiv M_1^1 M_2^2 - M_1^2 M_2^1 . \quad (5.40)$$

Weird, right? Determinants for higher-dimensional square matrices can be defined recursively by taking specific linear combinations of sub-matrices. There are silly rules for this. There are fancy words for the sub-matrices (minors) and the appropriate sign for summing together those determinants (cofactors). Look, it’s a mess. You should be able to calculate the determinant of a general 2×2 matrix because it’s easy. In case of national crisis, you should be able to calculate the determinant of a general 3×3 matrix after consulting a reference to double

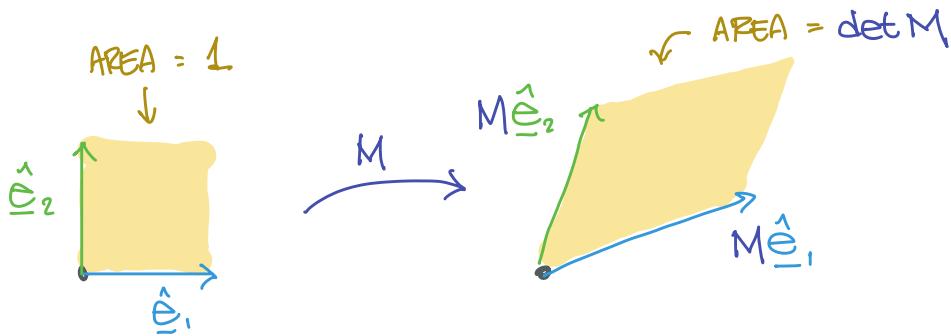


Figure 7: The transformation by M of a unit square with sides \hat{e}_i is a parallelogram with sides $M\hat{e}_i$. The determinant of M is the area of the parallelogram. In more than three dimensions the shape is called a parallelepiped, which is a fun word to say.

check the signs. However, at this stage we will not make a big deal about determinants. I think engineering classes make a big deal about determinants because they can help with solving systems of linear equations using matrices—but *that's not the kind of linear algebra we're doing in physics*.

Example 5.14. *The determinant of a diagonal matrix is simply the product of each of its diagonal elements.*

Determinants of a matrix are at least as curious as its trace because the determinant of a matrix turns out to also be unchanged under rotations. Furthermore, in multivariable calculus, the determinant shows up in the volume element. In fact, there is a formal definition of the determinant that comes from differential geometry where we use the determinant to rigorously define volumes of curved spaces. This sort of thing shows up in general relativity. One way of starting to see why this is plausible is the observation that the determinant of a matrix M is equal to the area of the parallelogram whose sides are given by the transformed basis vectors, $M\hat{e}_i$. We sketch this in two dimensions in Fig. 7. To prove this claim in two dimensions, we enclose the parallelogram into a box and start figuring out the area of various sub-shapes. Let A_r be the area in the two red triangles, A_b be the area in the two green triangles, and A_g be the area of the two blue boxes. These areas are—accounting for the factor of two because there are two of each shape:

$$A_r = M^1_1 M^2_1 \quad A_g = M^2_2 M^1_2 \quad A_b = 2M^1_2 M^2_1 . \quad (5.41)$$

The total area of the large box is

$$A_{\text{tot}} = (M^1_1 + M^1_2)(M^2_2 + M^2_1) \quad (5.42)$$

$$= M^1_1 M^2_2 + M^1_1 M^2_1 + M^1_2 M^2_2 + M^1_2 M^2_1 \quad (5.43)$$

$$= M^1_1 M^2_2 + A_r + A_g + \frac{1}{2} A_b . \quad (5.44)$$

The area of the parallelogram is simply the area of the large box minus the areas of the

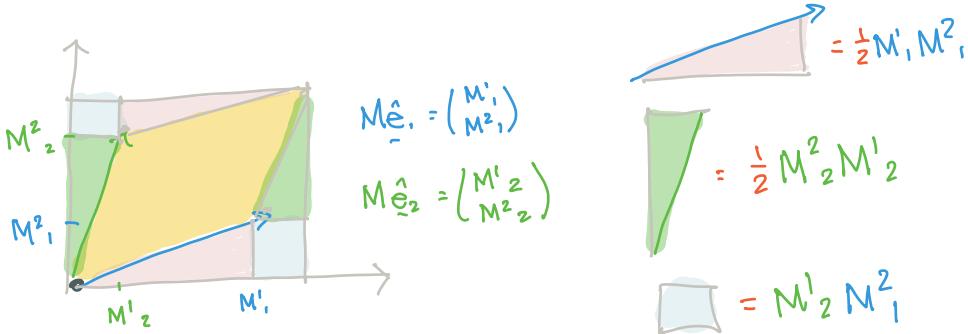


Figure 8: Graphical aid for understanding the area of the parallelogram whose sides are $M\hat{\mathbf{e}}_i$.

non-parallelogram shapes, $A_{\text{tot}} - A_r - A_g - A_b$:

$$A_{\text{parallelogram}} = M^1_1 M^2_2 - \frac{1}{2} A_b = M^1_1 M^2_2 - M^1_2 M^2_1 = \det M . \quad (5.45)$$

And so we have shown the result in a simple case: determinants are related to areas. Specifically, we see that the determinants are related to the areas (volumes) of parallelograms (parallelepipeds) with sides $M\hat{\mathbf{e}}_i$.

This interpretation of determinants as volumes²⁶ also motivates some further expressions for the determinant of the product of matrices and the determinant of the inverse of a matrix. If we perform subsequent transformations with two matrices, N and M , then the unit parallelepiped is first mapped to a parallelepiped of volume $\det N$. When we apply the second matrix M , this intermediate parallelepiped is mapped to a third parallelepiped whose volume is rescaled by $\det M$. The resulting volume is $(\det N)(\det M)$. We thus deduce that the determinant of a product of matrices is the product of the determinants of each matrix:

$$\det NM = (\det N)(\det M) . \quad (5.46)$$

Going further, because $MM^{-1} = \mathbb{1}$, we have

$$\det MM^{-1} = (\det M)(\det M^{-1}) = \det \mathbb{1} = 1 \quad \Rightarrow \quad \det M^{-1} = \frac{1}{\det M} . \quad (5.47)$$

We have been deliberately hand-wavy to motivate these relations for determinants. If we were to be more rigorous, they are rather tricky to prove, though you can do them by brute force by calculating components.

5.12 Matrices that aren't square: maps between vector spaces

You may be wondering why all of our linear transformations are square. After all, you have probably seen rectangular matrices where the number of rows does not equal the number of columns. For example

$$W = \begin{pmatrix} 1 & 2 & -6 & 7 \\ 6 & -3 & 3 & 0 \end{pmatrix} . \quad (5.48)$$

²⁶I now say ‘volumes’ instead of ‘area’ because what follows is general to any dimension of space.

Just by looking at this, you can tell that W acts on some kind of vector with four components and outputs a vector with two components. Assuming these vectors are part of \mathbb{R}^4 and \mathbb{R}^2 respectively, this means that W is a map from \mathbb{R}^4 to \mathbb{R}^2 , or

$$W : \mathbb{R}^4 \rightarrow \mathbb{R}^2 . \quad (5.49)$$

Non-square matrices are maps between different vector spaces. In this course we will have very little to say about non-square matrices. The type of linear algebra that comes up over and over again in physics are transformations from some vector space V to itself.²⁷

The one thing we have observed about non-square matrices is that the number of columns is the maximum dimensionality of the input space, and the number of rows is the maximum dimensionality of the output space.²⁸ The dimensionality tells us about how much information is in a vector. A three-dimensional vector space has “three numbers worth” of information, the three components relative to a basis. This means that:

1. If the number of columns is larger than the number of rows, then the output of the transformation may have less information than the input. For example, two different input vectors may map to the same output vector.
2. If the number of rows is larger than the number of columns, then input of the transformation may not be able to encode enough information to populate every possible vector in the output space. For example, there may be vectors \mathbf{v}_{out} in the output space that cannot be written as $W\mathbf{u}_{\text{in}}$.

There are some fancy words that are associated with these ideas. The **kernel**²⁹ of a transformation M is the set of all vectors \mathbf{v} such that $M\mathbf{v} = 0$. The **image** of a transformation is the set of all vectors \mathbf{v} that can be obtained by acting with the transformation M , that is: there is some \mathbf{u} such that $\mathbf{v} = M\mathbf{u}$.

Exercise 5.10. *Using the definition of a vector space, show that the kernel of M and the image of M are vector spaces.*

From this, we have argued that if $W : V_1 \rightarrow V_2$ is a linear transformation between two different vector spaces of different dimension, then

1. If the dimension of V_1 , $\dim V_1$, is larger than the dimension of V_2 , $\dim V_2$, then the kernel of W contains more than just the zero vector.

²⁷To be more technically precise: the maps tend to be between vector spaces V and V' that are isomorphic to one another, $V \cong V'$. This means that the vector spaces may be different, but there is a unambiguous translation between the two.

²⁸Why do we specify ‘maximum’? You could imagine a vector space that is two dimensional but happens to be written in a silly way as three-component vectors. For example, we could consider the space of vectors spanned by the x - and y -basis vectors but written as three component vectors in \mathbb{R}^3 . That space is two dimensional, but the vectors have three components—one of which is always zero.

²⁹The word kernel is really annoying here because it shows up in another place in the theory of distributions. Much to my chagrin, we will argue that distributions are an ingredient in the linear algebra of infinite dimensional vector spaces. This means that there is a place later in this course where the word *kernel* will come up again but will mean something completely unrelated.

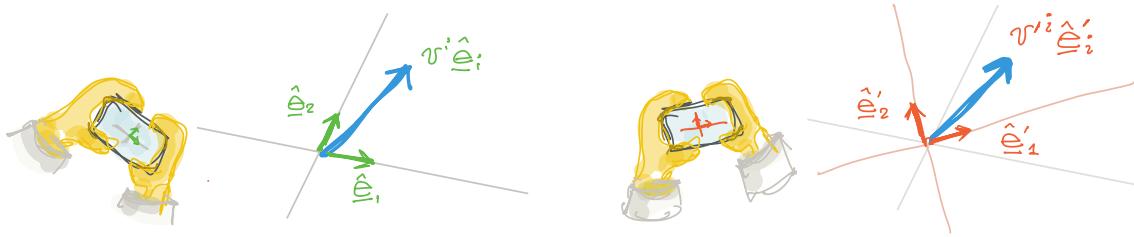


Figure 9: Example of a passive rotation. You are looking at a vector \mathbf{v} through your phone. The image on your phone has a natural orientation for a basis $\hat{\mathbf{e}}_i$ and the components of \mathbf{v} in this basis are v^i . You could also rotate your phone, leaving the vector \mathbf{v} untouched. The new basis vectors are $\hat{\mathbf{e}}'_i$; they are different from $\hat{\mathbf{e}}_i$ because your phone has changed orientation. With respect to the new basis, the same vector \mathbf{v} now has components v'^i .

2. Alternatively, if $\dim V_2 > \dim V_1$, then the image of W is a subspace of V_2 that does not contain all of V_2 .

6 Rotations

Rotations are a special class of transformations. We wait until Section 9 to really appreciate why rotations are special. For now, you may rely on your ordinary intuition for what it means to *rotate* something. If a vector is an arrow—*I know, this is the kindergarten picture of vectors*—then you rotate it by pinning its tail to the origin and then moving the arrowhead while leaving the length unchanged.

A more important notion of rotation is the idea of a change of basis between ‘nice’ basis vectors—a notion that we formalize when we have a dot (inner) product. The physical manifestation of this is the idea of a *passive* rotation. This corresponds in physics to a *change in reference frame*. We illustrate this in Fig. 9.

6.1 Rotations in Two-Dimensions

In two dimensions, there is only one class of rotation matrix. It has one parameter, θ , the amount that one is rotating. We conventionally measure rotations in the counter-clockwise direction. Rotations are linear transformations, and thus are represented by matrices. The form of a rotation matrix in two dimensions is³⁰

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} c_\theta & -s_\theta \\ s_\theta & c_\theta \end{pmatrix}, \quad (6.1)$$

where on the right we have introduced some shorthand for the trigonometric functions. The key relation of the trigonometric functions is

$$\sin^2 \theta + \cos^2 \theta = c_\theta^2 + s_\theta^2 = 1. \quad (6.2)$$

³⁰By convention we define a rotation by θ to be in the counter-clockwise direction.

The inverse of a rotation is simply a rotation in the opposite direction. This means that the inverse of the rotation matrix $R(\theta)$ is simply $R^{-1}(\theta) = R(-\theta)$. Because cosine is an even function and sine is an odd function, this means

$$R^{-1}(\theta) = \begin{pmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{pmatrix} = R(\theta)^T . \quad (6.3)$$

The last equality may be a little surprising, but it is of utmost significance. The inverse of a rotation is the transpose of the original rotation.

Exercise 6.1. Check component-by-component that $R(\theta)R^{-1}(\theta) = 1$ using the expressions above. One may directly show that for each of the four combinations of i and j ,

$$R(\theta)^i{}_k R^{-1}(\theta)^k{}_j = \delta_j^i . \quad (6.4)$$

Example 6.1. A physically motivated (and much simpler) way of writing the inverse rotation is to recall that to undo a counterclockwise rotation by angle θ , one does a clockwise rotation by θ . This is simply a counterclockwise rotation by angle $-\theta$. Thus

$$R^{-1}(\theta) = R(-\theta) = \begin{pmatrix} \cos(-\theta) & \sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{pmatrix} = R(\theta)^T . \quad (6.5)$$

We used the fact that sine is odd, $\sin(-\theta) = -\sin \theta$ and cosine is even, $\cos(-\theta) = \cos \theta$.

We have thus proven the existence of an inverse rotation by explicitly constructing the inverse of the rotation matrix. This makes sense: rotations are “nice” transformations that clearly do not lead to a loss of information. After all, the rotation of a vector is simply what you would see by looking at the vector and tilting your head a little. This remark will lead us to the idea of a *passive* transformation.

Exercise 6.2. Consider a point on the Cartesian plane (x, y) . Derive the rotation matrix (6.1) by determining the rotated coordinates (x', y') as a function of x, y , and the rotation angle θ . I suggest writing $x = r \cos \phi$ and $y = r \sin \phi$. Then use the fact that under a rotation, radius stays the same and angles add. Thus

$$x' = r \cos(\phi + \theta) \qquad \qquad y' = r \sin(\phi + \theta) . \quad (6.6)$$

Then invoke the angle addition formulae for trigonometric functions.

Exercise 6.3. Show that the determinant of the 2×2 rotation matrix is one.

6.2 Transformation of Vectors

What can we do with a rotation matrix? Why, rotate vectors, of course! In this respect, $R(\theta)$ is just like any other linear transformation. Here it is in matrix notation. A vector \mathbf{v} transforms into a vector $\mathbf{v}' = R(\theta)\mathbf{v}$ with components

$$\begin{pmatrix} v'^1 \\ v'^2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} \cos \theta v^1 - \sin \theta v^2 \\ \sin \theta v^1 + \cos \theta v^2 \end{pmatrix}. \quad (6.7)$$

Here it is in component notation:

$$v'^i = (R(\theta)v)^i = R(\theta)^i_j v^j. \quad (6.8)$$

We can observe something significant:

Key Idea 6.1. A generic component of a vector has an upper index, v^i . The component of the rotated vector under a rotation $R(\theta)$ has generic component $R(\theta)^i_j v^j$. That is: we take the rotation matrix and contract its second index (which is lower) with the upper index of the vector.

Thus far, rotations are just some linear transformation. We have not yet defined the features that make rotations special in Euclidean space. Indeed: we are not in Euclidean space yet, we are simply in \mathbb{R}^3 , the class of vector spaces described by three real numbers. Once we have a metric (dot product) then we can say that rotations preserve the dot product. A rotated vector does not change length under a rotation. Angles between vectors—which we saw in (3.14) are also defined by the dot product—are also unchanged under rotations.

6.3 Transformation of numbers

Numbers, also known as scalars, do not transform under rotations. Here's a useful phrase: we say that numbers are **invariant** under rotations. Here's a prescient way of saying the same thing a third time:

Key Idea 6.2. An object with no indices does not transform under rotations.

6.4 Transformation of Row Vectors

What about row vectors? Row vectors also transform under rotations. In matrix notation the transformation looks like this: $\underline{\mathbf{w}} \rightarrow \underline{\mathbf{w}'} = \underline{\mathbf{w}} R(\theta)^{-1}$,

$$\begin{pmatrix} w'_1 & w'_2 \end{pmatrix} = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (6.9)$$

In component notation:

$$w'_i = (wR(\theta)^{-1})_i = w_j (R(\theta)^{-1})^j_i = (R(\theta)^{-1})^j_i w_j. \quad (6.10)$$

One can of course write the components of $R(\theta)^{-1}$ in terms of those of $R(\theta)$, but it turns out to be convenient to keep track of when we have $R(\theta)$ versus $R(\theta)^{-1}$. Observe that by the usual rules of matrix multiplication, the row vector needs to be on the left of the matrix. That is, $\underline{\mathbf{w}} R(\theta)^{-1}$ makes sense, while $R(\theta)^{-1} \underline{\mathbf{w}}$ does not. However, on the right side of (6.10) we have done something slippery: we have written the *component* $(R(\theta)^{-1})^j_i$ to the *left* of the *component* w_j . This is totally allowed, since *components are just numbers*. They carry no vector-y (“tensorial”) structure themselves. Indeed, this is one of the benefits of component notation.

At this point you should be very perplexed: *why* does the row vector transform *oppositely* to the column vector? Thus far we have simply stated that “this is the way.³¹” However, consider the action of a row vector on a vector: $\underline{\mathbf{w}} \mathbf{v}$. This is simply a number, and as you know from Section 6.3, *numbers do not transform under rotations*. Is this consistent with $\underline{\mathbf{w}}$ and \mathbf{v} transforming? Let us quickly confirm. In matrix notation, we have:

$$\underline{\mathbf{w}} \mathbf{v} \rightarrow \underline{\mathbf{w}'} \mathbf{v}' = \underline{\mathbf{w}} R(\theta)^{-1} R(\theta) \mathbf{v} = \underline{\mathbf{w}} \mathbf{v}. \quad (6.11)$$

Here we have simply used $R^{-1}R = \mathbb{1}$. We can write the same thing in component notation;

$$w'_i v^i = (R(\theta)^{-1})^k_i R(\theta)^i_j w_k v^j = \delta_j^k w_k v^j = w_k v^k. \quad (6.12)$$

We used (5.30) and (5.8) to write the product of a matrix with its inverse in terms of the Kronecker δ , and then to use the Kronecker δ to collapse the lower j index into a lower k index. Let us be clear that we were able to ‘pull out’ the *components* of the rotation matrices because they are simply numbers. Further, let us be clear that the components of $R(\theta)$ and $R(\theta)^{-1}$ are contracted in specifically the right way to yield the Kronecker δ :

$$(R(\theta)^{-1})^k_i R(\theta)^i_j = \delta_j^k. \quad (6.13)$$

What we see is that the transformation law for row vectors, (6.10), is precisely what it needs to be so that the number $\underline{\mathbf{w}} \mathbf{v}$ is invariant under a rotation. Further, we observe something significant:

³¹<https://knowyourmeme.com/memes/this-is-the-way>

Key Idea 6.3. A generic component of a row vector has a lower index, w_i . The component of the rotated row vector under a rotation $R(\theta)$ has generic component $(R(\theta)^{-1})^j{}_i w_j$. That is: we take the inverse rotation matrix and contract its first index (which is upper) with the lower index of the vector.

6.5 Transformation of Matrices

Do matrices transform under a rotation? Matrices are linear maps from vectors to vectors, so we know that $M\mathbf{v}$ is a vector. That means that $M\mathbf{v}$ transforms the way any other vector transforms under a rotation $R(\theta)$:

$$M\mathbf{v} \rightarrow (M\mathbf{v})' = R(\theta) M\mathbf{v} \quad (M\mathbf{v})'^i = R(\theta)^i{}_j M^j{}_k v^k. \quad (6.14)$$

Note that $R(\theta)M \neq MR(\theta)$ because these are matrices. Similarly, on the right-hand side

$$R(\theta)^i{}_j M^j{}_k v^k \neq M^i{}_j R(\theta)^j{}_k v^k, \quad (6.15)$$

although it *is* true that (note the difference in indices compared to above)

$$R(\theta)^i{}_j M^j{}_k v^k = M^j{}_k R(\theta)^i{}_j v^k. \quad (6.16)$$

In the spirit of (6.11), we would like to write the transformed vector $(M\mathbf{v})'$ in terms of a transformed matrix M' acting on the transformed vector \mathbf{v}' . We define the transformed matrix M' this way:

$$M'\mathbf{v}' = R(\theta) M\mathbf{v} = R(\theta) MR(\theta)^{-1}\mathbf{v}'. \quad (6.17)$$

In component notation, this is:

$$(M')^i{}_\ell(v')^\ell = R(\theta)^i{}_j M^j{}_k (R(\theta)^{-1})^k{}_\ell(v')^\ell. \quad (6.18)$$

We have been very clever and chose to write the dummy indices on the left-hand side to match the dummy indices of the last contraction on the right hand side. We now claim that we can “cancel the $(v')^\ell$ ” on both sides of this equation. This can be a very subtle point the first time you see it. The crux of the argument is that (6.18) holds *for any* vector \mathbf{v}' , and thus the equality cannot depend on the components of \mathbf{v}' . That means that the equality is really an equality between the matrices acting on \mathbf{v}' . To appreciate this subtlety, please see Example 6.2 below.

Having performed the slick step of canceling out $(v')^\ell$, we have the expression for the components of the transformed matrix, M' :

$$(M')^i{}_\ell = R(\theta)^i{}_j M^j{}_k (R(\theta)^{-1})^k{}_\ell = R(\theta)^i{}_j (R(\theta)^{-1})^k{}_\ell M^j{}_k \quad M' = R(\theta)MR(\theta)^{-1}. \quad (6.19)$$

Observe that the rotation matrices in M' are precisely what they must be to ensure that the number $\underline{\mathbf{w}}M\mathbf{v}$ is invariant:

$$\underline{\mathbf{w}}M\mathbf{v} \rightarrow \underline{\mathbf{w}}'M'\mathbf{v}' = \underline{\mathbf{w}}R(\theta)^{-1}R(\theta) M R(\theta)^{-1}R(\theta)\mathbf{v} = \underline{\mathbf{w}}M\mathbf{v}. \quad (6.20)$$

We recognize two pairs of matrix-inverse matrix products. We could write out the long expression in terms of indices, but I fear we would run out of letters. What is catches our eye in (6.19) is the continuation of the string of ideas in Key Ideas 6.1, 6.2, and 6.3:

Key Idea 6.4. A generic component of a matrix has an upper index and a lower index, M^i_j . The generic component of the rotated matrix under a rotation $R(\theta)$ is $R(\theta)^i_k (R(\theta)^{-1})^\ell_j M^\ell_\ell$. We contract the upper index of M with the lower index of $R(\theta)$ and we contract the lower index of M with the upper index of $R(\theta)^{-1}$. The remaining indices of $R(\theta)$ and $R(\theta)^{-1}$ are the indices of $(M')^i_j$.

Example 6.2. Consider the case $M\mathbf{v} = N\mathbf{v}$ for the case $\mathbf{v} = \hat{\mathbf{e}}_1$. This sets two vectors equal:

$$\begin{pmatrix} M^1_1 & M^1_2 \\ M^2_1 & M^2_2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} M^1_1 \\ M^2_1 \end{pmatrix} \quad \begin{pmatrix} N^1_1 & N^1_2 \\ N^2_1 & N^2_2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} N^1_1 \\ N^2_1 \end{pmatrix}. \quad (6.21)$$

We see that the equality is satisfied if only the first columns of M and N match. The second columns could differ, say $M^1_2 \neq N^1_2$, and so we cannot conclude that $M = N$. However, if we say that $M\mathbf{v} = N\mathbf{v}$ for every vector \mathbf{v} , then we could also check $\mathbf{v} = \hat{\mathbf{e}}_2$ to find

$$\begin{pmatrix} M^1_1 & M^1_2 \\ M^2_1 & M^2_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} M^1_2 \\ M^2_2 \end{pmatrix} \quad \begin{pmatrix} N^1_1 & N^1_2 \\ N^2_1 & N^2_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} N^1_2 \\ N^2_2 \end{pmatrix}. \quad (6.22)$$

The combination of (6.21) and (6.22) means that $M = N$.

Exercise 6.4. Explain why it is sufficient to show $M\hat{\mathbf{e}}_i = N\hat{\mathbf{e}}_i$ for both basis vectors in Example 6.2 to prove the equality of M and N given the relation $M\mathbf{v} = N\mathbf{v}$ for any \mathbf{v} . How do you know that the equality holds for $\mathbf{v} \neq \hat{\mathbf{e}}_i$? Your argument should invoke the notion of linear combinations and linearity.

6.6 Transformation of Tensors

Why stop at two indices? A **tensor** is, in general, an object with indices associated with a vector space. The tensor can have any number of indices. Each index is either upper (column-like) or lower (row-like). A vector is a tensor with one upper index. A row vector is a tensor with one lower index. A matrix is a tensor whose first index is upper, and whose second index is lower. A number is a tensor with no indices. A tensor with three indices can be written as a cube of components in the same way that a matrix can be written as a table of components. Tensors with more than three indices are hard to write down in “matrix form.” It’s a good thing we’ve been spending all this time developing index notation, right?

Example 6.3. The **moment of inertia** tensor is a tensor with two indices, like a matrix. However, unlike a matrix, the moment of inertia tensor has two upper indices. This means it is rather different than a matrix. We now explain how they are different.

The significance of a tensor and its indices are that the indices tell us how the tensor transforms under rotations. This is the culmination of Key Ideas 6.1, 6.2, 6.3, and 6.4. We now *define* the generalization of these ideas to a tensor, which is a pretty good working definition for what a tensor is:

Key Idea 6.5. A tensor with p upper indices and q lower indices is called a **rank** (p, q) tensor. For convenience of writing components using ellipses, let us write all of the upper indices first and then all of the lower indices last—this does not have to be the case, but the following result will hold. Under a rotation $R(\theta)$, the components of a tensor T transform as

$$T^{i_1 \dots i_p}_{j_1 \dots j_q} \rightarrow (T')^{i_1 \dots i_p}_{j_1 \dots j_q} = \left[\prod_p R(\theta)^{i_p}_{k_p} \right] \left[\prod_q (R(\theta)^{-1})^{\ell_q}_{j_q} \right] T^{i_1 \dots i_p}_{j_1 \dots j_q}. \quad (6.23)$$

Example 6.4. The matrix T^{ij}_k transforms as

$$T^{ij}_k \rightarrow R(\theta)^i_{\ell} R(\theta)^j_m (R(\theta)^{-1})^n_k T^{\ell m}_n. \quad (6.24)$$

Exercise 6.5. The moment of inertia tensor has components I^{ij} with i and j taking values from 1 to 3. How does the moment of inertia tensor transform under a three-dimensional rotation R ? (A three dimensional rotation is described by three angles. We shall be lazy and suppress those arguments altogether.) Compare this transformation law to that of a matrix, M^i_j . This difference is why we say “moment of inertia tensor” and not ‘moment of inertia matrix.’

A large part of physics is written in tensor notation.³² This subsection finally gives a working definition for what a tensor is and hints at why it may be significant: it is an object that transforms in a prescribed way under rotations. Because rotations are related to changes in reference frame, tensors are objects where measuring the components in one frame allows

³²As a curious side note, there is a graphical way of writing tensors called birdtracks. You can learn more about it at <https://birdtracks.eu>, or in Roger Penrose’s book *Road to Reality*. The notation is quite fetching, but ultimately I have not found it to be significantly more useful than indices.

us to predict the components that are measured by other observers.

6.7 Rotations in Three-Dimensions

In \mathbb{R}^3 there are three different axes about which we can perform a rotation. These roughly correspond to three rotation matrices:

$$R_1(\theta_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{\theta_1} & -s_{\theta_1} \\ 0 & s_{\theta_1} & c_{\theta_1} \end{pmatrix} \quad R_2(\theta_2) = \begin{pmatrix} c_{\theta_2} & 0 & s_{\theta_2} \\ 0 & 1 & 0 \\ -s_{\theta_2} & 0 & c_{\theta_2} \end{pmatrix} \quad R_3(\theta_3) = \begin{pmatrix} c_{\theta_3} & -s_{\theta_3} & 0 \\ s_{\theta_3} & c_{\theta_3} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.25)$$

You can see how these break down into two-dimensional rotations along each respective axis. If that is not obvious, please do the following exercise.

Exercise 6.6. Write the transformation of a three-component vector \mathbf{v} with components v^i under each of the transformations in (6.25).

Additional rotations are produced by taking products of R_1 , R_2 , and R_3 by appropriate amounts. The properties of these rotations are outside the scope of our course, but their study is a big part of physics and is known as the *representation theory of Lie groups*. If you want to do a deeper dive into rotation matrices in three dimensions, I recommend Howie Haber's lecture notes for Physics 216 at UCSC³³.

All this is to say that there are more options for rotations in three dimensions. Given a particular rotation, R , then the tensor transformation rule in Key Idea 6.5 holds. In fact, one can start to piece together what a rotation in even higher dimensions looks like. You extend the list (6.25) to include rotations about every plane—where the plane is simply defined by the two components of the vector that are being mixed into each other. That gives you the list of rotations about each axis. A general rotation is a product of the rotation-about-a-given-axis. Now you know how to manipulate tensors in arbitrary dimensions. Not bad.

Exercise 6.7. One thing to notice is that the order in which you apply rotations matters. Explicitly perform the matrix multiplication on each side of the following non-equation to confirm this:

$$R_1(\theta_1)R_2(\theta_2) \neq R_2(\theta_2)R_1(\theta_1). \quad (6.26)$$

We say that the rotation matrices along different axes do not **commute** because the order in which they are applied matters. This is demonstrated in Fig. 10.

³³http://scipp.ucsc.edu/~haber/ph216/rotation_12.pdf

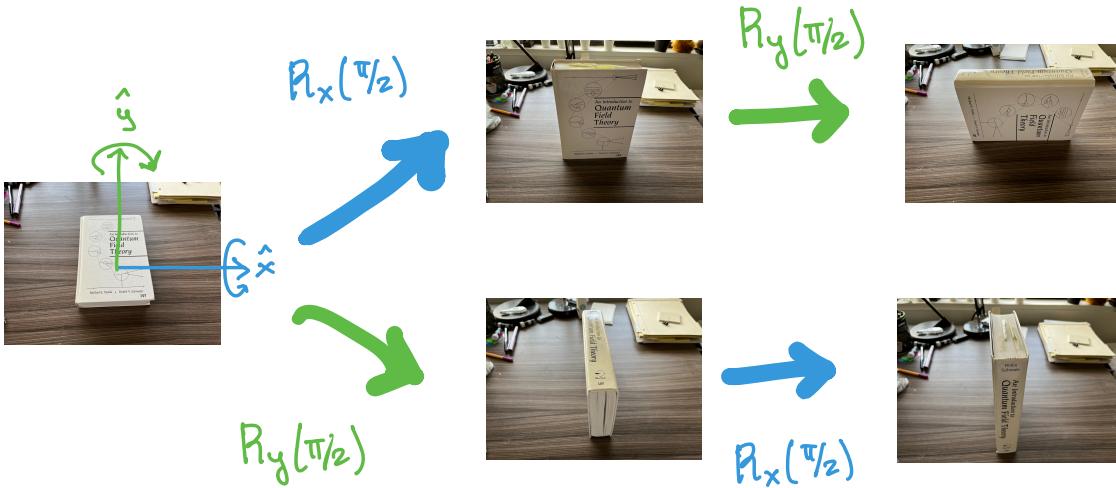


Figure 10: Example of the non-commutativity of rotations. Consider this copy of Peskin and Schröder's *Introduction to Quantum Field Theory*. Orient the \hat{x} axis in the horizontal direction along the desk and the \hat{y} axis in the vertical direction along the desk. Performing a $\pi/2$ rotation in the \hat{x} direction then in the \hat{y} direction (upper path) gives a different result than performing the rotations in the opposite order (lower path).

Exercise 6.8. Show that the determinant of each of the rotations in (6.25) is one. Argue that a general rotation in three dimensions, which is a product of these rotations, also has determinant equal to one. Argue further that the determinant of a rotation in any number of dimensions is equal to one.

6.8 Invariance of Trace and Determinant

The trace and the determinant of a matrix are invariant under rotations. This may sound trivial³⁴: they are numbers with no indices, and numbers with no indices do not transform under rotations. The relevance of the invariance of the trace and determinant are that they are numbers that are formed out of the components of a matrix that do not transform, even though the components of a matrix, M^i_j , do transform under a rotation.

Let us see that this is the case. Under a rotation, the trace of a matrix, M^i_j transforms as:

$$M^i_j \rightarrow (M')^i_i = R^i_k M^k_\ell (R^{-1})^\ell_i . \quad (6.27)$$

We have started writing R instead of $R(\theta)$ since we generalize to the case of higher-dimensional rotations with more than one parameters. Now we're going to do something sneaky. Let us move the $(R^{-1})^\ell_i$ factor around on the right-hand side. We made a big deal in Section 6.5 that we are allowed to do this *as long as we keep the indices the same*:

$$(M')^i_i = (R^{-1})^\ell_i R^i_k M^k_\ell = (R^{-1} R)^{\ell}_k M^k_\ell = \delta^{\ell}_k M^k_\ell = M^k_k . \quad (6.28)$$

³⁴“Trivial” is the mathematical translation of when I say ‘obvious.’

Did you see what we did? We noticed that because the last index of $(R^{-1})^\ell_i$ matches the first index of R^i_k , these two matrices are actually being multiplied in the expression for the trace. In other words, we have derived the following matrix-level expression:

$$\mathrm{Tr} RMR^{-1} = \mathrm{Tr} R^{-1}RM = \mathrm{Tr} M . \quad (6.29)$$

Exercise 6.9. Check that the above argument holds for the trace of any product of matrices, not necessarily rotations:

$$\mathrm{Tr} ABC = \mathrm{Tr} CAB = \mathrm{Tr} BCA . \quad (6.30)$$

We call this a *cyclic permutation* of the product ABC .

I hope you are starting to appreciate the utility of the index notation.

Moving on to the determinant, we appeal a couple of rules that we motivated in Section 5.11:

$$\det(MN) = (\det M)(\det N) \qquad \qquad \det(M^{-1}) = (\det M)^{-1} . \quad (6.31)$$

Then it is straightforward to read off that

$$\det M \rightarrow \det M' = \det(RMR^{-1}) = \det M \frac{\det R}{\det R} = \det M . \quad (6.32)$$

Appealing to our pictorial interpretation that the determinant is a volume, we remark that rotations do not change the volume of the parallelepiped formed by basis vectors. Rotating a mug does not change the capacity of the mug, though it may cause its contents to spill out.

6.9 Exponentiation of Generators

[**Flip:** Fill this in another time. See hand-written notes on the exponentiation of generators. Give an introduction to the representation of continuous groups. Maybe place this in an appendix.]

6.10 Active versus Passive

There is a concept that pops up here and there in physics: the idea of an **active** versus a **passive transformation**. An active transformation *is* a transformation. You have one vector, you apply a map to it, and then you have a totally different vector. By contrast, a passive transformation is one that is analogous to a *change in reference frame*. In a passive transformation, the vectors do not change—the observers change. For passive transformations, we almost exclusively consider rotations. In case you need a visual aid, we first brought up this idea of a passive rotation in (9). If you need a mnemonic to remember which is which, Fig. 11 may be a helpful meme.

In a passive transformation, the vector is the same but its *components* are changed because the rotated observer has different basis vectors. That is, to perform a passive transformation

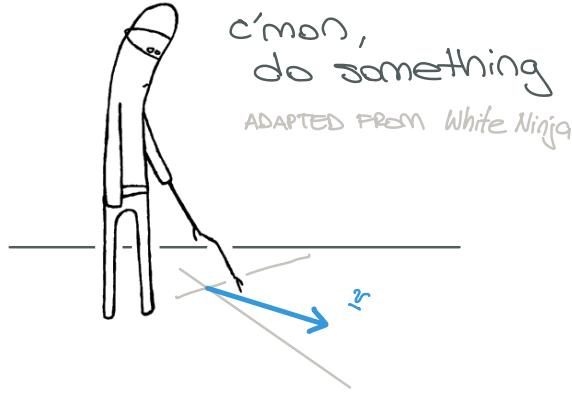


Figure 11: Nothing actually happens during a passive transformation. The basis vectors transform, but the components of the vector transform in such a way that $v'^i \hat{\mathbf{e}}'_i = v^i \hat{\mathbf{e}}_i$. Such a ‘transformation’ may come off as uninspiring, but this is the basis of symmetries in physics. From <https://knowyourmeme.com/memes/cmon-do-something>.

we *rotate on the basis vectors* $\hat{\mathbf{e}}_i \rightarrow \hat{\mathbf{e}}'_i$ and determine what the new components $(v')^i$ must be so that

$$\mathbf{v} = v^i \hat{\mathbf{e}}_i = (v')^i \hat{\mathbf{e}}'_i . \quad (6.33)$$

One way to determine this is to act on the basis vector with the identity matrix written as $\mathbb{1} = RR^{-1}$. We choose to do this in a way that acts on the basis vector:

$$\mathbf{v} = v^i \hat{\mathbf{e}}_j R^j_k (R^{-1})_i^k = [(R^{-1})_i^k v^i] [\hat{\mathbf{e}}_j R^j_k] . \quad (6.34)$$

On the right hand side we recognize that the second bracketed term is simply the rotation of an object with a lower index—remember Key Idea 6.3? Yes, all this time we had brought it up, but the basis vector $\hat{\mathbf{e}}_i$ has a lower index, and so it transforms like a lower-indexed object (row vector-like).³⁵ We thus define the rotated basis vectors:

$$\hat{\mathbf{e}}'_k = \hat{\mathbf{e}}_j R^j_k . \quad (6.35)$$

Then (6.34) tells us that we have a linear combination of these rotated basis vectors:

$$\mathbf{v} = (v')^k \hat{\mathbf{e}}'_k = (R^{-1})_i^k v^i \hat{\mathbf{e}}'_k \quad (v')^k = (R^{-1})_i^k v^i . \quad (6.36)$$

We recognize that we have rotated the basis vectors by a ‘forward’ rotation R , and so the *components* of a vector rotate in the opposite direction by the ‘backward’ rotation, R^{-1} . This is in contrast to an *active* rotation of \mathbf{v} where the basis stays the same, but we rotate the components of the vector by the forward rotation:

$$v^i \rightarrow (v')^i = \begin{cases} R^i_j v^j & \text{for an active transformation} \\ (R^{-1})_j^i v^j & \text{for a passive transformation} \end{cases} . \quad (6.37)$$

³⁵To be clear, the basis vector is *not* a row vector. That’s why $\hat{\mathbf{e}}_i$ is boldfaced: it carries all the abstract identity of the vector space. The basis vector just *happens* to have an index that tells you how it should transform under a rotation. That index happens to be lower.

Passive transformations are useful when we consider symmetries of systems. Each (continuous) symmetry has some analog of a ‘rotation’ matrix, even if the rotation is not necessarily in ordinary space, but in some more abstract space.

7 Tensors as (Multi-)Linear Maps: Tensors

We have informally introduced the idea that a tensor is an object with well defined transformation properties under rotations. A second perspective of tensors is that they are linear maps. This gives a slightly more rigorous definition for what all these indexed objects do. The general theme is that as linear maps, *tensors are born for index contraction*.

7.1 Products of Vector Spaces

First, a bit of review and notation. Recall that a **vector space**, V , is the maximal set of vectors formed by taking linear combinations of vectors. Given a basis, V is the collection of all linear combinations of those basis vectors. Now recall our jargon from Section 5.1: linear maps are functions that take as input any element of a space and returns as output a single element of another space.³⁶ A map is linear if

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) . \quad (7.1)$$

We can also consider functions that take more than one input. For example, a function could take two numbers and return their sum. We say that the input of this map is part of a *product space* $\mathbb{R} \times \mathbb{R}$. What we mean here is that \mathbb{R} is a vector space, and $\mathbb{R} \times \mathbb{R}$ is a vector space where each vector in the vector space is made up of two elements of \mathbb{R} . Does that sound familiar? This is simply what we’ve been calling \mathbb{R}^2 , which we now recognize is shorthand for $\mathbb{R} \times \mathbb{R}$. We can then imagine functions from $\mathbb{R} \times \mathbb{R}$ to numbers. We indicate this with the following type of notation:

$$f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \quad (7.2)$$

For example, $f(x, y) = \sqrt{x^2 + y^2}$ gives the distance of (x, y) from the origin. This kind of function is called **multi-linear** if it is linear in each of its arguments:

$$f(\alpha\mathbf{x} + \beta\mathbf{z}, \mathbf{y}) = \alpha f(\mathbf{x}, \mathbf{y}) + \beta f(\mathbf{z}, \mathbf{y}) \quad (7.3)$$

$$f(\mathbf{x}, \gamma\mathbf{y} + \delta\mathbf{w}) = \gamma f(\mathbf{x}, \mathbf{y}) + \delta f(\mathbf{x}, \mathbf{w}) . \quad (7.4)$$

Example 7.1. The function $f(x, y) = \sqrt{x^2 + y^2}$ is not multi-linear. The function $g(x, y) = x + 3y$ is multi-linear.

³⁶This always reminds me of rock octoroks in *The Legend of Zelda: Breath of the Wild*. If you feed little baddies rusty weapons, they will inhale them and spit out a clean weapon.

Example 7.2. Is the function $f(x, y) = 2x + 5y + 2$ multi-linear?

The moral of this sub-section is that you should not be intimidated by the **tensor product** notation for multiple copies of a vector space, $V \times V \times \cdots \times V = V^n$.

7.2 Linear Maps to Numbers: Row Vectors

Our humble row vectors are also linear transformations. The components of a row vector have lower indices, w_i . These lower indices are just looking for an upper index for contraction. Once we provide a row vector $\underline{\mathbf{w}}$ with an object with an upper index, $\mathbf{v} = v^i \hat{\mathbf{e}}_i$, then we can form the number $\underline{\mathbf{w}}\mathbf{v} = w_i v^i$. This tells us that a row vector $\underline{\mathbf{w}}$ defined by its components w_i is a linear map that takes in a vector and returns a number:

$$\underline{\mathbf{w}} : V \rightarrow \mathbb{R} . \quad (7.5)$$

From the rule of index contraction, we observe that this is a *linear map*:

$$\underline{\mathbf{w}}(\alpha \mathbf{v} + \beta \mathbf{u}) = w_i (\alpha v^i + \beta u^i) = \alpha w_i v^i + \beta w_i u^i = \alpha \underline{\mathbf{w}}\mathbf{v} + \beta \underline{\mathbf{w}}\mathbf{u} . \quad (7.6)$$

We deduce that lower indices are “born to contract” with the upper index of a vector. Lower indices say: *feed me an upper index and I will contract it; I will return a number that is linearly dependent on the upper indexed components*.

We say that these row vectors belong to a **dual vector space**, V^* . This V^* is the space of all row vectors. An element of V^* , say $\underline{\mathbf{w}}$, is a linear map from $V \rightarrow \mathbb{R}$.

Exercise 7.1. Argue that V^* is also a vector space. If $\underline{\mathbf{w}}$ and $\underline{\mathbf{q}}$ are row-vectors, then $\alpha \underline{\mathbf{w}} + \beta \underline{\mathbf{q}}$ is also a row vector with components

$$(\alpha \underline{\mathbf{w}} + \beta \underline{\mathbf{q}})_i = \alpha w_i + \beta q_i . \quad (7.7)$$

7.3 Tensors as Multi-Linear Maps

What about objects with multiple lower indices? If your tensor has q lower indices, then it can take as inputs q vectors and the output will be multi-linearly dependent on those inputs. Consider a tensor with two lower indices, g_{ij} . This is a linear map from $V \times V \rightarrow \mathbb{R}$. Once you tell me the coefficients g_{ij} , then I can arrange them into a machine that takes in two vectors \mathbf{v} and \mathbf{u} and spits out a number:

$$g : V \times V \rightarrow \mathbb{R} \quad g(\mathbf{v}, \mathbf{u}) = g_{ij} v^i u^j . \quad (7.8)$$

This is obviously multi-linear in its arguments.

What about an object with both lower and upper indices? We already have one example: matrices have one upper and one lower index. The lower index tells us that the matrix ‘eats’ a vector by contracting the matrix’s lower index with the vector’s upper index. Once you have done that contraction, you are left with an upper index. This means that the resulting object, $M(\mathbf{v}) = M\mathbf{v}$ with components $(M\mathbf{v})^i = M^i_j v^j$ is a vector. We thus repeat our earlier statement that matrices are linear maps from the vector space to itself,

$$M : V \rightarrow V . \quad (7.9)$$

What if we have a tensor with two upper indices and one lower index, say T^{ij}_k . Like a matrix, this object has a lower index and so it is a linear function on V . T has two upper indices, which means the output is in $V \times V$. We thus say that $T : V \rightarrow V \times V$. The components of $T\mathbf{v}$ are $(T\mathbf{v})^{ij} = T^{ij}_k v^k$.

Do you see how this game works? If a tensor T has p upper indices and q lower indices, then it is a map

$$T : V^q \rightarrow V^p , \quad (7.10)$$

where each of the q lower indices a linear vector argument for the tensor, and each of the p upper indices is a vector-like output. We should be a little careful about the meaning of ‘vector-like’ here. This is not to say that you return p different vectors. Instead, you have an object that transforms under rotation with p factors of the rotation matrix R , one for each upper index. Similarly, the each of the tensor’s lower indices transforms with a factor of the inverse rotation R^{-1} .

What if a tensor has only upper indices? One glib way of interpreting this is that you go from nothing to an object with upper indices. For example, let S^{ij} be a generic component of a tensor with two upper indices. Then I can say that

$$S : \mathbb{R} \rightarrow V \times V . \quad (7.11)$$

Here \mathbb{R} on the left hand side just means you can input an overall coefficient. In this sense, you input a number, say 3, and S gives you $3S^{ij}$, an object with two upper indices.

Wait a second, you think. If S^{ij} only has upper indices, I know how it can still be a linear function that outputs numbers. Instead of feeding it two vectors that each have one upper index, then why not feed it two row vectors that each have one lower index? After all, $w_i q_j S^{ij}$ is a number. We thus say that there is an equivalent way of thinking about S :

$$S : V^* \times V^* \rightarrow \mathbb{R} . \quad (7.12)$$

It appears that we simply took (7.11) and pulled the V ’s to the left-hand side, at the cost of each of them picking up a star. The rule now appears to be that for each of a tensor’s indices:

1. An upper index means that the tensor either has a vector-like index as an output, or it takes in a row vector as a linear input.
2. A lower index means that the tensor either has a row-vector-like index as an output, or it takes in a column vector as a linear input.

Example 7.3. Consider the tensor T^{ij}_k . This is equivalently:

$$T : V \rightarrow V \times V, \quad \mathbb{R} \rightarrow V \times V \times V^*, \quad V^* \times V \rightarrow V, \quad V^* \rightarrow V^* \times V^*. \quad (7.13)$$

What other ways can we interpret T as a multilinear map between products of V and V^* ?

Example 7.4. Consider the tensor T^{ij}_k . If we interpret this as a map from $V^* \times V^* \rightarrow V^*$, then we can feed T two row vectors and it outputs a row vector,

$$T(\underbrace{\mathbf{w}}_{\sim}, \underbrace{\mathbf{q}}_{\sim})_k = T^{ij}_k w_i q_j. \quad (7.14)$$

Be careful that it is implicit that $T(\mathbf{w}, \mathbf{q})$ means $T^{ij}_k w_i q_j$ and not $T^{ij}_k w_j q_i$. If there is ever ambiguity, it is up to you to explain to your audience exactly what you mean by a contraction. Either of the above interpretations is a valid linear map, but in general when your inputs include two or more of the same type of object, the order matters.

Exercise 7.2. In the previous example we showed that the order of arguments can be somewhat ambiguous when the arguments are of the same type. Show that the order does not matter when a tensor has **symmetric** indices, say $g_{ij} = g_{ji}$. Show that the order does matter when a tensor has **antisymmetric** indices, $h_{ij} = -h_{ji}$, but that the difference is an overall minus sign. Show that a tensor with a pair of indices of the same height may be written as the sum of a piece that is symmetric and a piece that is antisymmetric in those indices. That is, for any A_{ij} show that there exist symmetric g and anti-symmetric h such that:

$$A_{ij} = g_{ij} + h_{ij} \quad g_{ij} = g_{ji} \quad h_{ij} = -h_{ji}. \quad (7.15)$$

Argue that it does not matter if the two indices are both upper or both lower. Nor does it matter if there are other indices around.

Exercise 7.3. Show that a tensor that is symmetric with respect to two indices of the same height, say $g_{ij} = g_{ji}$, remains symmetric under a rotation. Do the same for antisymmetric pairs of indices of the same height.

The above digression of symmetric and antisymmetric indices may seem like an odd

tangent. However, we will soon see that the “additional mathematical ingredient” that allows us to define an inner product (dot product) is an object with two symmetric lower indices that we call a metric. And while it is outside the scope of this course, antisymmetric indices are the foundation of what is called an **alternating linear map**. These, in turn, show up in differential geometry (general relativity) as something called differential k -forms that play the role of infinitesimal volume elements.

Exercise 7.4. Let T be a tensor with p upper indices and q lower indices. If I feed this tensor p vectors and q row vectors by contracting the indices, the result is a number. How does this number transform under rotations?

7.4 Duality

The above discussion should lead you to realize that there is a Yin–Yang-like duality between upper and lower indices. Initially, we started this whole course by making a big deal about *vectors* and not row vectors. We had presented the *nouns* of linear algebra are described by components that have one upper index. Then we presented row vectors as linear maps from $V \rightarrow \mathbb{R}$, that is, row vectors are *verbs* that describe an action on the *nouns*. Then we realized that inasmuch as the lower index in w_i is an invitation to contract with the upper index of v^i , we could have said it in the other way. There is a vector space V^* whose components have a lower index. In this picture, the row vectors are the *nouns* and the column vectors are the *verbs* that represent linear maps $V^* \rightarrow \mathbb{R}$. The upper index in v^i is an invitation to contract with the lower index of w_i .

The point, of course, is that one upper index and one lower index can contract. It doesn’t matter which one you call the noun and which one you call the verb (linear map). They are each the noun of their own vector space, and relative to their vector space the other vector space is a ‘dual space’ of linear functions to \mathbb{R} .

Generalizing this, tensors T with p upper indices and q lower indices are multi-linear maps and objects in themselves. If you want a tensor with $(p - 1)$ upper indices and $(q - 1)$ lower indices, you could feed T one row vector and one column vector. Or you could feed it a matrix, which conveniently already has an upper and a lower index. Or you could just have one of the p upper indices contract with one of the q lower indices, taking a multi-linear version of the trace. Or, perhaps more perversely, you could feed it a tensor with $p' < p$ upper indices and $q' < q$ lower indices in such a way that *several* indices between the two tensors contract but you are left with a tensor with $(p - 1)$ upper indices and $(q - 1)$ lower indices.

7.5 A basis for the dual vector space

The above semi-philosophical sub-section was all to motivate the following. If V^* is a legitimate vector space in its own right, *what are its basis vectors*? Like $V = \mathbb{R}^3$, there are an infinite number of choices. But there is one “obvious” choice:

$$\hat{\mathbf{e}}^1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad \hat{\mathbf{e}}^2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \quad \hat{\mathbf{e}}^3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} . \quad (7.16)$$

The generalization to arbitrary dimensions is also obvious. Note that the basis row vectors have upper indices unlike the basis column vectors that have lower indices. This is so that a row vector is the contraction:

$$\underbrace{\mathbf{w}}_{\sim} = w_i \hat{\mathbf{e}}^i . \quad (7.17)$$

Exercise 7.5. How do the components w_i transform under a passive transformation where the basis row vectors are rotated by a rotation matrix R ?

These basis row vectors are each linear maps from $V \rightarrow \mathbb{R}$. For example, $\hat{\mathbf{e}}^2$ takes in a vector with components v^i and outputs v^2 . These basis row vectors are linear functions as far as vectors in V are concerned, but they are *simultaneously* a vector space in their own right. To these vectors in V^* , the elements of V are the linear functions.

As we have echoed several times now, the basis vectors carry the “vector-ness” of the objects. Thus the linear-function-ness of the V^* basis can be seen in its action on the V basis vectors:

$$\hat{\mathbf{e}}^i (\hat{\mathbf{e}}_j) \equiv \delta_j^i \quad \hat{\mathbf{e}}_j (\hat{\mathbf{e}}^i) \equiv \delta_j^i . \quad (7.18)$$

On the right-hand side we have also shown that the basis vectors in V are linear maps on the basis row vectors in V^* . When we’re lazy and there is no danger of ambiguity, we drop the parenthesis and just write $\hat{\mathbf{e}}^i \hat{\mathbf{e}}_j = \hat{\mathbf{e}}_j \hat{\mathbf{e}}^i$.

In this way, we can “derive” the rule for how a row vector acts on a column vector (or equivalently vice versa):

$$\underbrace{\mathbf{w}\mathbf{v}}_{\sim} = w_i \hat{\mathbf{e}}^i v^j \hat{\mathbf{e}}_j = w_i v^j \hat{\mathbf{e}}^i \hat{\mathbf{e}}_j = w_i v^j \delta_j^i = w_i v^i . \quad (7.19)$$

7.6 A basis for multi-linear maps

What is the basis of tensors? Let us start with matrices. Matrix components have one upper and one lower index. They are kind of like a column vector attached to a row vector, but *not acting on one another*.³⁷ To indicate that the two basis vectors are not acting on one another, we separate them with a **tensor product** symbol, \otimes . Thus we expand a matrix in the basis of tensor products:

$$M = M^i_j (\hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}^j) . \quad (7.20)$$

Let us see how this acts on a vector in order to “derive” the matrix-acting-on-a-vector rule that earlier we had simply stated as fact:

$$M\mathbf{v} = M^i_j (\hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}^j) v^k \hat{\mathbf{e}}_k = M^i_j v^k (\hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}^j) \hat{\mathbf{e}}_k . \quad (7.21)$$

³⁷Reminiscent of [https://en.wikipedia.org/wiki/PPAP_\(Pen-Pineapple-Apple-Pen\)](https://en.wikipedia.org/wiki/PPAP_(Pen-Pineapple-Apple-Pen)).

In the last line, we understand that $\hat{\mathbf{e}}^j$ acts on $\hat{\mathbf{e}}_k$ (or equivalently, $\hat{\mathbf{e}}_k$ acts on $\hat{\mathbf{e}}^j$). We recall that $\hat{\mathbf{e}}^j \hat{\mathbf{e}}_k = \delta_k^j$. Thus:

$$M\mathbf{v} = M_j^i v^k \delta_k^j \hat{\mathbf{e}}_i = M_j^i v^j \hat{\mathbf{e}}_i . \quad (7.22)$$

We can read the right hand side as a vector—since it is a linear combination of basis vectors $\hat{\mathbf{e}}_i$ —with components $M_j^i v^j$, which is of course the rule that we had simply posted earlier.

From this we generalize to say that a multilinear map with p upper indices and q lower indices has a basis:

$$\hat{\mathbf{e}}_{i_1} \otimes \hat{\mathbf{e}}_{i_2} \otimes \cdots \otimes \hat{\mathbf{e}}_{i_p} \otimes \underbrace{\hat{\mathbf{e}}^{j_1}}_{\sim} \otimes \underbrace{\hat{\mathbf{e}}^{j_2}}_{\sim} \otimes \cdots \otimes \underbrace{\hat{\mathbf{e}}^{j_q}}_{\sim} . \quad (7.23)$$

When acting on another tensor, we simply have to specify which of the $\hat{\mathbf{e}}_i$ and $\hat{\mathbf{e}}^j$ are acting on their corresponding (row-)vector components in the basis of the other tensor.

Example 7.5. *Taking a trace amounts to selecting one of the $\hat{\mathbf{e}}_i$ and one of the $\hat{\mathbf{e}}^j$ in the basis to contract with each other.*

8 Ket Notation

Does looking at basis elements like (7.23) give you a headache? The notation is partially to blame. s Quantum mechanics has a completely different notation for vectors. Rather than \mathbf{v} , we write $|v\rangle$. This is called a **ket**. The reason for this is that we write row vectors as $\underline{\mathbf{w}} = \langle w|$. We call this object a **bra**. Then the “matrix multiplication” $\underline{\mathbf{w}}\mathbf{v}$ from (4.2) is succinctly written $\langle w|v\rangle$. The pun is that this is a ‘bra–ket’ or a *bracket*. At the moment we do not have much use for bra and ket notation, but it is useful to establish this notation early on. It will soon be *very* convenient. To summarize:

$$\text{column vector: } \mathbf{v} = |v\rangle \quad \text{row vector: } \underline{\mathbf{w}} = \langle w| \quad (8.1)$$

There is no special notation for matrices, M . If you really want to be fancy, though, you can give matrices a hat and write them \hat{M} . This is really only important if you need to distinguish between a matrix and a number that happens to have a similar variable name.

Example 8.1. *As an example where the ‘hat’ notation comes in handy, recall the Schrödinger equation in (2.1):*

$$\hat{H} |\Psi\rangle = E |\Psi\rangle . \quad (8.2)$$

The \hat{H} has a hat because it is a matrix. The E does not have a hat because it is a number. Does this information give the equation more significance? Obviously $\hat{H} \neq E$ since these are two completely different classes of objects. The equation is an equality between two vectors, $|\Psi\rangle$. Evidently, when the matrix \hat{H} hits $|\Psi\rangle$, you end up with a new vector that

Notation	V element	V^* element	Contraction	Matrix	Identity
Classic	vector	row-vector			
	$\mathbf{v} = v^i \hat{\mathbf{e}}_i$	$\mathbf{w} = w_i \hat{\mathbf{e}}^i$	$\hat{\mathbf{e}}^i \hat{\mathbf{e}}_j = \delta_j^i$	$M = M_j^i \hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}^j$	$\mathbb{1}$
Quantum	ket	bra			
	$ v\rangle = v^i i\rangle$	$\langle w = w_i \langle i $	$\langle i j\rangle = \delta_j^i$	$M = M_j^i i\rangle \langle j $	$ i\rangle \langle i $

Table 2: Dictionary between ‘classic’ notation and ‘quantum’ (bra–ket) notation.

is simply a rescaling of $|\Psi\rangle$. We say that $|\Psi\rangle$ is an eigenvector of \hat{H} .

Table 2 provides a brief dictionary between the quantum bra–ket notation and the ‘classical’ notation that we had been using until this point. New notation can be intimidating³⁸, but sometimes different notation can simplify our expressions and, in so doing, clarify ideas.³⁹

Key Idea 8.1. *Bra–ket notation is nothing more than dressing up the same ideas above in different symbols and words. Every time you see the word ket, your brain should immediately connect it to vector. Every time you see a $\langle w|$, your brain should activate the same neurons as when you see \mathbf{w} . Do not let the novelty of the notation fool you into thinking there is something new going on. You already know everything in this section, you are just learning new ways to write it.*

8.1 Basis Bras and Kets

Kets are vectors. The basis kets are written like so:

$$\hat{\mathbf{e}}_i \equiv |i\rangle . \quad (8.3)$$

Just like that. The index is the name of the ket. This is potentially confusing. You lose any sense that the index is a lower index. You just have to know in your heart that a basis ket has a *lower* index so that it can contract with the *upper index* of the components of the ket:

$$|v\rangle = v^i |i\rangle . \quad (8.4)$$

Similarly, the basis bras have their index as their name. You have to know that the index of the basis bra is implicitly upper:

$$\hat{\mathbf{e}}^j \equiv \langle j| \qquad \qquad \qquad \langle w| = w_j \langle j| . \quad (8.5)$$

³⁸More than once in my young scientific life I had discarded a textbook because it used notation that I found unintuitive. One of my mentors gently scolded me that good science and good scientists are notation-independent.

³⁹Feynman diagrams, one of the most recognizable visualizations in physics, are essentially new notation for otherwise cumbersome mathematical expressions.

8.2 Action of bras on kets

The action of a bra on a ket is—by definition—the same as the action of a row vector on a column vector:

$$\langle w|v\rangle \equiv \underline{\mathbf{w}}\mathbf{v} = w_i v^i . \quad (8.6)$$

On the left hand side a bra and a ket combine to form a *bracket*. That's peak physics humor in 1939.⁴⁰ What is implicit here is that the basis bra and basis kets combine into a Kronecker δ :

$$\langle j|i\rangle = \delta_i^j . \quad (8.7)$$

This is the analog of (7.18), where we said that this operations connects the *abstract* row and column vectors to an object, δ_i^j , that tells us how the *components* of the row and column vectors combine in a contraction. In a loose sense, this definition *derives* the contraction rule for tensors.

8.3 Matrices and Tensors

Matrix components have one upper index and one lower index. This is a hint that their basis is some combination of a bra and a ket, as we already know from (7.20). We thus write:

$$M = M_j^i |i\rangle \otimes \langle j| = M_j^i |i\rangle \langle j| . \quad (8.8)$$

In the last step we realize that because of the shape of the bra and ket, there is *no ambiguity* that the “ket–bra” $|i\rangle \langle j|$ might be interpreted as a contraction. The contractions are only when the vertical parts of the bra and ket overlap, as in (8.7). We can extend our notation to general tensors as long as we take the convention where the upper indices all come first and the lower indices are last:

$$T = T^{i_1 \dots i_p}_{j_1 \dots j_q} |i_1\rangle \dots |i_p\rangle \langle j_1| \dots \langle j_q| . \quad (8.9)$$

If your tensor does not have all of the upper indices first, you can always define a new tensor which has identical components but reshuffled so that the components that are all upper index come first. When you act with one tensor onto another tensor you typically need to specify which bras and kets you intend to contract. Finally, if you are ever in doubt, you can always explicitly write the tensor product symbol \otimes .⁴¹

8.4 Ket Tricks

The identity matrix is blindingly obvious and will turn out to be rather significant for us soon:

$$\mathbb{1} = |i\rangle \langle i| = |1\rangle \langle 1| + |2\rangle \langle 2| + |3\rangle \langle 3| + \dots \quad (8.10)$$

⁴⁰Dirac, “A New Notation for Quantum Mechanics,” *Mathematical Proceedings of the Cambridge Philosophical Society* **35** issue 3, pages 416–418.

⁴¹Every time I do this I quietly hum the theme to the 90’s *X-Men* animated series.

Exercise 8.1. Remember that the ket is implicitly a lower indexed object and the bra is implicitly an upper indexed object, thus their repeated i index is contracted. It is worth taking time to understand this expression in as many ways as you are able to. Instead of ket-bras, please use row vectors and column vectors in \mathbb{R}^3 . Confirm that this $|i\rangle\langle i|$ acting on a three-vector gives the same three vector. What would happen if you dropped the last term and instead only wrote $|1\rangle\langle 1| + |2\rangle\langle 2|$? What is the effect of this not-quite identity matrix on a three-vector $\mathbf{v} = |v\rangle$ with components v^i ?

Another surprisingly useful expression is the trace of a matrix:

$$\text{Tr } M = \langle i|M|i \rangle . \quad (8.11)$$

Here we recognize again that the upper index of the bra is contracting with the lower index of the ket.

Exercise 8.2. Confirm (8.11) by inserting (8.8) for M .

9 Metric Spaces: Dot Products

A vector space V is a collection of abstract objects (vectors, kets) and their linear combinations. We defined related spaces based on the linear transformations on the vector space. This includes row vectors (bras) which live in a dual vector space V^* , and all tensor products of these spaces. We saw that matrices—and tensors in general—can be understood as linear transformations between these spaces.

Now we turn the idea of a **dot product**, which we introduced for Euclidean three-space in (3.36). The dot product took in two vectors and returned a number. This means that it is a multi-linear map from $V \times V \rightarrow \mathbb{R}$. The generalization of the dot product is called the **inner product** and the tensor associated with it is called the **metric**. We write the inner product with a bracket notation that looks (intentionally) similar to a ket acting on a bra:

$$\langle \cdot, \cdot \rangle = g_{ij} : V \times V \rightarrow \mathbb{R} . \quad (9.1)$$

The nomenclature of using g_{ij} for the metric is inherited from relativity. The nomenclature with the bracket notation is inherited from quantum mechanics. There is one other defining characteristic of the metric: it is *symmetric* in its indices:

$$g_{ij} = g_{ji} . \quad (9.2)$$

The metric is a special tensor that we have to specify. A vector space with a metric is called a **metric space** or and **inner product space**. \mathbb{R}^3 is a vector space. \mathbb{R}^3 with a metric—say, the Euclidean metric in (3.36)—is a metric space.

Key Idea 9.1. *The difference between a vector space and a metric space is that in a metric space, you can raise/lower indices (transpose or Hermitian conjugate for spaces with complex numbers) and measure length and angles between vectors. There can be choices of metrics on the same vector space.*

The most common example is N -dimensional **Euclidean space**, a metric space where the components of the metric g_{ij} are simply that of the $N \times N$ identity matrix:

$$g_{ij} = \text{diag}(1, \dots, 1) = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & & 1 \end{pmatrix}. \quad (9.3)$$

Key Idea 9.2. *Let me be clear: g_{ij} is not a “matrix” in the sense that acting with it on a vector does not produce a vector. g_{ij} are the components of a tensor that is a symmetric^a map from $V \times V \rightarrow \mathbb{R}$. However, because g_{ij} has two indices, we may visually arrange the components in a $N \times N$ table, where N is the dimension of the vector space. If you want you can call this table a matrix, though in these notes we have taken a more strict definition. In general—but specifically in Section 9.9 where this comes up—when we write something of the form*

$$A_{ij} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (9.4)$$

all we mean is that $A_{11} = a$, $A_{12} = b$, and so forth. We can then talk about index contraction as “matrix multiplication” as long as consecutive indices are contracted:

$$A_{ij}B^j_k = (AB)_{ik} = \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix} \begin{pmatrix} B^1_1 & B^1_2 \\ B^2_1 & B^2_2 \end{pmatrix}. \quad (9.5)$$

If consecutive indices are not contracted, then you need to arrange things a bit more carefully. Specifically:

$$A^k_i B_{kl} C^\ell_j \neq \begin{pmatrix} A^1_1 & A^1_2 \\ A^2_1 & A^2_2 \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix} \begin{pmatrix} B^1_1 & B^1_2 \\ B^2_1 & B^2_2 \end{pmatrix}. \quad (9.6)$$

But if we define^b $(A^T)_i^k = A^k_i$, then we may write

$$A^k_i B_{kl} C^\ell_j = (A^T)_i^k B_{kl} C^\ell_j = \begin{pmatrix} A^1_1 & A^2_1 \\ A^1_2 & A^2_2 \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix} \begin{pmatrix} B^1_1 & B^1_2 \\ B^2_1 & B^2_2 \end{pmatrix} \quad (9.7)$$

$$= \begin{pmatrix} (ABC)^{11} & (ABC)^{12} \\ (ABC)^{12} & (ABC)^{22} \end{pmatrix}. \quad (9.8)$$

This is a modest point that can cause some confusion later on.

^aMeaning $g_{ij} = g_{ji}$ or equivalently $g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v})$.

^bThe transpose swaps the order of the indices of a matrix. What about the height of the indices? For now let us maintain the height so that an the first, upper index of A becomes the second, upper index of A^T . We give a more formal definition of transpose in Section 9.12.

Metrics do two things for us:

1. They allow us to define the length of a vector, $|\mathbf{v}|$. In so doing, they allow us to define a unit length (*normalized*) vector, $\hat{\mathbf{v}}$.
2. They allow us to define the angle between two vectors, $\cos \theta = \langle v, w \rangle$.

Together, they allow us to define *nice* basis vectors that are unit length and orthogonal to one another. We say that such a basis is **orthonormal**.

9.1 Example Polar Coordinates

The Euclidean space metric is too boring. It is essentially the identity element and it seems like there's much ado about nothing. Let me motivate the idea of a metric by reviewing polar coordinates. This is the simplest form of what are called **curvilinear coordinates**. These are simply coordinates grids that are not straight lines, the way that the Cartesian x - y plane is a grid. Polar coordinates replace (x, y) with (r, θ) . Here r is the distance from the origin and θ is the angle from the positive x -axis. This example is surprisingly subtle. The take-away is that the metric in polar coordinates is

$$g_{ij} = \begin{pmatrix} g_{rr} & g_{r\theta} \\ g_{\theta r} & g_{\theta\theta} \end{pmatrix} = \begin{pmatrix} 1 & \\ & r^2 \end{pmatrix}. \quad (9.9)$$

The bottom right component is not one, but rather the radial coordinate. If this is your first time seeing this, please feel free to *skim* this section. We will present the example, then dig a bit deeper to discuss the significance. You may want to postpone reading those details until after reviewing these notes once through.

9.1.1 What does this mean

The surprising piece in (9.9) is that $g_{\theta\theta} = r^2$. This means that the metric changes as you move further away from the origin of the coordinate space. This is in contrast to the Euclidean metric whose components are proportional to the identity no matter where you happen to be.

This should not surprise us. We know that the same *angular displacement* $\Delta\theta$ corresponds to a different traversed circumference depending on the radius. If you are close to the origin, then a displacement of $\Delta\theta$ is not particularly far. If you are far from the origin, then a displacement of $\Delta\Theta$ can be huge, see Fig. 12.⁴² From a physics perspective, you could have guessed this from pure dimensional analysis: $\Delta\theta$ is dimensionless. If you want your metric to output a squared distance, then you need something with dimensions of length—the only plausible candidate is r^2 . This is illustrated nicely in the *Calvin and Hobbes* comic in Fig. 13.

⁴²This sometimes shows up in the “paradox” (it is not a paradox) that “shadows travel faster than the speed of light.”

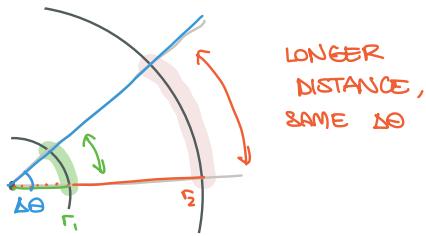


Figure 12: The meaning of the $g_{\theta\theta} = r^2$ is related to the idea that the distance traversed along a circle for a displacement $\Delta\theta$ depends on the radius of the circle.

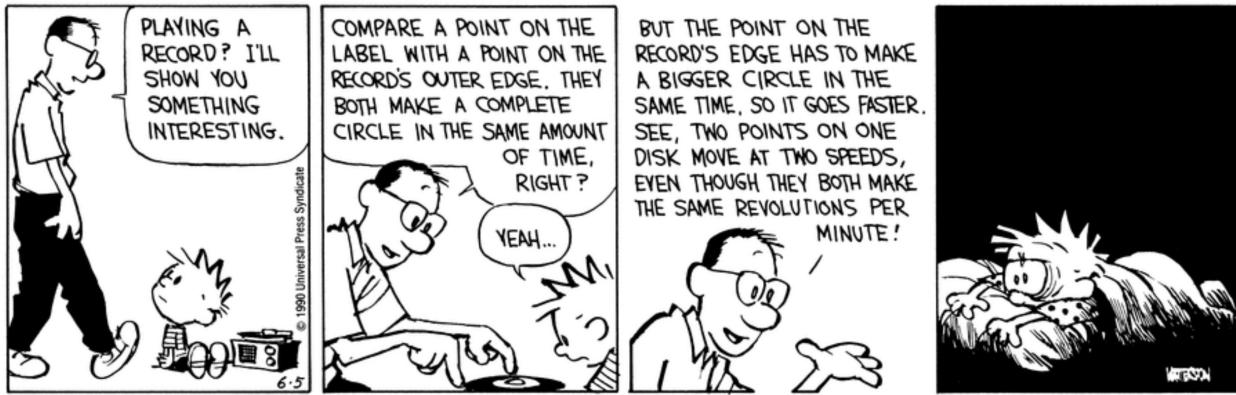


Figure 13: *Calvin and Hobbes*, “Record Player,” by Bill Watterson, 5 June 1990. Image from <https://www.gocomics.com/calvinandhobbes/1990/06/05>. If you have never heard of a record player than look it up.

Example 9.1. While this is not really relevant for linear algebra, there are a few neat ways in which the difference between angular and linear velocity show up. One is the mechanical integrator^a, and another is differential steering^b.

^ahttps://www.youtube.com/watch?v=s-y_lnzWQjk

^b<https://www.youtube.com/watch?v=yYAw79386WI>

Exercise 9.1. Consider the vector $\mathbf{v} = 3\hat{\mathbf{e}}_r - (\pi/4)\hat{\mathbf{e}}_\theta$ located at the point $(r, \theta) = (2, \pi/6)$. Find the length of this vector, $|\mathbf{v}|$ using the polar coordinate metric (9.9). Observe that for this problem we have to specify the location of the vector; this corresponds to defining where the ‘zero’ of the vector space is relative to the coordinate space. This suggests that there’s more structure here than simply a metric space—we discuss this in Section 9.1.4.

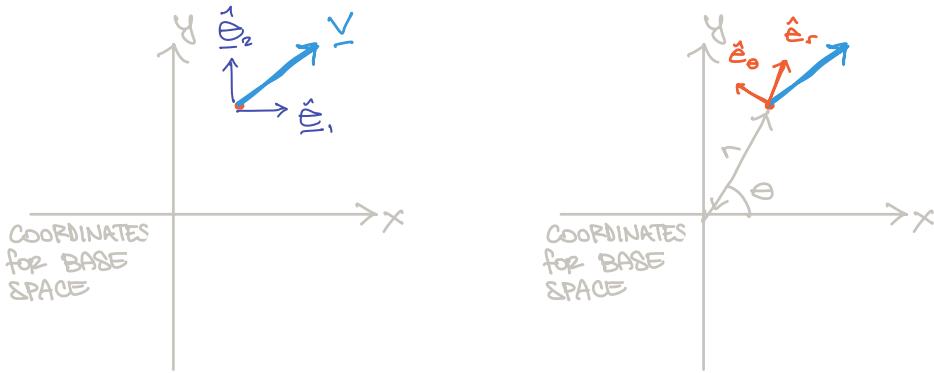


Figure 14: Consider a point in \mathbb{R}^2 and some vector \mathbf{v} that represents the velocity of a particle at that point. We can write \mathbf{v} in the usual Cartesian basis (left), or in the polar basis (right). Both bases are orthonormal, but they are oriented differently.

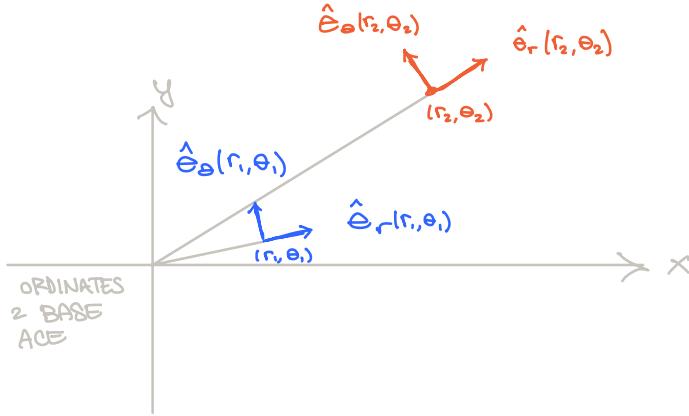


Figure 15: The basis vectors are a basis for potential velocities of a particle at that point. The orientation of the polar coordinate basis vectors depends on the point at which we are using them.

9.1.2 What else do we notice?

The metric depends on the location of the vector—please make sure to do Exercise 9.1. This is in contrast to the ordinary Euclidean metric, which is the same no matter where you are in coordinate space. Another difference is that the orientation of the polar basis vectors is different compared to the Cartesian basis vectors that always point in the x - and y -directions, see Fig. 14. In fact, the orientation of the polar basis vectors is different depending on where in the coordinate space you happen to be looking, see Fig. 15. Note that the length and orthogonality of the polar coordinate basis vectors do not change, just their relative orientation.

When a metric that differs from the identity⁴³ we should suspect that we may be in a *curved space*. For example, the surface of the Earth is a curved space whose metric is spherical

⁴³By which we mean a metric whose components are different from the components of the identity matrix. See Key Idea 9.2.

coordinates is

$$g_{ij} = \begin{pmatrix} g_{\theta\theta} & g_{\theta\phi} \\ g_{\phi\theta} & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} r^2 & \\ & r^2 \sin^2 \theta \end{pmatrix}, \quad (9.10)$$

where r is the radius of the Earth. However, it is not true that all ‘funny looking metrics’ imply that the underlying space is curved. In the polar coordinate example of this subsection, the space itself is the same two-dimensional Euclidean space that we are used to. All we have done is changed coordinates. We see that a funny looking metric may mean that the space is curved, or it may mean that our coordinate system is curved (that is, *curvilinear*).

Exercise 9.2. *Prove (9.10). You may need to review your trigonometry and draw the sphere carefully. Observe that this is a two-dimensional metric space for the surface of the sphere. There is no radial vector. Feel free to derive the three-dimensional spherical coordinate metric that includes the radial direction.*

9.1.3 The polar basis

Let us check the orthonormality of the polar basis at some point, $(r, \theta) = (3, \pi/4)$. The fact that $\hat{\mathbf{e}}_r$ and $\hat{\mathbf{e}}_\theta$ are orthogonal should be obvious: one points in the radial direction and one points in the polar direction—these are orthogonal directions by definition. Another way of seeing this is that the metric is diagonal.

Exercise 9.3. *Confirm that if a metric is diagonal, then the basis vectors are obviously orthogonal.*

The length of the $\hat{\mathbf{e}}_r$ vector is also normalized since $\langle \hat{\mathbf{e}}_r, \hat{\mathbf{e}}_r \rangle = g_{rr} = 1$. What about the length of \mathbf{e}_θ ? This is a vector corresponding to a shift of one unit in the θ direction. We have

$$\langle \mathbf{e}_\theta, \mathbf{e}_\theta \rangle = g_{\theta\theta} = r^2. \quad (9.11)$$

Well that’s unusual. If we are at $r = 3$, then $\langle \mathbf{e}_\theta, \mathbf{e}_\theta \rangle = 9$. This means that the normalized basis vector is

$$\hat{\mathbf{e}}_\theta = \frac{\mathbf{e}_\theta}{r}. \quad (9.12)$$

This makes sense because a ‘shift of one unit in the θ direction’ is a much larger step the further away you are from the coordinate origin, as we saw in Fig. 12.

9.1.4 The tangent space

This example of polar coordinates is helpful because it is a simple case of a funny-looking metric that should feel familiar to us. However, it is also a very dangerous example because it

can lead to some misconceptions. In fact, you should have noticed that something funny was going on here. The metric (9.9) depends on the ‘position’ of the vector—where ‘base’ of the vector located.

This notion of ‘position’ should make you feel uncomfortable. Positions depend on a coordinate system. We remarked before that there is no such thing as a “position vector” and that “position space” is not a vector space—see Exercise 3.1. Positions depend on a choice of origin, whereas in a vector space there is no ‘origin,’ only the zero vector. The zero vector is independent of the choice of basis vectors. There is no such thing as ‘translating’ a vector in a vector space.⁴⁴ What is happening here is that there is a separate space called the **base space or manifold**.⁴⁵ We call the base space M , in our present example $M \simeq \mathbb{R}^2$.

The base space is not a vector space. It has an origin. You can choose different coordinates that have a different origin—for example, you could shift your Cartesian coordinates to the right by one unit. You can choose curvilinear coordinates like polar coordinates. All this is to say that a given *point* on the base space, p , has many different representations depending on your coordinate system. Let’s just call it p , but to some people p corresponds to a specific value of (x, y) , to others its a value of (r, θ) .

Now imagine a particle at position p . Maybe that particle is moving. It has a velocity. The space of possible velocity vectors *is a vector space*. The vector space is called the tangent space at p , which math-y folks call $T_p M$. This is a special two-dimensional vector space that *looks a lot like* the base space, but it is not. The velocity vectors obey the rules for a vector space: you can take linear combinations, there is a well defined zero vector, and so forth. In particular, the vectors are independent of any particular coordinate system of the base space. They have their own basis, which we can draw as a set of vectors whose base is “at” p . This is illustrated in Fig. 16. We can determine a basis for each tangent space by using the coordinates. The polar basis comes from considering the directions that infinitesimal changes in r and θ correspond to at the point p .⁴⁶

The metric is defined for *each* tangent space. There are an infinite number of tangent spaces, one for each point p in the manifold. The polar coordinate metric (9.9) tells us the following. Suppose you have a point p with a tangent space $T_p M$. What is the appropriate metric for this space if you are using the polar basis vectors? Take (9.9) and plug in the value of (r, θ) corresponding to p . In this case the metric is independent of θ . Suppose you are at a point 2 units from the origin and $\pi/6$ from the x -axis, $(r, \theta) = (2, \pi/6)$. Then the metric for the tangent space at this point is

$$g_{ij}|_p = \begin{pmatrix} 1 & \\ & r^2 \end{pmatrix} \Big|_{r=2} = \begin{pmatrix} 1 & \\ & 4 \end{pmatrix}. \quad (9.13)$$

⁴⁴You can add vectors, but one should not think about this as translation because there is a more subtle idea of how to translate a vector across a curved surface. The word ‘translate’ is being used differently, but to avoid potential confusion there is no need to talk about vector addition within a vector space as ‘translation.’

⁴⁵This comes with some mathematical assumptions about topology, but for our present purposes just assume that the manifold is something like a sheet of graph paper whose grid lines gives us a coordinate system.

⁴⁶Do you see how this now interfaces with multivariable calculus? The marriage of calculus and geometry is called differential geometry.

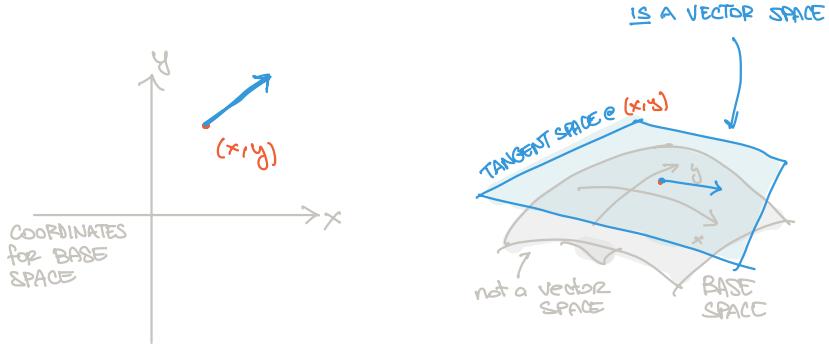


Figure 16: The base space is “coordinate space.” These are positions. Positions are not vectors and the base space is not a vector space. Positions depend on the coordinate system. At each position (x, y) , there is a vector space called the tangent space at (x, y) . This *is* a vector space and the vectors can be thought of as possible velocities of a particle at (x, y) . In the picture on the right hand side we have ‘peeled’ the base space away from the tangent space to make it clear that they are different, even though they both look like sheets of paper. There is no assumed curvature in the base space. There is a different tangent space for each position. The combination of the infinite number of tangent spaces over the base space is known as a tangent space bundle.

If you are at a different point p' with $r = 3$, then the metric for $T_{p'}M$ is instead

$$g_{ij}|_{p'} = \begin{pmatrix} 1 & \\ & r^2 \end{pmatrix} \Big|_{r=3} = \begin{pmatrix} 1 & \\ & 9 \end{pmatrix}. \quad (9.14)$$

In this way, once you specify the point p , you have an unambiguous metric for $T_p M$

Example 9.2. *The distinction between the tangent space at a point and the base space is especially significant when the base space is curved. Let us return again to the example of a sphere, say the Earth. If you have a particle constrained to move on the surface of the Earth, then its velocity vector points in a plane that is tangent to its position—hence the name ‘tangent space’ for the associated vector space. Suppose you are on the equator and want to move towards the north pole. Then your velocity vector is formally in the z -direction, which coincides with the $-\theta$ -direction, where θ is the polar angle. However, we note that this vector is tangent to the equator and does not ‘follow’ the curvature of the Earth.*

Example 9.3. *A trajectory in a base space M is a curve $\gamma(t)$: this maps each value of a ‘time’ parameter t to a position on the space. This could be, for example, the trajectory of a particle in a curved space. How do vectors at some tangent space $T_{\gamma(t)}M$ relate to the nearby tangent space $T_{\gamma(t+\varepsilon)}M$? In other words, how do we transport vectors from one tangent space to a nearby tangent space? This is called **parallel transport** and is*

one of central mathematical questions en route to the formalism of general relativity.

9.2 Length

The length of a vector $\mathbf{v} = |v\rangle$ is $|\mathbf{v}|$ and is given by the inner product of the vector with itself:

$$|\mathbf{v}|^2 = \langle v, v \rangle = g_{ij}v^i v^j . \quad (9.15)$$

Exercise 9.4. The length-squared of a vector is also called its **norm**. Confirm that with the Euclidean metric in \mathbb{R}^3 this simply corresponds to the distance from the origin.

Sometimes we get lazy and will write $v = |\mathbf{v}|$ when there is no danger of ambiguity.

Key Idea 9.3. Usually physical spaces have a metric where the length of every vector is positive, with the exception of the zero vector that has length zero. We call say that the metric is **positive-definite**. When the metric over a manifold^a is positive definite, we say that the manifold is Riemannian.^b In relativity, there is a relative sign between the spatial and temporal components of the metric. This means that there are vectors that have negative norm. We then say that the manifold is pseudo-Riemannian.

^aManifold? See Section 9.1.4.

^bThere are a few other reasonable assumptions along the lines of differentiability.

Example 9.4. In special relativity and using the sign conventions that I like as a particle physicist, the metric is

$$g_{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} . \quad (9.16)$$

Using these signs, any purely space-like vector $v^\mu = (0, v^1, v^2, v^3)$ has a negative norm:

$$v^2 = v^\mu v_\mu = g_{\mu\nu} v^\mu v^\nu = -(v^1)^2 - (v^2)^2 - (v^3)^2 . \quad (9.17)$$

Exercise 9.5. In special relativity there is a class of vectors with zero norm even though the vectors themselves are non-zero.

We can define **normalized** vectors by dividing them by their norm. I choose to decorate these vectors with a hat, say $\hat{\mathbf{v}}$ or $|\hat{v}\rangle$. These *unit* vectors have norm one:

$$\hat{\mathbf{v}} = |\hat{v}\rangle = \frac{\mathbf{v}}{|\mathbf{v}|} = \frac{|v\rangle}{|v|} = \frac{\mathbf{v}}{\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}} = \frac{|v\rangle}{\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}} . \quad (9.18)$$

9.3 Orthogonality

The metric allows us to define the cosine of the angle between two normalized vectors as follows:

$$\cos \theta \equiv \langle \hat{v}, \hat{w} \rangle . \quad (9.19)$$

This means that the angle between two not-necessarily normalized vectors follows from the bi-linearity of the inner product:

$$\cos \theta \equiv \left\langle \frac{\mathbf{v}}{|v|}, \frac{\mathbf{w}}{|w|} \right\rangle = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{|v||w|} . \quad (9.20)$$

The metric only tells us the cosine of the angle, thus we do not define a sense of whether this angle is positive or negative. In other words, there is not a sense of *orientation* in the metric.

Example 9.5. What is the angle between the following kets:

$$|v\rangle = |1\rangle + |2\rangle \quad |w\rangle = 3|1\rangle ? \quad (9.21)$$

We can start by writing the normalized kets:

$$|\hat{v}\rangle = \frac{1}{\sqrt{2}} (|1\rangle + |2\rangle) \quad |\hat{w}\rangle = |1\rangle . \quad (9.22)$$

Plugging these into the expression for the cosine of the angle between these kets

$$\cos \theta = \langle \hat{v}, \hat{w} \rangle = \frac{1}{\sqrt{2}} \langle 1, 1 \rangle + \frac{1}{\sqrt{2}} \langle 2, 1 \rangle = \frac{1}{\sqrt{2}} . \quad (9.23)$$

Taking the positive value of inverse cosine gives $\theta = \pi/4$. Draw this in \mathbb{R}^2 if the result is not immediately obvious.

Example 9.6. Consider a point in polar coordinates, (r, θ) . The component along the x -axis is simply $r \cos \theta$. One way to think about this is that it is the projection of the point onto the x -axis. In the same way, the inner product of a vector $|v\rangle$ with a unit vector $|\hat{w}\rangle$ gives the component of $|v\rangle$ along the axis defined by $|\hat{w}\rangle$:

$$\langle v, \hat{w} \rangle = |v| \cos \theta . \quad (9.24)$$

9.4 Orthonormal Basis

When I ask freshmen what a vector is, they often reply that it is a quantity that has a *magnitude* and a *direction*. This boils down to the statement:

$$|v\rangle = |v| |\hat{v}\rangle . \quad (9.25)$$

The length $|v|$ is the magnitude, and the unit vector $|\hat{v}\rangle$ only encodes a direction since its length has been normalized. What is significant about this is that we can only define a vector as a “*magnitude and a direction*” if you are not only in a vector space, but a *metric space* where you have defined the metric.

Armed with a metric, we can be more picky about the sets of vectors that are sensible basis vectors. In addition to the requirement that a basis is a *minimal set* of vectors that *span* the vector space, we can use our metric to require that the basis is **orthonormal**. This mean that each basis vector is

1. Normalized, so that $\langle \hat{e}_i, \hat{e}_i \rangle = 1$, and
2. Orthogonal, so that $\langle \hat{e}_i, \hat{e}_{i \neq j} \rangle = 0$.

In other words,

$$\langle \hat{e}_i, \hat{e}_j \rangle = \langle i, j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \quad (9.26)$$

When our basis is orthonormal, then we notice that the inner product is similar to the action of a row vector on a column vector:

$$\langle i, j \rangle = \langle i | j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \quad (9.27)$$

This relation looks silly: what’s a comma versus a vertical bar between friends? But there’s something deeper going on here. The expression $\langle i, j \rangle$ contains two basis column vectors (kets), whereas the expression $\langle i | j \rangle$ contains a basis row vector (bra) and a basis column vector (ket). Evidently when we use our metric to define an orthonormal basis, this whole bra-ket notation starts to hint at its mathematical underpinnings.

From this point on we assume that all of our bases for metric spaces are orthonormal. Having come this far, it would simply be uncouth to do otherwise.

9.5 The tensor structure of the metric

The metric has two lower indices, so we may write it in full tensor notation as

$$g = g_{ij} \langle i | \langle j | . \quad (9.28)$$

When acting on a pair of vectors, we have

$$g(\mathbf{v}, \mathbf{w}) = g_{ij} \langle i | \langle j | v^k | k \rangle w^\ell | \ell \rangle = g_{ij} v^k w^\ell \langle i, k \rangle \langle j, \ell | = g_{ij} v^k w^\ell \delta_k^i \delta_\ell^j = g_{ij} v^i w^j . \quad (9.29)$$

In the above contraction we assumed that the first basis bra $\langle i |$ in g hits the basis ket $|k\rangle$ in $|v\rangle$, and the second basis bra $\langle j |$ in g hits the basis ket $|\ell\rangle$ in $|w\rangle$. By the symmetry of g_{ij} , it does not matter if we had chosen instead to contract the $\langle i |$ with $|\ell\rangle$.

Exercise 9.6. Show that $g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v})$ when $g_{ij} = g_{ji}$.

9.6 Machine to create row vectors

What we started to see in Section 9.4 is that the metric is related to the transmogrification of column vectors (kets) into row vectors (bras). Equivalently, it takes components that have an upper index and turn them into components that have a lower index.

One way to see this is that the metric is a map $g : V \times V \rightarrow \mathbb{R}$. We read this as: feed g two vectors, and it will spit out a number. As a function, the metric has two slots:

$$g(_, _) = \langle _, _ \rangle . \quad (9.30)$$

What if we *pre-load* one of those slots with a vector, $\mathbf{v} = |v\rangle$? Just go ahead and drop a vector into one of the slots—it doesn't matter which, since g is symmetric:

$$g(\mathbf{v}, _) = \langle v, _ \rangle . \quad (9.31)$$

What is this metric-with-a-pre-loaded-vector? Well, there is still one open slot. That open slot takes in a vector. Once you provide that vector, $g(\mathbf{v}, _)$ outputs a number. In fact, $g(\mathbf{v}, _)$ is a linear function of that last slot. Thus $g(\mathbf{v}, _)$ is a *linear* machine that takes a vector and returns a number. This is *precisely* what we defined to be a *row vector* (bra) in Section 7.2. Evidently a metric takes a vector and outputs a machine-that-eats-vectors, i.e. a row vector: $g : V \rightarrow V^*$. Of course, this is completely equivalent to our earlier identification:

$$g : V \times V \rightarrow \mathbb{R} \iff g : V \rightarrow V^* . \quad (9.32)$$

What are the components of this new bra that we have created out of a ket, $|v\rangle$? Let's work it out using the machinery in (9.29):

$$g(\mathbf{v}, _) = \langle v, _ \rangle = g_{ij} \langle i | \langle j | v^k | k \rangle = g_{ij} v^j \langle i | \equiv v_i \langle i | \equiv \langle v | . \quad (9.33)$$

Do you see what we have done? We have *defined* a new component that has a lower index, $v_i \equiv g_{ij} v^j$. Of course, this is precisely what is prescribed by the index contraction rules we dictated at the beginning of this class. Here we see *why* that rule makes sense: in the third equality we see that the basis tensor is a bra, $\langle i |$. This means that the object must be a bra (row vector). The components of a bra have lower indices.

Key Idea 9.4. The $\langle v|$ is defined with respect to metric pre-loaded with $|v\rangle$. If we were sticklers, we would have called this $\langle g(v, _)|$ so that nobody gets confused that somewhere out in the world there is a row vector that had nothing to do with $|v\rangle$ but that we had irresponsibly labelled with the same name. Because the metric shows up so often, it is a convention that we do not explicitly write it when there is no ambiguity. Thus we understand that there only ever was a ket $|v\rangle$, and the thing that we are now calling $\langle v|$ comes from the metric “lowering the index” of v^i . We call this lowered index v_i , which we understand to actually mean the sum $g_{ij}v^j$. That contraction with one index of the metric is what we mean by “lowering an index.”

You can use the metric to lower *any* upper index. For example, if you have a tensor with p upper indices and q lower indices, then the metric can convert one of those upper indices into a lower index:

$$g_{i_p j_{q+1}} T^{i_1 \dots i_p}_{\quad j_1 \dots j_q} \equiv T^{i_1 \dots i_{p-1}}_{\quad j_1 \dots j_{q+1}} . \quad (9.34)$$

We had to *choose* one of the indices to contract. We have left this choice implicit in the above expression. We have also opted to simply write the components of the tensor rather than writing out all the basis kets and bras.⁴⁷ On the right-hand side the index structure of the component $T^{i_1 \dots i_{p-1}}_{\quad j_1 \dots j_{q+1}}$ is ‘new’ in the sense that no such object existed until we decided to hit $T^{i_1 \dots i_p}_{\quad j_1 \dots j_q}$ with the metric. Rather than calling this ‘new’ tensor T' or gT , we simply let the indices do the talking. If you see a tensor with a familiar name, but some of the indices have fallen compared to where you were expecting them, then you should *understand* that they have been lowered by contraction with the metric.

Exercise 9.7. Suppose you have a metric on a two dimensional vector space,

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (9.35)$$

What are the components of the row vector v_i in terms of the components the column vector v^i ? What are the components of the two-lower-index tensor M_{ij} in terms of the components of the matrix M^i_j ?

Recalling the relation (9.27), we now understand the *tautology*⁴⁸

$$\langle v, w \rangle = \langle v | w \rangle \quad (9.36)$$

⁴⁷If this is not obvious, then please write the full tensors with their tensor product basis.

⁴⁸This is how mathematicians say “more obvious than obvious,” which itself is a phrase that I picked up from Tony Zee.

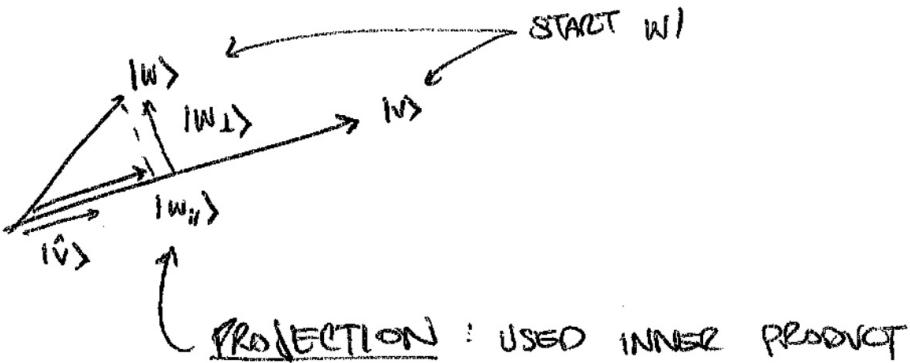


Figure 17: Projection of a vector $|w\rangle$ onto the axis of another vector, $|v\rangle$.

9.7 Example: Projection Matrices

Suppose you have a vector $|v\rangle = 4|1\rangle + 2.5|2\rangle$. What is the component along the x -axis, the one given by the basis vector $|1\rangle$? *Simple!* That's $v^1 = 4$. That's what the linear combination of basis vectors means. Now suppose you have a different vector $|w\rangle$. What is the component of $|w\rangle$ along the axis of $|v\rangle$? See Fig. 17. Furthermore, what is the linear transformation that *projects* any vector $|x\rangle$ onto a vector $|x_{\parallel}\rangle$ that is parallel to $|v\rangle$?

Our strategy is to separate the ket $|w\rangle$ into a piece that is parallel to $|v\rangle$ and a piece that is perpendicular to $|v\rangle$. We will call these pieces:

$$|w\rangle = |w_{\parallel}\rangle + |w_{\perp}\rangle . \quad (9.37)$$

The defining characteristic of $|w_{\parallel}\rangle$ is that $\langle \hat{w}_{\parallel}, \hat{v} \rangle = 1$. That is: the angle between the unit vectors is zero—they are parallel. This means that

$$|w_{\parallel}\rangle \propto |\hat{v}\rangle . \quad (9.38)$$

What is the proportionality? This is simply the length of $|w\rangle$ along the $|\hat{v}\rangle$ axis, which we know from Example 9.6 is the inner product:

$$|w_{\parallel}\rangle = \langle w, \hat{v} \rangle |\hat{v}\rangle = |\hat{v}\rangle \langle \hat{v}, w \rangle . \quad (9.39)$$

In the last line we decided to be silly: since the inner product of two kets is simply a number, we can place it to the right of the ket $|\hat{v}\rangle$. Let us say that this is for aesthetic reasons for now. Then let us recall (9.36) which related an inner product with $|v\rangle$ to a contraction with $\langle v|$. We thus write:

$$|w_{\parallel}\rangle = |\hat{v}\rangle \langle \hat{v}|w\rangle = (|\hat{v}\rangle \langle \hat{v}|) |w\rangle . . \quad (9.40)$$

Here we recognize that the terms in parenthesis are simply the form of a *matrix*, a linear transformation that eats a ket and returns a ket, see (8.8). Evidently $|\hat{v}\rangle \langle \hat{v}|$ is the matrix that takes $|w\rangle$ and returns $|w_{\parallel}\rangle$, the component of $|w\rangle$ along $|\hat{v}\rangle$. In this language, it is somewhat *obvious* that this is what $|\hat{v}\rangle \langle \hat{v}|$ does. However, we have not written the components of $|\hat{v}\rangle \langle \hat{v}|$ in terms of our standard basis. To do that, one simply expands each of $|\hat{v}\rangle$ and $\langle \hat{v}|$ in the standard basis. One may invoke linearity to read off the components of the projection matrix, $P = P_j^i |i\rangle \langle j|$, as seen in the following examples.

Example 9.7. First a simple warm up. Consider the vectors

$$|v\rangle = 2.5 |1\rangle \quad |w\rangle = 3 |1\rangle + 4 |2\rangle . \quad (9.41)$$

We want to find $|w_{\parallel}\rangle$. First we identify the unit vector in the $|v\rangle$ direction, which is simply $|\hat{v}\rangle = |1\rangle$. As a matrix, the projection operator $P = |\hat{v}\rangle \langle \hat{v}|$ is

$$|\hat{v}\rangle \langle \hat{v}| = |1\rangle \langle 1| = \begin{pmatrix} 1 & \\ & 0 \end{pmatrix} . \quad (9.42)$$

In the last step we have matched the components to $P = P_j^i |i\rangle \langle j|$. Applying this to $|w\rangle$ gives

$$|\hat{v}\rangle \langle \hat{v}| w\rangle = |1\rangle \langle 1| (3 |1\rangle + 4 |2\rangle) = 3 |1\rangle . \quad (9.43)$$

Example 9.8. Here is a less-simple example. Consider the vectors

$$|v\rangle = \sqrt{3} |1\rangle + |2\rangle \quad |w\rangle = 3 |1\rangle + 4 |2\rangle . \quad (9.44)$$

We want to find $|w_{\parallel}\rangle$. As a hint, the coefficients of $|v\rangle$ were chosen to be reminiscent of a 30-60-90 degree triangle. The unit vector in the $|v\rangle$ direction is

$$|\hat{v}\rangle = \frac{\sqrt{3}}{2} |1\rangle + \frac{1}{2} |2\rangle . \quad (9.45)$$

As a matrix, the projection operator $P = |\hat{v}\rangle \langle \hat{v}|$ is

$$|\hat{v}\rangle \langle \hat{v}| = \left(\frac{\sqrt{3}}{2} \langle 1| + \frac{1}{2} \langle 2| \right) \left(\frac{\sqrt{3}}{2} |1\rangle + \frac{1}{2} |2\rangle \right) \quad (9.46)$$

$$= \frac{3}{4} |1\rangle \langle 1| \frac{\sqrt{3}}{4} |2\rangle \langle 1| \frac{\sqrt{3}}{4} |1\rangle \langle 2| \frac{1}{4} |2\rangle \langle 2| \quad (9.47)$$

$$= \frac{1}{4} \begin{pmatrix} 3 & \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} . \quad (9.48)$$

Observe that the determinant of $|\hat{v}\rangle \langle \hat{v}|$ is zero. We will see that this is a feature of projection matrices that “throw out” information. In this case, $|\hat{v}\rangle \langle \hat{v}|$ has thrown out information about the perpendicular direction, $|w_{\perp}\rangle$.

Exercise 9.8. Draw the vectors $|v\rangle$ and $|w\rangle$. You should recognize that $|w\rangle$ is $\pi/6$ (30 degrees) counterclockwise of $|v\rangle$. This means that if we rotated the system by this amount in the clockwise direction, the above example should reduce to projection onto the $|1\rangle$ axis, which was our “easy” example above. The appropriate rotation matrix is

$$R = \begin{pmatrix} \cos(-\pi/6) & -\sin(-\pi/6) \\ \sin(-\pi/6) & \cos(-\pi/6) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix}. \quad (9.49)$$

Further, recall how a matrix transforms under a rotation. If you do not remember, please go back and review Section 6.5. Go ahead and do it. I’ll wait here. Yes, it is that important. We see that the components of $P = |\hat{v}\rangle \langle \hat{v}|$ transform as

$$P_j^i \rightarrow (P')_j^i = R_k^i P_\ell^k (R^{-1})_\ell^j. \quad (9.50)$$

Do not forget that the inverse of a rotation is simply its transpose. Show that the only non-zero component of P' is $(P')_1^1 = 1$. Thus P' is indeed a projection onto the x -axis.

Key Idea 9.5. The ket-bra notation for matrices gives a nice representation of the unit matrix:

$$\mathbb{1} = \sum_i |i\rangle \langle i|. \quad (9.51)$$

In quantum mechanics we say that we insert a complete set of states. All that means is that we’re multiplying by one. Keep this in mind for later. The significance of this “multiply by one” operation is that it is true no matter what basis $|i\rangle \langle i|$ is written in. We will use “multiplication by one” judiciously to convert from one basis to another.

Example 9.9. Consider the following bases:

$$|1'\rangle = \frac{1}{\sqrt{2}} |1\rangle - \frac{1}{\sqrt{2}} |2\rangle \quad |2'\rangle = \frac{1}{\sqrt{2}} |2\rangle + \frac{1}{\sqrt{2}} |1\rangle. \quad (9.52)$$

This is clearly a rotation by $\pi/4$. What are the components of $|v\rangle = |1\rangle + |2\rangle$ in the primed basis? The answer should be obvious either from inspection or by drawing a picture. Let us see how to calculate this by “multiplying by one”:

$$|v\rangle = \mathbb{1} |v\rangle = (|1'\rangle \langle 1'| + |2'\rangle \langle 2'|) (|1\rangle + |2\rangle). \quad (9.53)$$

Now we use the overlaps:

$$\langle 1' | 1 \rangle = \frac{1}{\sqrt{2}} \quad \langle 1' | 2 \rangle = \frac{-1}{\sqrt{2}} \quad \langle 2' | 1 \rangle = \frac{1}{\sqrt{2}} \quad \langle 2' | 2 \rangle = \frac{1}{\sqrt{2}} . \quad (9.54)$$

This gives

$$|v\rangle = \frac{1}{\sqrt{2}} (|1'\rangle - |2'\rangle + |1'\rangle + |2'\rangle) = \frac{2}{\sqrt{2}} |1'\rangle . \quad (9.55)$$

As expected from their definitions, $|v\rangle$ is along the $|1'\rangle$ direction.

9.8 Inverse Metric

The metric has an inverse, g^{-1} . If we think of the metric as a machine that lowers indices, the inverse metric raises indices. This should already tell you that the index structure of the inverse metric is $(g^{-1})^{ij}$. By virtue of being an inverse, we expect that acting with the metric and then its inverse—or the reverse order—returns you to the original tensor:

$$g_{ik}(g^{-1})^{kj} = (g^{-1})^{jk} g_{ki} = \delta_i^j . \quad (9.56)$$

Note that $(g^{-1})^{ij}$ is symmetric because it inherits this symmetry from g . Because we are *lazy practical*, we recognize that there is never any ambiguity about whether we are using g or g^{-1} . If you know a tensor T and then you see some component that looks like the components of T except one of the indices is upper rather than lower, then you know that that index was raised by contracting g^{-1} . Similarly, if an index that you expected to be upper turns up lowered, then you know that the index was lowered by contracting with g . Thus we usually do not bother writing g^{-1} , leaving it implicit unless there is some ambiguity.

In fact, we can go further. In the rare cases when we *do* explicitly write out the inverse metric, we go ahead and drop the -1 . This is because we can tell from the index structure whether we are using g or g^{-1} . So we now set the convention:

$$g^{ij} \equiv (g^{-1})^{ij} . \quad (9.57)$$

Example 9.10. The following is a metric and inverse-metric pair:

$$g_{ij} = \begin{pmatrix} 1 & \\ & 3 \end{pmatrix} \quad g_{ij} = \begin{pmatrix} 1 & \\ & 1/3 \end{pmatrix} . \quad (9.58)$$

As always, we are writing the components of these tensors. Do not confuse these as “matrices,” see Key Idea 9.2.

9.9 Isometries: A Fancy Name For Rotations

Section 6 highlighted the significance of rotations in \mathbb{R}^2 and \mathbb{R}^3 . What was not evident was *why* rotations are so important and where they came from mathematically. Now that we have defined the metric, we can define rotations in Euclidean space and their generalization to any metric space.

Let R be an invertible transformation—not necessarily a rotation. The inner product of two vectors is $g(\mathbf{v}, \mathbf{w})$. Under the transformation the vectors each transform as $\mathbf{v} \rightarrow \mathbf{v}' = R\mathbf{v}$ and $\mathbf{w} \rightarrow \mathbf{w}' = R\mathbf{w}$. The metric—a two component tensor with lower indices—also transforms, $g_{ij} \rightarrow g'_{ij}$. How does it transform? Recalling how we derived the transformation of matrices in Section 6.5, let us require that $g(\mathbf{v}, \mathbf{r}) = g'(\mathbf{v}', \mathbf{w}')$:

$$g'_{ij}(v')^i(w')^j = g_{k\ell} v^k w^\ell = g_{k\ell} (R^{-1})^k_i (v')^i (R^{-1})^\ell_j (w')^j . \quad (9.59)$$

Peeling off the $(v')^i$ and $(w')^j$ on both sides⁴⁹ gives

$$g'_{ij} = g_{k\ell} (R^{-1})^k_i (R^{-1})^\ell_j = (R^{-1})^k_i g_{k\ell} (R^{-1})^\ell_j . \quad (9.60)$$

For simplicity, define $\bar{R} = R^{-1}$. We shall now define the properties that will make \bar{R} a rotation—if \bar{R} is a rotation then its inverse R is also a rotation. Let us write (9.60) as “matrix multiplication,” by which we mean that consecutive indices are contracted. As we exhorted in Key Idea 9.2, this does not mean that g has the tensorial structure of a matrix (it does not), only that we are writing the components as a 2×2 array such that the index contractions aligns with the grade-school notion of matrix multiplication:

$$g'_{ij} = (\bar{R}^T)_i^k g_{k\ell} \bar{R}_j^\ell \quad (9.61)$$

$$= \begin{pmatrix} \bar{R}^1_1 & \bar{R}^2_1 \\ \bar{R}^1_2 & \bar{R}^2_2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} \bar{R}^1_1 & \bar{R}^1_2 \\ \bar{R}^2_1 & \bar{R}^2_2 \end{pmatrix} . \quad (9.62)$$

In the first rotation matrix we used $(\bar{R}^T)_1^2 = \bar{R}^2_1$. We say that the transformation R is an **isometry** if $g'_{ij} = g_{ij}$.

Let us consider the case when g_{ij} is the Euclidean metric. This means that the components g_{ij} are the components of the identity matrix, $\mathbb{1}$. Then then plugging in $g'_{ij} = g_{ij}$ into (9.61) gives a matrix equation for the components of \bar{R} :

$$\mathbb{1} = \bar{R}^T \mathbb{1} \bar{R} = \bar{R}^T \Rightarrow \bar{R}^{-1} = \bar{R}^T . \quad (9.63)$$

This is the *definition* of a rotation: it is an isometry of Euclidean space. Even though we have shown that $\bar{R} = R^{-1}$ is a rotation, we know that this imposes that the transformation on vectors, $R = \bar{R}^{-1}$, is also a rotation since it is simply a transformation in the opposite direction as \bar{R} .

⁴⁹Recall that “peeling off” is not a valid mathematical operation. What we really mean is that because the equality holds for any v' and w' , the prefactors must match. Note that we labelled our dummy indices in a convenient way to do this “peeling off.”

Exercise 9.9. Show that for \mathbb{R}^2 with the Euclidean metric, the condition (9.63) imposes a generic form

$$\bar{R} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (9.64)$$

It may help to remember that $\cos^2 \theta + \sin^2 \theta = 1$.

Isometries are special transformations. Because they leave the metric unchanged, they also preserve the orthonormality of a basis. Under an isometry, an orthonormal basis is transformed into another orthonormal basis. We will make ample use of this: there is often a “most convenient basis” in which the action of a given matrix simplifies tremendously. Physically, these isometries correspond to a shift in observer. It just goes to show: sometimes the easiest way to solve a problem is to change your perspective.

Exercise 9.10. Show that if $\hat{\mathbf{e}}_i$ is an orthonormal basis, then its transformation under an isometry R is still an orthonormal basis.

Exercise 9.11. Considering the Minkowski metric for (1+1)-dimensional spacetime:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (9.65)$$

We have used μ and ν indices as is conventional for relativity. Following this convention of R , we write Λ for the isometries of Minkowski space. Show that the following forms of Λ satisfy the isometry condition, $\Lambda^T \Lambda = \mathbb{1}$:

$$\Lambda = \begin{pmatrix} \cosh \eta & \sinh \eta \\ \sinh \eta & \cosh \eta \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta \\ \gamma\beta & \gamma \end{pmatrix}, \quad (9.66)$$

where $2 \cosh \eta = e^\eta + e^{-\eta}$, $2 \sinh \eta = e^\eta - e^{-\eta}$, and $\gamma^2 = (1 - \beta^2)^{-1/2}$. It may help to know that $\cosh^2 \eta - \sinh^2 \eta = 1$. The isometry Λ is called a Lorentz transformation.

Key Idea 9.6. Isometries are the generalization of rotations. These are defined by transformations that leave the metric unchanged. In what follows, we will use the word ‘isometry’ and ‘rotation’ interchangeably. These generalized rotations form a set of transformations that let us go from one orthonormal basis to another.

Here's a nice little aside. The isometries that we have met so far all have some continuous transformation parameter. Rotations in \mathbb{R}^2 are defined by a rotation angle θ . In more Euclidean dimensions there are more rotation angles. These rotation angles can take any real value. Even the boosts have a continuous parameter: this is either the rapidity, η , or the velocity in units of the speed of light, $\beta = v/c$. Compositions of isometries are simply another isometry: a rotation by θ_1 and then θ_2 is a rotation by $(\theta_1 + \theta_2)$. A few "beyond the scope of this class" comments are worth highlighting since they will show up in your physics life:

1. Other than the simplest cases, the order in which one composes isometries matters. For an explicit example, see Fig. 10.
2. The *way* in which the order of isometries matters reflects an underlying mathematical structure of the isometries.
3. Isometries are *symmetries* of a physical system.
4. The theory of symmetries with continuous parameters and its application to physical systems is called the representation theory of Lie groups.

9.10 Discrete isometries

One of the notable features of rotations is that you can rotate by any continuous amount, θ . You can rotate by $\theta = \pi/2$, or $\theta = \pi/2 + \varepsilon$ for, say, $\varepsilon = .0001$. Thus the rotation matrix parameterizes an infinite number of possible rotations. Because there is a continuous parameter for the amount by which one performs a rotation, we say that this is a continuous isometry. There are, however, often *discrete* isometries where there is simply a handful of matrices under which the metric is invariant. In Euclidean two dimensional space, one such matrix is

$$P = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}. \quad (9.67)$$

Under P , the second component of a vector (in the standard basis) changes sign. Because the metric transforms with two factors of P —here we use $P^\dagger = P$ —it remains unchanged:

$$P^T g P = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}. \quad (9.68)$$

Exercise 9.12. Suppose that the components of the metric in \mathbb{R}^2 are not diagonal. That is, $g_{12} = g_{21} \neq 0$. Is P an isometry of this metric space?

The total set of isometries is every combination of rotations and discrete transformations. We often say that this is a **group**. One may write a generic isometry as a product of rotations and discrete transformations.

Example 9.11. The following transformation is a combination of a discrete transformation P and a rotation:

$$\tilde{R} = PR = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix} = \begin{pmatrix} c & -s \\ -s & -c \end{pmatrix}. \quad (9.69)$$

This matrix does not look like a rotation. The diagonal elements differ by a sign, while the off-diagonal elements are the same. If we did not know about discrete isometries, we would have said that this matrix does not map one orthonormal basis to another. In fact, it does.

Exercise 9.13. Compare the rotation-and-discrete transformation in (9.69) to a transformation that is simply a rotation. For concreteness, pick a rotation angle $\theta = \pi/4$ so that $c = s = 1/\sqrt{2}$. Apply the rotations R and \tilde{R} to the standard basis. How do the “rotated” bases differ?

In three dimensions this notion shows up as *parity*. Here are two ways to represent the parity transformation in three dimensions:

$$P = \begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix} \quad P = \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}. \quad (9.70)$$

The reason why there are different ways to represent this discrete symmetry is that the two are related by a rotation.

Exercise 9.14. Describe—in words or drawing if that is simpler—the rotation that relates the two parity matrices above. It may help to think about what each representation of P does to the standard basis vectors.

You can tell that the two are related by a rotation if you think about all that “right hand rule” stuff you learned when you first did angular momentum in mechanics. The right hand rule enforced a certain *orientation* of your basis vectors. The positive x , y , and z directions are chosen to have a particular orientation: if you curl the fingers of your right hand from the x -axis to the y -axis, then your thumb is pointing in the z direction. You can rotate your hand in any way, but this orientation stays the same. Similarly, if you are contrarian and decided to define your axes by a “left hand rule,” then (1) several of your angular momentum-related results are off by a minus sign, and (2) there is no way to rotate your left hand to make it match the orientation of your right hand. This notion is pervasive in nature in three spatial dimensions: fundamental particles have a spin that defines an orientation that is either left- or right-handed. Even protein can have a *chirality* where identical sequences of amino acids may

spiral in one direction or another—and biological life on earth can only use one orientation and not the other.

By manipulating objects (like your right hand) you have an intuitive sense of rotations in three-dimensions. What may be less obvious is that you also have a simple tool for understanding what a parity transformation does: look in the mirror. If you raise your right hand, then the image of you in the mirror raises their left hand. If you do physics with a right-handed coordinate system, then your mirror image will do physics in a left-handed coordinate system. You can bet that the image of you is composed of proteins with the opposite chirality and subatomic particles spinning in the opposite orientation.

Exercise 9.15. *Explain why your image in a mirror exchanges left and right but does not swap up and down.^a*

^aThis is a question that was posed in my freshman physics class. It was not the first time that I heard the question, but I was surprised that it is a mathematical question in the realm of physics. I have since come to appreciate the significant role that parity plays in many aspects of physics. I hope that soon you will as well.

9.11 The meaning of indices

Having rigorously defined isometries as “generalized rotations” for a given metric, we can now reflect on the significance of our index notation. We proposed that the index pattern of a tensor’s components indicates how the components transform under rotations. This is generalized to the statement that the indices of an object indicate how it transforms under isometries. Because isometries preserve the metric, they naturally map between different orthonormal bases. These orthonormal basis represent different ‘reference frames’ in which a physical system is described. By rotating (isometry-ing) to a convenient basis, one can often convert a challenging problem into a much simpler problem.

It is a bit of a joke in physics that there is a standard definition for what a tensor is:

A tensor is an object that transforms like a tensor.

This is not an obviously useful definition, though it does highlight the significant piece for physicists: transformation laws. Mathematicians often find this definition silly. The ‘proper’ definition of a tensor is that it is a multi-linear map between vector spaces. Further, mathematicians will scoff at our emphasis on isometries (rotations): they will say *hey, you, physicists: don’t you see that tensors will transform ‘like a tensor’ for any invertible transformation? Why are you so obsessed with isometries?*

Exercise 9.16. *Confirm what our mathematician friend is saying. Suppose you have a general tensor with p upper indices and q lower indices.^a Further, suppose that your vectors transform under an invertible transformation M , $|v\rangle \rightarrow M|v'\rangle$. Under this transformation, argue that row vectors transform with M^{-1} , and further that the general*

tensor transforms as in (6.23) with the rotation matrix $R(\theta)$ replaced by the general invertible matrix M . You may want to follow the rough steps in Section 6.5 where we motivated the transformation law for matrices.

^aYou can always start some specific index structure and see how the result generalizes.

So why are we making a big deal about isometries? Isometries physically correspond to a change in observer. As physicists, we would like to deduce the fundamental laws of nature—laws that we expect to be independent whether we are oriented along a particular direction in space or in a particular reference frame. Isometries are transformations that translate the physical quantities that one observer measures to the physical quantities that another observer measures. The physical quantities are encoded in the components of tensors and the fact that those components transform indicate what a different observer would measure.

In this sense, it is true that you could do an arbitrary invertible transformation⁵⁰ and this all holds. However, a transformation that is not an isometry will change the metric. This means the transformation maps an orthonormal basis to a non-orthonormal basis.⁵¹ This means that the transformed basis is not meaningful as the basis vectors that another physical observer may measure—at least not a sensible physical observer with a sensible basis.

9.12 Transpose, Adjoint

In our discussion so far, we have made use of a ‘colloquial’ definition of the transpose of a matrix: $(M^T)^i_j = M_j^i$. What dictated the index height? Thus far we have been glib about this because we have not had to use a careful definition of the transpose. Now that we have a metric, we can be more rigorous. The **adjoint** of a linear transformation A is defined to be A^\dagger such that

$$\langle A^\dagger \mathbf{v}, \mathbf{w} \rangle \equiv \langle \mathbf{v}, A \mathbf{w} \rangle . \quad (9.71)$$

Example 9.12. By the symmetry of the metric we could have equivalently written

$$\langle A^\dagger \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A \mathbf{w} \rangle = \langle \mathbf{w}, A^\dagger \mathbf{v} \rangle = \langle A \mathbf{w}, \mathbf{v} \rangle . \quad (9.72)$$

The point of the metric is not that you’re moving the matrix A from one slot to the other, it is that you are changing from a transformation A that acts on \mathbf{v} to a transformation A^\dagger that acts on the other input vector \mathbf{w} defined such that the inner product is the same.

⁵⁰Invertibility is important to make sense of the transformation of lower-indices.

⁵¹Convince yourself that the condition of invertibility ensures that the orthonormal basis is at least mapped to a basis.

When A is a real matrix, $A^\dagger \equiv A^T$. What does this mean? Write out the components of each side:

$$\langle A^\dagger \mathbf{v}, \mathbf{w} \rangle = w^k g_{ki} (A^\dagger)_\ell^i v^\ell \quad \langle \mathbf{v}, A \mathbf{w} \rangle = v^\ell g_{\ell j} A^j_k w^k . \quad (9.73)$$

Now set these two inner products equal to each other gives

$$g_{ki} (A^\dagger)_\ell^i = g_{\ell j} A^j_k \quad (9.74)$$

Contracting both sides with the inverse metric g^{km} gives

$$(A^\dagger)^m_\ell = g_{\ell j} A^j_k g^{km} . \quad (9.75)$$

This relation *defines* the adjoint. Observe that when the metric is diagonal (say for the Euclidean metric), the first index of A^\dagger is a contraction of the second index of A and the second index of A^\dagger is a contraction of the first index of A . In this sense, the order of the indices have swapped.

Ah! But we already have a familiar operation that swaps the order of indices: the transpose, which we defined for [Euclidean] \mathbb{R}^N in (5.37) to be $(M^T)_j^i = M_i^j$. Let us give a formal definition: When the vector space is real, the adjoint is the **transpose** of a matrix. We can read this off of (9.75) when we use⁵² $g_{ij} = g^{ij} = 1$.

What about the height of the indices? In the non-rigorous definition of the transpose, there is a potential ambiguity on the height of the indices. If $M = M^i_j |i\rangle \langle j|$:

$$M^T \stackrel{?}{=} \begin{cases} (M^T)_j^i |j\rangle \langle i| & \text{or} \\ (M^T)_i^j |i\rangle \langle j| \end{cases} . \quad (9.76)$$

This ambiguity came up in (9.8), where we glibly wrote $(A^T)_i^i = A_i^i$, and apologized that this was not quite formally correct—especially since the fact that the “first upper then second lower” index pattern appears to be violated. We now see that the index structure of the transpose of a matrix is preserved: the transpose of a matrix is a matrix. That is to say: a matrix has its first index upper and the second index lower. The transpose has the same pattern.

We see how this happens: the first (upper) index of A becomes the second index of A^T . However, we want the second index of A^T to be *lower*. How do we manage this? The metric gives us precisely the tool to lower indices. We see that this is exactly what happens in (9.75): we use the metric and its inverse to ensure that the heights of the A^T indices match that of an ordinary matrix. This had to be since our definition (9.75) has A^T acting on a vector, so we expect A^T to be an “ordinary” linear transformation.

⁵²Please understand what the following statement means. Normally it is dangerous to write equations where the indices on the left-side do not match the indices on the right-side. What we’re saying here is that the *components* of g and g^{-1} —but not necessarily the tensor structure (upper or lower)—match the components of the identity. Again, we refer back to the subtle point in (9.2).

Exercise 9.17. In your own words, articulate the difference between a matrix A , its inverse A^{-1} , and its transpose A^T . What if A is a rotation?

Exercise 9.18. Show that taking the adjoint twice returns you to the original matrix, $(A^\dagger)^\dagger = A$.

Example 9.13. The transpose/adjoint also gives us a way to understand why isometries act differently on column vectors (ket) versus row vectors (bra). The metric lets us convert a column vectors into row vectors, as we saw in Section 9.6. Start with a vector $|v\rangle$ with components v^i . Under a rotation, these components become $(v')^i = R^i_j v^j$. From the transformed ket $|v'\rangle = (v')^i |i\rangle$ we invoke the metric to convert this into a bra:

$$\langle v'| = \langle v', __ \rangle = \langle __, v' \rangle \equiv v'_i \langle i| . \quad (9.77)$$

In the last line we have implicitly used the metric: we have a definition for $(v')^i$, so the meaning of $(v')_i$ is understood to be

$$(v')_i = g_{ij} R^j_k v^k . \quad (9.78)$$

What are the components of $(v')_i$ relative to v_i ? To do that we can insert $\mathbb{1}$ in a clever way:

$$g^{mn} g_{n\ell} = \delta_\ell^m \quad (9.79)$$

so that

$$(v')_i = g_{ij} R^j_k g^{kl} g_{lm} v^m = (g_{ij} R^j_k g^{kl}) v_\ell = v_\ell (R^\dagger)_i^\ell , \quad (9.80)$$

where again we have implicitly defined $v_\ell = g_{\ell m} v^m$. That's just what metrics do. That's what they do.^a In the last equality we recognize the definition of the adjoint, (9.75). In our case of a real metric space, of course, this is simply $R^\dagger = R^T$. This confirms for us that the objects with lower indices transform by contracting with R^\dagger , which we identified in (9.63) to be the inverse of the rotation R . So there you have it: objects with lower indices transform with $R^{-1} = R^\dagger$.

We can do a sanity check for this. The inner product is a number and should not transform under rotations. We can see what happens:

$$\langle v, w \rangle \rightarrow \langle Rv, Rw \rangle = \langle v, R^\dagger R w \rangle = \langle v, w \rangle . \quad (9.81)$$

where we used $R^\dagger R = \mathbb{1}$.

^a<https://www.youtube.com/watch?v=Akq0xeu-RHE>

9.13 Complex metric spaces

Complex vector spaces It is time to embrace complex numbers. A complex vector space is one where all of the numbers that we previously assumed were real can now be complex: $\mathbb{R} \rightarrow \mathbb{C}$. This means that a linear combination of vectors $\alpha\mathbf{v} + \beta\mathbf{w}$ may have complex coefficients, $\alpha, \beta \in \mathbb{C}$.

Quick refresher of complex numbers Recall that complex numbers may be written either in Cartesian form, $z = x + iy$ or in polar form $z = re^{i\theta}$. The defining features is that $i^2 = -1$. The complex conjugate of a complex number is the complex number you get from changing $i \rightarrow -i$:

$$(x + iy)^* = x - iy \quad re^{i\theta} = re^{-i\theta}. \quad (9.82)$$

Metrics on complex spaces Things become a bit more interesting when we complexify a metric space. We have to update our rules a bit. Previously we assumed that the metric is symmetric, $\langle v, w \rangle = \langle w, v \rangle$. This turns out to only be true for real spaces. When we have a complex vector space, the metric must satisfy symmetry with a complex conjugation:

$$\langle v, w \rangle = \langle w, v \rangle^*. \quad (9.83)$$

That is a little unusual. At least means that the norm (length squared) of a complex vector is real:

$$|v|^2 = \langle v, v \rangle = \langle v, v \rangle^*. \quad (9.84)$$

This helps justify the notation $|v|$ for the magnitude of a vector.

What about the adjoint of a transformation? For a real vector space, the adjoint is simply the transpose. For a complex vector space, the adjoint is the transpose *and* a complex conjugation on each component:

$$(A^\dagger)_j^i = g^{ik} (A^\ell_k)^* g_{\ell j}. \quad (9.85)$$

In terms of components:

$$\begin{pmatrix} A_1^1 & A_1^1 \\ A_2^1 & A_2^1 \end{pmatrix}^\dagger = \begin{pmatrix} (A_1^1)^* & (A_2^1)^* \\ (A_1^2)^* & (A_2^2)^* \end{pmatrix}. \quad (9.86)$$

Example 9.14. Here's the adjoint of an explicit matrix:

$$\begin{pmatrix} 3.5 + 2i & 2.3 + 7i \\ 8.3 + 5i & 5 - 2i \end{pmatrix}^\dagger = \begin{pmatrix} 3.5 - 2i & 8.3 - 5i \\ 2.3 - 7i & 5 + 2i \end{pmatrix}. \quad (9.87)$$

It is also useful to consider the adjoint of a product of matrices, AB . We can see that

$$\langle AB\mathbf{v}, \mathbf{w} \rangle = \langle B\mathbf{v}, A^\dagger \mathbf{w} \rangle = \langle \mathbf{v}, B^\dagger A^\dagger \mathbf{w} \rangle. \quad (9.88)$$

This tells us that

$$(AB)^\dagger = B^\dagger A^\dagger. \quad (9.89)$$

Example 9.15. What do isometries look like in a complex vector space? Let us suppose that we have the Euclidean metric (even though the space is now complex):

$$g_{ij} = \text{diag}(1, 1, \dots, 1) = \mathbb{1}. \quad (9.90)$$

Let U be an isometry in this space; this is a conventional name whose meaning will be clear soon. The condition for an isometry in this space is that a transformation $|v\rangle \rightarrow U|v\rangle$ preserves the metric. Observe that because the metric is not symmetric, but rather symmetric-up-to-a-conjugate, it is useful to write the inner product using brackets rather than simply index contraction. We have:

$$\langle U\mathbf{v}, U\mathbf{w} \rangle = \langle \mathbf{v}, U^\dagger U\mathbf{w} \rangle \stackrel{?}{=} \langle \mathbf{v}, \mathbf{w} \rangle. \quad (9.91)$$

This tells us that the condition for U to be an isometry is that $U^\dagger U = \mathbb{1}$. Matrices that satisfy this are called **unitary** matrices, thus justifying the name U .

Example 9.16. The usual rotation matrices (obviously!) satisfy the unitary condition. Thus rotations are a simple example of a unitary matrix. There are more complicated unitary matrices, for example in two complex dimensions:

$$U(\theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \quad (9.92)$$

is a transformation that rephases the first $v^1 \rightarrow e^{i\theta}v^1$ and second $v^2 \rightarrow e^{-i\theta}v^2$ components of a vector by opposite phases. It should be obvious that $U^\dagger U = \mathbb{1}$.

Exercise 9.19. A bonus observation: show that (9.92) can be understood as a Taylor expansion with respect to a matrix T :

$$U(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} T^n \quad T = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \quad (9.93)$$

The matrix T is called the **generator** of the isometry U .

Exercise 9.20. Show that the generator of rotations is

$$R(\theta) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} T^n \quad T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (9.94)$$

9.14 Gram–Schmidt

The metric allows us to determine whether a basis is orthonormal or not. We *like* orthonormal bases. Sometimes life gives us garbage bases that are not orthonormal. What are we to do? The Gram–Schmidt procedure is a way to use the metric to take any basis and convert it into an orthonormal basis. We already met the key ideas in Section 9.7. Here we focus on the two dimensional case. Refer back to Fig. 17: we have two vectors, $|v\rangle$ and $|w\rangle$. How do we turn these into two orthonormal basis vectors? The general procedure is:

1. Grab one vector from the pile of garbage basis vectors. Normalize this vector and stick it in your bag of nice basis vectors.
2. Grab another vector from the pile. Do whatever you need to do to make it fit as a nice basis vector with the other nice basis vectors in your bag. This means that you have to find what new ‘direction’ the vector identifies relative to the existing basis vectors, and then you have to normalize the component of the vector along that direction. That normalized vector is perpendicular to the other basis vectors and can now be placed in the bag of nice basis vectors.
3. Repeat the previous step until you have a complete basis.

We can start with $|v\rangle$ and simply normalize it: $|\hat{v}\rangle = \hat{\mathbf{e}}_1$ is the first basis vector. The next step is to pull out the perpendicular part of $|w\rangle$. In Section 9.7 we saw how to do this: simply decompose

$$|w\rangle = |w_{\parallel}\rangle + |w_{\perp}\rangle \quad |w_{\parallel}\rangle = \langle \hat{v} | \langle \hat{v} | w \rangle. \quad (9.95)$$

This gives us $|w_{\perp}\rangle$ in terms of two vectors $|w\rangle$ and $|w_{\parallel}\rangle$ that we know. We can thus normalize this and that gives us the second vector: $|\hat{w}_{\parallel}\rangle = \hat{\mathbf{e}}_2$. If you have more vectors, then continue

until you are out. When you identify the perpendicular part, $|u_{\perp}\rangle$, you have to make sure that $|u_{\perp}\rangle$ is perpendicular to *all* of the earlier basis vectors:

$$\langle \hat{\mathbf{e}}_1, u_{\perp} \rangle = \langle \hat{\mathbf{e}}_2, u_{\perp} \rangle = 0 . \quad (9.96)$$

Exercise 9.21. *Would you have the same basis vectors if you started with $|w\rangle$ and then took the perpendicular part of $|v\rangle$?*

10 Special Relativity: Minkowski Space

Minkowski spacetime is the metric space for special relativity. The Minkowski metric is similar to the Euclidean metric, except that the time-like dimensions have a negative sign relative to the space-like dimensions. We say that a spacetime is (3+1)-dimensional when there are three dimensions of space and one dimension of time.⁵³

10.1 Conventions

We call vectors in (3+1)-dimensional vector space **four-vectors** to distinguish them from three-dimensional vectors. We typically label our vectors using zero as the first index⁵⁴:

$$v^{\mu} = \begin{pmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{pmatrix} . \quad (10.1)$$

We are faced with a conundrum: often we will need to consider both *spatial* three-vectors and *spacetime* four-vectors. A standard convention is to label three-vectors in the ‘old school’ way, $\mathbf{v} = v^i \hat{\mathbf{e}}_i$ and use se lower-case Roman indices for spatial indices: $i = 1, 2, 3$ corresponding to the x -, y -, and z -directions.

How, then, do we label spacetime four-vectors? Physicists specializing in relativity have their own notations. However, I am just a poor particle physicist⁵⁵, and so I stick to the convention in this field where we never write four-vectors in index-free notation.⁵⁶ That means we will write a four-vector v^{μ} with $\mu = 0, 1, 2, 3$. We may even refer to this as “the four-vector v^{μ} ,” even though this is not a four-vector, it is a *component* of a four-vector⁵⁷.

⁵³Which dimensions are space and which dimensions are time are mathematically a matter of convention. The physically significant fact is that there is a relative sign. It is rather unusual to have more than one time dimension. There are some cases where you may use two time dimensions as a crutch—look up anti-de Sitter space—but I have never had the occasion to consider more than two dimensions of time.

⁵⁴This is a historical convention. In the past, people would write $x^4 = ict$ to give the relative minus sign in the norm of a four-vector with a Euclidean metric. However, this makes it seem like we are secretly in a complex vector space.

⁵⁵<https://arxiv.org/abs/1012.4515>

⁵⁶Some folks just love indices too much.

⁵⁷https://en.wikipedia.org/wiki/The_Treachery_of_Images

There are two choices for the metric:

$$g_{\mu\nu} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \quad \text{or} \quad g_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}. \quad (10.2)$$

These are known as the mostly-minus (or “west coast”) convention and the mostly-plus (or “east coast”) convention respectively. One of the most diabolically frustrating open debates in physics is which of these equivalent metrics is more convenient. Some physicists like the mostly-minus metric because the norm of the four-momentum of a physical particle is positive. Others like the mostly-plus metric because the spatial components match the Euclidean metric. Whichever one you use, you pay the price somewhere else. As a particle physicist, I impose the mostly-minus metric.

Exercise 10.1. *Does the mostly-minus metric screw things up for ordinary rotations? Argue that it does not by considering \mathbb{R}^3 with a ‘minus-Euclidean’ metric where $g_{ij} = -\mathbb{1}$. The norm of a non-zero three-vector is now always negative, $|v|^2 < 0$. However, we can simply understand that $\langle \mathbf{v}, \mathbf{v} \rangle$ gives a negative number whose absolute value is identified with the norm.*

There’s one more convention that is useful in this business: **natural units**. This has nothing to do with linear algebra, but it does make our lives significantly easier. For our purposes, natural units is the following:

$$c = 1. \quad (10.3)$$

Here c is the speed of light, which usually has units of length divided by time. In natural units, we use the universal constant c to convert between the two. Astronomers are already used to this: they measure distances in light years: the distance that a photon travels in one year, $d = c(1 \text{ year})$. In natural units, we would just say ‘year.’ In particle physics we also set Planck’s constant $\hbar = 1$ which lets us convert between energy and time, but for our purposes it is sufficient to leave c implicit.

Natural units make our notation easier because we want the components of a four-vector to have the same units. It is silly to talk about a displacement four-vector Δx^μ if Δx^0 has units of seconds and Δx^i has units of meters. This is even more perverse when we do a ‘rotation’ between space and time (called a *boost*). If we rotate in space, we know what it means when $\Delta x'^1 = \cos \theta \Delta x^1 + \sin \theta \Delta x^2$. This is because every term has the same dimensions. What does it mean to write $\Delta x'^0 = a \Delta x^0 + b \Delta x^1$ when Δx^0 is in seconds and Δx^1 is in meters? (We sort out the coefficients a and b below.) So one usually writes

$$\Delta x^\mu = \begin{pmatrix} c \Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}. \quad (10.4)$$

Maybe you don't mind that factor of c in there? Things get annoying again when we write down four-momenta, whose time-like component is an energy. In order for the energy of a particle to have the same units as the three-momentum, we need to divide by c . Thus we write

$$p^\mu = \begin{pmatrix} E/c \\ p^1 \\ p^2 \\ p^3 \end{pmatrix}. \quad (10.5)$$

In order to avoid all of these factors of c —you can always replace them at the end of a calculation by requiring dimensional consistency—we simply set $c = 1$ so that

$$\Delta x^\mu = \begin{pmatrix} \Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \quad p^\mu = \begin{pmatrix} E \\ p^1 \\ p^2 \\ p^3 \end{pmatrix}. \quad (10.6)$$

Finally, for our purposes we can focus on (1+1)-dimensional Minkowski space where we only include one dimension of space. This amounts to considering motion only along one spatial direction. We will write the vectors of (1+1)-dimensional Minkowski space as follows:

$$v^\mu = \begin{pmatrix} v^0 \\ \mathbf{v} \end{pmatrix}, \quad (10.7)$$

where \mathbf{v} in (1+1)-dimensions is simply a number. However, the notation above also generalizes nicely if we treat \mathbf{v} as a three-vector.

10.2 Principles of special relativity

The speed of light is constant.

That's it. You can find lots of great expositions on special relativity that derive the implications of the above observation.⁵⁸ To learn this properly, you should also look up the experimental history that lead to the deduction that the speed of light is constant. You should also appreciate how the mathematical structure of Minkowski space was intuited/invented/proposed to explain the above observation. Then you should build some facility on a graphical picture of how special relativity works.⁵⁹ In this section, we very briefly highlight how special relativity is a manifestation of a metric space that is not our usual Euclidean space.

Because we are using units where $c = 1$, the maximum velocity is 1. For historical reasons, we refer to the relative velocity between frames as $\beta = v/c$. In natural units, $\beta = v$ and $|\beta| \leq 1$.

⁵⁸See, e.g. https://www.feynmanlectures.caltech.edu/I_15.html

⁵⁹I highly recommend *Very Special Relativity: An Illustrated Guide*, Sander Bais.

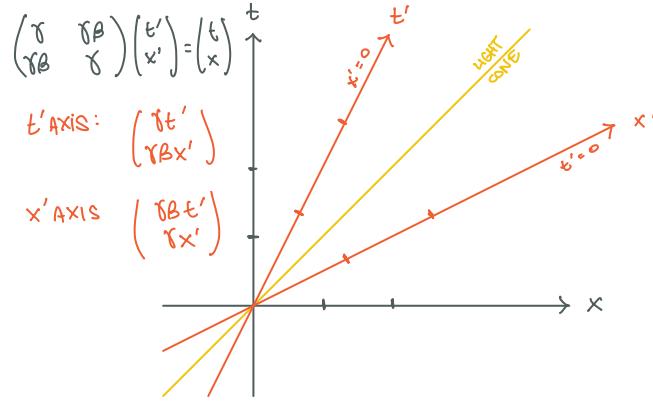


Figure 18: Your coordinate system (t, x) and the coordinate system of a skateboarder moving past you with velocity β . The skateboarder's coordinate system (t', x') appears as the red lines in your coordinate system. The yellow line is the trajectory of light called the light cone.

10.3 Minkowski Space and its Isometries

In Exercise 9.11 you showed⁶⁰ that the isometries of $(1+1)$ -dimensional Minkowski space are the Lorentz transformations Λ satisfying $\Lambda^T \Lambda = \mathbb{1}$. You showed that the following two parameterizations satisfy this condition:

$$\Lambda = \begin{pmatrix} \cosh \eta & \sinh \eta \\ \sinh \eta & \cosh \eta \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta \\ \gamma\beta & \gamma \end{pmatrix}, \quad (10.8)$$

These ‘rotations’ between space and time are called **boosts**. In $(d+1)$ -dimensional Minkowski space with d -dimension of space, you also have rotations in the spatial dimensions. Thus d -dimensional rotations are a subset of the Lorentz transformations. The parameter $\beta = v/c$ is the relative velocity between the frames of the two observers.

Suppose you have a nice coordinate system,⁶¹ (t, x) . You do not observe yourself to be moving⁶², but suppose further that you see someone else skateboarding past you with some velocity v . The skateboarder has their own coordinate system, (t', x') . They observe their coordinate system the same way you observe your coordinate system. A good question is: what does their coordinate system look like *in* your coordinate system? How do the coordinates of a vector change as one boosts from the skateboarder’s perspective to your perspective?

To plot the skateboarder’s axes on your axis, simply recall that the t' axis is determined by all points with $x' = 0$. Similarly, the x' axis is all points with $t' = 0$. We use the Lorentz

⁶⁰You did do this exercise, right?

⁶¹I’m getting myself into trouble again because coordinates are not vectors. In this section what we really mean is that we all agree on an origin of our coordinate systems and we measure the displacements from that origin, $\Delta t = t - 0$ and $\Delta x = x - 0$. These displacements are valid vector spaces.

⁶²We assume the absence of any acceleration. Note that you do not observe yourself to be moving even if you are sitting in a car at constant velocity.

transformation to determine what these points map to:

$$\begin{pmatrix} t(t') \\ x(t') \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta \\ \gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} t' \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma t' \\ \gamma\beta x' \end{pmatrix} \quad (10.9)$$

$$\begin{pmatrix} t(x') \\ x(x') \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta \\ \gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} 0 \\ x' \end{pmatrix} = \begin{pmatrix} \gamma\beta x' \\ \gamma x' \end{pmatrix}. \quad (10.10)$$

We plot these lines in Fig 18. We recognize that because $\gamma \geq 1$, the length of a unit tick in the t' direction as seen in our reference frame has t -component $\gamma t' > t'$. This means that a unit step in the t' direction—which we draw as a tick mark—is stretched out relative to a unit step in the t direction! Of course, this is only ‘stretched out’ because we are used to looking at plots using the Euclidean metric: the Minkowski norm remains the same. The same observation holds for a unit step in the x' direction.

We also draw a diagonal line that we label the light cone. This is the trajectory of a particle moving at the speed of light. As you increase the relative velocity β between the two observers, the axes of one observer relative to the other approach the light cone.

Exercise 10.2. Show that the t' and x' axes approach the light cone as $\beta \rightarrow \infty$.

10.4 Example: Muon Decay

Cosmic rays can hit the upper atmosphere and produce muons about 10 kilometers above the surface of the Earth. These muons are highly relativistic, with a velocity of $\beta = 0.9999$. See Fig. 19. We know from laboratory experiments that the lifetime of a muon *at rest* is 2 microseconds. Based on the simple estimate $d = c\tau \approx 600$ meters, we would not expect any muons to reach the surface of the Earth. However, not only do large cosmic ray telescopes have dedicated muon detectors, but you can make your own citizen science muon detector⁶³. What gives?

Here are the facts:

- Both the observer on Earth and the muon agree that their relative velocity is $|\beta| = 0.9999$.
- The muon’s lifetime is known in the muon’s rest frame.
- The distance from the surface of the Earth to the upper atmosphere is known in the Earth’s rest frame.

With the understanding that boosts will mix up space and time, we can only determine the distance or time that the muon travels by calculating in the same reference frame.

First let us consider calculating everything in the Earth’s reference frame. This is shown on the left of Fig. 20. This means we have to take the muon’s lifetime in the muon’s rest frame and convert it into a lifetime in the Earth’s frame. In the muon’s frame the lifetime is

⁶³<https://muonpi.org>

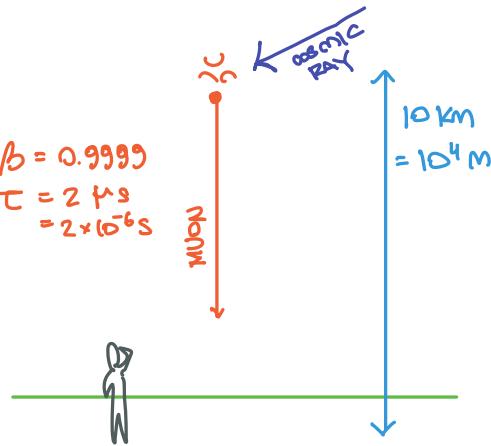


Figure 19: Very technical sketch of a muon produced in the upper atmosphere heading towards earth with velocity $\beta = 0.9999$. A stationary muon has a lifetime of 2×10^{-6} seconds before it decays. The muon is produced 10^4 meters above the surface of the earth. The speed of light is $c = 3 \times 10^8$ meters per second. If we estimate the distance traveled as $d = c\tau$, then the muon travels about 600 meters before decaying. We would then conclude that no muons reach the surface of the earth. What's wrong with this argument?

simply

$$\Delta x'^\mu = \begin{pmatrix} \tau \\ 0 \end{pmatrix} \quad \Delta x^\mu = \begin{pmatrix} \gamma\tau \\ \gamma\beta\tau \end{pmatrix}, \quad (10.11)$$

where we have noticed that the muon lifetime is a displacement along the t' axis used our results in (10.10). We find that the lifetime in the Earth's frame is actually larger than the lifetime in the muon rest frame: $t = \gamma\tau$. There is also a spatial component, but this is no surprise: in the Earth frame the muon is moving, so when the muon decays it is in a different position.

How much is the muon's lifetime *time dilated*? We plug β into the expression for $\gamma = (1 - \beta^2)^{-1/2}$. We find⁶⁴ that to one significant figure, $\gamma = 100$. This means that the time for the muon decay in the Earth's frame is 2×10^{-4} seconds, which means that it travels approximately $d = 60$ km, which is larger than the distance from the upper atmosphere to the surface. As a sanity check: this is exactly the value in the spatial component of Δx^μ .

Great: so one story is that relativistic muons have lifetimes that are much larger than their at-rest lifetime. This means that to the observer on earth, the muon simply lives longer than we would expect from measurements of muons at rest. There are, however, (at least) two sides to every good story. What does the muon see?

In the muon's rest frame, the muon *knows* that it goes *kaput* in 2 microseconds. It sees the surface of the Earth approaching it with velocity $\beta = 0.9999$. By now you can guess that must change: the measurement in the Earth's frame—that the height of the upper atmosphere is 10 km—must be different in the muon's frame. And in fact, the muon must measure the

⁶⁴There's a cute trick here: $\beta^2 = (1 - \epsilon)^2 \approx 1 - 2\epsilon$ by Taylor expanding in the small number $\epsilon = 10^{-4}$.

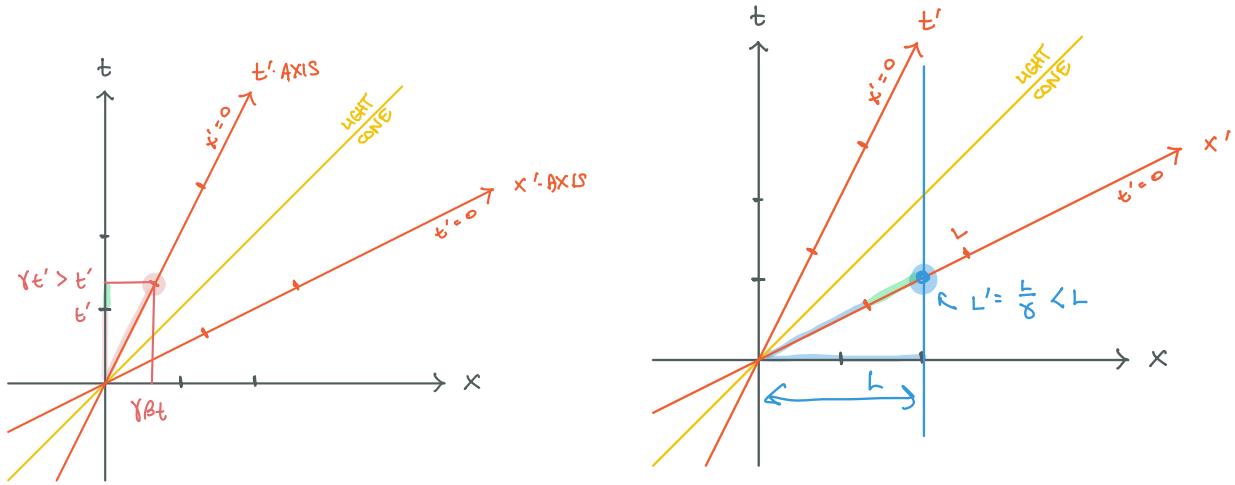


Figure 20: Left: the muon’s lifetime is time-dilated in the Earth’s frame. Right: the distance from the surface of the Earth to the upper atmosphere is length contracted in the muon’s frame.

distance of the rapidly approaching Earth to be much smaller than 10 km. How does this *length contraction* work?

We show this on the right side of Fig. 20. Note that now we have a measurement in the Earth frame (a vertical line denoting a fixed distance) that we want to project onto the muon frame (red axes). In the Earth frame, we denote the distance by two unit ticks in the spatial direction. In the muon frame, this line intersects the x' axis with *less* than two ticks.

Exercise 10.3. Using the Lorentz transformation laws, show that the distance from the muon to the surface of the Earth at the moment of the muon’s creation is $L' = L/\gamma$ where $L = 10 \text{ km}$ is the distance in the Earth frame.

Exercise 10.4. Rephrase everything in this example in terms of the isometries of Minkowski space.

10.5 Example: What does someone else measure?

Momentum is also a four-vector in relativity. The components of the momentum four vector are in (10.6), which we write succinctly as

$$p^\mu = \begin{pmatrix} E \\ \mathbf{p} \end{pmatrix} . \quad (10.12)$$

If we measure the four-momentum of a particle in our laboratory, then we know that the energy of the particle is $p^0 = E$. That’s by definition of what the first component of the

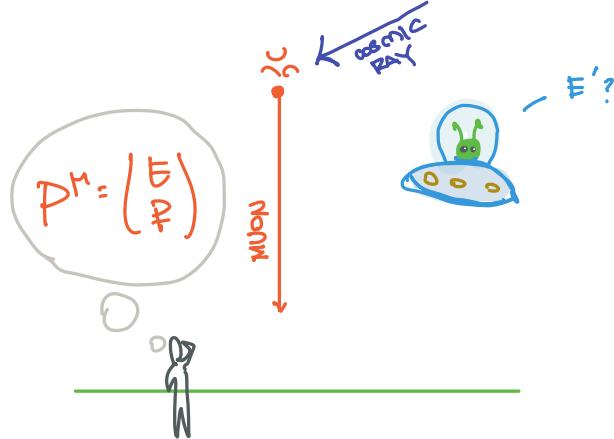


Figure 21: You measure the four-momentum of a particle. One of the components is the energy of the particle. What is the energy that an alien moving at some velocity β relative to you measures?

four-momentum is—we did not actually do any work to do that.

In special relativity there is another object called the four-velocity. In our rest frame, our four velocity is

$$v^\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \quad (10.13)$$

This literally means that when we are at rest, we are moving one second per second in the time direction. Objects moving relative to us have four velocities that are Lorentz transformations of the v^μ above.

We notice that we can write the energy of a particle in a way that uses the inner product:

$$\langle v, p \rangle = p \cdot v = g_{\mu\nu} p^\mu v^\nu . \quad (10.14)$$

Of course, all this does is pick out the p^0 component, as we knew it had to. However, unlike writing p^0 , the inner product $p \cdot v$ has no indices. It is a pure number and so it does not transform under Lorentz transformations.

At this point you wonder if we are simply reciting random facts that we have developed. Consider the following scenario illustrated in Fig. 21. While you have measured the energy E of a particle, you notice an alien traveling relativistically with velocity β relative to you. The alien has sophisticated equipment to measure the particle energy, and you know that the alien measures a different energy E' relative to what you measure. How can you determine what that energy is?

One way to do this is to calculate the full Lorentz transformation between your frame and the alien frame. It turns out, however, that we can use the four-velocity as a useful trick. All objects with mass have a four-velocity equal to (10.13) in their rest frame. This means that the alien measures the particle to have energy $v_{\text{alien}} \cdot p_{\text{alien}}$, where the subscript ‘alien’ means that these are all calculated in the alien’s frame.

We now remember that $v_{\text{alien}} \cdot p_{\text{alien}}$ is a number. It does not matter what frame we



Figure 22: Left: an electron moves near a current. The current has no net electric charge. In the absence of magnetism, we expect the electron to move in a straight line. Right: if we boost into the electron frame, the particles in the current moving in the opposite direction are length-contracted. This means that the charge density increases. The stationary electron now feels a net electric force. This implies that something is missing in the picture on the left.

calculate it in. Thus it is equivalent to the same dot product measured in our frame:

$$E' = v_{\text{alien}} \cdot p_{\text{alien}} = v \cdot p , \quad (10.15)$$

where the right-hand side is the alien four-velocity and the particle four-momentum as measured in our frame. The alien four-velocity is simply a Lorentz transformation of (10.13). More practically, it is something that you can measure in your own frame.

Exercise 10.5. Rephrase everything in this example in terms of the inner product in Minkowski space. Bonus if you use the word ‘projection.’

10.6 Example: the electromagnetic field strength

Another place where special relativity rears its head is in electrodynamics. Electricity and magnetism are two manifestations of the same electromagnetic phenomenon. This is illustrated in Fig. 22. If you did not know about magnetism, you would find a paradox when you consider a charged particle moving along a current. In one frame, the current is an equal number of positive and negative charges moving in opposite directions.⁶⁵ If you boost to the external charged particle’s rest frame, length contraction forces one species of the current particles to increase their charge density relative to the other species. This creates a net electric field that acts on the stationary external particle. Without magnetism, the first frame is missing this additional force.

The electric and magnetic fields are unified in the *electromagnetic field strength*, which is a two-index tensor:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E^x & -E^y & -E^z \\ E^x & 0 & -B^z & B^y \\ E^y & B^z & 0 & -B_x \\ E^z & -B^y & B^x & 0 \end{pmatrix} . \quad F_{\mu\nu} = \begin{pmatrix} 0 & E^x & E^y & E^z \\ -E^x & 0 & -B^z & B^y \\ -E^y & B^z & 0 & -B_x \\ -E^z & -B^y & B^x & 0 \end{pmatrix} . \quad (10.16)$$

⁶⁵I suppose more realistically the negative charges move while the positive charges stay put—that does not change the conclusion.

We see that electricity and magnetism are unified in that their components mix into one another under a Lorentz transformation.

Exercise 10.6. Confirm that $F_{\mu\nu} = g_{\mu\alpha}g_{\nu\beta}F^{\alpha\beta}$ with the (3+1)-dimensional Minkowski metric $g_{\mu\nu}$.

Exercise 10.7. Find the components of $F'^{\mu\nu}$ under a boost along the z -direction.

10.7 Example: simultaneity

One of the key ideas in special relativity is that we sacrifice our notion of simultaneity for objects that are not observed at the same spacetime point. We may reuse our diagram in Fig. 20. Consider two different points on the x' axis. These are simultaneous with respect to the primed observer: they both occur at $t' = 9$. However, these two points obviously do *not* have the same t coordinate to the unprimed observer. This observation helps clear up several apparent paradoxes that may show up in relativity. More importantly, it completely upends our notion of causality.

One of the unwritten-but-understood laws of physics is that the cause precedes the effect. I have to drop my mug before I hear the sound of the ceramic shattering. This notion is imperiled if in some other reference frame someone else would have heard the shattering before they observe the mug being dropped. One deduction of this is that the laws of physics should be *local* in spacetime. A consequence of this observation is that the laws of physics should be written with derivatives, as we will briefly discuss in Section 14.5.

10.8 Example: discrete isometries

The Minkowski metric inherits the same parity isometry that we discussed in Sec 9.10. Its matrix form is

$$P = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \quad (10.17)$$

There is a second discrete transformation called **time reversal**:

$$T = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}. \quad (10.18)$$

This one sounds rather dramatic, doesn't it?⁶⁶ It should be clear that T indeed reverses the direction of time. This, however, does not mean that time reversal is something we can do.⁶⁷ Our understanding of causality breaks if we allow time reversal. However, mathematically time reversal is a clear isometry of the Minkowski metric.

In fact, there is something ‘deep’ to say that the classical laws of physics are time reversal invariant. If you run a video backwards, everything that happens obeys the laws of physics. The sign of the gravitational force may swap, but the dynamics of such an “anti-gravity” force law follows Newtonian mechanics. Entropy may decrease rather than increase, but there is no sense in which the microscopic transition from one configuration to the next would violate any laws of microphysics.⁶⁸ Though, to be fair, we also do not know how to take a right-handed person and do a parity transformation on them to turn them left-handed.

In order to restrict to sensible isometries, we say that valid observers in special relativity are those that are related by isometries that are *connected to the identity*. This means that one may write the isometry with respect to a continuous parameter and that for some value of that parameter, the isometry is simply the identity matrix. In this way, we restrict our physical isometries to those that maintain the direction of time.

11 Central Dogmas

We interrupt our regularly scheduled broadcasting⁶⁹ to take stock of the “big picture” that we are slowly developing. In this section, we review the salient lessons from the first part of this course and present some “practical perspectives” on how to think about the machinery developed. We close by introducing the idea of eigenvectors and eigenvalues.

11.1 Vectors are not columns

Vectors are not columns of numbers. The numbers are the *components* of a vector: we write these with indices. The components just tell you how to write the vector as a linear combination of basis vectors:

$$|v\rangle = v^i |i\rangle = v^i \hat{\mathbf{e}}_i . \quad (11.1)$$

All of the ‘meaning’ of the vector is encoded in the basis vectors. In fact, we have *deliberately* avoid stating what a vector ‘means.’ This meaning is abstract because we want it to be general. Maybe your basis vectors are possible velocities of a particle at a point in space.

⁶⁶One of my mentors in graduate school once told the story that he was preparing a lecture on particle physics at the pub. When his friends asked him what the topic of the lecture would be, he said “time reversal.” The bar crowd suddenly grew silent until the bartender quietly asked: “... we can do that?”

⁶⁷As I write this the video game *The Legend of Zelda: Tears of the Kingdom* was recently released. Time reversal is a major part of the game’s mechanic. Before this, there was the ground-breaking independent computer game *Braid* that pioneered this mechanic. The latter has an additional connection to physics in that the story is largely understood to be a parable about the development of atomic weapons.

⁶⁸This is to say that the “arrow of time” from statistical mechanics is not a statement about microscopic interactions nor is it a statement about what is not *possible*, only about what is increasingly *improbable*.

⁶⁹This is a phrase from my childhood. Television stations would interrupt an ongoing show to present ‘breaking news’ that they felt all of their viewers needed to see.

Maybe the basis vectors are units of something you are counting. Maybe the basis vectors are Fibonacci sequences, or the primary colors. The point is that each of these very different vector spaces are described by the same basic machinery. The vectors live in a vector space, V , which carries this abstraction.

If we generalize this a bit: if you have done a bit of programming⁷⁰, then you may wonder how a tensor is different from an array. An array is a (hyper-)grid of numbers. These are what are encoded by the components of a tensor, say T^{ij}_k . However, a tensor is more than that: it comes with a transformation law for how these components change under an isometry (rotation) that lets us say “an observer in a different basis would measure the components of the tensor to be $(T')^{ij}_k$.

11.2 Everything is a vector space, everything is a linear map

We adopt the convention where vectors have upper indices. This allowed us to introduce row vectors as funny objects with lower indices. We then saw that these two objects are dual to each other:

1. The row vectors are linear functions of the column vectors.
2. The column vectors are linear functions of the row vectors.

In both cases, they are linear functions that output numbers. We called the vector space of row vectors the dual space, V^* . The above statements are:

1. Row vector, $\underline{w} : V \rightarrow \mathbb{R}$
2. Column vector, $\mathbf{v} : V^* \rightarrow \mathbb{R}$

We then realized that we could have linear functions that go from tensor products of vector spaces. That is, linear functions that take some number of row vectors and some number of column vectors and return a number. A tensor with p upper indices (vector-like indices) and q lower indices (row-vector like indices) is a linear map that takes in p row vectors and q column vectors to output a number. In this point of view, tensors are simply linear transformations.

11.3 What the inner product brings

The inner product is a special function that takes two vectors and spits out a number. We may thus write it as a tensor with two lower indices; this tensor is called the metric. A **metric space** or equivalently **inner product space** is a vector space with a metric. The metric and its inverse allow us to raise and lower indices. This gives us a way to create row vectors from column vectors.

We identified a set of linear transformations from $V \rightarrow V$ that leave the metric invariant. These are called **isometries**. Under an isometry, $g'_{ij} = g_{ij}$. Isometries are generalizations of rotations: under an isometry, an orthonormal basis is transformed into another orthonormal basis. From a physics perspective, an isometry takes us from one reference frame to another. That is: a physical quantity may be tensorial. The components of the tensor has significance

⁷⁰In this century, *every* physicist should be able to program.

to a particular observer with a particular basis. A different observer has a different basis that are related to the first basis by an isometry. By performing this transformation, we can determine what the second observer measures.

It also afforded us a notion of the **adjoint** of a matrix, A^\dagger which we define using the inner product:

$$\langle A\mathbf{v}, \mathbf{w} \rangle \equiv \langle \mathbf{v}, A^\dagger \mathbf{w} \rangle . \quad (11.2)$$

The adjoint may seem like a curiosity, except that it defines our notion of a *nice* matrix. A *nice* matrix is one that is self-adjoint: $A^\dagger = A$. From the above definition, we show that the adjoint is equivalent to the Hermitian conjugate:

$$(A^\dagger)^i_j = g_{aj}(A^a_b)^*g^{bi} , \quad (11.3)$$

that is: the transpose with a complex conjugate. We call self-adjoint matrices **Hermitian**. This is the generalization of a symmetric matrix. These matrices are nice because they can be diagonalized by an isometry, and the basis in which they are diagonal will be so convenient that we devote all of Section 12 to it.

11.4 Basis Dependence

When we write the components of a tensor, we assume some basis. In the ‘high school’ picture of vectors as columns of numbers, this basis takes the form

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \dots \quad \hat{\mathbf{e}}_N = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} . \quad (11.4)$$

We colloquially call this the **standard basis**. However, there is nothing holy about this basis. It just happens to be the basis of one observer. The underlying physics of the system described by tensors is *basis independent*. Mathematicians address this by eschewing bases whenever they can. Often this means that physicists reading a mathematics textbook will be amazed at (1) how clean their equations look and (2) how incomprehensible those equations are when you want to figure out what they look like to a specific observer.

Nature is basis-independent. Sometimes we can be sneaky about this: we can derive a result in a particular basis, then write that same result in a way that is basis-independent⁷¹ so that the basis-independent result is true in any basis. We demonstrated this approach in special relativity in Section 10.5.

In what follows, we will contrast the “standard basis” with other convenient bases that come up. In particular, we will be especially fond of the *eigenbasis* where a given (Hermitian) matrix is diagonal. The general strategy we follow will be to

1. We have a matrix M in the standard basis.

⁷¹For example: write the final result in a way where all indices are contracted.

Vector Space	\mathbb{R}	\mathbb{C}	∞ -dimensional
Vector/ket	$\mathbf{v} = v\rangle$	$\mathbf{v} = v\rangle$	f
Basis vector	$\hat{\mathbf{e}}_i = i\rangle$	$\hat{\mathbf{e}}_i = i\rangle$	$\hat{e}_i(x)$
			$\hat{e}_p(x)$
Components	$v^i \in \mathbb{R}$ $\mathbf{v} = v^i \hat{\mathbf{e}}_i = v^i i\rangle$	$v^i \in \mathbb{C}$ $\mathbf{v} = v^i \hat{\mathbf{e}}_i = v^i i\rangle$	$f^i, \tilde{f}(p) \in \mathbb{C}$ $f(x) = f^i \hat{e}_i(x)$ $f(x) = \int dp \tilde{f}(p) e_p(x)$
Row vector/bra	$\mathbf{w} = w_i \hat{\mathbf{e}}^i = w_i \langle i $	$\mathbf{w} = w_i \hat{\mathbf{e}}^i = w_i \langle i $	distribution, e.g. $\delta(x)$
Matrix	$A = A^i_j i\rangle \langle j $	$A = A^i_j i\rangle \langle j $	operator, e.g. $\frac{d^2}{dx^2}$
Inner Product	$\langle v, w \rangle = g_{ij} v^i w^j$ $\langle v, w \rangle = \langle w, v \rangle$	$\langle v, w \rangle$ $\langle v, w \rangle = \langle w, v \rangle^*$	$\langle f, g \rangle = \int dx f^*(x) g(x)$ $\langle f, g \rangle = \langle g, f \rangle^*$
Adjoint	Transpose $(A^T)^i_j = g_{jk} A^k \ell g^{\ell i}$	Hermitian Conjugate $(A^\dagger)^i_j = [(A^T)^i_j]^*$	Integration by parts e.g. $(d/dx)^\dagger = -(d/dx)$
Self-adjoint	Symmetric $A^T = A$	Hermitian $A^\dagger = A$	Sturm–Liouville $\mathcal{O}^\dagger = \mathcal{O}$
	\mathbb{R} Eigenvalues	\mathbb{R} Eigenvalues	\mathbb{R} Eigenvalues
	\perp Eigenvectors	\perp Eigenvectors	\perp Eigenvectors
Isometry, e.g.	Rotations, Boosts	Unitary Matrices	Change of variable

Table 3: Terms and notation in real, complex, and infinite-dimensional vector spaces.

2. We find a special basis (eigenbasis) where M takes a convenient form. We rotate to this convenient basis.
3. In the convenient basis, we do whatever manipulations we need to do to solve our problem. Maybe we need to take an inverse. Maybe it is an exponential. Whatever it is, this is the step where we “do the physically significant” mathematical step.
4. The result that we get is not always easy to interpret. So we have to rotate back to the standard basis. This gives a full solution to our problem in the basis that the problem was originally posed.

11.5 A multitude of conventions

It is potentially confusing that physicists use multiple conventions to describe what are essentially the same ideas in linear algebra. Table 3 makes some of these connections explicit.

11.6 The pantheon of nice matrices

Mathematicians like generality. If there is a true statement for a class of matrices, they are not content with stating the truth—they want to either extend it to all matrices, or otherwise clearly articulate every necessary assumption. This is sometimes caricatured as an obsession with marginal cases. Physicists have the luxury of not having to worry about that: our matrices are all typically “nice.” In this section, we try to give a definition for what *nice* means.

Gold tier The nicest matrices are multiples of the identity matrix, $\mathbb{1}$. These matrices are diagonal. If you rotate them, they don’t change. The inverse of these matrices is simply the inverse of the coefficient, $(\alpha\mathbb{1})^{-1} = \alpha^{-1}\mathbb{1}$. These matrices turn out to be *too* nice. The matrix is proportional to the identity no matter what basis you are in. In fact, these matrices are effectively not even matrices: they’re simply numbers.

Exercise 11.1. Show that if $M = \alpha\mathbb{1}$, then the rotation of M is still M :

$$RMR^\dagger = M . \quad (11.5)$$

Silver tier The next-nicest matrices are diagonal matrices

$$\hat{M} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} . \quad (11.6)$$

We make the following additional assumptions:

1. The diagonal elements are all real. This is a physical bias: the elements of a diagonal matrix typically represent observable quantities. Anything we can measure is a real number.
2. None of the diagonal values are zero. This is because a zero value would discard information. If $\lambda_i = 0$, then $\hat{M}|v\rangle$ is the same no matter what the value of v^i is. This further implies that \hat{M} is not invertible since \hat{M}^{-1} cannot “know” what the i^{th} component of $\hat{M}^{-1}\mathbf{v}$ should be.

Diagonal matrices are a sign that you are using the *right* basis. When acting on the i^{th} basis vector, we simply pick up λ_i :

$$\hat{M}|i\rangle = \lambda_i|i\rangle . \quad (11.7)$$

This pattern is so nice that we call this basis an *eigenbasis*. The diagonal element λ_i is called an *eigenvalue* and the corresponding basis element $|i\rangle$ is called an *eigenvector*. We will specifically be interested in the case where all of the eigenvectors λ_i are real, even when we

are in a complex vector space. Another nice feature of diagonal matrices is that their inverse is easy:

$$\hat{M} = \begin{pmatrix} \lambda_1^{-1} & & & \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ & & & \lambda_N^{-1} \end{pmatrix}. \quad (11.8)$$

Exercise 11.2. Show that if \hat{M} is a diagonal matrix (11.6), then we can make sense of the exponential of this matrix by generalizing the Taylor expansion of the exponential:

$$e^{\hat{M}} = \begin{pmatrix} e^{\lambda_1} & & & \\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_N} \end{pmatrix}. \quad (11.9)$$

Argue that you can also take the logarithm of a diagonal matrix.

Exercise 11.3. Show that if \hat{M} is a diagonal matrix then the following identity holds:

$$\text{Tr } \ln \hat{M} = \ln \det \hat{M} \quad (11.10)$$

This relation turns out to be true even if we replace \hat{M} by a matrix that is not diagonal.

Bronze tier Self-adjoint (Hermitian) are matrices that can be diagonalized. This means that we can alternatively define them to be rotations of diagonal matrices⁷²:

$$M = R\hat{M}R^\dagger. \quad (11.11)$$

Self-adjoint matrices have some nice features that we will exploit in Section 12. Most of the matrices of physical significance are Hermitian. That also means that physicists spend a lot of time *diagonalizing* Hermitian matrices to find the right basis where they are diagonal. As a caveat, we will usually assume that the eigenvalues are all non-zero.⁷³ We will see below that matrices of this type have a clear interpretation: they rescale the vector space along a set of orthogonal axes—the rescaling factor can differ along each axis, and the axes may be different from the standard basis.

⁷²For complex matrices, replace ‘rotation’ with the appropriate isometry. This is usually a unitary matrix, U , that satisfies $U^\dagger U = \mathbb{1}$.

⁷³A zero eigenvalue means that the matrix is not invertible. This is most clearly seen in the basis where the matrix is diagonal.

Exercise 11.4. The **commutator** of two matrices A and B is

$$[A, B] = AB - BA . \quad (11.12)$$

We say that two matrices commute if $[A, B] = 0$. Which of the gold, silver, and bronze tier of matrices commute?

Isometries Isometries are a totally different class of matrices. Isometries are linear transformations that preserve the metric. This means that they transform orthonormal bases into other orthonormal bases. When we diagonalize a Hermitian matrix, we use the isometries of our metric space to change basis. In this sense, we should think about isometries as a transformation between basis vectors.

Projections Projections are *not nice* matrices, except when they are useful. Most *non-invertible* matrices are projections. They are non-invertible because they throw out information. A true projection would be a matrix $|\hat{v}\rangle\langle\hat{v}|$ that acts on a vector by singling out the component along a particular direction. We thus remove projections from our definition of a ‘nice’ matrix. Nice matrices are invertible. Of course, projections can be useful—so we single them out here as a second special class of linear transformation.

Trash What about all the other kinds of matrices? All of our matrices are assumed to be square. These square matrices may be infinite dimensional, but they are square. The uses of non-square matrices are few and far between. Matrices that are not Hermitian, isometries, or projections are also rare. Once in a while there will be a physical matrix that is not one of the gold/silver/bronze medalists nor one of the special classes (isometry, projection). My opinion on this is that young physicists should focus on the central dogma of nice matrices and the corresponding story of eigenvalues and eigenvectors. There is plenty more to learn about linear algebra, but those are things you can learn as you need them.⁷⁴

Example 11.1. An example of a *not-nice* matrix in particle physics are the Yukawa matrices that determine how much the Higgs boson talks to matter particles. When you find such a matrix, you can then take the time to learn the extent to which such a matrix can be diagonalized. The Yukawa matrices are a reason why particle physicists know about singular value decomposition.

In this course and in our physics lives, we almost exclusively focus on the *nice* matrices acting on *nice* metric spaces.

⁷⁴Every young physicist should understand the eigenvalue problem and how it shows up in many guises in their studies—that is the focus of these notes. The linear algebra of matrices that are not ‘nice’ shows up more rarely, and any physics student who deeply understands the ‘central dogma’ can self-study the gaps as needed.

12 Eigensystem

The German prefix *eigen-* means “proper.” The sense in which this is proper is analogous to our notion of niceness in Section 11.6. Consider the following expression for a matrix M :

$$M |\xi\rangle = \lambda |\xi\rangle . \quad (12.1)$$

Here M is a matrix and λ is a number. We say that this is an eigenvalue equation because M acts on a particular vector, $|\xi\rangle$, and the result is a rescaling of that vector, $\lambda |\xi\rangle$. The vector $|\xi\rangle$ is called an **eigenvector** of M with **eigenvalue** λ . The matrix M acts on the eigenvector $|\xi\rangle$ the way that a diagonal matrix acts on a basis vector.

Exercise 12.1. Show that if $|\xi\rangle$ is an eigenvector of M with eigenvalue λ , then it is also an eigenvector of M^{-1} with eigenvalue λ^{-1} .

12.1 Self Adjoint (Hermitian) Matrices

The eigenvalue problem for Hermitian (self-adjoint) matrices is particularly compelling. The following two facts show why:

1. The eigenvalues of Hermitian matrices are real.
2. The eigenvectors of Hermitian matrices are orthogonal.

Exercise 12.2. Prove the above statements.

Exercise 12.3. We say that the eigenvectors of Hermitian matrices are orthogonal, but not necessarily orthonormal. This is because the eigenvalue condition does not guarantee a normalization of the eigenvector. Show that if $|\xi\rangle$ is an eigenvector of M with eigenvalue λ , then $\alpha |\xi\rangle$ is also an eigenvector of M with eigenvalue λ . In the same vein, argue that we can always choose eigenvectors that are normalized so that the eigenvectors of a Hermitian matrices can be chosen to be an orthonormal basis.

12.2 Finding Eigenvalues: Characteristic Equation

Suppose you have a Hermitian matrix M . This means that there is a nice basis where M is diagonal, which we indicate with a hat: \hat{M} . In this *eigenbasis*, each basis vector $|\xi_i\rangle$ has an associated eigenvalue λ_i . Suppose that none of the eigenvalues are zero so that \hat{M} is nice and invertible. The diagonal elements of \hat{M} are simply these eigenvalues, see (11.6). This means

that the matrix $\hat{M} - \lambda_i \mathbb{1}$ is diagonal with (at least) one zero element: the the i^{th} element along the diagonal:

$$\hat{M} - \lambda_i \mathbb{1} = \begin{pmatrix} (\lambda_1 - \lambda_i) & & & \\ & (\lambda_2 - \lambda_i) & & \\ & & \ddots & \\ & & & 0 \\ & & & & \ddots \\ & & & & & (\lambda_N - \lambda_i) \end{pmatrix}. \quad (12.2)$$

We can exploit this curious fact using the observation in Example 5.14: the determinant of a diagonal matrix is simply the product of its diagonal elements. This means that

$$\det(\hat{M} - \lambda \mathbb{1}) = 0 \quad (12.3)$$

whenever λ is an eigenvalue of \hat{M} . This would be a great way to find the eigenvalues of diagonal matrix, \hat{M} , if it weren't for the fact that this is a silly task: if \hat{M} is diagonal, then you can simply *read off* the eigenvalues from the diagonal elements. It would be *much* more useful if we had an expression like (12.3) that we could use to find the eigenvalues of non-diagonal matrices.

It turns out that we can do this. Recall (5.46): the determinant of a product of matrices is the product of the determinants. We motivated this from the observation that the determinant is the area of a parallelogram in two dimensions—though the result generalizes hyper-volumes of *parallelepipeds* in any number of dimensions. From this observation, we also found that the determinant of an inverse matrix is the inverse of the determinant (5.47). This means that

$$\det(R\hat{M}R^\dagger) = \det R \det \hat{M} \det R^\dagger = \det R \det \hat{M} \frac{1}{\det R} = \det \hat{M}. \quad (12.4)$$

Here we used the fact that the adjoint of a rotation⁷⁵ is its inverse. Using the fact that $R\mathbb{1}R^\dagger = \mathbb{1}$, we may thus rotate (12.3) to a basis where \hat{M} is not diagonal:

$$0 = \det(\hat{M} - \lambda \mathbb{1}) = \det[R(\hat{M} - \lambda \mathbb{1})R^\dagger] = \det(R\hat{M}R^\dagger - \lambda \mathbb{1}) = \det(M - \lambda \mathbb{1}). \quad (12.5)$$

This is the **characteristic equation** for the eigenvalues of M . This is incredibly powerful: suppose you have a nice, invertible, self-adjoint matrix M . The matrix is not diagonal and so it is not obvious what its eigenvalues are. The characteristic equation gives you a polynomial in λ that you can solve to determine the values of λ that are eigenvalues of M . For an N -dimensional vector space, you will have an N -dimensional polynomial. Because the matrix is self-adjoint, we know that the eigenvalues are real and so there will be N real roots to the characteristic equation.

⁷⁵And again, this generalizes to complex matrices with a Euclidean metric to a unitary matrix. In full generality, this is an isometry.

Example 12.1. Solving the characteristic equation for 2×2 matrices is straightforward.^a Suppose you have a real, symmetric (i.e. self-adjoint in a real vector space) matrix:

$$M = \begin{pmatrix} M_{11}^1 & M_{12}^1 \\ M_{21}^2 & M_{22}^2 \end{pmatrix} = \begin{pmatrix} a & c \\ c & b \end{pmatrix}. \quad (12.6)$$

The characteristic equation (12.5) is

$$\det(M - \lambda \mathbb{1}) = \begin{pmatrix} (a - \lambda) & c \\ c & (b - \lambda) \end{pmatrix} = (a - \lambda)(b - \lambda) - c^2 = 0. \quad (12.7)$$

This is a quadratic equation with roots^b

$$\lambda_{\pm} = \frac{(a + b)}{2} \pm \frac{\sqrt{(a + b)^2 - 4(ab - c^2)}}{2}. \quad (12.8)$$

These are the eigenvalues of M . You can check that the argument of the square root is always non-negative in accordance with our expectation that the eigenvalues are real. To see this, we note that as $|c|$ becomes larger the argument becomes more positive. For $c = 0$, the argument becomes $(a - b)^2$, whose square root is $|a - b|$. Note that in the $c = 0$ case we find that the eigenvalues are precisely a and b , since in that case M is a diagonal matrix.

^aThe method is straightforward for $N \times N$ matrices as well, but the determinant becomes more tedious to calculate by hand.

^bThere may come a time in your life where you can rattle off facts about the zeros of Bessel functions or wax poetic about the normalization of the Planck mass... only to find that you cannot for the life of you remember the correct factors of two in the quadratic equation. Faced with the choice of re-deriving it in a margin or looking it up on Google, you decide the latter is faster. Then puzzled undergraduates students come into your office and glance at your screen in bewilderment: you're teaching us this advanced stuff and *this* is what is in your search history?

12.3 Finding Eigenvectors

Once you have the eigenvalues of a matrix, you can find the eigenvectors using the eigenvalue equation for each eigenvalue λ_i :

$$M |\xi\rangle = \lambda |\xi\rangle. \quad (12.9)$$

This is a set of N equations for the N components of $|\xi\rangle$. Let us examine a simple example:

Example 12.2. Consider the 2×2 matrix M with the eigenvalues λ_{\pm} :

$$M = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \quad \lambda_{\pm} = -2 \pm 1. \quad (12.10)$$

You should take a moment to derive those eigenvalues using the characteristic equation. Let us find the eigenvector, $|\xi_+\rangle$ corresponding to the eigenvalue $\lambda_+ = -1$. We would like to solve for the components

$$|\xi_+\rangle = \begin{pmatrix} x \\ y \end{pmatrix}. \quad (12.11)$$

We are relaxing some of our notation for clarity: the above equation really means $|\xi_+\rangle = x|1\rangle + y|2\rangle$ with respect to the standard basis in which we defined M . Plugging this into the eigenvalue equation:

$$\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda_+ \begin{pmatrix} x \\ y \end{pmatrix}. \quad (12.12)$$

The equation for the first component is

$$-2x + y = -x, \quad (12.13)$$

from which we deduce $x = y$. This tells us that

$$|\xi\rangle \propto \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (12.14)$$

Great! The second equation should give us the overall normalization of the vector, right? Let us check:

$$x - 2y = -y. \quad (12.15)$$

Huh: this gives us the same equation: $x = y$. Is this unusual? No! We already expressed that the eigenvectors are defined up to an overall normalization: if $|\xi_+\rangle$ is an eigenvector, then so is $\alpha|\xi_+\rangle$. So it is no surprise that the second equation is redundant with the first: the eigenvalue condition is simply not enough to determine the normalization.

In our lives as physicists, we will always want to work with normalized eigenvectors. This normalization, $\langle \xi_+, \xi_+ \rangle = 1$, fixes the components of $|\xi_+\rangle$:

$$|\xi_+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (12.16)$$

And there you go!

Exercise 12.4. Find the eigenvector $|\xi_-\rangle$ of M in (12.12) with eigenvalue $\lambda_- = -3$.

For an N -dimensional vector space, the eigenvalue equation gives N equations for the N components of an eigenvector. It will always be the case that one of these equations

will be redundant: it will not contribute any additional information about the eigenvector components. This is because the overall normalization of the eigenvector is not set by the eigenvector equation. One simply replaces this redundant equation with the normalization condition $\langle \xi, \xi \rangle = 1$.

You may have noticed that in my examples, I try to keep all of my numbers integers. Physicists ‘in the field’ also try to keep their numbers nice—often by picking convenient units. However, despite our best efforts, there you will find that your upper division textbooks are full of funny factors of $\sqrt{2}$ or even (2π) . These factors are usually unavoidable consequences of normalizing eigenvectors.

Exercise 12.5. *If you have to find the eigenvalues and eigenvectors for matrices larger than 3×3 , you may want to use a computer algebra system. Find your favorite computer algebra system (e.g. Mathematica or SciPy) and learn how to find the eigenvectors and eigenvalues of a 3×3 matrix.*

Exercise 12.6. *Suppose you had to diagonalize a 3×3 matrix by hand. Go through the process of finding the eigenbasis, even if you do not “do the math.” How many eigenvalues are there? What does the characteristic equation look like? Is it still a polynomial? If so, what order? Are the roots of the characteristic equation all real, or could there be complex solutions? How do you solve for the eigenvectors? How many equations are there for how many components? How do you use the normalization condition?*

Example 12.3. *It is curious that the Standard Model of particle physics contains matrices up to 3×3 . This shows up in both the gauge group (definition of forces) and the number of generations of matter (so called Yukawa matrices). This is just about the threshold of what a graduate student can work through by hand. At least three Nobel prizes in particle physics came from diagonalizing 2×2 matrices. Two of these diagonalizations are so famous that the rotation angles have names attached to them: the Weinberg angle (for which Steven Weinberg won the prize in 1979) and the Cabibbo angle (for which Cabibbo did not win the prize, but the physicists who generalized the result to 3×3 matrices did). The Weinberg angle measures how the photon and Z-boson mass eigenstates are related to a basis of underlying forces. The Cabibbo angle measures the mixing between the first two generations of quarks. A cousin of the Cabibbo angle for neutrinos was the subject of the 2015 Nobel prize, though to the best of my knowledge this angle has no special name.*

12.4 Rotating to the Eigenbasis

“Finding the eigenvectors” really means finding the components of the eigenvectors in the standard basis. This defines the rotation between the eigenbasis and the standard basis. Let us see how this works explicitly.⁷⁶

For simplicity, we work with a two-dimensional vector space. The final result will be general. Suppose we have an eigenbasis for some nice matrix M . That is we know the components of two eigenvectors ξ_A^i with respect to the two standard basis elements:

$$|\xi_I\rangle \equiv |I\rangle = \xi_I^1 |1\rangle + \xi_I^2 |2\rangle \quad |\xi_{II}\rangle \equiv |II\rangle = \xi_{II}^1 |1\rangle + \xi_{II}^2 |2\rangle . \quad (12.17)$$

Here’s what’s going on with our notation. Our standard basis vectors are $|1\rangle$ and $|2\rangle$. We want to write the eigenbasis in a similar way, so instead of $|\xi_I\rangle$ we write $|I\rangle$. We use Roman numerals to index the eigenbasis to disambiguate from the standard basis. Do not confuse I with a index variable: we use $A = I, II$ as a variable that indexes the eigenbasis. We write the components of the eigenbasis written in the standard basis as ξ_A^i , where $i = 1, 2$ indexes the standard basis. You notice that ξ_A^i is an object with one upper index and one lower index... this smells like a matrix. The indices i and A index the same vector space, but with respect to different bases. So far we are agnostic about whether we write ξ_A^i or ξ_A^i . Because the indices are different, there is never a danger of ambiguity. We will choose a convenient definition shortly.

What can we do with (12.17)? Let us suppose that we have a vector v written in the eigenbasis. We know the components v^A , and we would like to find the components in the standard basis. We can simply plug in (12.17):

$$|v\rangle = v^A |A\rangle = v^A \xi_A^i |i\rangle = \xi_A^i v^A |i\rangle \equiv v^i |i\rangle . \quad (12.18)$$

Here we have chosen to write the indices of ξ_A^i in a particular order, ξ_A^i . This makes $\xi_A^i v^A$ a standard “matrix multiplication” contraction of indices. We identify the components v^i in the standard basis as a transformation of the components v^A in the eigenbasis. This means that ξ_A^i is the rotation from the eigenbasis to the standard basis:

$$v_{\text{std.}}^i = \xi_A^i v_{\text{eig.}}^A \equiv R(\text{std.} \rightarrow \text{eig.})_A^i v_{\text{eig.}}^A . \quad (12.19)$$

Note that in doing this, it was critical that the eigenvectors are normalized. Otherwise this transformation is not a pure rotation, but a rotation and a rescaling.

Key Idea 12.1. *The components of the (normalized) eigenbasis written in the standard basis are simply the components of the rotation matrix from the eigenbasis to the standard basis. The rotation from the standard basis to the eigenbasis is simply the inverse (Hermitian conjugate or transpose) of this rotation.*

⁷⁶This section generalizes to any change of basis. We present it here because in physics, nearly all of your change of basis will be between the basis in which your problem is posed (the standard basis) and the basis in which it is most simply solved (the eigenbasis).

We can go further and relate this rotation to my favorite operation: multiplication by the identity. The trick is as follows:

$$\mathbb{1} = |i\rangle\langle i| . \quad (12.20)$$

Suppose you have a vector $|v\rangle$ whose components you know in the eigenbasis. Another way of writing (12.18) is to multiply by one:

$$|v\rangle = v^A |i\rangle\langle i| A = \langle i| A \rangle v^A |i\rangle . \quad (12.21)$$

We identify $\langle i| A \rangle$ as the rotation from the eigenbasis $|A\rangle$ to the standard basis $|i\rangle$,

$$\langle i| A \rangle = R^i{}_A = \xi^i{}_A . \quad (12.22)$$

We remember that R takes a vector's components in the eigenbasis and returns the vector's components in the standard basis. As a mnemonic, the height of the indices tell you that R can contract with an upper A index (which we use to denote eigencomponents) and returns an upper i index (which we use for standard basis components). Of course, the A and i indices both run over the dimension of the same vector space: we have simply chosen a different notation to make it more clear how this rotation matrix is meant to be used.⁷⁷

Another way that this is often written is that one can insert $\mathbb{1} = R^\dagger R$ where R^\dagger acts on the basis $|A\rangle$ and R acts on the components of a vector, v^A . This is precisely what we meant by a *passive transformation* in Section 6.10—albeit with $R \Leftrightarrow R^\dagger$. Let us see this more explicitly. Rather than using $\mathbb{1} = |i\rangle\langle i|$, we use the components of $\mathbb{1} = R^\dagger R$:

$$\delta_B^A = (R^\dagger R)^A{}_B = (R^\dagger)_i{}^A R^i{}_B . \quad (12.23)$$

Observe that R^\dagger has an upper A index and lower i index. This is because $R^\dagger = R^{-1}$, so that it takes in a vector in the standard basis and returns a vector in the eigenbasis.⁷⁸ Now we can insert this into the expansion of $|v\rangle$ in the eigenbasis:

$$|v\rangle = \delta_B^A v^A |A\rangle = (R^\dagger)_i{}^A R^i{}_B v^B |A\rangle = R^i{}_B v^B [(R^\dagger)_i{}^A \hat{\mathbf{e}}_i] = R^i{}_B v^B |i\rangle = v^i |i\rangle . \quad (12.24)$$

We have used the fact that the basis vector is a lower-indexed object that transforms with R^\dagger under a rotation that acts on the vector space as R . This means that $(R^\dagger)_i{}^A |A\rangle = |i\rangle$. We write $v^i = R^i{}_B v^B$, the component of $|v\rangle$ in the standard basis.

Key Idea 12.2. *Because finding the eigenvectors of a matrix corresponds to finding the components of the rotation matrix between the standard basis and eigenbasis, we often refer to this whole eigen-procedure as **diagonalizing** a matrix.*

⁷⁷Nothing stops you from contracting the eigenbasis A index with a vector's components in the standard basis. However, the way we have constructed R is that such a contraction has no significance: it does not rotate to the eigenbasis and is effectively a random rotation.

⁷⁸Again: we know that R^\dagger is meaningful as a rotation that takes components in the standard basis and returns the components in the eigenbasis. You could use R^\dagger to act on the components in *any* basis, but in general the vector that comes out is not meaningful from the perspective of the eigenvalue problem.

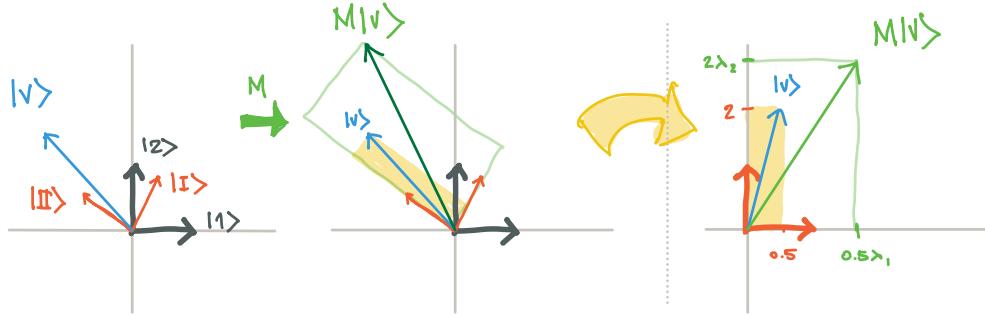


Figure 23: Standard basis (black) and eigenbasis (red) vectors and the vector $|v\rangle$ (blue) in (12.26). Under a transformation M , the vector is mapped to $M|v\rangle$ (green) which is a rescaling of the components of $|v\rangle$ along the eigenbasis directions (shown as a green box). Right: the transformation as seen by an eigen-observer's frame.

■

12.5 Intuition for a linear transformation in the eigenbasis

Passing to the eigenbasis gives a clear understanding of what a Hermitian linear transformation does. We sketch this in Figs. 23 and 24. For that example, we consider a matrix M with eigenvectors $|I, II\rangle$ and corresponding eigenvalues $\lambda_{I,II}$:

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \quad \lambda_{I,II} \approx 3.6, 1.4 . \quad |I\rangle, |II\rangle \approx \begin{pmatrix} 0.5 \\ 0.9 \end{pmatrix}, \begin{pmatrix} -0.9 \\ 0.5 \end{pmatrix} . \quad (12.25)$$

Figure 23 shows a vector

$$|v\rangle = 0.5|I\rangle + 2|II\rangle \approx -1.4|1\rangle + 1.5|2\rangle \quad (12.26)$$

under the transformation M ,

$$M|v\rangle = 0.5\lambda_I|I\rangle + 2\lambda_{II}|II\rangle \approx -1.4|1\rangle + 3|2\rangle . \quad (12.27)$$

What we see is that in the eigenbasis, M simply *rescales* the components of $|v\rangle$ along the eigenbasis directions.

Figure 23 shows a unit ball in the eigenbasis and how it is deformed under M . The surface of the ball should be understood as endpoints of a set of possible vectors. Under M , this surface is deformed to the green ellipse. In the standard basis, this ellipse is skewed by the rotation between the two bases.

Key Idea 12.3. Let us revisit our pantheon of nice matrices that we presented in Section 11.6. The gold medal (proportional to identity) matrices were too simple: they behave like numbers. The silver medal matrices are those matrices that are already in an eigenbasis. The bronze medal matrices are the kinds of matrices that you will most likely

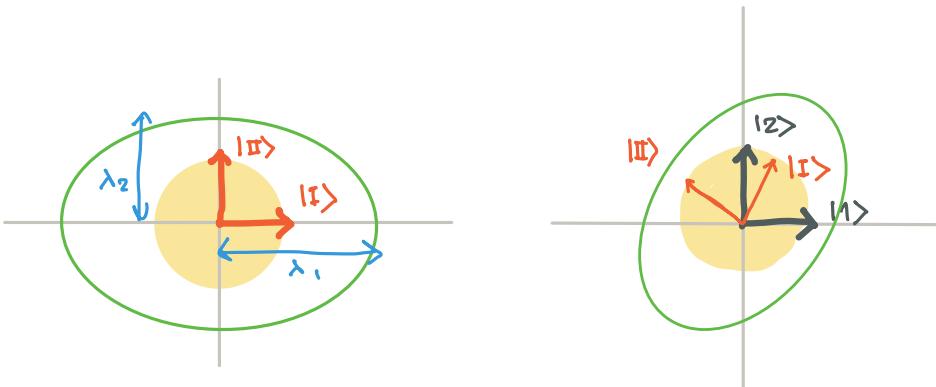


Figure 24: Eigenbasis directions are shown in red. The surface of the yellow ball represents all vectors of unit length. The green ellipse represents the deformation of this surface by acting on each of those vectors by M . We see that the eigenbasis components of these vectors are simply rescaled by the eigenvalues of M . Right: Same figure, but drawn relative to the standard basis.

have to deal with “in the field.” These matrices that are diagonalizable by a rotation, but that are not yet in an eigenbasis. The meaning of these kinds of transformation is simply a rescaling along a set of orthogonal axes—with independent rescaling along each direction.

You can also start to imagine what a generic ‘trash’ matrix does. If the matrix is non-invertible, it is some sort of projection. It throws out information. If the matrix is invertible but not diagonalizable by a rotation, then it means that it is not simply rescaling along a set of orthogonal directions. For example, it may be a rescaling along a set of non-orthogonal directions.

Exercise 12.7. Trash matrices that are invertible but not diagonalizable by a rotation can usually still be diagonalized. You can look up the procedure, it is called a singular value decomposition. The matrix can be made diagonal not by a rotation $M_{\text{trash}} \rightarrow RM_{\text{trash}}R^\dagger$, but by “half rotation” by two different rotation matrices: $M_{\text{trash}} \rightarrow SM_{\text{trash}}R^\dagger$, where $S \neq R$. One way to find these matrices is to diagonalize the Hermitian matrices $N = M^\dagger M$ and $L = MM^\dagger$. Show that N and M are indeed Hermitian and that they are diagonalized by S and R .

12.6 Degenerate Eigenvalues

Sometimes a matrix will have degenerate eigenvalues. This means that $\lambda_i = \lambda_{i+1}$ for some pair of eigenvalues. We continue to assume that the eigenvalues are all non-zero, as required for invertible (nice!) matrices. Degenerate eigenvalues can lead to some curiosities when applying the ‘standard procedure’ above.

Diagonal Matrix

Let us start with the simplest case where we have a diagonal 3×3 matrix, \hat{M} , with a pair of degenerate eigenvalues⁷⁹:

$$\hat{M} = \begin{pmatrix} \lambda_1 & & \\ & \lambda_d & \\ & & \lambda_d \end{pmatrix} \quad \lambda_d \neq \lambda_1 . \quad (12.28)$$

It should be obvious that the eigenvalues are $\lambda_{1,d}$ and that we can choose eigenvectors $|1\rangle$, $|2\rangle$, $|3\rangle$. That is: the standard basis is an eigenbasis. After all, that's what it means for a matrix to be diagonal in a basis.

Just for the sake of argument, let us try to 'derive' the eigenvectors. The first eigenvalue equation is

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_d & \\ & & \lambda_d \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda_1 \begin{pmatrix} x \\ y \\ z \end{pmatrix} . \quad (12.29)$$

We can solve this to find that x can be anything and $y = z = 0$ because $\lambda_1 \neq \lambda_d$. The normalized choice is that the first eigenvector is $|1\rangle$. Curiously, second and third eigenvalue equations are the same:

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_d & \\ & & \lambda_d \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda_d \begin{pmatrix} x \\ y \\ z \end{pmatrix} . \quad (12.30)$$

We can see that $x = 0$, but there appears to be no constraints on y or z other than that they eigenvectors are normalized. How are we supposed to pick them? Let us take stock of the situation: We know that the standard basis vectors $|2\rangle$ and $|3\rangle$ work as eigenvalues. Why doesn't the eigenvalue equation (12.30) just tell us that? If we stuck in *any* values for y and z , then it appears that (12.30) is still satisfied! Curious!

The degeneracy of the second and third eigenvalues has created a degeneracy in how we define our eigenvectors. One way of thinking about this is that the lower 2×2 block of \hat{M} is a 2×2 matrix proportional to the identity. This means if we did a rotation on that block, it does not change:

$$\bar{R} \hat{M} R^\dagger = \hat{M} \quad \bar{R} = \begin{pmatrix} 1 & & \\ & \cos \theta & -\sin \theta \\ & \sin \theta & \cos \theta \end{pmatrix} . \quad (12.31)$$

In the same way, it does not matter how we orient our $|2\rangle$ and $|3\rangle$ basis vectors in this plane as long as (1) they are orthogonal to each other and (2) they are normalized. Any rotation of these two vectors into each other is not only a legitimate basis (which is always true), it is still a legitimate eigenbasis of \hat{M} , see Fig. 25:

$$|2'\rangle = \cos \theta |2\rangle - \sin \theta |3\rangle \quad |3'\rangle = \sin \theta |2\rangle + \cos \theta |3\rangle . \quad (12.32)$$

⁷⁹That means that the eigenvalues are the same, not that they are somehow immoral.

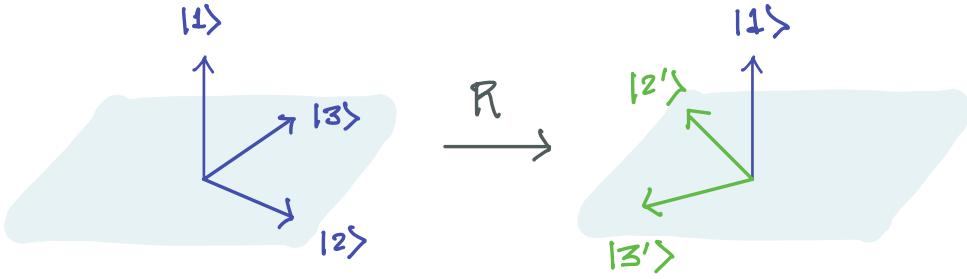


Figure 25: Example of a rotation of a subspace while leaving one direction in the full space unchanged.

Hermitian Matrix

So far so good. What happens with general Hermitian matrices with degenerate eigenvalues? We can reduce this case to the case of a diagonal matrix and deduce that the degeneracy in the eigenbasis appears in this more general case as well. To start, we remember that any Hermitian matrix M can be written as a rotation of a diagonal matrix, \hat{M} :

$$M = R\hat{M}R^\dagger . \quad (12.33)$$

Suppose that M has eigenvalues, λ_1 and λ_d , where the latter eigenvalue has “multiplicity two.” This is our fancy way of saying that there are two eigenvalues that are λ_d . This just means that \hat{M} takes the form in (12.28).

From Section 12.4, we know that the eigenvectors are

$$|A\rangle = (R^\dagger)_A^i |i\rangle . \quad A = \{\text{I, II, III}\} . \quad (12.34)$$

Exercise 12.8. Verify (12.34). One way is to follow the method in Section 12.4. Another way is to use the fact that $M = R\hat{M}R^\dagger$ and to notice that \hat{M} acts on the eigenbasis. At any rate, make sure you are comfortable with this step.

In the eigenbasis, we know that there is a degeneracy where we can rotate by \bar{R} in (12.31). That means:

$$M = R\bar{R}\hat{M}\bar{R}^\dagger R^\dagger . \quad (12.35)$$

But we see that $R' = R\bar{R}$ is simply a rotation from the eigenbasis to the standard basis. Alternatively, $(R\bar{R})^\dagger = \bar{R}^\dagger R^\dagger$, is the rotation from the standard basis to *an* eigenbasis. We have used the rule for the Hermitian conjugate of a product of matrices, (9.89).

We have discovered that both R^\dagger and R'^\dagger are rotations from the standard basis to valid eigenbases. This means that for any lower 2×2 rotation \hat{R} , we still have an eigenbasis with the same eigenvalues.

Key Idea 12.4. When there is a degeneracy in the eigenvalues of a matrix, there is a degeneracy in the definition of the corresponding eigenvectors. A better word to use is symmetry: if two eigenvalues are the same, then there is a symmetry that allows us to rotate the corresponding eigenvectors between themselves while preserving the eigenbasis.

Lifting the Degeneracy

It is often the case that when we have a degenerate eigenvalue, there is a small effect that distinguishes the two eigenvalues. In fact, you may already know examples of this in a different guise: the fine structure of hydrogen. To good approximation, the hydrogen atom's three lowest energy states are a lowest-energy 1S state and two excited 2P states. The energies of these states are eigenvalues of a Hamiltonian ("energy matrix"). The two 2P states correspond to electrons with opposite spins. However, the orbital angular momentum of the electrons creates a little magnetic field that causes one of these degenerate 'eigenstates' to have a little more energy and the other to have a little less energy.

Let us see this in action, albeit for a simpler example. Suppose that in addition to the matrix \hat{M} , you have another matrix, \hat{N} that also is diagonal and also is degenerate. (This latter features is not necessary, but helps illustrate our point nicely.)

$$\hat{N} = \begin{pmatrix} \rho_d & & \\ & \rho_d & \\ & & \rho_3 . \end{pmatrix} \quad \rho_d \neq \rho_3 . \quad (12.36)$$

Observe that the degeneracy in \hat{N} is in the upper two basis vectors, while the degeneracy in \hat{M} is in the lower two basis vectors. While the eigenbasis of either \hat{N} or \hat{M} has some freedom to be rotated, there is only one eigenbasis that simultaneously leaves \hat{N} and \hat{M} diagonal.

Exercise 12.9. Show that the rotation (12.31) that represents the degeneracy in \hat{M} will turn \hat{N} into a non-diagonal (but Hermitian) matrix. Argue that conversely, the rotation that represents the degeneracy in \hat{N} sill render \hat{M} non-diagonal (but Hermitian).

If we happen to have these two matrices, then it should be obvious that there is a *best* basis in which we render both of them diagonal. This basis is unique. In fact, because \hat{M} and \hat{N} are diagonal in the standard basis, then it is *only* in the standard basis that these two matrices are simultaneously diagonal. Let that sink in: there is no longer any degeneracy in what the eigenvectors are or what their eigenvalues are. They are the following:

1. $|1\rangle$ that has eigenvalue λ_1 under \hat{M} and ρ_d under \hat{N}
2. $|2\rangle$ that has eigenvalue λ_d under \hat{M} and ρ_d under \hat{N}
3. $|3\rangle$ that has eigenvalue λ_d under \hat{M} and ρ_3 under \hat{N} .

In fact, we can use the eigenvalues to *label* our eigenvectors:

$$|1\rangle \equiv |\lambda_1, \rho_d\rangle \quad |2\rangle \equiv |\lambda_d, \rho_d\rangle \quad |3\rangle \equiv |\lambda_d, \rho_3\rangle . \quad (12.37)$$

Thus, for example, $M|\lambda_d, \rho_d\rangle = \lambda_d|\lambda_d, \rho_d\rangle$ or $N|\lambda_d, \rho_3\rangle = \rho_3|\lambda_d, \rho_3\rangle$. In physics, you will find ‘kets’ with lists of eigenvalues. These simply mean that you are in a system that has many matrices that are simultaneously diagonal. The eigenvectors of that system are being labelled by the simultaneous eigenvalues with respect to these matrices.

Example 12.4. *When we give the electron states in the hydrogen atom funny names, those funny names are precisely labelling the eigenvalues of simultaneously diagonalizable matrices.*

Diagonalize all the matrices?

The procedure so far seems to reduce to the following observation: everything is good in an eigenbasis. Why don’t we just diagonalize *all the matrices* and work in some mega eigenbasis? Maybe we never have to worry about degenerate eigenvalues because there will always be enough diagonal matrices—like M and N in the previous subsection—to lift all the degeneracies and define a unique eigenbasis.

This is not generally possible. The reason is that the rotation that diagonalizes one matrix is, in general, not the rotation that diagonalizes another.

Exercise 12.10. *Here’s a real-world example from quantum mechanics. Consider the spin of an electron. The spin can be measured along any axis. Let us consider the z -axis and the x -axis. The spin measurement is encoded in the eigenvalues of the following spin matrices:*

$$\hat{S}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \hat{S}_x = \frac{\hbar}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (12.38)$$

Please check that both of these matrices have eigenvalues $\pm\hbar/2$. The first matrix is already in an eigenbasis, while the second is not. Show that diagonalizing \hat{S}_x will cause \hat{S}_z to be not-diagonal.

Comment: in quantum mechanics, measurements are associated with Hermitian matrices. The eigenvalues of the Hermitian matrix encodes the possible values that one may measure. Here we see that the spin of the electron is $\pm 1/2$ along some measurement axis.

Exercise 12.11. *Vectors in quantum mechanics are possible states of a system. A state that is an eigenvector of a Hermitian matrix is one where there is no uncertainty in the value of a measurement. If a state is a linear combination of eigenstates, $|\psi\rangle = a|I\rangle + b|II\rangle$, then the probability of measuring a particular eigenvalue is the squared absolute value of its eigenvector coefficient. That is: given a state $|I\rangle$, the probability of measuring eigenvalue λ_I is $|a|^2$.*

We assume that the states are always normalized, so $|a|^2 + |b|^2 = 1$. In Example 12.10, suppose you started with a state $|\psi\rangle = |\lambda_z = +\hbar/2\rangle$. That is, the eigenstate with \hat{S}_z eigenvalue $\lambda_z = \hbar/2$. Write this in the eigenbasis of \hat{S}_x . Show that the state of definite spin in the z -direction has maximally uncertain spin in the x -direction. That means that the probabilities of measuring spin in the x -direction to be either $\pm\hbar/2$ is 50%.

Maybe that's fair enough. If every matrix could be diagonalized then we would never deal with non-diagonal matrices. But now we have a more pressing question. Suppose you have two different Hermitian matrices, M and N . You know that they can both be diagonalized, but you do not know if they can be *simultaneously* diagonalized. This is a critical question. There are two possibilities:

1. The two matrices can be simultaneously diagonalized. They are different only because they have different eigenvalues. Then there is a rotation R such that

$$M = R\hat{M}R^\dagger \quad N = R\hat{N}R^\dagger , \quad (12.39)$$

where \hat{M} and \hat{N} are diagonal. The rotation R^\dagger is the transformation to go from the standard basis to the eigenbasis.

2. The two matrices cannot be simultaneously diagonalized. They may even have the same eigenvalues (in general they will not), but there is no basis where both M and N are diagonal. In other words:

$$M = R\hat{M}R^\dagger \quad N = S\hat{N}S^\dagger \quad S \neq R . \quad (12.40)$$

You could simply diagonalize both matrices and see what happens. Unfortunately, as algorithmic as we have made the procedure of diagonalization—that is, finding the eigenvectors—it is still kind of a pain to do this to every single matrix we meet. Is there perhaps an easier way to see whether or not two matrices are simultaneously diagonalizable?

Example 12.5. *We have a hint from an observation in Section 11.6. If you have two diagonal matrices, \hat{M} and \hat{N} , then obviously the two matrices are simultaneously diagonalizable. However, the multiplication of diagonal matrices reduces to the multiplication of the diagonal elements as numbers. This means that*

$$\hat{M}\hat{N} = \hat{N}\hat{M} \quad \Rightarrow [\hat{M}, \hat{N}] = 0 . \quad (12.41)$$

Here we have used the definition of the **commutator**, $[A, B] = AB - BA$. We say that two matrices **commute** if their commutator vanishes. We have shown that diagonal matrices

commute.

If you are already in the basis where two matrices are diagonal, then the question is moot. The answer to whether they are simultaneously diagonalizable is solved. The observation that $[\hat{M}, \hat{N}] = 0$ for simultaneously diagonalizable matrices holds for any pair of Hermitian matrices. Let us see why. Let us take our diagonal matrices and rotate them by some matrix R to go to any other basis where they are non-diagonal but Hermitian. Then we have matrices

$$M = R\hat{M}R^\dagger \quad N = R\hat{N}R^\dagger . \quad (12.42)$$

We can write the diagonal matrices in terms of their rotated form and plug this into the commutator:

$$0 = [R^\dagger MR, R^\dagger NR] = R^\dagger MRR^\dagger NR - R^\dagger NRR^\dagger MR = R^\dagger(MN - NM)R = R^\dagger [M, N] R . \quad (12.43)$$

Here we have used $RR^\dagger = \mathbb{1}$. We can multiply both sides by R from the left and R^\dagger on the right—or we may simply argue that rotating a matrix cannot cause it to vanish—to find that the commutator of the matrices vanishes:

$$[M, N] = 0 . \quad (12.44)$$

Critically, this is true in *any* basis, whether or not M and N are diagonal in that basis. Further, you can show that this is true *only* when M and N are simultaneously diagonalizable.

If M and N are not simultaneously diagonalizable, then we could *try* to massage their commutator into a commutator of diagonal matrices:

$$[M, N] = [R\hat{M}R^\dagger, S\hat{N}S^\dagger] = R\hat{M}R^\dagger S\hat{N}S^\dagger - S\hat{N}S^\dagger R\hat{M}R^\dagger . \quad (12.45)$$

We see that $R^\dagger S$ and $S^\dagger R$ do not equal to $\mathbb{1}$ and so we cannot massage this into a form where we get $[\hat{M}, \hat{N}]$.

Key Idea 12.5. *Matrices are simultaneously diagonalizable if they all commute with one another, $[M, N] = 0$. When this is the case, there is a single eigenbasis where all the matrices are diagonal. If some of the matrices have degenerate eigenvalues, it is possible that other matrices can lift those degeneracies. In this way, it is convenient to label eigenvectors by their unique combination of eigenvalues with respect to the set of commuting matrices.*

When matrices do not commute, they cannot be simultaneously diagonalized. In quantum mechanics, two observables that do not commute are observables that you cannot simultaneously measure. This is the algebraic manifestation of quantum uncertainty.

12.7 Inverting Nice Matrices

The power of all of this eigenstuff is that we can solve problems of the following form

$$M |v\rangle = |s\rangle , \quad (12.46)$$

where we are given a nice matrix M and the output vector $|s\rangle$. We are asked to solve for $|v\rangle$. The solution is obviously $|v\rangle = M^{-1} |s\rangle$. Because M is ‘nice,’ we know it is invertible—after all, it is a Hermitian matrix with non-zero eigenvalues. For small and simple matrices, you can simply solve $M^{-1}M = \mathbb{1}$ component-by-component. For moderately difficult matrices you can plug this into a computer. However, we will soon concern ourselves with the problem of *infinite dimensional* vector spaces. In this limit, the problem (12.46) becomes a differential equation

$$\mathcal{O}f(x) = s(x) , \quad (12.47)$$

where \mathcal{O} is a differential operator like $-(d/dx)^2$. The vector $|v\rangle$ is now written as a function f and the discrete component index i of v^i has become a continuous argument x in $f(x)$. You are familiar with these differential equations: they show up all the time in physics.

Example 12.6. *Newton’s law of motion is*

$$m \frac{d^2}{dt^2} \mathbf{x}(t) = -\nabla V[x] , \quad (12.48)$$

where V is the potential of the force.

For this infinite-dimensional limit, the best way to solve (12.47) is to rotate into a basis of *eigenfunctions* and use the fact that the action of \mathcal{O} simplifies greatly in this basis.

It is thus instructive to walk through the process of solving (12.46) using a rotation to the eigenbasis. It may seem silly that we do things this way for a 2×2 matrix, but this example will make it clear how everything works. To this end, let us solve (12.46) with the following values:

$$M = \begin{pmatrix} 9 & -\sqrt{3} \\ -\sqrt{3} & 11 \end{pmatrix} \quad v^i = \begin{pmatrix} x \\ y \end{pmatrix} \quad s^i = \begin{pmatrix} a \\ b \end{pmatrix} . \quad (12.49)$$

We choose explicit numbers for the components of M so that we can write out explicit eigenvalues and derive explicit eigenvectors. We would like to find the components v^i in the standard basis, which we have labelled as x and y . For the source, s^i , we simply assume some *a priori* known values a and b . In fact, our solution will depend linearly on a and b so that we will have solved the problem of inverting this matrix M for *any* source.

Characteristic equation First we solve for the eigenvalues of M . The characteristic equation is

$$0 = \det(M - \lambda \mathbb{1}) = (9 - \lambda)(11 - \lambda) - 3 \quad \lambda_{\text{I,II}} = 8, 12 . \quad (12.50)$$

Eigenvectors The eigenvectors come from solving $M|A\rangle = \lambda_A|A\rangle$ for $A = \text{I}, \text{II}$. The general form of this equation is

$$9q - \sqrt{3}s = \lambda_A q \quad -\sqrt{3}q + 11s = \lambda_B s , \quad (12.51)$$

where the standard basis components of $|A\rangle$ are s and q . Of course, only one of these equations is sufficient to specify the relation between the two components of the eigenvector $|A\rangle$: the other is redundant.⁸⁰ This fixes the “direction” of the eigenvector. We find that the eigenvectors are proportional to

$$|\text{I}\rangle \propto \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \quad |\text{II}\rangle \propto \begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix} . \quad (12.52)$$

Since this is a two dimensional space, one could have simply used orthogonality to deduce $|\text{II}\rangle$ from $|\text{I}\rangle$. Normalizing the eigenvectors according to $\langle A|A\rangle = 1$ gives us

$$|\text{I}\rangle = \frac{\sqrt{3}}{2}|\text{1}\rangle + \frac{1}{2}|\text{2}\rangle \quad |\text{II}\rangle = \frac{1}{2}|\text{1}\rangle - \frac{\sqrt{3}}{2}|\text{2}\rangle . \quad (12.53)$$

Rotating to the eigenbasis Now that we have the components of the eigenbasis in terms of the standard basis, $\langle i|A\rangle$, we know all of the components of the rotation matrix between the two bases, as we saw in Section 12.4. If you’re like me, you always get confused about whether these components are the rotation from the standard basis to the eigenbasis, or from the eigenbasis to the standard basis. Do we arrange them as rows or as columns? It is worth taking a moment to sort that out without looking at Section 12.4. For the sake of pedagogy, let us do this a bit more tediously but straightforwardly. Let us invert (12.53) to write the standard basis vectors in terms of the eigenbasis vectors.⁸¹ By taking appropriate linear combinations, we find that

$$|\text{1}\rangle = \frac{\sqrt{3}}{2}|\text{I}\rangle + \frac{1}{2}|\text{II}\rangle \quad |\text{2}\rangle = \frac{1}{2}|\text{I}\rangle - \frac{\sqrt{3}}{2}|\text{II}\rangle . \quad (12.54)$$

You might object that (12.54) does not look like much of a rotation. You may recall that this is the form of a rotation with a discrete isometry, (9.69). What has happened is that we have changed the relative orientation⁸² of the basis vectors. Or, in other words, maybe we should have chosen a different overall sign for one of the eigenvectors in (12.53). No matter! We press forward.

Inserting (12.54) into our expression for $|v\rangle$, we find the components of $|v\rangle$ in the eigenbasis:

$$|v\rangle = x|\text{1}\rangle + y|\text{2}\rangle = \frac{1}{2}(\sqrt{3}x + y)|\text{I}\rangle + \frac{1}{2}(x - \sqrt{3}y)|\text{II}\rangle \equiv x'|\text{I}\rangle + y'|\text{II}\rangle . \quad (12.55)$$

Similarly, the components of the source $|s\rangle$ are

$$|s\rangle = \frac{1}{2}(\sqrt{3}a + b)|\text{I}\rangle + \frac{1}{2}(a - \sqrt{3}b)|\text{II}\rangle \equiv a'|\text{I}\rangle + b'|\text{II}\rangle . \quad (12.56)$$

⁸⁰Be sure you are clear *why* the second equation has to be redundant. The eigenvalue equation is true even for $|A\rangle \rightarrow \alpha|A\rangle$, so it cannot determine the normalization of the eigenvectors.

⁸¹This is, of course, a rotation.

⁸²I say this somewhat colloquially because the term ‘orientation’ has a more formal meaning.

The primed variables are the components in the eigenbasis. Now that our kets are written as linear combinations of eigenvectors, the action of M is straightforward. Further, the inverse is simply

$$M^{-1} = \begin{pmatrix} 1/\lambda_I & 0 \\ 0 & 1/\lambda_{II} \end{pmatrix}. \quad (12.57)$$

This means that the equation $M |v\rangle = |s\rangle$ is

$$\begin{pmatrix} \lambda_I & 0 \\ 0 & \lambda_{II} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a' \\ b' \end{pmatrix}, \quad (12.58)$$

from which we straightforwardly deduce

$$x' = \frac{a'}{\lambda_I} \quad y' = \frac{b'}{\lambda_{II}}. \quad (12.59)$$

This formally solves the problem, though we should be courteous and state the result in the standard basis.⁸³

Rotating back to the standard basis We have already written transformation back to the standard basis, (12.53). It is straightforward to plug this into our solution in the eigenbasis. For the sake of pedagogy, let us try to write this a a rotation. We find that

$$\frac{1}{2} \begin{pmatrix} \sqrt{3} & 1 \\ 1 & -\sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a'/\lambda_I \\ b'/\lambda_{II} \end{pmatrix}, \quad (12.60)$$

where we have explicit expression for a' and b' in terms of the original source components a and b in (12.56). We recall that this matrix is not a rotation, but rather a rotation and a discrete isometry. We may simplify the problem by multiplying both sides by the discrete isometry

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (12.61)$$

to to “undo” the discrete isometry and recover a rotation matrix

$$\frac{1}{2} \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a'/\lambda_I \\ -b'/\lambda_{II} \end{pmatrix} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} a'/\lambda_I \\ -b'/\lambda_{II} \end{pmatrix}, \quad (12.62)$$

On the right-hand side we have used the fact that the inverse of a rotation is simply its transpose.

13 Quantum Mechanics

[**Flip:** Skipping typing this for the moment. I think this brief introduction to quantum should go before the eigenvalue section, see my hand written notes. It lets us appeal to the spin-1/2 system as an example.]

⁸³This is similar to the etiquette that you should re-rack your weights in the gym.

14 Function Space

[Flip: I am here]

14.1 Functions as Vectors

Functions are vectors in a vector space that we call *function space*. I am pretty sure that mathematicians do not call it that, but we use this terminology anyway. This should be *obviously* true. Consider two polynomial functions f and g . Maybe

$$f(x) = x^3 \quad g(x) = x + 2 , \quad (14.1)$$

though the precise forms do not matter. Clearly you can take a linear combination of the two and the result is also a polynomial function:

$$\alpha f(x) + \beta g(x) = \alpha x^3 + \beta x + 2\beta . \quad (14.2)$$

This is a totally valid function that we could call $(\alpha f + \beta g)(x)$.

In this example, it does not matter at all that the functions/vectors are themselves linear maps from $\mathbb{R} \rightarrow \mathbb{R}$.⁸⁴ All that matters is that the function/vectors have components that add linearly when we take linear combinations. Of course, the components are components *with respect to a basis*. Let us restrict to the class of polynomials up to degree N , so these are functions of the form

$$f(x) = a^0 + a^1 x + a^2 x^2 + \cdots + a^N x^N . \quad (14.3)$$

Okay, the indices are a little annoying. The components a^N have an upper index because they are components of a vector. The monomials x^n are actual powers of the variable x . Evidently, these $(N + 1)$ monomials are our basis for polynomial space.⁸⁵

$$|i\rangle = \hat{\mathbf{i}} = x^i \quad i \in \{0, 1, 2, \dots, N\} . \quad (14.4)$$

Note that we have not said anything about whether this basis is orthonormal or not—we have not yet defined a metric for this polynomial space.

It is clear that the an N -component column vector encodes the information of a function:

$$|f\rangle = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ \vdots \\ a^N \end{pmatrix} \quad |f\rangle = \begin{pmatrix} b^0 \\ b^1 \\ b^2 \\ \vdots \\ b^N \end{pmatrix} . \quad (14.5)$$

⁸⁴This should sound familiar. The dual vector space V^* of row vectors are simultaneously a “vector space in itself” and a space of functions from $V \rightarrow \mathbb{R}$.

⁸⁵The vector space is $(N + 1)$ -dimensional because we include the constant, x^0 .

Further, one may take linear combinations of these columns to encode linear combinations of functions:

$$|\alpha f + \beta g\rangle = \begin{pmatrix} \alpha a^0 + \beta b^0 \\ \alpha a^1 + \beta b^1 \\ \alpha a^2 + \beta b^2 \\ \vdots \\ \alpha a^N + \beta b^N \end{pmatrix}. \quad (14.6)$$

We can take N to be as large as we want. In fact, if we go to $N \rightarrow \infty$, we can reproduce any function that admits a Taylor expansion. As physicists we will take this limit cavalierly. There are a few subtleties that show up that we will point out, but we will not let ourselves get bogged down in mathematical detail. Welcome to the world of infinite-dimensional vector spaces. We call these spaces **Hilbert spaces**.

A judicious choice of basis for function space can make our lives much easier. This is a central theme of this course and justifies the beastiary of special functions that you meet in graduate school. Let us ignore the subtleties of defining a function space for now—we get to this soon enough. The following example gives a taste of what it means to work with functions as vectors.

Exercise 14.1. *Here is a cute two-dimensional function space that gives us a shortcut to calculate a particular indefinite integral. Consider a two dimensional vector space spanned by the functions*

$$|f_1\rangle = f_1(x) = e^{ax} \cos bx \quad |f_2\rangle = f_2(x) = e^{ax} \sin bx, \quad (14.7)$$

where a and b are constants. Forget orthonormality or boundary conditions for this problem. The derivative d/dx is a linear operator that acts on this space. Write down the derivative as a 2×2 matrix in the above basis, D .

Invert D in the usual way that you learned to invert 2×2 matrices during your childhood^a. Call this matrix D^{-1} .

Now stop and think: the inverse of a derivative is an indefinite integral^b. Thus acting with D^{-1} on the vector $|f_1\rangle$ should be understood as an integral of $f_1(x)$. Show that, indeed,

$$D^{-1}|f_1\rangle = \int dx e^{ax} \cos bx. \quad (14.8)$$

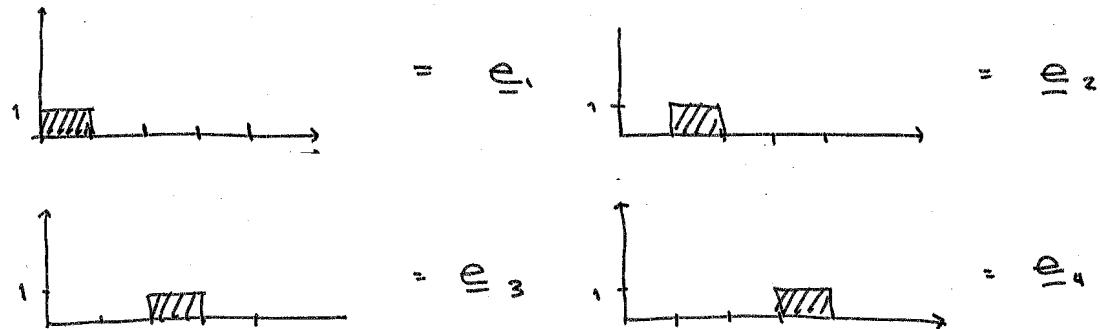
Feel free to use Mathematica to do the indefinite integral on the right-hand side. Pat yourself on the back if you can do it without a computer.

^aStuck? Here's a life pro tip: <http://bfy.tw/KG2Z>

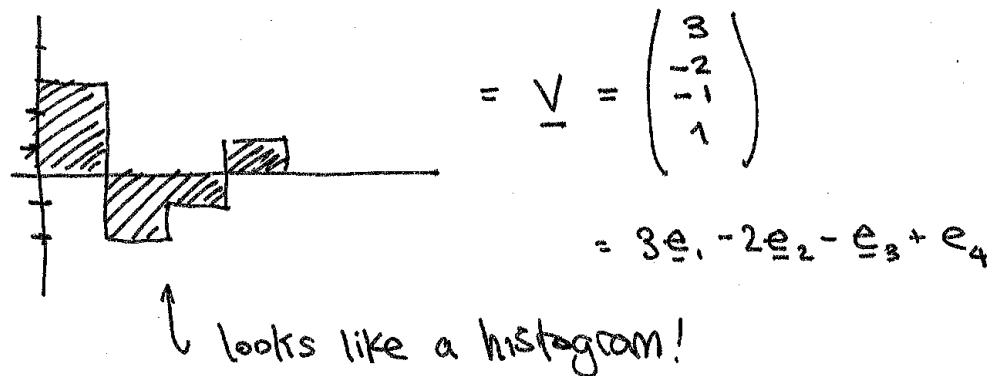
^bIgnore the constant term.

14.2 Histogram Space

Here is a funny vector space that we call *histogram space*. The basis vectors are:

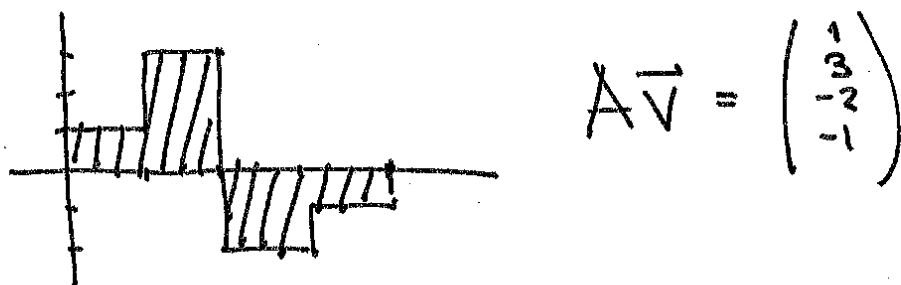


This is a basis for a histogram over unit bins from $x = 0$ to $x = 4$. A vector in this space is, for example:



looks like a histogram!

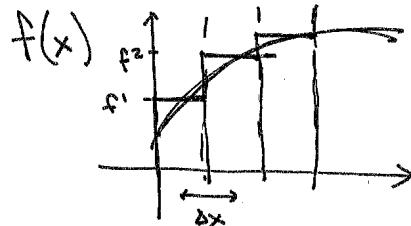
We can perform a linear transformation A on \vec{v} which outputs another vector. Let us say it is this:



Exercise 14.2. From the image above, can you derive what A is?

The answer to the above exercise is *no*. Please make sure you convince yourself why: there are many different transformations that convert to old histogram into the new histogram. The matrix A is 4×4 and thus has 16 entries that we need to define. The matrix equation $\mathbf{Av} = \mathbf{w}$ for known vectors \mathbf{v} and \mathbf{w} encodes only four equations.

The power of this admittedly strange vector space is that we can think of these histograms as approximations of continuous functions:



This discretized function space does not formally admit an $N \rightarrow \infty$ limit. The primary reason has to do with convergence, which I always thought was not the fun part of calculus. With the caveat that this space and this basis is really silly, we push forward to get some intuition for what an infinite-dimensional vector space looks like.

14.3 Derivatives as Linear Transformations

Our discretized function space allows us to define a forward derivative⁸⁶:

$$|f'\rangle = \frac{1}{\Delta x} \begin{pmatrix} f^2 - f^1 \\ f^3 - f^2 \\ \vdots \\ f^{i+1} - f^i \\ \vdots \end{pmatrix}. \quad (14.9)$$

This is familiar if you have ever had to manually program a derivative into a computer program. Note that the right-hand side is a linear transformation of \mathbf{f} . In other words, we may write a matrix D_+ so that

$$|f'\rangle = D_+ |f\rangle. \quad (14.10)$$

⁸⁶One could have also defined a backward derivative where $(f')^i \sim f^i - f^{i-1}$. Note that you *cannot* try to make this symmetric by defining a ‘centered’ derivative like $(f')^i \sim f^{i+1/2} - f^{i-1/2}$ because there’s no such thing as a fractional index. If you tried to write $(f')^i \sim f^{i+1} - f^{i-1}$ you’re making a worse approximation. If you’re like me, the fact that there’s some asymmetry in how we define the first derivative is deeply unsettling. There’s something to this intuition!

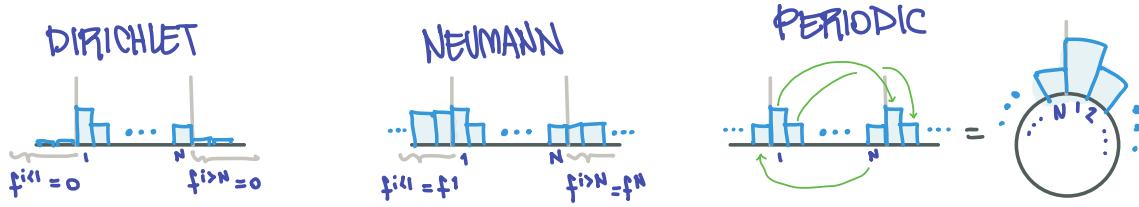


Figure 26: Some of our favorite boundary conditions. These are not only physically motivated, but will be important for defining Hermitian differential operators.

One problem is apparent: what happens at the ‘bottom’ of the vector? What is the last component of the derivative, \mathbf{f}'^N ? Formally, this is

$$(\mathbf{f}')^N = \frac{1}{\Delta x} (f^{N+1} - f^N) \quad (14.11)$$

but now we have no idea what f^{N+1} is. That was never a component in our vector space. There is no $\mathbf{e}_{(N+1)}$ basis vector.

We could also define a backward derivative, D_- ,

$$|\mathbf{f}'\rangle = D_- |\mathbf{f}\rangle = \frac{1}{\Delta x} \begin{pmatrix} \vdots \\ f^2 - f^1 \\ \vdots \\ f^i - f^{i-1} \\ \vdots \\ f^N - f^{N-1} \end{pmatrix}. \quad (14.12)$$

The question is now on the opposite boundary: what is the first component of $D_- |\mathbf{f}\rangle$?

This demonstrates an important lesson that we’ll need when we move more formally to function spaces:

Boundary conditions are part of the definition of the function space.

That was so important that I put the whole damn sentence in boldface and set it in the middle of the line. The significance of boundary conditions may be a bit surprising—but think of this as part of the definition of which functions we allow into our function space. There are three nice choices for boundary conditions, which we illustrate in Fig. 26.

Dirichlet The simplest option is a **Dirichlet boundary condition**. Here we assert that the function vanishes for all of the bins below $i = 1$ and above $i = N$. This choice corresponds to the following definition of what happens to all functions outside the domain of the function space:

$$f^{i>N} = f^{i<1} = 0. \quad (14.13)$$

This solves the problem of the derivative on the last component of D_+ and the first component of D_- :

$$(f')_{+, \text{Dir.}}^N = \frac{f^{N+1} - f^N}{\Delta x} = \frac{-f^N}{\Delta x} \quad (f')_{-, \text{Dir.}}^1 = \frac{f^1 - f^0}{\Delta x} = \frac{f^1}{\Delta x} . \quad (14.14)$$

Example 14.1. With Dirichlet boundary conditions, the matrix expression for the forward/backward derivative D_\pm is

$$(D_+)_{\text{Dir.}} = \frac{1}{\Delta x} \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \\ & & & & & -1 \end{pmatrix} \quad (14.15)$$

$$(D_-)_{\text{Dir.}} = \frac{1}{\Delta x} \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \quad (14.16)$$

Observe that neither of these matrices is Hermitian in the conventional finite dimensional sense.

Neumann Another choice for boundary conditions is that the function remains constant below $i = 1$ and above $i = N$. This amounts to enforcing that the derivative vanishes at the boundaries. We call these **Neumann boundary conditions**:

$$f^{i<1} \equiv f^i \quad f^{i>N} \equiv f^N . \quad (14.17)$$

This implies that

$$(f')_{+, \text{Neu.}}^N = \frac{f^{N+1} - f^N}{\Delta x} = 9 \quad (f')_{-, \text{Neu.}}^1 = \frac{f^1 - f^0}{\Delta x} = 0 . \quad (14.18)$$

Example 14.2. With Neumann boundary conditions, the matrix expression for the forward/backward derivative D_{\pm} is

$$(D_+)_\text{Neu.} = \frac{1}{\Delta x} \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \\ & & & & & 0 \end{pmatrix} \quad (14.19)$$

$$(D_-)_\text{Neu.} = \frac{1}{\Delta x} \begin{pmatrix} 0 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \quad (14.20)$$

Periodic Alternatively, we could impose **periodic boundary conditions**. This is the assumption that the boundary to the right of $i = N$ is equivalent to the boundary to the left of $i = 1$. Periodic boundary conditions amount to wrapping the x -axis into a circle. Older folks sometimes call this *Asteroids* boundary conditions:

$$f^i = f^{i+kN} \quad k \in \mathbb{Z}. \quad (14.21)$$

This gives

$$(f')_{+, \text{Per.}}^N = \frac{f^{N+1} - f^N}{\Delta x} = \frac{f^1 - f^N}{\Delta x} \quad (f')_{-, \text{Per.}}^1 = \frac{f^1 - f^0}{\Delta x} = \frac{f^1 - f^N}{\Delta x}. \quad (14.22)$$

Example 14.3. With Periodic boundary conditions, the matrix expression for the forward/backward derivative D_{\pm} is

$$(D_+)_\text{Per.} = \frac{1}{\Delta x} \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & -1 & 1 & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \\ 1 & & & & & -1 \end{pmatrix} \quad (14.23)$$

$$(D_-)_\text{Per.} = \frac{1}{\Delta x} \begin{pmatrix} 1 & & & & -1 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \quad (14.24)$$

Look at those funny non-zero elements in the far corners.

Periodic boundary conditions show up *all* the time in physics. Sometimes they show up in obvious places, like the Brillouin zone of a crystal lattice. Other times they show up in not-so-obvious places like the boundary conditions of the known universe.

Exercise 14.3. We don't know anything about the universe outside the Hubble radius. Why do you think it would be reasonable in a physical model to assume that it has periodic boundary conditions? Hint: what would happen to the x -momentum of an asteroid in the classic arcade game Asteroids if the game did not have periodic boundary conditions?

Mixed boundary conditions You may also impose a Dirichlet boundary condition on one boundary and a Neumann boundary condition on the other boundary.

Exercise 14.4. Write out the matrix expressions for the forward and backward derivatives D_{\pm} with mixed boundary conditions with a Dirichlet boundary at $i = 0$ and a Neumann boundary at $i = N + 1$.

Second derivative The second derivative may be defined symmetrically:

$$(f'')^i = \frac{(f^{i+1} - f^i) - (f^i - f^{i-1})}{\Delta x^2}. \quad (14.25)$$

You may pontificate about the reason why the first derivative does have a symmetric discretization while the second derivative does.

Example 14.4. We may write the second derivative in matrix form:

$$\begin{pmatrix} \ddots & & & \\ & 1 & -2 & 1 & & \\ & & 1 & -2 & 1 & & \\ & & & 1 & -2 & 1 & \\ & & & & & & \ddots \end{pmatrix}. \quad (14.26)$$

We omit writing zeros and remain agnostic about the boundary conditions which affect the first and last rows of the matrix. Observe that the second derivative matrix is Hermitian, whereas all of our first derivative matrices are not Hermitian.

Like the first derivative, one must impose boundary conditions on the second derivative operator.

Key Idea 14.1. The choice of boundary conditions is part of the definition of the function space. Every differential operator in the function space must obey the same boundary conditions.

Exercise 14.5. Write out the matrix form of the forward third derivative:

$$(D_+^3 f)^i = \frac{1}{\Delta x} [(D^2 f)^{i+1} - (D^2 f)^i]. \quad (14.27)$$

Observe that the third derivative matrix populates elements further away from the diagonal. This is a manifestation of the non-locality of higher derivatives. In a Taylor expansion, higher derivative terms probe the behavior of function at successively further points, even though each term is evaluated at a specific point.

Example 14.5. Do you remember the first time you saw how integration by parts worked? I still remember the first time someone showed me that $(fg)' = f'g + g'f$, so we thus have $f'g = (fg)' - g'f$. Integrate both sides over the argument and use the fundamental theorem of calculus to evaluate the definite integral over a total derivative. You can also prove this in ‘histogram space,’ i.e. for finite differences. Let us do this the ponderous

way to see how it works:

$$\sum_i \Delta x (D_+ f)^i g^i = \sum_i (f^{i+1} - f^i) g^i \quad (14.28)$$

$$= (f^2 g^1 - f^1 g^1) + (f^3 g^2 - f^2 g^2) + \cdots + (f^{N+1} g^N - f^N g^N) \quad (14.29)$$

$$= -f^1 g^1 + (f^2 g^1 - f^2 g^2) + (f^3 g^2 - f^3 g^3) + \cdots + f^{N+1} g^N \quad (14.30)$$

$$= f^{N+1} g^N - f^1 g^1 - \sum_i f^i (g^{i-1} - g^i) \quad (14.31)$$

$$= f^{N+1} g^N - f^1 g^1 - \Delta x \sum_i f^i (D_- g)^i , \quad (14.32)$$

All we have done in the intermediate steps is regrouped pairs of terms that have the same f^i coefficient. Observe that integrating by parts over the forward derivative on f leaves the backward derivative on g . Also observe that we have recovered the usual boundary term. Because the initial integral is a forward derivative, the forward boundary is not symmetric.

Exercise 14.6. If $x(t)$ is the position of a particle of mass m , then its kinetic term is $\dot{m}x(t)/2$. One way to derive the equation of motion for a theory is to integrate this term by parts, see Section 14.15. Let us see how this works. Rather than \dot{x}^2 , let us consider $(D_+ f)(D_- g)$. We write f and g to make it easier to keep track of terms, but for the specific case of the kinetic term we know $f = g = x(t)$.

The integral over time corresponds to a sum over the time slice (histogram) index i of f and g :

$$\int dt \dot{x}(t)^2 \rightarrow \sum_i \Delta x (D_+ f)^i (D_- g)^i . \quad (14.33)$$

Show that this sum takes the form

$$\sum_i \Delta x (D_+ f)^i (D_- g)^i = - \sum_i f^i (D^2 g)^i + f^{N+1} (g^N - g^{N-1}) - f^1 (g^1 - g^0) . \quad (14.34)$$

I was sloppy with the limits of the sum above. What values does i range over in that sum? Partial answer: Here's the pattern:

$$D_+ f D_- g = f^2 g^2 - f^2 g^0 - f^1 g^1 + f^1 g^0 \quad (14.35)$$

$$+ f^3 g^2 - f^3 g^1 - f^2 f^2 + f^2 g^1 \quad (14.36)$$

$$+ f^4 g^3 - f^4 g^2 - f^3 g^3 + f^3 g^2 \quad (14.37)$$

$$+ \cdots \quad (14.38)$$

$$+ f^{N+1} g^N - f^{N+1} g^{N-1} - f^N g^N + f^N g^{N-1} . \quad (14.39)$$

14.4 Derivatives in other function space bases

[**Flip:** This section can be skipped on a first reading.]

There are other ways to write a discrete basis of functions. Here is a natural one for functions that are up to second-order polynomials:

$$\mathbf{e}_{(0)} = 1 \quad \mathbf{e}_{(1)} = x \quad \mathbf{e}_{(2)} = x^2 . \quad (14.40)$$

Let's sidestep questions about orthonormality for the moment. Clearly linear combinations of these basis functions can produce any quadratic function:

$$f(x) = ax^2 + bx + c \Rightarrow \mathbf{f} = \begin{pmatrix} c \\ b \\ a \end{pmatrix} . \quad (14.41)$$

The derivative operator has an easy representation in this space:

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} . \quad (14.42)$$

We can see that

$$D\mathbf{f} = \begin{pmatrix} b \\ 2a \\ 0 \end{pmatrix} \quad D^2\mathbf{f} = \begin{pmatrix} 2a \\ 0 \\ 0 \end{pmatrix} \quad D^3\mathbf{f} = 0 . \quad (14.43)$$

The last line is, of course, the realization that the third-derivative of a quadratic function vanishes. Feel free to attach mathy words to this like *kernel*.⁸⁷ One hint why this basis may be annoying is that the second derivative matrix is not Hermitian, unlike in (14.26).

There are other bases that we may use for function space. A particularly nice one that we will use over and over is the Fourier basis, which we usually refer to as *momentum space*. The basis vectors are things like sines, cosines, or oscillating exponentials. These do not vanish for any power of D .

14.5 Locality

Notice that in the histogram basis, the derivative matrix D is sparse: it is zero everywhere away from the diagonal. The only non-zero elements on the i^{th} row are around the $(i \pm 1)^{\text{th}}$ column. Higher powers of D sample further away, but the non-zero elements are always clustered near the diagonal.

This is simply a notion of **locality**. Remember the Taylor expansion:

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \dots . \quad (14.44)$$

⁸⁷Unfortunately ‘kernel’ is also used when discussing probability distributions. To the best of my knowledge the two usages are completely independent.

If we think about the histogram as a discretization of a continuous function, then it is clear what the higher derivatives are doing. Given a function $f(x) = \mathbf{f}$, one might like to know about the function around some point x_0 corresponding to some index i . That is: $f^i = f(x_0)$. If you'd like to learn more about the function around that point, one can express the derivative at x_0 . Thus $D\mathbf{f}$ says something about the slope, $D^2\mathbf{f}$ says something about the curvature, and so on. Because each successive power of D samples terms further away from f^i , you can tell that these higher order terms are learning about the function further and further away from x_0 .

Now think about the types of differential equations that you have encountered in physics. They often include one or two derivatives. You hardly ever see three, four, or more derivatives⁸⁸. There is a reason for this: at the scales that we can access experimentally, nature appears to be local. Our mathematical models of nature typically have locality built in⁸⁹. Physics at one spacetime point should not depend on spacetime points that are far away.

We mentioned this in Section 10.7 when we thought about special relativity. Recall the idea of causality—the idea that A causes B therefore A must have happened *before* B . One of the key results in special relativity is that causality can be tricky if two events do not occur at the same spacetime point. More carefully, A can only cause B if there is a timelike separation of the appropriate sign. If we want to build causal theories of nature, then the dynamics at x_0 should not rely on what is happening at x_1 , a finite distance away.⁹⁰

The assumption that nature is local leads to physical *dynamics* that are described in terms of derivatives. Newton's law is

$$m\ddot{\mathbf{x}}(t) = -\nabla V . \quad (14.45)$$

On the left hand side is a second derivative in time. The right hand side is simply a function—nevermind that it is a spatial gradient of a potential. Most of our equations of motion are written in a differential form:

$$\mathcal{O}f(x) = s(x) , \quad (14.46)$$

where \mathcal{O} is a **differential operator** that is a matrix in function space built out of derivatives, $f(x)$ encodes something about the state of the system, and $s(x)$ is a source for the dynamics. For Newton's force law, $\mathcal{O} = (d/dt)^2$ and $s = -\nabla V/m$. The variable is often time since we care about the time evolution of a system and our dynamical laws tell us the ‘small change’ in a state after a ‘small time evolution.’ More generally, it can be a spacetime position coordinate.

14.6 Differential Operators

Linear transformations on function space are differential operators. In principle you can imagine linear transformations that are not differential operators, for example a finite trans-

⁸⁸With some thought, it may also be clear why spatial derivatives typically appear squared.

⁸⁹A recent counterexample “A Jewel at the Heart of Quantum Physics,” by Natalie Wolchover in *Quanta Magazine* (2013).

⁹⁰This is different from saying that information cannot propagate from x_0 to x_1 ; such propagation could come from some causal excitation of the electromagnetic field traveling every infinitesimal distance between the two positions. This is reminiscent of the classical Zeno's paradox.

lation. However, because our models of nature are typically *local* and *causal*, the linear transformations that we obtain from physical models are differential operators⁹¹.

Let's write a general differential operator as:

$$\mathcal{O} = p_0(x) + p_1(x) \frac{d}{dx} + p_2(x) \left(\frac{d}{dx} \right)^2 + \dots \quad (14.47)$$

where the $p_i(x)$ are polynomials. Sometimes we will write this as \mathcal{O}_x to make it clear that the argument of the polynomials is x and the variable with which we are differentiating is x .

Exercise 14.7. Explain why (14.47) is a linear operator acting on function spaces.

Exercise 14.8. A confused colleague argues to you that (14.47) cannot possibly be ‘linear.’ Just look at it, your colleague says: the functions $p_i(x)$ are polynomials—those aren’t linear! There are also powers of derivatives—how is that possibly linear? Explain to your colleague why the $p_i(x)$ does not have to be linear nor is one restricted to finite powers of derivatives for the operator \mathcal{O} to be a linear operator acting on function space.

Technically (14.47) is called a **formal operator** because we haven’t specified the boundary conditions of the function space. Recall in our discretized ‘histogram space’ in Section 14.2 that we had to be careful about how to define the derivative acting on the boundaries of the space. A differential operator along with boundary conditions is called a **concrete operator**.

14.7 Inner Product

Let us return to finite-dimensional histogram space. As long as we are finite dimensional, this is a perfectly legitimate vector space. We would like to endow this vector space with a metric. Perhaps the Euclidean metric? As we will see shortly, it is useful to define a metric that is proportional to the Euclidean metric by Δx , the discretization size of our histogram space:

$$g_{ij} = \Delta x \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \quad (14.48)$$

This gives a definition for components with a lower index:

$$\langle f | = f_i \langle i | \quad f_i = \Delta x f^i. \quad (14.49)$$

⁹¹This is not to say that finite transformations are somehow not permitted. The dynamics that govern our models of nature, however, only dictate how information is transmitted infinitesimally in space and time. Propagation forward in time by some finite interval is described by the exponentiation of infinitesimal forward time translations. This is, of course, why the time-translation operator in quantum mechanics is $e^{i\hat{H}t}$, where the Hamiltonian H is described as a local function with perhaps one or two derivative operators.

The factor of Δx makes it seem like the ‘row vector’ or bra $\langle f |$ is something that is born to be integrated over. *As long as N is finite*, the dual basis $\langle i |$ is defined as usual, $\langle i | j \rangle = \delta_j^i$. The inner product of two histograms is

$$\langle f, h \rangle = f^i h^j g_{ij} = f^i h_i \Delta_x = \sum_i f^i h^i \Delta x . \quad (14.50)$$

If we think of $f_i = f(x_i)$ and $h^i = h(x_i)$ then the above sum is simply the discretization of the integral

$$\langle f, h \rangle = \int dx f(x) h(x) . \quad (14.51)$$

For complex functions, we generalize this a bit to include a complex conjugate on the first argument:

$$\langle f, h \rangle = \int dx f^*(x) h(x) . \quad (14.52)$$

This is called the L2 inner product. Observe that this definition satisfies the metric condition for complex vector spaces (9.83), $\langle f, h \rangle = \langle h, f \rangle^*$.

Example 14.6. *Wave functions in 1D quantum mechanics obey this norm. For an infinite domain, we typically restrict to square-integrable functions meaning that $|f|^2$ goes to zero fast enough at $\pm\infty$ so that the integral $\langle f, f \rangle$ is finite.*

Sometimes the inner product is defined with respect to a **weight** function $w(x)$:

$$\langle f, g \rangle_w = \int dx w(x) f^*(x) g(x) . \quad (14.53)$$

There’s nothing mysterious about inner products with weights. They typically boil down to the fact that one is not using Cartesian coordinates. We will assume unit weight until we go to higher spatial dimensions⁹².

Example 14.7. *Have you met the Bessel functions? If not, you’re in for a treat in your electrodynamics course. The Bessel functions satisfy a funny orthogonality relation with weight $w(x) \sim x$ because they show up as the radial part of a solution when using polar coordinates. When you separate variables, $d^2x = r dr d\theta$, we see that the measure over the radial coordinate r carries a weight r .*

⁹²My dissertation focused on theories of extra dimensions. I also noticed that my weight increased in my final year of graduate school as I spent most of my time writing about extra dimensions and eating cafe pastries.

14.8 Adjoint

This looks very promising. We have an inner product that can pave the way to taking the infinite dimensional limit of our histogram space. We can now make more precise statements about derivatives. We use the term **differential operator** (or simply operator) to mean some linear transformation built out of derivatives that act as matrices in function space. There are other ‘matrices’ that are not differential operators, but these tend to be *nonlocal* and hence do not often show up in physical theories. We can now formalize the sense in which some differential operators are self-adjoint (Hermitian) and others are not. Recall that a self-adjoint operator satisfies (9.71)

$$\mathcal{O}^\dagger = \mathcal{O} \quad \text{where} \quad \langle \mathcal{O}^\dagger f, h \rangle \equiv \langle f, \mathcal{O} h \rangle . \quad (14.54)$$

Recall that we have to assume some boundary conditions for the space. Writing this out with respect to our integral expression for the inner product:

$$\int dx [\mathcal{O}^\dagger f(x)]^* h(x) = \int dx f(x)^* [\mathcal{O} h(x)] . \quad (14.55)$$

The strategy is as follows: given an inner product (integral) over f^* and g where there is some stuff (\mathcal{O}) acting on g , can we re-write this as an integral with no stuff acting on g and some *other* stuff acting on f^* ? If so, then the ‘other stuff’ is the adjoint \mathcal{O}^\dagger .

Example 14.8. What is the adjoint of the derivative operator, $\mathcal{O} = d/dx$? Assume an interval $x \in [a, b]$ and Dirichlet boundary conditions, $f(a) = f(b) = 0$. There’s a simple way to do this: integrate by parts.

$$\int dx f(x) \left[\frac{d}{dx} g(x) \right] = - \int dx \left[\frac{d}{dx} f(x) \right]^* g(x) + [f^*(x)g(x)]_a^b = \int dx \left[-\frac{d}{dx} f(x) \right]^* g(x) . \quad (14.56)$$

From this we deduce that

$$\left(\frac{d}{dx} \right)^\dagger = -\frac{d}{dx} . \quad (14.57)$$

We will be especially interested in **self-adjoint** (Hermitian) operators for which

$$\mathcal{O}^\dagger = \mathcal{O} . \quad (14.58)$$

This is, as we mentioned for the finite-dimensional case, because self-adjoint operators are *nice*: they have real eigenvalues and orthogonal eigenvectors. Since most physical values are real eigenvalues of some operator, one may expect that the differential operators that show up in physics are typically self-adjoint.

Exercise 14.9. We saw above that the derivative operator is not self-adjoint. Show that the complex differential operator $\mathcal{O} = i(d/dx)$ is self-adjoint. Hint: what is the momentum operator in quantum mechanics?^a

^a<https://aapt.scitation.org/doi/abs/10.1119/1.9932>

Example 14.9. Consider $\mathcal{O} = -(d/dx)^2$ defined on the domain $x \in [0, 1]$ with the boundary conditions $f(0) = f(1) = 0$. Is this operator self-adjoint? We want to check of $\langle f, \mathcal{O}g \rangle = \langle \mathcal{O}f, g \rangle$. We have one trick: integration by parts. Let's see how this works.

$$\langle f, \mathcal{O}g \rangle = - \int dx f^*(x) \frac{d^2}{dx^2} g(x) . \quad (14.59)$$

This is compared to

$$\langle \mathcal{O}f, g \rangle = - \int_0^1 dx \left[\frac{d^2 f}{dx^2} \right]^* g(x) \quad (14.60)$$

$$= - \left(\frac{df}{dx} \right)^* g(x) \Big|_0^1 + \int_0^1 dx \left(\frac{df}{dx} \right)^* \frac{dg}{dx} \quad (14.61)$$

$$= f^*(x) \frac{dg}{dx} \Big|_0^1 - \int_0^1 dx f^*(x) \frac{d^2 g}{dx^2} \quad (14.62)$$

$$= - \int_0^1 dx f^*(x) \frac{d^2 g}{dx^2} . \quad (14.63)$$

And so we see that indeed $(-d^2/dx^2)^\dagger = -d^2/dx^2$.

Exercise 14.10. In the previous example, what is the significance of the overall sign of the operator? Hint: the sign doesn't matter, it's because we typically think of $-\partial^2$ and its higher-dimensional derivatives as the square of the momentum operator.

Example 14.10. The *eigenfunctions* f_n of $-d^2/dx^2$ defined on $x \in [0, 1]$ with Dirichlet boundary conditions are simply

$$f_n(x) = A_n \sin(n\pi x) \quad \lambda_n = n^2\pi^2 , \quad (14.64)$$

where λ_n is the associated eigenvalue and A_n is some normalization that. These eigenfunctions are orthonormal in the following sense:

$$\langle f_n, f_m \rangle = \int_0^1 dx \sin(n\pi x) \sin(m\pi x) = \frac{A_n A_m}{2} \delta_{nm}, \quad (14.65)$$

from which we deduce that the normalization is $A_n = \sqrt{2}$. That's basically all there is to know about Fourier series.

Exercise 14.11. What would change if we had instead assumed Neumann boundary conditions? What if we had assumed periodic boundary conditions? What about anti-periodic boundary conditions?

Exercise 14.12. A function $g(x)$ defined on an interval $x \in [0, 1]$ with Dirichlet boundary conditions can be written with respect to the Fourier basis (14.64). In ket notation, the n^{th} component of g with respect to this basis is

$$g^n = \langle f_n | g \rangle. \quad (14.66)$$

Confirm that this is precisely what you know from Fourier series. In other words, we can decompose $g(x)$ as

$$g(x) = \sum_n \langle f_n | g \rangle f_n(x). \quad (14.67)$$

With a clean definition of self-adjointness, we can now define ‘nice’ differential operators are those operators that are self-adjoint. This is the function space equivalent of symmetric matrices in real, finite dimensional vector spaces and Hermitian matrices in complex, finite dimensional vector spaces. In the histogram basis we have some semblance of Hermiticity, but we saw that the $N \rightarrow \infty$ limit is a little tricky. $i(d/dx)$ is a self-adjoint operator in function space, but its discretization, the forward/backward derivative D_{\pm} , is not Hermitian because the two non-zero elements on each row are cannot be evenly spaced about the diagonal.

The following observations carry over from finite-dimensional vector space:

1. The eigenvectors of a self-adjoint differential operator are orthogonal and may be chosen to be orthonormal.
2. The eigenvalues of a self-adjoint differential operator are real.

The proof of these assertions are the same since they depend on general properties of the inner product. There is an additional feature: the eigenvalues are ordered. Function space is infinite

in two senses: first the domain may be infinite, $x \in [-\infty, \infty]$. Second, the number of points in any finite interval is also infinite: $\Delta x \rightarrow 0$. Rather than having to discover an infinite number of eigen-stuff, we will often have *one* eigenvalue equation with multiple enumerable solutions and whose eigenvalues are enumerated with respect to these solutions.

Example 14.11. Consider the solution to the function space eigenvalue problem in Example (14.10). The eigenvalues are $\lambda_n = n^2\pi^2$ with corresponding eigenfunctions proportional to $\sin(n\pi x)$. The integer n enumerates the solution and plays role of the index of the eigenbasis.

Finally, we remind ourselves that we have restricted our study to ‘matrices’ in function space that are differential operators. It is entirely possible to conceive of a matrix that is not a differential operator. The discussion of locality in Section 14.5 motivates that *physical* theories are local and hence the dynamical equations of motion are written with respect to differential operators. Similarly, matrices are usually uniform across all of space: this is because of the physical principle of translation invariance. The laws of physics in one location are the same laws as they are at another location in both space and time. This does not mean that the *source* of a physical effect is uniform, only that the *dynamics* are uniform.

At this point one should review Section 12.7. The general form of an eigenfunction problem in physics is that we are given some *equation of motion*

$$\mathcal{O}f(x) = s(x) \quad (14.68)$$

that describes what happens to a *state* $f(x)$ —perhaps the position of a particle—under some *dynamics* described by the differential operator \mathcal{O} whose effects are sourced by $s(x)$. If \mathcal{O} is self-adjoint, then we can diagonalize it with respect to an eigenbasis and solve this problem trivially. Most of the named special functions that show up in physics are only name-worthy because they are eigenfunctions of differential operators that show up in physics.

14.9 Cracks in the scaffolding as we go to the continuum

It may be clear that something goes wrong as we formally pass from the histogram basis to a continuum basis. That is, when we take the $N \rightarrow \infty$ limit.⁹³

The notion of building up continuous functions $f(x)$ from the histogram basis above turns out to be problematic from the point of view of mathematical rigor. You can read more about the problems in Appel, *Mathematics for Physicists* Chapter 9.1 (“Insufficiency of Vector Spaces”), or Hassani, *Mathematical Physics*, Chapter 7 (“Hilbert Spaces”). As long as histogram space is finite dimensional, you can treat it as \mathbb{R}^N and even endow it with an inner product—say the Euclidean inner product. If you want to be fancy, at this level the metric space in the $N \rightarrow \infty$ limit is called a *pre-Hilbert* space.

One problem that shows up is that the norm of a vector is now an infinitely long series. If we had not inserted the Δx in our definition of the metric (14.48), then the norm of a

⁹³There are two $N \rightarrow \infty$ limits: the first one is the infinite number of points in any finite interval. The second is the infinite size of a space whose boundaries go to $\pm\infty$.

function in histogram space is

$$|f|^2 - \langle f, f \rangle = \sum_{i=1}^{\infty} (f^i)^2 . \quad (14.69)$$

How do you normalize a vector that has $f^i = 1$ for all values of i ? Enforcing $|f|^2 = 1$ seems to imply that each $f^i \rightarrow 0$. This is very fishy. Fortunately, inserting the Δx in the metric seems to help:

$$|f|^2 = \sum_{i=1}^{\infty} (f^i)^2 \Delta x \rightarrow \int dx |f(x)|^2 . \quad (14.70)$$

If the interval on which our function is defined has length $L = x_{\max} - x_{\min}$, then we may enforce that $\Delta x = L/N$ so that the discrete sum makes sense as $N \rightarrow \infty$. On the right-hand side we take the continuum limit. Functions for which the integral over $|f(x)|^2$ are well-defined⁹⁴ are known as *square integrable*, or L2.

Let us now try to make sense of the basis bras and kets. We know from Section 9.6 that the inner product gives us a natural way to define bras from kets. In the finite dimensional limit, we had (14.49) $\langle f |$ with components $f_i \equiv \Delta x f^i$. The factor of $\Delta x \rightarrow dx$ makes it clear that the bras of function space are *distributions*: functions that are born to be integrated over.⁹⁵ The inner product tells us that the bra $\langle f |$ is

$$\langle f | = \langle f, _ \rangle = \int dx f(x)^* _ . \quad (14.71)$$

Just feed it a function (ket) and perform the integral. We have left the limits of integration implicit, but we should remember that these are integrals over the domain of the function. The definition of this domain and the boundary conditions are part of the definition of the function space.

Example 14.12. Let $\langle p | = \int_0^1 dx p(x) _$ be the probability function for some variable $x \in [0, 1]$. The expectation value for some function $|g\rangle = g(x)$ is

$$\langle g \rangle_p \equiv \langle p, g \rangle = \int_0^1 dx p(x) g(x) . \quad (14.72)$$

The histogram basis $|x\rangle$ satisfies

$$\langle x | y \rangle = \langle x, y \rangle = 0 \quad \text{if } x \neq y . \quad (14.73)$$

⁹⁴This means that they are non-zero and do not diverge.

⁹⁵In probability theory we write the probability distribution of a continuous variable to be $p(x) dx$. The interpretation is that the probability to find $x \in (a, b)$ is the integral $\int_a^b dx p(x)$. The value of $p(x)$ by itself does not mean anything: it is the probability that x is *exactly* some value, say $x = 1.00000$ and not $x = 1.00001$.

However, in the *infinite dimensional* (continuum function) limit, we can no longer say that $\langle x|y \rangle = 1$ if $x = y$. This is a consequence of the unusual normalization we noticed in (14.70). Let us assume that the domain of the function space is $x \in [0, N\Delta x]$. In the histogram basis where space is discretized, we may to define the basis functions as follows:

$$|i\rangle = e_i(x) \equiv \mathcal{N} \begin{cases} 0 & \text{if } x < (i-1)\Delta x \\ 1 & \text{if } (i-1)\Delta x \leq x \leq i\Delta x \\ 0 & \text{if } i\Delta x < x \end{cases} . \quad (14.74)$$

What is the value of the normalization, \mathcal{N} ? We would like $\langle i, f \rangle = f^i = f(x_i = i\Delta x)$. Given our normalization of the inner product, this means

$$\langle i, f \rangle = \int dx e_i(x) f(x) = \mathcal{N} \Delta x f(x_i) . \quad (14.75)$$

We have been a little sneaky: we have defined our histogram space as a stepwise functions on a continuous space so that we could use the integral representation of the inner product. No matter. The result is that $\mathcal{N} = 1/\Delta x$. However, we want to take the $\Delta x \rightarrow 0$ limit, which means that the normalization \mathcal{N} of the basis ket $|i\rangle$ blows up. Oh dear.

We can gain more insight on this by using the **completeness** relation. Earlier we called this *multiplication by one*: $\mathbb{1} = \sum_i |i\rangle \langle i|$. For a continuous function space this sum becomes an integral:

$$\int dx |x\rangle \langle x| = \mathbb{1} . \quad (14.76)$$

where we have again left the integration limits implicit. If we act on a function with this representation of the identity, we have:

$$|f\rangle = \int dx |x\rangle \langle x| f\rangle . \quad (14.77)$$

The significance becomes more clear if we apply $\langle y|$ to both sides:

$$f(y) = \langle y|f\rangle = \int dx \langle y|x\rangle \langle x|f\rangle = \int dx \langle y|x\rangle f(x) . \quad (14.78)$$

This tells us that whatever $\langle y|x\rangle$ is, it must be a distribution such that when it is integrated over f it picks out the the value of f at position y . Let us *define* this object $\delta(x - y) \equiv \langle y|x\rangle$:

$$\langle y|x\rangle \equiv \delta(x - y) \quad \int dx \delta(x - y) f(x) \equiv f(y) . \quad (14.79)$$

This object is called the Dirac δ -function and is the function space analog of the Kronecker δ . We have assumed that δ -function only depends on the difference $|x - y|$. In principle it could have depended on x and y separately, but we know that it is only non-zero when $x = y$. Furthermore, the δ -function is normalized:

$$\int dx \delta(x) = 1 , \quad (14.80)$$

assuming that the integration region includes $x = 0$. Note that $\delta(0)$ is divergent. It seems to be infinite, reflecting the $\Delta x \rightarrow 0$ limit that worried us below (14.75). This means that the δ -function is not really an admissible “function.” It is not a basis vector. $\delta(x - y)$ is simply a way to make sense of the object $\langle y|x \rangle$. Because this object always shows up under an integral, we never have to worry about what it means at an individual point. In other words: if you ever have a δ -function without a corresponding integral, you have almost certainly done something wrong.

Exercise 14.13. *How do you make sense of the derivative of a δ -function? Hint: integrate by parts.*

Key Idea 14.2. *The issue of a normalized basis in an infinite dimensional vector space is the difference between the Kronecker δ and the Dirac δ -function. The Kronecker δ is always finite: it is zero most places, and one in a small window where two discrete indices match. The Dirac δ -function is zero almost everywhere, except for a single point (vanishing width) where it is formally infinite.*

14.10 Why histogram space is dumb

[**Flip:** This section can be skipped on a first reading.]

There is a more formal sense in which the $N \rightarrow \infty$ limit is challenging. This is the notion of convergence. Imagine a *sequence* of vectors. This is just a list of vectors. Let us label these vectors as $|v\rangle_{(i)}$, where i indexes vectors in the sequence. For example,

$$|v\rangle_{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad |v\rangle_{(2)} = \begin{pmatrix} 1 \\ 2^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} |v\rangle_{(3)} = \begin{pmatrix} 1 \\ 2^{-1} \\ 2^{-2} \\ \vdots \\ 0 \end{pmatrix}. \quad (14.81)$$

We have chosen this sequence such it has a limiting form, $|v\rangle_{(\infty)}$

$$|v\rangle_{(\infty)} = \left. \begin{pmatrix} 2^{-0} \\ 2^{-1} \\ 2^{-2} \\ \vdots \\ 2^{-N} \end{pmatrix} \right|_{N \rightarrow \infty}. \quad (14.82)$$

We say that the sequence $\{|v\rangle_{(i)}\}$ is a **Cauchy sequence** that limits to $|v\rangle_{(\infty)}$ if

$$\lim_{i \rightarrow \infty} |v\rangle_{(i)} \rightarrow |v\rangle_{(\infty)}, \quad (14.83)$$

by which we mean that the norm $\|v\rangle_{(i)} - \|v\rangle_{(\infty)}\| \rightarrow 0$. The example sequence in (14.81) has finite support⁹⁶, and so each element is normalizable. If we want to hold on to this notion of normalizability, then we may want to take our infinite-dimensional vector space and restrict to only those vectors with finite support. However, the vector $\|v\rangle_{(\infty)}$ clearly does not have finite support. We say that the space of finite support vectors is not **complete** because there are Cauchy sequences that do not contain their limits.

Exercise 14.14. (*From Appel, Section 9.1*) Show that the sequence (14.81) is a Cauchy sequence that converges to (14.82). As an intermediate step, show that

$$\|v\rangle_{(i)} - \|v\rangle_{(\infty)}\| = \sum_{j=i}^{\infty} \frac{1}{4^j} = \left(\frac{4}{3}\right) \frac{1}{4^i} \rightarrow 0 . \quad (14.84)$$

In the penultimate step you can use the expression for a geometric series.

The observation here is that infinite-dimensional metric spaces are not always complete. When they are complete, the metric space is called a **Hilbert space**. The great thing about Hilbert spaces is that we can treat them as $N \rightarrow \infty$ limits of finite dimensional vector spaces.

14.11 Legendre Polynomials: salvaging the polynomial basis

Consider functions on the domain $x \in [-1, 1]$ with the standard function space inner product, (14.52). We now have a few reasons to think that the basis of monomials (14.4) is silly. First, they do not obey an obviously meaningful boundary condition: the functions are 1 at $x = 1$ and ± 1 at $x = -1$ depending on whether the index of the basis function is even or odd. It is also clear that linear combinations of these basis functions do not satisfy the same boundary conditions. Secondly, this basis behaves poorly under derivatives. Derivatives act as projection operators onto successively smaller subspaces. Finally, the monomials are not orthonormal in any meaningful sense. For example, in the basis

$$\langle n, m \rangle = \int_{-1}^1 dx x^{n+m} = \frac{x^{n+m-1}}{n+m} \Big|_{x=-1}^{x=1} = \begin{cases} 0 & \text{if } n+m-1 \text{ is odd} \\ \frac{2}{n+m} & \text{if } n+m-1 \text{ is even} \end{cases} . \quad (14.85)$$

The Legendre polynomials are a basis of polynomial functions on the domain $x \in [-1, 1]$ with the standard function space inner product, (14.52) that are at the very least orthogonal.⁹⁷ The first few are listed here:

$$P_0(x) = 1 \quad P_1(x) = x \quad P_2(x) = \frac{1}{2}(3x^2 - 1) \quad P_3(x) = \frac{1}{2}(5x^3 - 3x) . \quad (14.86)$$

⁹⁶Here the support of a vector means the non-zero elements.

⁹⁷They have the feature that $(1-x^2)f(x)$ obeys Dirichlet boundary conditions.

Exercise 14.15. Let $|n\rangle$ be the n^{th} Legendre polynomial, $P_n(x)$. Confirm that $\langle 2, 1 \rangle = 0$. Show that $\langle 2, 2 \rangle$ is not zero, however it is also not one. Derive what the appropriate normalization needs to be imposed on the first four Legendre polynomials to make them orthonormal.

Exercise 14.16. Use the Gram–Schmidt procedure to derive $P_5(x)$, up to normalization.

14.12 Completeness in Function Space

We rarely have much to say about the unit matrix in linear algebra. However, much like when we discussed units, we can squeeze a lot out of inserting the identity in our mathematical machinations. In order to help with translate this to function space, let's review how it works in finite dimensional vector spaces. The unit matrix is $\mathbb{1}$ and may be written:

$$\mathbb{1} = \sum_i |i\rangle\langle i| , \quad (14.87)$$

where $|i\rangle$ and $\langle j|$ are basis (dual-)vectors.

Exercise 14.17. Take a moment and convince yourself that (14.87) is true and obvious. It may be helpful to explicitly write out $|i\rangle\langle j|$ as a matrix.

Exercise 14.18. Suppose you have a two-dimensional Euclidean vector space. Show that (14.87) is true for the basis

$$|1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad |2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (14.88)$$

$$\langle 1| = \frac{1}{\sqrt{2}} (1 \ 1) \quad \langle 2| = \frac{1}{\sqrt{2}} (1 \ -1) . \quad (14.89)$$

In fact, (14.87) defines what it means that a set of basis vectors is **complete**. You can write any vector $|v\rangle$ with respect to the basis $|i\rangle$ —the components are simply

$$v^i = \langle i|v\rangle \quad (14.90)$$

so that

$$|v\rangle = \sum_i |i\rangle\langle i|v\rangle , \quad (14.91)$$

which we recognize as nothing more than ‘multiplying by the identity.’

What does completeness look like in function space? Let $e_{(n)}(x)$ be a set of basis functions. The basis is **complete** if

$$\sum_n [e_{(n)}(x)]^* e_{(n)}(y) = \delta(x - y) . \quad (14.92)$$

Compare this *very carefully* with the completeness relation (14.87). The sum over i in the finite-dimensional case has been relabeled into a sum over n in the function space—this is just my preference⁹⁸. The $\mathbb{1}$ has been replaced by a Dirac δ -function, $\delta(x - y)$. Let’s confirm that this makes sense. The *multiply by one* completeness relation (14.91) in function space is

$$|g\rangle = \sum_n |e_{(n)}\rangle \langle e_{(n)}|g\rangle \quad \langle e_{(n)}|g\rangle = \int dy [e_{(n)}(y)]^* g(y) . \quad (14.93)$$

We have deliberately changed the name of the integration variable to y to avoid confusion; since this variable is integrated over it’s simply a *dummy variable* and it doesn’t matter what we name it—the quantity $\langle e_{(n)}|g\rangle$ is independent of y because y is integrated over⁹⁹. Writing this out explicitly as functions:

$$g(x) = \sum_n \left[\int dy e_{(n)}^*(y) g(y) \right] e_{(n)}(x) . \quad (14.94)$$

The factor in the square brackets is simply $\langle e_{(n)}|g\rangle$, which is just a *number*—it has no functional dependence on x . If this seems unusual, please refer back to Example 14.10 and Exercise 14.12.

By the way, you’ll often hear people (perhaps even me) say that the Dirac δ function is not strictly an *function* but rather a **distribution**—this means that it only makes sense when it is integrated over. As physicists we’ll sometimes be sloppy and talk about physical quantities that could be Dirac δ -functions. There is *never* an appropriate, measurable physical quantity that is described by a $\delta(x)$. Anything with a $\delta(x)$ is an object that was meant to be integrated over. When you imagine that a point charge density is a δ -function, this is only because you will eventually integrate over it to determine the total charge. If you ever calculate a *measurable* quantity to be $\delta(x)$ check your work. If you ever find $\delta(x)^2$, then go home, it’s past your bed time.

Example 14.13. One can vaguely motivate the δ -function as the unit matrix by appealing to the ‘histogram basis’ of discretized function space. In an ordinary finite-dimensional vector space, unit matrix can be written as

$$\mathbb{1} = |1\rangle\langle 1| + |2\rangle\langle 2| + \dots = \sum_{i,j} \delta_i^j |i\rangle\langle j| . \quad (14.95)$$

⁹⁸I think this is because we will deal with complex functions and I want to avoid using i as an index. But if we’re being honest, it’s just become a habit.

⁹⁹By the way, this should ring a bell from our summation convention. When an upper and lower tensor index are contracted, the resulting object behaves as if it didn’t have those indices: $A^i_j v^j$ behaves as a vector with one upper index.

The Dirac δ -function in histogram space is analogous to

$$\delta(x - x') \rightarrow \delta_x^{x'} |x\rangle\langle x'| , \quad (14.96)$$

where x and x' are discrete bins on the right-hand side. Thus for a discretized function $f = f(x_1)|x_1\rangle + f(x_2)|x_2\rangle + \dots$, one has

$$\int dy \delta(x - y) f(y) = f(x) \longrightarrow \sum_j \delta_{x_j}^{x_i} |x_i\rangle\langle x_j|f\rangle = f(x_i)|x_i\rangle . \quad (14.97)$$

14.13 Orthonormality in Function Space

One should contrast the notion of completeness of a basis this with that of **orthonormality** of the basis. Orthonormality is the statement that

$$\langle i|j\rangle = \delta_i^j . \quad (14.98)$$

Completeness has to do with the ‘outer product’ $|i\rangle\langle i|$ while orthonormality has to do with the ‘inner product’ $\langle i|i\rangle = \langle i, i\rangle$. The function space generalization of orthonormality is¹⁰⁰

$$\langle e_{(n)}|e_{(m)}\rangle = \int dx e_{(n)}^*(x)e_{(m)}(x) = \delta_{nm} . \quad (14.99)$$

Exercise 14.19. Why does (14.99) have a Kronecker δ with discrete indices when (14.92) has a Dirac δ ? Please make sure you can answer this; it establishes the conceptual foundation of the analogy between finite- and infinite-dimensional vector spaces.

For the completeness relation, we sum over the same eigenfunction label n for a function and its conjugate evaluated at different continuous positions. For the orthonormality relation, we integrate over the positions of two different eigenfunction indices, n and m .

Do not confuse the eigenfunction label with the index of a vector. If this is confusing, please refer back to Exercise 14.18. You may be stuck thinking about basis vectors in the Cartesian basis—this is the analog of thinking about basis functions in the ‘histogram basis’ of Section 14.2. What we want to do is generalize to more convenient bases, like the eigenfunctions of differential operators (e.g. the Fourier basis for $-d^2/dx^2$).

¹⁰⁰If you’re a purist, you’ll note that δ_{nm} should really be written as δ_m^n because the dual basis vector has an upper index. While this may be true, I’m making the present notational choice because the object that we would call $\tilde{e}^{(n)}$ really does contain $e_{(n)}^*(x)$, the complex conjugate of $e_{(n)}(x)$.

Example 14.14. In the case of a finite interval, say $x \in [0, 1]$, the space of functions on this interval is continuous. In fact, we wrote a nice eigenbasis of $-d^2/dx^2$ on this space assuming Dirichlet boundary conditions. The basis consists of a discrete but infinite number of eigenfunctions. The discrete index, n , corresponded to the wave number (or momentum). The continuous ‘index,’ x , corresponded to a position-space location. This index is continuous because there is a continuum of positions x in the finite interval $[0, 1]$. If we extended the interval to the infinite real line, $\mathbb{R} = [-\infty, \infty]$, then the discrete spectrum of eigenfunctions—that is, the discrete separation of wave numbers—also becomes a continuum. Here the discrete spectrum of ‘particle in a box’ states turns into a continuum of plane waves, e^{ipx} .

A sufficiently large box $[-L, L]$ is approximately the same as an infinite interval. Of course, ‘large L ’ is a dimensionful statement. What we really want to say is that if we are probing dynamics on a scale much smaller than L , then we should expect the discrete spectrum of states to be so close to each other that it is well approximated by the continuum of plane waves.

You can twist this around in the other direction and wonder if spacetime were not actually continuous but rather composed of discrete points with some spacing on the order of the Planck length. Our continuum formalism of general relativity should be valid as long as we do not ask questions about length scales comparable to the separation between discrete points.

14.14 Fourier Series

Recall that the **Fourier Series** is a basis of sinusoidal functions over a finite domain. One changes variables from position x to momentum or wave-number p . For an interval from $x \in [0, 1]$, the Fourier series a function $f(x)$ with Dirichlet boundary conditions is:

$$f(x) = \sqrt{\frac{2}{L}} \sum_{p \in \mathbb{Z}} \tilde{f}_p \sin(\pi p x / L) . \quad (14.100)$$

The eigenfunctions $e_p(x) = \sqrt{2/L} \sin(\pi p x / L)$ satisfy the boundary conditions and are normalized so that $\langle e_p, e_p \rangle = 1$. The data of the function $f(x)$ is encoded in an infinite number of coefficients, \tilde{f}_p .

Fourier transforms are a basis of sinusodial functions—conventionally written as complex exponentials—over an *infinite* domain. Our conventions for Fourier transforms are summarized in Appendix ??.

In statistics and machine learning, this is related to the idea of a **kernel**. Kernels are the ‘heart’ of an integral transformation that takes a function $f(x)$ to a transformed function $\tilde{f}(p)$ by

$$f(x) \rightarrow \tilde{f}(p) \equiv \int dp f(x) K(x, p) \quad (14.101)$$

with the appropriate limits of integration. Observe that the transformed function \tilde{f} is formally a function of a different variable.

Exercise 14.20. What is the kernel of the Fourier transform using our conventions?

14.15 The Variational Principle

One of the fundamental principles in physics is the **principle of least action**. This is the statement that classically, the trajectory of a system is governed by the minimum of a functional called the action, S . A functional is a “function of functions.” So suppose we care about the position of a particle $\mathbf{x}(t)$ in some potential. The position is itself a function of time, and the action is a function of $\mathbf{x}(t)$. You may be more familiar with the action as the time integral of the Lagrangian,

$$S = \int dt L[\mathbf{x}, \dot{\mathbf{x}}] , \quad (14.102)$$

where we use the standard convention where we treat the position and velocity as independent variables in phase space. In quantum mechanics, this same object appears as a weight for possible trajectories.¹⁰¹

The usual method Perhaps you are familiar from analytical mechanics that we can ‘vary’ the Lagrangian to derive the equation of motion. For example, the standard treatment goes something like this. The Lagrangian of a classical particle is the kinetic minus the potential energy,

$$L = \frac{m}{2} \dot{\mathbf{x}}^2 - V[\mathbf{x}] . \quad (14.103)$$

When we treat \mathbf{x} and $\dot{\mathbf{x}}$ as independent variables, the infinitesimal change (variation) of the action S with respect to infinitesimal perturbations $\delta\mathbf{x}$ and $\delta\dot{\mathbf{x}}$ is

$$\delta S = \int dt \left[\frac{\partial L}{\partial \mathbf{x}} \delta \mathbf{x} + \frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right] . \quad (14.104)$$

Now we ‘remember’ that $\dot{\mathbf{x}}$ is actually the time derivative of $\mathbf{x}(t)$. This means we can integrate the second term by parts to remove the time derivative in the $\delta\dot{\mathbf{x}}$. What we’re really saying here is that there is *one* variation we’re doing: the function $\mathbf{x}(t)$. So if the path changes by $\mathbf{x} \rightarrow \mathbf{x} + \delta\mathbf{x}$, then the time derivative of the path changes by

$$\dot{\mathbf{x}} \rightarrow \dot{\mathbf{x}} + \frac{d}{dt}(\delta\mathbf{x}) \equiv \dot{\mathbf{x}} + \delta\dot{\mathbf{x}} . \quad (14.105)$$

We assume some boundary conditions—the position of the particle at some initial and some final times—so that the boundary terms in the integration by parts vanishes. We are thus left with (suppressing the limits of integration)

$$\delta S = \int dt \left[\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{x}}} \right] \delta \mathbf{x} . \quad (14.106)$$

¹⁰¹I recommend looking up Feynman’s phasor description of the sum over histories to see how the action in quantum mechanics leads to the principle of least action in classical mechanics.

Because this is true for any variation $\delta \mathbf{x}$ about a path $\mathbf{x}(t)$ that minimizes S , we find that the equation of motion is

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{x}}} = 0 . \quad (14.107)$$

For the single particle Lagrangian above, this indeed gives

$$m \ddot{\mathbf{x}} = V'[\mathbf{x}] . \quad (14.108)$$

The path integral perspective Our discretized function space—for all of its flaws—is helpful here for an alternative picture of the equation of motion that is often invoked in field theory. Let us treat everything as ‘histogram space.’ For simplicity, let us stick to one spatial dimension, leaving the generalization as an exercise. We would like to determine what is the function, $|x\rangle = x(t)$ that minimizes $S[x]$. We may write S as a sum:

$$S = \sum_i \Delta t L[x^i, \dot{x}^i] \quad |x\rangle = x^i |i\rangle \quad |\dot{x}\rangle = \dot{x}^i |i\rangle , \quad (14.109)$$

where we are in a ‘histogram basis’ in time slices. The basis ket $|i\rangle$ is a unit ‘histogram’ at time $t = t_0 + i\Delta t$ for some initial time t_0 . Further, we know that

$$|\dot{x}\rangle = D |x\rangle , \quad (14.110)$$

where D is the matrix representation of the time derivative, d/dt .¹⁰² To be explicit, let us assume a harmonic oscillator potential. The Lagrangian is a local function of time. We can write it—somewhat awkwardly—as an object with an index:

$$L^i = \frac{m}{2} \dot{x}^i \dot{x}^i - V^i \quad V^i = \frac{k}{2} x^i x^i , \quad (14.111)$$

where we understand that there is no sum over i . Recalling that our histogram space metric is proportional to Δt , it is in fact more straightforward to directly write the quadratic part of the action:

$$S = \frac{m}{2} g_{ij} (Dx)^i (Dx)^j - \frac{k}{2} g_{ij} x^i x^i . \quad (14.112)$$

We will see that the *quadratic* part of the action gives a linear equation of motion. This means that higher-order terms (e.g. trilinear terms like $(x^i)^3$) must be treated separately.¹⁰³

¹⁰²In principle we should specify the forward or the backward derivative, D_\pm . We should also remember that $D^2 = D_\mp D_\pm$. However, everything that we say here is applicable to the continuum limit, so if you are a stickler for formalism, you can treat these steps as an analogy to the continuous case.

¹⁰³Often we treat these as a small correction to the quadratic action. This amounts to saying that everything is a perturbation on a harmonic oscillator. The perturbations can be treated systematically using a technique called Feynman diagrams.

Exercise 14.21. Confirm that (14.112) is the discretization of the action for a simple harmonic oscillator. For example, show that in the continuum limit you recover the simple harmonic oscillator action.

Recalling Exercise 14.6, we may integrate the derivative term by parts to obtain

$$S = -\frac{m}{2}g_{ij}x^i(D^2x)^j - \frac{k}{2}g_{ij}x^ix^i \equiv -\frac{q}{w}x^iQ_{ij}x^j , \quad (14.113)$$

where we have defined the ‘quadratic term’ operator

$$Q_{ij} = m g_{ik}(D^2)_j^k + k g_{ij} = Q_{ji} . \quad (14.114)$$

Exercise 14.22. Confirm that Q_{ij} is symmetric, $Q_{ij} = Q_{ji}$.

Exercise 14.23. What is the support of (14.114)? That is, what is the condition on i and j to find a non-zero Q_{ij} ?

The principle of least action posits that the classical trajectory $x(t)$ —encoded in the discretized x^i coefficients—is the configuration that extremizes S . To do this, we vary each x^i independently. This is what the functional variation $\delta x(t)$ means in the continuous limit. The extremization corresponds to the N equations

$$\frac{dS}{dx^1} = 0 \quad \frac{dS}{dx^2} = 0 \quad \dots \quad \frac{dS}{dx^N} = 0 . \quad (14.115)$$

where relevant, these are subject to the boundary conditions. Of course, the x^i are not independent: we know that the kinetic term relates the displacement at time i to the neighboring displacements at time $i \pm 1$. Rather than solving each of these one by one, we can solve the equation for a generic x^i :

$$-\frac{dS}{dx^i} = \frac{d}{dx^i} \left[\frac{1}{2}Q_{ii}(x^i)^2 + \sum_{j \neq i} \frac{1}{2} (Q_{ij}x^i x^j + Q_{ji}x^j x^i) \right] \quad (14.116)$$

$$= Q_{ii}x^i + \sum_{j \neq i} Q_{ij}x^j . \quad (14.117)$$

where there is no sum over repeated i indices and we have explicitly written the sum over $j \neq i$. In the last term we have used $Q_{ij} = Q_{ji}$. Expanding this expression gives

$$0 = -\frac{dS}{dx^i} = -2\frac{m}{\Delta t^2}x^i + kx^i + \frac{m}{\Delta t^2}x^{i+1} + \frac{m}{\Delta t^2}x^{i-1} = g_{ij} [m(D^2x)^j + kx^j] . \quad (14.118)$$

This gives us the discretized equation of motion:

$$(D^2x)^j = -kx^j , \quad (14.119)$$

which we recognize is the discretized Hooke's law $m\ddot{x} = -kx$.

What is significant about this approach is that we never invoked the standard Euler–Lagrange trick of first treating x and \dot{x} as independent variables and then ‘remembering’ that they are related by a time derivative when it was convenient. We took a more direct path where we independently varied each piece of the path in time, x^i , while treating a differential operator Q_{ij} as a matrix. We noticed that the x^i are *coupled* by the kinetic term, which tells us that wiggles in x^i will affect the variation of $x^{i+\pm 1}$. In quantum mechanics, this approach is known as the path integral formulation.

A Things we did not get to

- Tensor densities
- Symmetries and Lie Groups
- Other types of integral transforms
- Hamiltonian Mechanics and the co-tangent space
- Rotations in the $\partial x'/\partial x$ notation.
- non-coordinate bases

Acknowledgments

PT thanks the students of Physics 17 (Spring 2022, Spring 2023) for their feedback and patience. PT is supported by a NSF CAREER award (#2045333).