

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	2
2	Motivation	3
3	Basics	4
3.1	Pre-conceptions	4
3.2	Real Three-Vectors	5
3.3	Vectors and Numbers	6
3.4	Notation: Indices	7
3.5	Arithmetic and linear combinations	7
3.6	Rescaling: multiplication by a number	8
3.7	Linear Combination and Span	9
3.8	Basis vectors: an illustrative example	9
3.9	Basis vectors, formally	12
3.10	The meaning of column vectors	14
3.11	Operations that are not (yet) allowed	15
3.12	Euclidean three-space	16
3.13	Length in Euclidean three-space	17
3.14	Angles in Euclidean three-space	18
4	Index Notation and Summation Convention	19
5	Ket Notation	22
6	Matrices and Linear Transformations	22
6.1	Jargon	22
6.2	Linear transformations	23
6.3	Linear Transformations in One Dimension	24
6.4	Linear Transformations in Two and Three Dimensions	24
6.5	The trivial transformation: identity matrix	24
6.6	Composition of linear transformations	26
6.7	Linearity and Basis Vectors	27
6.8	Linear transformations as maps acting on basis vectors	30
6.9	Inverse transformations	30
6.10	Matrices that aren't invertible: Projections	32
6.11	Funny things you can do to matrices	33
6.12	Matrices that aren't square: maps between vector spaces	34

7	Rotations	36
7.1	Rotations in Two-Dimensions	36
7.2	Transformation of Vectors	37
7.3	Transformation of Row Vectors	37
7.4	Transformation of Matrices	37
7.5	Transformation of Tensors	37
7.6	Rotations in Three-Dimensions	37
7.7	Invariance of Trace and Determinant	37
7.8	Exponentiation of Generators	37
7.9	Symmetry and Indices	37
7.10	Active versus Passive	37
8	Central Dogma I: vectors are not columns	38
9	The incredibly un-mathematical idea of niceness	38
10	Central Dogma II: the eigensystem	38
11	Linear Maps to Numbers: Row Vectors	38
12	Metric Spaces: Dot Products	38
12.1	In contrast to Row Vectors	38
12.2	Machine to create row vectors	38
12.3	Length	38
12.4	Orthogonality	38
12.5	Rotations, once again	38
12.6	The meaning of indices	38
13	More General (Multi-)Linear Maps: Tensors	38
14	Special Relativity: Minkowski Space	38
15	Interlude: an abundance of names and notation	38
16	Nice and Not-So-Nice Transformations	38
16.1	Rotations	38
16.2	“Nice” Transformations	38
16.3	Projections	38

1 Logistics

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.

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

2 Motivation

Here are three incredibly significant equations in physics:

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

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

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.

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

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.

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

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 $\mathcal{O}(\epsilon)$ term. Integration boils down to summing trapezoids whose angle-bits are given by the first derivative of a function... the linear component.

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.”⁵*

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.

⁵I 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.

	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.

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.⁶

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 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.⁷ 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

⁶The 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.

⁷Formally, 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.

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:

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

We see that a boldfaced Roman letter, \mathbf{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 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:

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

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.⁸ The components of $\alpha\mathbf{v}$ are

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

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*

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

⁸“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 is \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.⁹ 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 .

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:

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

⁹You may want to convince yourself that this satisfies the requirements of vector space arithmetic.

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):¹⁰

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

The last two equations mean exactly the same thing. Now here’s something cute: we can treat the two-component array¹¹ 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”

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

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

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

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 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.¹² All of this begs for a re-definition.

¹⁰We will soon see that there is nothing holy about (3.12).

¹¹I’m trying not to call it a vector.

¹²Though 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.

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.

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* statements¹³:

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} .

¹³This means: if they are not immediately apparent, stop and think about it to make sure you understand.

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 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.¹⁴*

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*

¹⁴There are some great YouTube videos on this. Here are a few: <https://www.youtube.com/watch?v=xAoljeRJ3lU>, <https://www.youtube.com/watch?v=AS10HMMW873s>, <https://www.youtube.com/watch?v=99v96TL-tuY>.

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.¹⁵ 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 .

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 bases¹⁶ is so important that we should really emphasize the mathematical edifice that we have reverse-engineered¹⁷

¹⁵This 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 $N \rightarrow \infty$.

¹⁶The plural of ‘basis’ is ‘bases,’ pronounced *bay-sees*, just like the plural of ‘axis’ is ‘axes’ pronounced *axe-sees*.

¹⁷Do 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.

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 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.¹⁸

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**.

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

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

¹⁸If 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.

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.¹⁹ 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.)

The **cross product** is a different story. You may remember the cross product from such hits as²⁰ 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 different²¹ 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.²²

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

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

¹⁹It is dangerous to use the phrase “position vector,” see Exercise 3.1.

²⁰<https://tvtropes.org/pmwiki/pmwiki.php/Main/YouMightRememberMeFrom>

²¹The technical meaning of ‘completely different’ is *orthogonal*, which we define below with the help of the metric.

²²That 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.

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**, $\hat{\mathbf{v}}$:

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

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

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

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:²³ 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

²³There 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.

$\tilde{\mathbf{w}}$ on a column vector \mathbf{v} . Nevermind the formal definition of “row vector” as opposed to “column vector.” Let’s just write it out in components where it is obvious for \mathbb{R}^3

$$\tilde{\mathbf{w}} = (w_1 \quad w_2 \quad w_3) \quad \mathbf{v} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \quad \tilde{\mathbf{w}}\mathbf{v} = w_1v^1 + w_2v^2 + w_3v^3 . \quad (4.1)$$

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 notice that we chose to write the components of $\tilde{\mathbf{w}}$ 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:

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

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.

Example 4.1. *Matrices M have the following index structure: M^i_j . 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:*

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

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:

$$M^i_j v^j = M^i_1 v^1 + M^i_2 v^2 + M^i_3 v^3 . \quad (4.4)$$

Example 4.2. *From the above example, you can then excuse the glib statement: “the vector $M^i_j v^j$.” As we explained above, $M^i_j v^j$ 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.*

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

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

These have index structure v^i , w_i , and M^i_j respectively. Note that the first index of a matrix is the row and the second is the column, thus $M^1_2 = 2$ while $M^2_1 = 4$. Calculate the following: $(wM)_2$, $(Mv)^1$, $(MM)^1_2$. Here MM is understood to be the square of the matrix M , $(M^2)^i_j = M^i_k M^k_j$.

Example 4.3. It should be clear that

$$w_i M^i_j = w_1 M^1_j + w_2 M^2_j + w_3 M^3_j = M^1_j w_1 + M^2_j w_2 + M^3_j w_3 = M^i_j w_i. \quad (4.6)$$

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

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

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 $w_i M^i_j = M^i_j w_i$ 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 vector-y-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 “vector-y” 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 Ket Notation

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 $\tilde{\mathbf{w}} = \langle w|$. We call this object a *bra*. Then the “matrix multiplication” $\tilde{\mathbf{w}}\mathbf{v}$ 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:

$$\text{column vector: } \mathbf{v} = |v\rangle \qquad \text{row vector: } \tilde{\mathbf{w}} = \langle w| \qquad (5.1)$$

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

$$\hat{H}|\Psi\rangle = E|\Psi\rangle . \qquad (5.2)$$

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

6 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.

6.1 Jargon

Let us introduce some jargon here. Rather than formal definitions, we give practical “physicist’s” definitions.²⁴ 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. 1.

²⁴If 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.

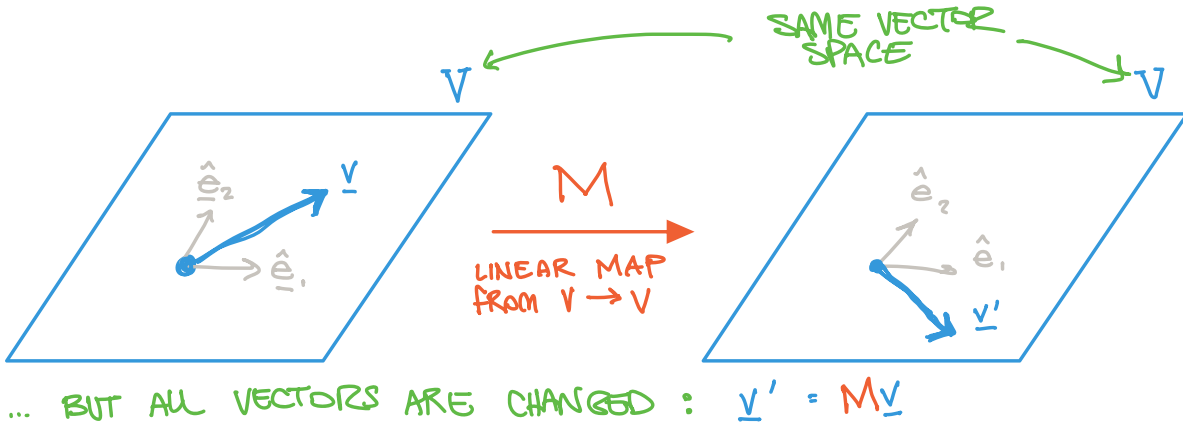


Figure 1: Example of a linear transformation. The transformation M (for matrix) takes a vector \mathbf{v} and turns it into a different vector that we call \mathbf{v}' . This new vector is related to \mathbf{v} by the matrix $\mathbf{v}' = M\mathbf{v}$. Every vector is transformed under the map M . This is an *active transformation* where all vectors transform, but the basis ($\hat{\mathbf{e}}_i$) stays fixed.

Example 6.1. The dot product is a **function** that takes in two vectors and outputs a number.

Example 6.2. The magnitude is a **map** that takes a vector and returns a number.

Example 6.3. A function $f(x, y) = z$ is not a map because it takes two inputs.

Example 6.4. 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.

Example 6.5. 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 6.6. 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).

6.2 Linear transformations

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

$$f(\alpha x) = \alpha f(x) \qquad f(x + y) = f(x) + f(y) . \qquad (6.1)$$

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:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \qquad (6.2)$$

Exercise 6.1. Show that equations (6.1) and (6.2) imply one another. Going from (6.1) to (6.2) is slightly more tricky than the other way around.

6.3 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 6.2. Show that $f(x) = mx + b$ does not satisfy the definition of linearity, (6.2). 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 6.3. 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 spaces²⁵ if are 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.

6.4 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

$$(Mv)^i = M^i_j v^j . \quad (6.3)$$

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.²⁶ 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.²⁷

6.5 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 :

$$\mathbb{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (6.4)$$

²⁵https://en.wikipedia.org/wiki/Affine_space

²⁶I 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.

²⁷This should remind you of dimensional analysis. If you have different dimensions on two sides of an equation, you have done something wrong.

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 :

$$\delta_j^i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \quad (6.5)$$

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_{ij} , 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_j^i 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

$$\delta_j^i = \delta_{ij}^i = \delta_i^j . \quad (6.6)$$

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.²⁸

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

$$(\mathbb{1}\mathbf{v})^i = \delta_j^i v^j = \delta_1^i v^1 + \delta_2^i v^2 + \delta_3^i v^3 . \quad (6.7)$$

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

$$\delta_j^i v^j = v^i . \quad (6.8)$$

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.

Exercise 6.4. *Show that this works for lower indices as well. Consider the multiplication of two matrices, $M\mathbb{1}$. Show that*

$$M_j^i \delta_k^j = M_k^i , \quad (6.9)$$

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

²⁸Instead, 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 6.5. *Even though we have not introduced any objects with more complicated index structure, suppose you have the following object:*

$$R^\mu_{\nu\rho\sigma} . \quad (6.10)$$

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

$$\delta^\mu_\alpha \delta^\nu_\beta \delta^\rho_\gamma \delta^\sigma_\tau R^\mu_{\nu\rho\sigma} . \quad (6.11)$$

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.

6.6 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

$$f \circ g(\mathbf{x}) \equiv f(g(\mathbf{x})) \quad \Leftrightarrow f \circ g(\mathbf{x}) \quad = MN\mathbf{x} \quad (6.12)$$

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

$$(MN)^i_j = M^i_k N^k_j . \quad (6.13)$$

Example 6.7. *One thing that we notice from (6.13) is that the index structure makes it clear that the order of matrix multiplication matters.*

$$(MN)^i_j = M^i_k N^k_j \neq N^i_k M^k_j = (NM)^i_j . \quad (6.14)$$

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:

$$(MN)^i_j = M^i_k N^k_j = N^k_j M^i_k . \quad (6.15)$$

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

$$(NM)^i_j = N^i_k M^k_j = M^k_j N^i_k . \quad (6.16)$$

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 6.8. 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 6.6. 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:

$$\hat{M} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \text{diag}(2, 7, 4) . \quad (6.17)$$

Show that the product of diagonal matrices is also diagonal whose components are the products of the diagonal components. You can use (6.13) armed with the knowledge that $(\hat{M})^i_j = 0$ if $i \neq j$. Then you must show that $(\hat{M}\hat{N})^i_j = 0$ for $i \neq j$. Finally, check the values of the diagonal elements to show that

$$\hat{M}\hat{N} = \text{diag}(M^1_1 N^1_1, M^2_2 N^2_2, M^3_3 N^3_3) . \quad (6.18)$$

6.7 Linearity and Basis Vectors

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

Exercise 6.7. Calculate the following linear transformations by a matrix M on two different vectors, \mathbf{v} and \mathbf{w} :

$$M\mathbf{v} = \begin{pmatrix} 3 & 2 & 1 \\ -1 & 4 & -2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = ? \quad M\mathbf{w} = \begin{pmatrix} 3 & 2 & 1 \\ -1 & 4 & -2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} = ? \quad (6.19)$$

Exercise 6.8. That was tedious, wasn’t it? What if I asked you to calculate three more...

$$M \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = ? \quad M \begin{pmatrix} 0 \\ 1 \\ 7 \end{pmatrix} = ? \quad M \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix} = ? \quad (6.20)$$

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 (6.20) are **linear combinations** of \mathbf{v} and \mathbf{w} . What you should notice is that

$$\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = \mathbf{v} + \mathbf{w} \qquad M \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} = M(\mathbf{v} + \mathbf{w}) = M\mathbf{v} + M\mathbf{w} . \quad (6.21)$$

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 (6.20).

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

$$\mathbf{u} = a\mathbf{v} + b\mathbf{w} \quad (6.22)$$

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

$$M\mathbf{u} = a(M\mathbf{v}) + b(M\mathbf{w}) . \quad (6.23)$$

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 (6.23) 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 6.9. *The two approaches above are equivalent when the basis vectors are*

$$\mathbf{b}_1 = \hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathbf{b}_2 = \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \quad (6.24)$$

This should be clear since

$$M\hat{\mathbf{e}}_1 = \begin{pmatrix} M^1_1 \\ M^2_1 \end{pmatrix} \qquad M\hat{\mathbf{e}}_2 = \begin{pmatrix} M^1_2 \\ M^2_2 \end{pmatrix} . \quad (6.25)$$

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 be settle into thinking that the 2×2 of numbers somehow *is* the matrix.

Example 6.10. 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 $M\hat{\mathbf{e}}_1$.

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

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \quad \cdots \quad \hat{\mathbf{e}}_N = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} . \quad (6.26)$$

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:

$$\mathbf{v} = v^i \hat{\mathbf{e}}_i \quad \mathbf{v} \text{ “=” } \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^N \end{pmatrix} . \quad (6.27)$$

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,

$$M\hat{\mathbf{e}}_i = (M\hat{\mathbf{e}}_i)^k \hat{\mathbf{e}}_k \quad M^k_i \equiv (M\hat{\mathbf{e}}_i)^k . \quad (6.28)$$

Be very careful with this. What we mean is the N -dimensional generalization of (6.25), 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^i_j that is fully general.

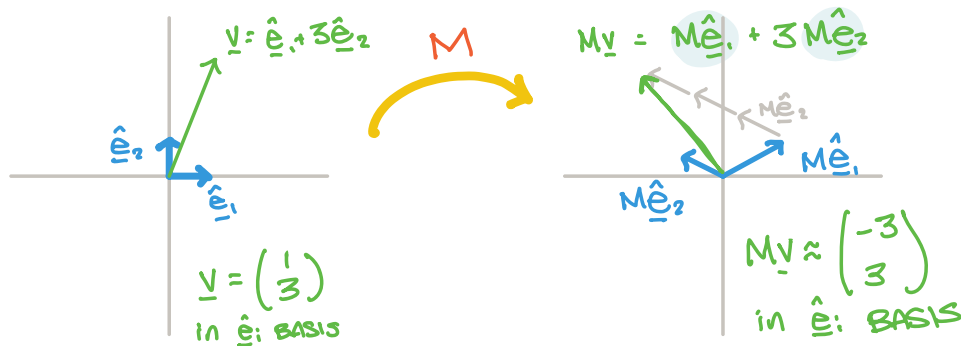


Figure 2: M is a linear transformation from $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. If we know the action of M on the basis vectors $\hat{\mathbf{e}}_i$ of \mathbb{R}^2 , then we know the action of M on any vector \mathbf{v} .

6.8 Linear transformations as maps acting on basis vectors

The relation between the components of linear transformations and basis vectors is elucidated in Fig. 6. 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 $\hat{\mathbf{e}}_i$:

$$\mathbf{v} = v^i \hat{\mathbf{e}}_i M \mathbf{v} = v^i M \hat{\mathbf{e}}_i. \quad (6.29)$$

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 (6.25).
- What is nice about (6.29) 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.

6.9 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 $(M^{-1})^i_j$. The defining property of the matrix inverse

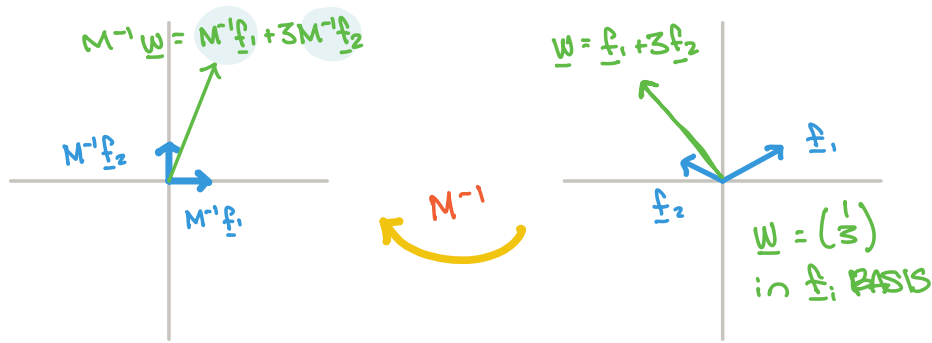


Figure 3: Let M is a linear transformation from $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, as we drew in Fig. 6. Now assume that you have a vector \mathbf{w} that is the linear combination of some basis vectors \mathbf{f}_i , see the right-hand side of the sketch. If the \mathbf{f}_i can be obtained as the transformation by M of some other vectors—the basis vectors we called $\hat{\mathbf{e}}_i$ in Fig. 6—then the inverse transformation M^{-1} acts as a map “back” to the original basis.

is $M^{-1}M = \mathbb{1}$. That means that $M^{-1}(M\mathbf{v}) = \mathbf{v}$ for any vector \mathbf{v} . The matrix inverse simply undoes the transformation M . This is shown in Fig. 3.

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

$$M(M^{-1}) = (M^{-1})M = \mathbb{1} \quad M_k^i (M^{-1})_j^k = (M^{-1})_k^i M_j^k = \delta_j^i. \quad (6.30)$$

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 6.9. Convince yourself that (6.30) 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?

Example 6.11. The inverse of a diagonal matrix is easy:

$$M = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} \quad M^{-1} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \quad (6.31)$$

We have used the earlier observation in Exercise 6.6 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

$$M\mathbf{v} = \mathbf{s} \quad \Leftrightarrow (\text{dynamics})(\text{state}) = (\text{source}). \quad (6.32)$$

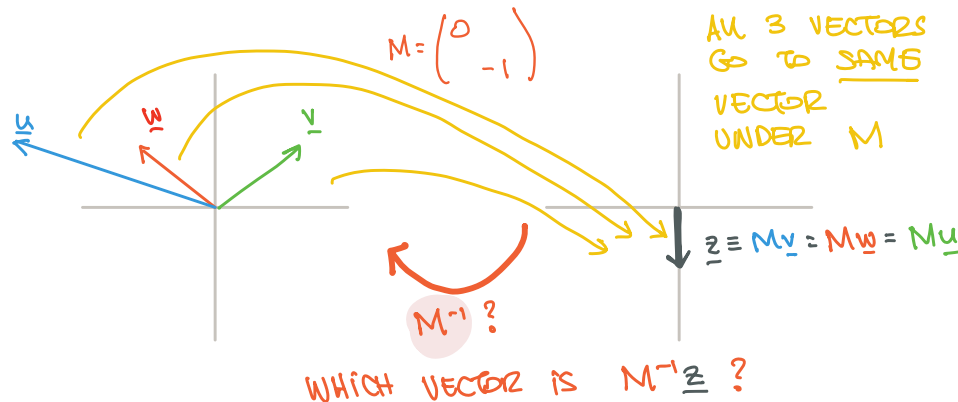


Figure 4: M is a non-invertible matrix. We notice that there’s no problem applying M to vectors. The odd thing is that there are multiple vectors— \mathbf{v} , \mathbf{w} , and \mathbf{u} in the picture—that all turn into the same vector, $M\mathbf{v} = M\mathbf{w} = M\mathbf{u} \equiv \mathbf{z}$. The 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.

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.

6.10 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. 4. 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

$$(M^{-1})\mathbf{z} = (M^{-1})M\mathbf{v} = \mathbf{v} \quad (M^{-1})\mathbf{z} = (M^{-1})M\mathbf{w} = \mathbf{w} . \quad (6.33)$$

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. 4. The map M is essentially only using the information about the second component of its input vector and 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

$$\mathbf{v} = \begin{pmatrix} 0 \\ v^2 \end{pmatrix} . \quad (6.34)$$

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.²⁹

Example 6.12. *One of the most curious puzzles in fundamental physics over the last half century is the black hole information paradox.³⁰ 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.*

6.11 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

$$(M^T)^i_j = M^j_i . \quad (6.35)$$

Example 6.13. *The following two matrices are transposes of each other.*

$$M = \begin{pmatrix} 9 & -3 & 2 \\ 2 & 5 & 7 \\ 1 & 0 & 3 \end{pmatrix} \quad M^T = \begin{pmatrix} 9 & 2 & 1 \\ -3 & 5 & 0 \\ 2 & 7 & 3 \end{pmatrix} . \quad (6.36)$$

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 us take (6.35) as the definition of something we can ‘do’ to matrices.

²⁹I 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.

³⁰<https://www.quantamagazine.org/the-most-famous-paradox-in-physics-nears-its-end-20201029/>

Trace

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

$$\text{Tr } M = M^i_i = M^1_1 + M^2_2 + \cdots . \quad (6.37)$$

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

$$\det M \equiv M^1_1 M^2_2 - M^1_2 M^2_1 . \quad (6.38)$$

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 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 6.14. *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.

6.12 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

$$W = \begin{pmatrix} 1 & 2 & -6 & 7 \\ 6 & -3 & 3 & 0 \end{pmatrix} . \quad (6.39)$$

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

$$W : \mathbb{R}^4 \rightarrow \mathbb{R}^2 . \quad (6.40)$$

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.³¹

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.³² 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**³³ 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 6.10. *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.
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 .

³¹To 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.

³²Why 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.

³³The 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.

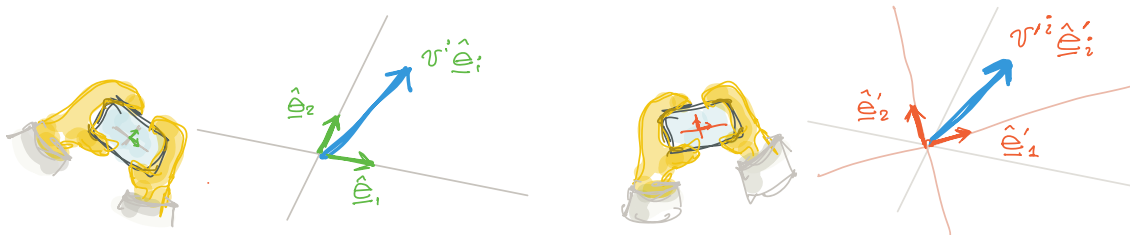


Figure 5: Example of a passive rotation. You are looking at a vector \mathbf{v} through your phone. The image on your phone has a natural orientation for a basis $\hat{\mathbf{e}}_i$ and the components of \mathbf{v} in this basis are v^i . You could also rotate your phone, leaving the vector \mathbf{v} 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 .

7 Rotations

Rotations are a special class of transformations. We wait until Section 12 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 kindergarden 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. 5.

7.1 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 is

$$R(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{pmatrix}, \quad (7.1)$$

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

$$\sin^2 \theta + \cos^2 \theta = c_\theta^2 + s_\theta^2 = 1. \quad (7.2)$$

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

$$R^{-1}(\theta) = \begin{pmatrix} c_\theta & -s_\theta \\ s_\theta & c_\theta \end{pmatrix} = R(\theta)^T. \quad (7.3)$$

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

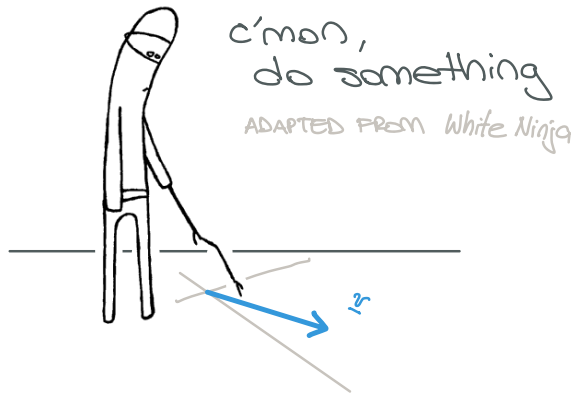


Figure 6: Nothing actually happens during a passive transformation. The basis vectors transform, but the components of the vector transform in such a way that $v^i \hat{\mathbf{e}}'_i = v^i \hat{\mathbf{e}}_i$. Such a ‘transformation’ may come off as uninspiring, but this is the basis of symmetries in physics. From <https://knowyourmeme.com/memes/cmon-do-something>.

Exercise 7.1. Check component-by-component that $R(\theta)R^{-1}(\theta) = \mathbb{1}$ using the expressions above. One may directly show that

$$R(\theta)^i_k R^{-1}(\theta)^k_j = \delta^i_j \quad (7.4)$$

For each of the four combinations of i and j .

7.2 Transformation of Vectors

7.3 Transformation of Row Vectors

7.4 Transformation of Matrices

7.5 Transformation of Tensors

7.6 Rotations in Three-Dimensions

7.7 Invariance of Trace and Determinant

7.8 Exponentiation of Generators

7.9 Symmetry and Indices

7.10 Active versus Passive

Passive is a symmetry. Active is an actual “do something”

- 8 Central Dogma I: vectors are not columns
- 9 The incredibly un-mathematical idea of niceness
- 10 Central Dogma II: the eigensystem
- 11 Linear Maps to Numbers: Row Vectors
- 12 Metric Spaces: Dot Products
 - 12.1 In contrast to Row Vectors
 - 12.2 Machine to create row vectors
 - 12.3 Length
 - 12.4 Orthogonality
 - 12.5 Rotations, once again
 - 12.6 The meaning of indices
- 13 More General (Multi-)Linear Maps: Tensors
- 14 Special Relativity: Minkowski Space
- 15 Interlude: an abundance of names and notation
- 16 Nice and Not-So-Nice Transformations
 - 16.1 Rotations
 - 16.2 “Nice” Transformations
 - 16.3 Projections

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).