Flip Tanedo

Physics 231: Methods of Theoretical Physics

A mathematical methods course for first-year graduate students in physics & astronomy

Lecture Notes

September 28, 2022

Abstract

Physics 231 is a crash course on mathematical methods necessary to succeed in the first-year physics graduate curriculum at UC Riverside. The focus is how to solve differential equations using Green's functions. This version is revised for Fall 2022. Last Compiled: September 28, 2022

Contents

1   *Mathematical Methods*

This is a *crash course* on the mathematical toolkit necessary for graduate courses in electrodynamics, quantum mechanics, and statistical mechanics. The emphasis is physical intuition rather than mathematical rigor. Let us be clear: as a student you are *expected* to be as mathematically rigorous as your discipline requires. Fortunately, there are plenty of excellent textbooks pitched at various levels of rigor and you can find the one most appropriate for you. This course is meant to complement those references, not to replace them.[1]

Unfortunately for you, it is unlikely that the choice of topics in this course will be either necessary or sufficient for your training. If anyone asks, you should say that the theme of this course is to solve the types of linear differential equations that will show up in your physics coursework (*ugh! Boring!*). The actual choice of topics is meant to highlight big, unifying themes in mathematical physics sprinkled with topics of current research significance.

1.1   *Green's functions and this course*

Our goal is to solve linear differential equations:

$$\mathcal{O}f(x) = s(x) . \tag{1.1}$$

In this equation, $\mathcal{O}$ is a *differential operator* that encodes some kind of physical dynamics[2], $s(x)$ is the *source* of those dynamics, and $f(x)$ is the system's physical *response* that we would like to determine. The solution to this equation is:

$$f(x) = \mathcal{O}^{-1}s(x) . \tag{1.2}$$

[2]A **differential operator** is just something built out of derivatives that can act on a function. The differential operator may contain coefficients that depend on the variable that we are differentiating with respect to; for example, $\mathcal{O} = (d/dx)^2 + 3x\,(d/dx)$.

This statement is trivial and deeply unsatisfying. We will think carefully about what $\mathcal{O}^{-1}$ actually means and how to calculate it. $\mathcal{O}^{-1}$ is the **Green's function** for the differential operator $\mathcal{O}$.

Exercise 1.1.1. *Consider the differential operator $\mathcal{O} = (d/dx)^2 + 3x\,(d/dx)$. A colleague tells you that $(d/dx)^2$ is squared, therefore it is not a linear operator. Explain why the colleague is mistaken.*

To make sense of $\mathcal{O}^{-1}$, we appeal to linear algebra. A linear transformation—
that is, a **matrix**—A acts on a vector $\boldsymbol{v}$ to give equations like

$$A\boldsymbol{v} = \boldsymbol{w}\,, \tag{1.3}$$

whose solution is

$$\boldsymbol{v} = A^{-1}\boldsymbol{w}\,. \tag{1.4}$$

In this course, we think of linear differential operators $\mathcal{O}$ as infinite-dimensional
matrices. $\mathcal{O}^{-1}$ is the inverse of this matrix. At this point, you should feel a bit
nervous because you remember that inverse of a 3×3 matrix is tricky... say
nothing of the *infinite* dimensional limit. We remember, however, that calculus
is infinite-dimensional linear algebra. Complex analysis extends the real line to
the complex plane. In so doing, the *analytic structure* of our theories offer both a
method to calculate challenging integrals and their own physical significance.

Our simple example through this journey is the humble harmonic oscillator.
In subsequent chapters we extend this to higher-dimensions and, indeed, to
infinite dimensions. Along the way we reflect on the nature of all of these infini-
ties we bandy about. We close by connecting our study of Green's functions to
statistics, non-perturbative methods, and the burgeoning connections of physics
to machine learning.

1.2   *This is not what I expected*

This is a course in mathematical methods for *physicists*. We do not solve *every*
class of differential equation that is likely to pop up in your research careers—
that would be a course on mathematical methods for *engineers*. (I recommend
Carl Bender's 2011 lectures at PSI for an insightful course along those lines.[1])
Instead, we methodically dissect a physically motivated example—the harmonic
oscillator—to emphasize how we think about mathematical problems.

We weave together ideas that are not often connected explicitly in undergrad-
uate physics courses: linear algebra, differential equations, complex analysis,
statistics. I expect that you have had some formal training in these topics so
that we may focus on the interconnections between these ideas and how those
interconnections come up over and over again in our study of nature.

[1] Carl Bender. *Mathematical Physics - Lecture 1*. Nov. 2011. URL: https://pirsa.org/11110040.

Do not be surprised if we only mention Bessel functions in passing. Do not think less of our efforts if we do not calculate Wronksians or go beyond a single Riemann sheet. As graduate students, it is *your* responsibility to be able to grab your favorite reference to apply mathematics as needed to *your* research. *This* course is about the larger narrative that is not often shared explicitly in those books. It is about that which makes physicists employable in Silicon Valley while simultaneously terrible at splitting the bill at a restaurant.

Example 1.2.1. *Our cavalier attitude towards mathematical rigor should not make you think that mathematical rigor is not necessary. For a nice, visual example, see "How to lie using visual proofs" by 3Blue1Brown.[2]*

1.3   *The non-mathematical idea of mathematical niceness*

I find it useful to appeal to the notion of a **nice** mathematical situation. This is not a formal idea. In fact, it is one many things that mathematicians find ridiculous about me. But as a physicist, the concept of mathematical *niceness* is remarkably helpful.

The physical systems that we spend the most time thinking about are all *nice*. While our mathematical cousins spend years proving every exceptional case to a theorem, we tend to be happy to push onward as long as mathematical results are true for the *nice* cases. In fact, nature often admits an *approximately* nice mathematical description.[1]

Nice mathematical models make tidy predictions. Then we can Taylor expand about these nice predictions to make better predictions. We sometimes chant *perturbation theory* out loud several times in case someone watching us does not think we are being rigorous enough.[2]

This is not to say that nature cares at all about our physical models. Every once in a while, we *do* have to worry about the exceptional cases because our models fail to accommodate what is *actually* happening in nature.[3] Those scenarios are the most exciting of all: that is when our mathematical formalism grabs us by the collar and says, *listen to me—something important is happening!* This often happens when a calculation tells us that a physical result is infinite.

Exercise 1.3.1. *Consider the potential that an electron feels in the hydrogen atom:*

$$V(r) = -\frac{\alpha}{r} \ . \tag{1.5}$$

[1]This is not because nature is kind, but rather because we are only clever enough to build simple theories. What is important to appreciate as a student is *why* simple theories can so nearly approximate nature.

[2]Sometimes our Taylor expansions have zero radius of convergence. "*E pur si muove,*" as Galileo would say.

[3]Full disclosure? Your Ph.D will likely depend on finding a clever solution to one of these cases.

As the electron–proton separation goes to zero, r → 0, the potential goes to infinity. Classical electrodynamics is telling us that something curious is happening. What actually happens? (And why didn't you ask this question when you were in high school?)

We focus on *nice* functions and *nice* operators and *nice* boundary conditions, and so forth. We often only need the *nice* math to make progress on our *nice* physical models. It is worth spending our time learning to work with these *nice* limits. Leave the degenerate cases to the mathematicians for now. Eventually, you will find yourself in a situation where physics demands *not nice* mathematics. In that case—and only when the physics demands it—you will be ready to poke and prod at the mathematical curiosity until the underlying *physics* reason for the not-niceness is apparent. All this is to say: if you object to this course because we do not start with proofs about open sets or convergence, then you are missing the point of an education in physics.

1.4   *The unbearable arrogance of physicists*

Sometimes we, as physicists, have a reputation for arrogance. The most generous interpretation is that we must have some Promethean *chutzpah* to seek to comprehend/invent/discover an underlying mathematical organizing principle for the universe. On the other side of this is the damaging ways in which scientists can mistreat each other in academia. Somewhere in between are footnotes poking fun at mathematicians, or being a bit of a bore at parties. But these are lecture notes about mathematical physics; it is up to you to figure out how to be the best version of you-as-physicist-and-human-being that you can realize.[1]

[1]There are plenty of excellent pieces to reflect on this. For example, *The Disordered Cosmos*, *The Only Woman in the Room*, and *Beamtimes and Lifetimes*.

What is within the scope of a set of lecture notes on mathematical physics is the apparent arrogance about the way we speak and write about technical ideas. In particular, the arrogance of phrases like *it is obvious that…*. In colloquial conversation, these phrases are smug or aggressive: *look how smart I am that I comprehend this so easily!*[2] In a graduate course, however, this sentiment means something very specific with pedagogical value. When these notes say that something is *obvious* or *clear*, what we really mean is the following:

[2]In my youth I would often curse at my textbooks: *If it's so obvious, then why don't you explain it, you lazy asshole!*

> If you think about this idea with a certain perspective, then the idea is self-evident in a way that is illuminating. However, there are many other perspectives in which the idea is unclear. If you find

the idea unclear, do *not* assume that the idea itself is esoteric or that
you are somehow deficient. Instead, take a step back to see if the
idea is a natural consequence of a different approach.

Any time that these notes refer to being *obvious*, it is a checkpoint. When something is not obvious, it is an invitation to reflect and perhaps backtrack a bit. In fact, when something is not obvious, it is an *opportunity* to understand the idea more deeply—for now you have seen how the idea can become apparently complex, but you are being assured that in this complexity there is a underlying simple organizing principle. What can be more tantalizing than that?

There is a complementary idea that students have a secret superpower that they can exercise in the classroom. It is terrifying to ask questions in public—after all, what if your peers decide that you must be *stupid*? There is didactic armor against this. Whenever you are confused, and at the first appropriate moment after you are confused, raise your hand and phrase your question as follows:

> *Is it obvious that ...* ?

Linguistically, this is a trick of the passive voice: it removes *you* from the query. It does not insist that you are incapable of comprehending something, it simply asks if there is some intuitive understanding that you want to make sure you do not miss. After all, developing your physics intuition is one of the goals of your first-year graduate courses. Any self-respecting instructor will respond sympathetically, either:

⋄ *no*, it is not obvious. Perhaps then you work through the idea carefully. Or,

⋄ *yes*, it is obvious—but only when we remember some previous key step, which your instructor should then highlight.

Either way, the result is wisdom rather than risking how you look in front of your peers.

By the way, *how you look in front of your peers* is not a good reason to do anything. It is almost as bad as not asking questions because you do not want to look stupid in front of your advisor. Here is some free advice: your adviser knows *exactly* how stupid you are. Most likely your advisor does not think you are stupid, but if you are convinced that you are stupid, then rest assured that

your advisor knows this and has still chosen to invest their time into you. Make the most of this time: ask questions.

1.5   *Footnotes*

[1] *Pale Fire*, V. Nabokov.

I tend to be verbose with foot/margin-notes.[1] There are a few types of marginalia:

- ◇ References, so you do not have to flip back to a bibliography at the end of the document.

- ◇ Enigmatic hints of how ideas interconnect. This document's main narrative focuses on first-year graduate physics, but because so many of the ideas here reappear in more advanced topics, I cannot help but mention a few of them in passing. These notes seem more mysterious than pedagogical to those who have not studied those topics; please take this as an invitation to dig deeper into the topic if it excites you.

- ◇ Miscellaneous examples or observations that are not germane to the specific topic in the main text, but are worth highlighting for the eager student.

- ◇ Miscellaneous personal reflections with no pedagogical value other than to remind the reader that the human being writing these notes was once a beginner student as well.

1.6   *One last piece of advice*

It took me way too long to appreciate the crucial significance of homework and exercises in learning physics. Your job in your Ph.D is to answer questions where no previous answer had ever existed in the history of humanity. You will be guided by your advisor and your mentors, but you will be the discover-er of truth. This is a tall order, something like completing a marathon or climbing Everest. And like those physical feats, the only way to succeed in your intellectual pursuit is to *train*. And the best training we have in physics are practice problems. These are problems that are crafted to hone your skills. They are examples that are assured to be *solvable*[2] and with a framework (like a course with peers) to guide you through the challenges. Do not squander the opportunity to

[2] The fact that they are solvable does not mean that you are entitled to a solution set other than the solution that you earn by deriving it yourself.

train. My undergraduate advisor used to say, *you should do every problem in the book—but especially the ones that you cannot do.*

2   *Physics versus Mathematics*

Let us make one point clear:

$$\text{Physics} \neq \text{Mathematics} . \qquad (2.1)$$

This is a truth in many different respects[1]:

⬦ Physicists are rooted in experimental results. Even theorists? *Especially* theorists.

⬦ Physicists Taylor expand to their hearts' content. Even when it is sometimes not mathematically valid.[3]

⬦ Physicists pick a basis, use coordinates, and decorate every tensor with indices. Equations in physics appear intimidating because of the indices decorating our variables. Ironically, physicists are often intimidated by mathematics because of the conspicuous absence of any indices..

⬦ Physicists seek truths about *this* universe.

⬦ Physicists have a fast and loose relationship to the concept of infinity and the related concept of the infinitesimal—on this, I recommend Jim Holt's essay "The Dangerous Idea of the Infinitesimal."[4] At the same time, many of our tools seem to *beg* questions about the infinite.

My friends, we are not doing mathematics.

2.1   *The most important binary relation*

When we write equations, the symbol that separates the left-hand side from the right-hand side is a binary relation. We use binary relations like $=$ or $\neq$. Sometimes to make a point we write $\cong$ or $\equiv$ or $\doteq$ to mean something like 'definition' or 'tautologically equivalent to' or some other variant of *even more equal than equal.*

As physicists the most important binary relation is none of those things[2]. What we usually care about is $\sim$.[3] The symbol $\sim$ tells us how how something *scales.* If I double a quantity on the right-hand side, how does the quantity on

[1] The astronomer Fritz Zwicky would perhaps call this a *spherical truth*; no matter how you look at it, the statement is still true.

[3] John Baez. *Struggles with the Continuum (Part 6).* 2016. URL: https://johncarlosbaez.wordpress.com/2016/09/21/struggles-with-the-continuum-part-6/.

[4] J. Holt. *When Einstein Walked with Gödel: Excursions to the Edge of Thought.* Farrar, Straus and Giroux, 2018. ISBN: 9780374717841.

[2] I thank Yuval Grossman teaching me this.

[3] I use this the same way as $\propto$, which is completely different from 'approximately,' $\approx$.

Figure 2.1.
Mathematical symbol fight
from XKCD. [5] CC BY-NC 2.5
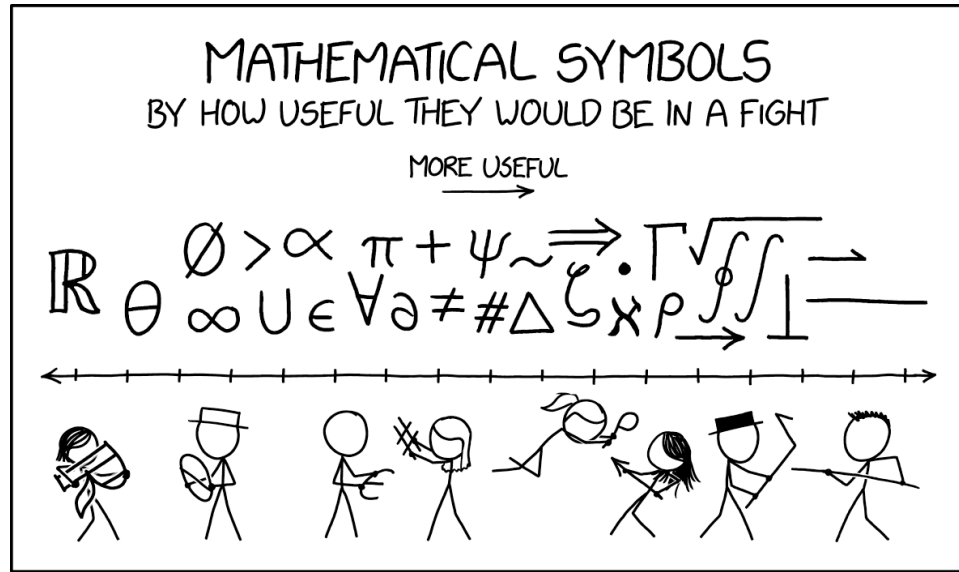
the left-hand side scale? Does it depend linearly? Quadratically? Non-linearly? The answer encodes something important about the underlying physics of the system. The symbol ~ the reason why *imagine the cow is a sphere* is a popular punchline in a joke about physicists.

Implicit in this discussion is the pragmatic policy that we will not care about stray factors of 2 in this class. As my adviser used to say, if you are worried about a factor of 2, then you have addition homework to figure out that factor of 2.[1]

[1]That being said, you are reading these notes and find an error, do let me know about it.

2.2   *Units*

There is another way in which physics is different from mathematics. It is far more prosaic. *Quantities in physics have units.* We do deal in simply numbers, we deal with kilograms, electron volts, meters. It turns out that dimensional analysis is a big part of what we do as physicists.

Exercise 2.2.1. *Explain, in words, why the quantity* $\sin(3\text{ cm})$ *is absolute nonsense in any context. What about* $\exp(2\text{ kg})$?

3 *Dimensional Analysis*

You may be be surprised how far one can go in physics by thinking deeply about dimensional analysis. Here we only get started. To take the next step, you may read more about the Buckingham Pi theorem or applications in physics. I recommend any of the following:

⬦ John H. Evans. "Dimensional Analysis and the Buckingham Pi Theorem". In: *American Journal of Physics* 40.12 (1972), pp. 1815–1822. DOI: 10.1119/1.1987069. URL: https://doi.org/10.1119/1.1987069

⬦ R. W. Robinett. "Dimensional analysis as the other language of physics". In: *American Journal of Physics* 83.4 (2015), pp. 353–361. DOI: 10.1119/1.4902882. URL: https://doi.org/10.1119/1.4902882

⬦ Fredrick Olness and Randall Scalise. "Regularization, renormalization, and dimensional analysis: Dimensional regularization meets freshman E&M". in: *American Journal of Physics* 79.3 (2011), pp. 306–312. DOI: 10.1119/1.3535586. URL: https://doi.org/10.1119/1.3535586

⬦ Paul M. Stevenson. "Dimensional Analysis in Field Theory: An Elementary Introduction to Broken Scale Invariance and the Renormalization Group Equations". In: *Annals Phys.* 132 (1981), p. 383. DOI: 10.1016/0003-4916(81)90072-5.

Dimensional analysis is simply the idea that by keeping track of the units of physical quantities, we can learn quite a bit about how those quantities must show up in our physical laws.

3.1 *Converting Units*

Imagine that you have three apples. This is a number (three) an a unit (apple). The meaning of the unit depends on what you're using it to measure. For example, if apples are \$1 each, then you could use an apple as a unit of currency. The way to do this is to simply *multiply by one*:

$$(3 \text{ apples}) \times \left(\frac{\$\, 1}{\text{apple}} \right) = \$3 \,. \tag{3.1}$$

We have used the fact that the exchange rate is simply the statement that

$$1 \text{ apple} = \$1 \qquad \Longrightarrow \qquad 1 = \frac{\$1}{1 \text{ apple}} \ . \qquad (3.2)$$

You can do a similar thing for [kilo-]calories or any other conversion rate.

All that matters is that the conversion factor is a constant. The constants of nature make very good 'exchange rates.' For example, high-energy physicists use **natural units**:

$$\hbar = c = 1 \ . \qquad (3.3)$$

At face value, this does not make sense. $\hbar$ has units of action, c is a speed, and 1 is dimensionless. In more conventional units,[1]

$$c = 3 \times 10^{10} \text{ cm/s} \qquad\qquad \hbar = 10^{-34} \text{ kg m}^2/s \ . \qquad (3.4)$$

However, because nature gives us a *fundamental* unit of action and a *fundamental* unit of speed, we may use them as conversion factors (exchange rates). If $c = 1$, then

$$1 \text{ s} = 3 \times 10^{10} \text{ cm} \ . \qquad (3.5)$$

This connects a unit of time to a unit of distance. By measuring time, the constant c automatically gives an associated distance. The physical relevance of the distance is tied to the nature of the fundamental constant: one second (or 'light-second') is the distance that a photon travels in one second. Observe that this only works because c is a constant.

3.2   Quantifying units

We use the notation that a physical quantity Q has **dimension** $[Q]$ that can be expressed in terms of units of length, mass, and time:

$$[Q] = L^a M^b T^c \ . \qquad (3.6)$$

[1]For the most part, we are happy with one significant figure in this course.

The dimension is the statement of the powers a, b, and c. You may want to also include units of, say, electric charge. Sticklers may pontificate about whether electric charge formally carries a new unit or not.

Example 3.2.1. *What are the units of force? We remember that $F = ma$, so*

$$[F] = [m][a] = M \times LT^{-2} = L^1 M^1 T^{-2} \, . \tag{3.7}$$

Exercise 3.2.1. *What are the units of the fine structure constant?*

When working in **natural units**, $c = 1$ means that units of length and time are the same and $\hbar = 1$ means that units of time and energy (mass) are inversely related. In natural units, one simply writes $[Q]$ to mean the mass-dimension of a quantity. To revert back to conventional units, one simply multiplies by appropriate factors of $1 = c$ and $1 = \hbar$.

Example 3.2.2. *What are the units of force in natural units? From* (3.7) *we multiply by one to convert length and time into mass dimensions:*

$$[F] = [c^{-3}\hbar F] = M^2 \, . \tag{3.8}$$

In natural units we say $[F] = 2$. Recall that energy and mass have the same dimension, which you may recall from the Einstein relation $E^2 = m^2 c^4 + p^2 c^2$.

3.3 *Dimensional analysis at work*

Sanity Check

The simplest use of dimensional analysis is to check your work. The following expression is obviously wrong:

$$1 + (3 \text{ cm}) \, . \tag{3.9}$$

This does not make sense. You cannot sum terms with different dimensions. Similarly, $\sin(3 \text{ cm})$ does not make sense. What about $e^{5 \text{ cm}}$? This doesn't make sense because

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots \tag{3.10}$$

Since each term comes with a different power of x, the argument of the exponential must be dimensionless.

Example 3.3.1. *As pointed out by Matta et al.[1], this argument is not quite correct. Each term in the Taylor expansion of a function $f(x)$ maintains the dimensions of $f(x)$, as is obvious when written out carefully:*

$$f(x_0 + \Delta x) = f(x_0) + \left.\frac{df}{dx}\right|_{x_0} \Delta x + \frac{1}{2}\left.\frac{d^2f}{dx^2}\right|_{x_0} \Delta x^2 + \cdots . \tag{3.11}$$

The units of every dx^2 in the 'denominator' of $d^{(n)}f/dx^n$ is canceled by the units in Δx^n, no matter what the dimensions of Δx are. The real issue is that for many functions, $f(x_0)$ is simply not defined for dimensionful arguments. This is certainly true for trigonometric functions. For the exponential, one may fall back to the limit definition:

$$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n , \tag{3.12}$$

where it is now an issue of different terms having different dimensions. Note that the right-hand side is not a Taylor expansion. The exponential definition above is handy because it makes sense even when x is a matrix or operator.

Exercise 3.3.1. *Sometimes you may think it is useful to keep track of radians (or degrees) as a dimensionful quantity. This, by the way, is a slippery slope because then you may want to think of π as some unit of circles... whatever that means. Following the exercise above, show that (1) each term in the Taylor expansion of $f(x) = \sin(x)$ has the same dimensions, and (2) that there is no issue with trigonometric functions being defined as having 'dimensionful' arguments in this way.*

Exercise 3.3.2. *Consider the energy spectrum of light emitted from some constant source—a distant star, the ongoing annihilation of dark matter in the galactic center, or a high-intensity laser. The spectrum encodes how many photons are emitted per unit time. We can plot this spectrum as a curve on a graph. We can even normalize the curve so that it integrates to one photon. This means we only care about the distribution of energy, not the absolute amount. The horizontal axis of such a plot is the photon energy. What are the units of the vertical axis?*

Solving problems

Here is a common problem in introductory physics. Assume you have a pendu-
lum with some sufficiently small initial displacement θ_0. What's the period, τ of
the pendulum? We draw a picture like Fig 3.1.

From dimensional analysis, we know that the period has dimensions of time,
$[\tau] = T$. The problem gives us a length $[\ell] = L$ and the gravitational accelera-
tion, $[g] = LT^{-2}$. Note that $[\theta_0] = 1$ is dimensionless. This means that the only
way to form a quantity with dimensions of time is to use $g^{-1/2}$. This leaves us
with a leftover $L^{-1/2}$, which we can fix by inserting a square root of ℓ:

Figure 3.1.
 Sketch of a simple pendu-
lum.

$$\tau \sim g^{-1/2} \ell^{1/2} \ . \tag{3.13}$$

If we want to be fancy, we can make this an equal sign by writing a function of
the other dimensionless quantities in the problem:

$$\tau = f(\theta_0) \sqrt{\frac{\ell}{g}} \ . \tag{3.14}$$

[**Flip**: To do: include problems from R.W. Robinett *American Journal of Physics* **8**3, 353 (2015);
https://doi.org/10.1119/1.4902882.]

Scaling

A key theme in physics is scaling relations. We present a somewhat contrived
example of how this works adapted from section 11 of V. I. Arnold's *Mathe-
matical Methods of Classical Mechanics*.[1]. Suppose you have some static, central
potential $U(r)$. Maybe it's some planet orbiting a star.

[1]This is one of my favorite
differential geometry textbooks
because it is disguised as a
book on mechanics.

Figure 3.2.
 A orbital trajectory, $r_0(t)$.

The force law gives:

$$m\ddot{r} = -\frac{\partial U}{\partial r} \ . \tag{3.15}$$

Suppose we are given a solution, $r_0(t)$. Perhaps this is a trajectory that is exper- imentally verified. Dimensional analysis gives us a way to scale this solution into other solutions. For example, let us scale time by defining a new variable t':

$$t \equiv \alpha t' . \tag{3.16}$$

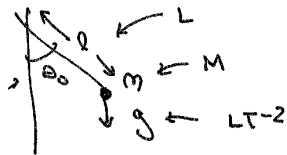Because the potential is static, then only the left-hand side of the force law changes. Even though the right-hand side formally has dimensions of time, T^{-2}, it does not transform because those units are carried in a constant, perhaps G_N, not a $(d/dt)^2$ like the left-hand side. The left-hand side of the force law gives:

$$m \left(\frac{d}{dt} \right)^2 r_0(t) = m\alpha^{-2} \left(\frac{d}{dt'} \right)^2 r_0(\alpha t') . \tag{3.17}$$

This begs us to define a new mass $m' = m\alpha^{-2}$ so that

$$m' \left(\frac{d}{dt'} \right)^2 r_0(\alpha t') = -\frac{\partial U}{\partial r_0} . \tag{3.18}$$

What this tells us is that we may define a new trajectory, $r_1(t') \equiv r_0(\alpha t')$, which is a solution in the same potential that traces the same trajectory but at α times the speed and with mass m'. Changing labels $t' \rightarrow t$ for a direct comparison:

$$m' \left(\frac{d}{dt} \right)^2 r_1(t) = -\frac{\partial U}{\partial r_1} , \tag{3.19}$$

which is indeed[1] (3.15) with a new mass m' and a trajectory $r_1(t') \equiv r_0(\alpha t')$. For example, if $\alpha = 2$, then $r_1(t)$ traces the same trajectory at double the velocity with one fourth of the mass.

Exercise 3.3.3. *I missed something in the example above. In order for a planet of mass m' to have trajectory $r_1(t')$, what is the mass of the star compared to the original mass $M_\star$?*[2]

Example 3.3.2. *Business-y people like to quantify effort using words like 'person– hour' or 'person–years.' This is the idea that a 10 person–hour task would take 10 people one hour to complete, or one person 10 hours to complete, or 5 people two hours to complete, etc. As you can see, this choice of units implies that effort has a linear scaling in both the number of people and the amount of time needed. Any-*

[1]We were able to swap r_0 with r_1 simply because U only de- pends on the position.

[2]Thanks to Eric Zhang (2021) for pointing this out.

one who has worked on a group project knows that this linear scaling is bullshit.
Frederick Brooks reflects on this in the 1974 essay, "Myth of the Man–Month."[10]

[10]Frederick P. Brooks. *The myth-ical man-month – Essays on Software-Engineering*. Addison-Wesley, 1975.

Error Estimates

This section is based on a lovely *American Journal of Physics* article by Craig
Bohren.[11] Let us go back to another high school physics problem: we drop a ball
of mass m from height h. See Fig. 3.3. The task is to find the time t_0 for the ball
to hit the ground.

[11]Craig F. Bohren. "Dimensional analysis, falling bodies, and the fine art of not solving differential equations". In: *American Journal of Physics* 72.4 (2004), pp. 534–537. DOI: 10.1119/1.1574042. URL: https://doi.org/10.1119/1.1574042.

You can integrate the force equation to get

$$t_0 = \sqrt{\frac{2h}{g}} \ . \tag{3.20}$$

This is the *exact* answer *within our model* of the system. The model made sev-
eral assumptions: the mass is a point mass, the gravitational acceleration is
constant at all positions, there is no air resistance, etc. In fact, we *know* that if
we do an experiment, our result will almost certainly *not* be t_0. All we know is
that t_0 is probably a good approximation of the actual answer. What we would
like to to know is: *how good of an approximation is it?*
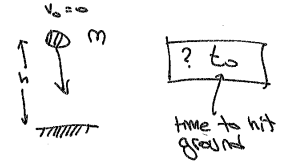
Figure 3.3.
Dropping a ball of mass m.

One way to check this is to do the next-to-leading order (NLO) calculation,
taking into account a more realistic model and then compare to t_0. Of course,
"more realistic" is also code for "more complicated." Take a moment to appreci-
ate that doing this is *stupid*. Why do we need to do a *hard* calculation to justify
doing an *easy* one? If we are going to do the hard calculation anyway, what was
the point of ever doing the easy one?

What we really want is an error *estimate*. The error is

$$\varepsilon = \frac{t_1 - t_0}{t_0} \ . \tag{3.21}$$

This is a dimensionless quantity that determines how far off t_0 is from a more
realistic calculation, t_1. Ideally we should not actually have to do much work to
estimate t_1.

Let us assume that we are not completely nuts and that we are in a regime
where the error is small[1]. Then the error is a function of some dimensionless
parameters, ξ, in the system. We define these ξ so that as $\xi \to 0$, $\varepsilon(\xi) \to 0$. In

other words, the approximation gets better as the ξ are made smaller. By Taylor expansion:

$$\varepsilon(\xi) = \varepsilon(0) + \varepsilon'(0)\xi + \mathcal{O}(\xi^2) . \tag{3.22}$$

By assumption, $\varepsilon(0) = 0$ and $\mathcal{O}(\xi^2)$ is small. We can then make a reasonable *assumption* that the dimensionless value $\varepsilon'(0)$ is $\mathcal{O}(1)$. This tells us that the error goes like $\varepsilon(\xi) \sim \xi$.

By the way $\mathcal{O}(1)$ is read "order one" and is fancy notation for the order of magnitude. Numbers like 0.6, 2, and π are all $\mathcal{O}(1)$. A number like 4π, on the other hand, is $\mathcal{O}(10)$. The assumption that a dimensionless number is $\mathcal{O}(1)$ is reasonable. When nature gives you a dimensionless parameter that is both (a) important and (b) very different from $\mathcal{O}(1)$, then there's a good chance that it's trying to tell you something about your model. Good examples of this are the cosmological constant, the strong CP phase, and the electroweak hierarchy problem.[1]

Here is how it works in practice. One effect that we miss in our toy calculation of t_0 is that the earth is round with radius R. This means that assuming a constant g is an approximation. We have two choices for a dimensionless parameter ξ:

$$\xi = \frac{h}{R} \qquad \text{or} \qquad \xi = \frac{R}{h} . \tag{3.23}$$

There is an obvious choice: $\xi = h/R$, because we know that as h is made smaller (drop the ball closer to the ground) or R becomes bigger (larger radius of Earth) then the constant g approximation gets better. We thus expect that the corrections from the position-dependence of g go like $\mathcal{O}(h/R)$.

Exercise 3.3.4. *Check by explicit calculation that the correction to the constant g approximation is linear in h/R. Start by writing the force law for a point source of at distance r = R + h from the center of the Earth. Taylor expand to find a second order differential equation that is difficult to solve:*

$$\ddot{h} = \frac{-g}{\left(1 + \frac{h}{R}\right)^2} . \tag{3.24}$$

[1] There are also 'bad' examples. The ratio of the angular size of the moon to the angular size of the sun is unity to very good approximation. This is quite certainly a coincidence. Our universe appears to be in an epoch where the density of matter, radiation, and dark energy all happen to be in the same ballpark. Our cosmological models imply that this is purely a coincidence. It would be very curious if this were not the case. As an exercise, you can critically explore the use of the anthropic principle in physics.

Taylor expand to reduce this to an equation of the form

$$\frac{d^2q}{ds^2} = -1 + 2q ,$$
(3.25)

Here we define the natural dimensionless variables, $q = h/R$ and $s = (g/R)^{1/2}\,t$. If the choice of s is not obvious, please do everything in terms of t and then observe that one can conveniently absorb a factor of g/R into dimensionless time variables.[1] Plug the dimensionless differential equation into Mathematica or your favorite symbolic solver to obtain

[1]You should find an equation of the form $\ddot{q} = -(g/R)(1 - \cdots)$.

$$q(s) = c_1 e^{\sqrt{2}s} + c_2 e^{-\sqrt{2}s} + \frac{1}{2} .$$
(3.26)

Argue that the initial condition $\dot{h}(t)\big|_{t=2} = 0$ implies that the coefficients satisfy $c_1 = c_2$ so that you can combine the exponentials into a hyperbolic cosine. Show that one obtains:

$$\frac{2q(s) - 1}{2q(0) - 1} = \cosh\left(\sqrt{2}s\right) .$$
(3.27)

Argue why you can Taylor expand the right-hand side about small argument; that is, explain why $s \ll 1$. (Hint: use $h \ll R$.) Perform the Taylor expansion of the hyperbolic cosine to find that the leading correction to the fall time is

$$s_1 = \frac{2q_0}{1 - 2q_0} .$$
(3.28)

The zeroth order approximation was $s_0 = (g/R)^{1/2} t_0 = \sqrt{2q_0}$. Calculate $(s_1 - s_0)/s_0$ to confirm that this is $\mathcal{O}(h/R)$.

Bonus: Allometry

There is a fun topic called **allometry**. This is basically dimensional analysis applied to biology. A typical example is to consider two people who have roughly the same shape but different characteristic lengths, ℓ and L, Fig. 3.4.
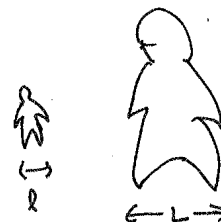
Exercise 3.3.5. *If both people exercised at the same rate, which one loses more absolute weight? By how much? Let us assume that weight loss is primarily from the conversion of organic molecules into carbon dioxide.*

Figure 3.4.
Two mathematically similar people.

Exercise 3.3.6. *David Hu won his first IgNobel prize for determining that mammals take about 21 seconds to urinate, largely independently of their size[1]. Can you use dimensional analysis to argue why this would be the case? It may be helpful to refer to the paper[12]. As you read it, figure out which terms are negligible (and in what limits), identify the assumptions of the mathematical model (scaling of the bladder and urethra), and prove the approximate scaling relation. Make a note to yourself of which steps were non-trivial and where one may have naively mis-modeled the system. By the way, David Hu won a second IgNobel prize for understanding wombats' cubical poop.*

[1]I learned about this in his excellent popular science book, *How To Walk on Water and Climb Up Walls*.

[12]Patricia J. Yang et al. "Duration of urination does not change with body size". In: *Proceedings of the National Academy of Sciences* 111.33 (2014), pp. 11932–11937. DOI: 10.1073/pnas.1402289111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1402289111.

The above exercise on mammalian urination is a good example of *modeling*. As physicists, we must identify and make a mathematical model for the most salient features of a problem. We must also be able to quantify the error from neglecting sub-leading contributions. As a rough model for scaling purposes, we can ignore viscosity and surface tension effects on human-sized mammals. For much smaller mammals, these effects become larger—the authors of the study note that mice tend to urinate droplets—in which case one can ignore the 'inertial' $\frac{1}{2}\rho v^2$ term in Bernoulli's equation. For human-sized mammals, we may assume that steady state urination is given by Bernoulli's equation:

$$P + \rho g h = \frac{1}{2}\rho v^2 \, , \tag{3.29}$$

where P is the pressure from the bladder, h is the column height of the urethra, ρ is the mass density of urine, and v is the velocity of the urine at the end of the urethra. Let us simplify to the condition where urination is purely driven by gravity—that is, the bladder does not exert any additional pressure, $P = 0$. You can now show that the total urination time scales like the mass of the mammal to the one-sixth power, $\tau \sim M^{1/6}$. That is, the urination time has a very weak scaling dependence on how massive the mammal is.

Exercise 3.3.7. *In August 2021, Ezra Klein interviewed Dr. Céline Goudner about the COVID-19 variant.[13] In the interview, Klein cited the statement that the Delta variant has $\mathcal{O}(1000)$ times the viral load than prior COVID strains. Goudner then interprets this in the following way: if the CDC defined 'close contact' for prior strains as 15 minutes of being indoors with an infected invdividual without a mask, then the equivalent 'close contact' time for the Delta variant is around* one second. *What scaling assumptions go into that estimate? Some of these assumptions are*

[13]Ezra Klein. *Transcript: Ezra Klein interviews Céline Gounder.* Aug. 2021. URL: https://www.nytimes.com/2021/08/06/podcasts/transcript-ezra-klein-interviews-celine-gounder.html.

not obvious to me: for example, parts of the respiratory have a fractal-like structure that would lead me to suspect fractal scaling dimensions for surface area. Remark: *Just because you know dimensional analysis, that does not make you a medical, healthcare, or public policy expert.*[1]

The following exercises draw from an article by Nicole Meyer-Verneta and Jean-Pierre Rospars in the American Journal of Physics[14] and the references therein.

Exercise 3.3.8. *Estimate the expected velocity of an All Terrain Armored Transport (AT-AT)[2] of characteristic height L. You can assume that the walking behavior is based on a pendulum.* Answer: $v \sim \sqrt{Lg}/2\pi$.

Exercise 3.3.9. *Based on the density ρ, the force-per-cross-sectional area σ, and the maximum rate of energy consumption per unit mass b, one may estimate the 'sprint' velocity of an animal of length L. This sprint velocity is conveniently described with respect to the dimensionless 'body lengths per time,' v_{spr}/L.*

Remarkably, for over 20 orders of magnitude in animal length L, the value of v_{spr}/L is within an order of magnitude of 10/sec:

Nicole Meyer-Vernet and Jean-Pierre Rospars
doi.org/10.1119/1.4917310
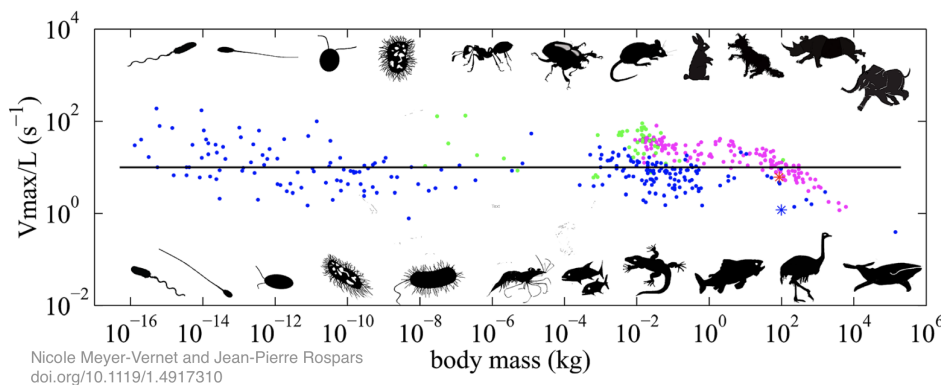
Argue from dimensional analysis that $v_{\mathrm{spr}}/L \sim b\rho/\sigma$. (This is the easy part.) It turns out that there are simple physical principles for each of these terms to be roughly constant for all life on Earth (this is the more subtle part); see the article for a discussion.

Exercise 3.3.10. *The height of trees. How does the maximum height of a tree, L scale with the diameter of its cross section, d? For an argument that $L \sim d^{3/2}$, see*

Figure 3.5:
Image from Meyer-Verneta and Rospars. [14]

[1] https://www.jstor.org/stable/24949846

[15] Robert D. Howe and Richard E. Kronauer. "Thomas McMahon: A Dedication in Memoriam". In: *Annual Review of Biomedical Engineering* 3.1 (2001). PMID: 11712277, pp. xv–xxxix. DOI: 10.1146/annurev.bioeng.3.1.0. URL: https://doi.org/10.1146/annurev.bioeng.3.1.0.

Thomas McMahon's article "The Mechanical Design of Trees" in Scientific American *volume 233 (1975)*[1]. *McMahon was the first to propose a physical explanation for the observed scaling law that the metabolic rate of an animal scales like the characteristic size to the 3/4 power. A nice bibliography of his work can be found in* Annual Review of Biomedical Engineering.[15]

LINEAR ALGEBRA

4 *Finite-Dimensional Linear Algebra*

4.1 *Yet another review of linear algebra*

Linear algebra is part of our physics DNA. So why should we patronize ourselves with yet another review of linear algebra? We want to understand Green's functions as a matrix inverse. The 'matrix' in question is the differential operator $\mathcal{O}$ in (1.1). The identification boils down to the following:

$$\text{differential operator} = \infty\text{-dimensional matrix} . \tag{4.1}$$

This is a poetic equal sign; for example, not every infinite dimensional matrix is a differential operator.[1] You may know matrices as a block of numbers that act on columns of numbers—*vectors*—to produce another vector. If differential operators are matrices, are vectors on which they act? These matrices act on a space of functions, which turns out to be a vector space:

$$\text{function space} = \infty\text{-dimensional vector space} . \tag{4.2}$$

Again, the equal sign is poetic. Do not be intimidated by terminology like *function space*; this is just an abstract place where functions live. Just recall back to your intuition from 3D Euclidean vector space, $\mathbb{R}^3$: any 3-vector $\boldsymbol{v}$ lives in the vector space $\mathbb{R}^3$. If we transform $\boldsymbol{v}$ by a linear transformation A, you get a new vector $\boldsymbol{w} = A\boldsymbol{v} \in \mathbb{R}^3$ that is also in the vector space.

Weird things can happen when we extend our intuition from finite things to infinite things[2], but for this course we try to draw as much intuition as we can from finite dimensional linear algebra to apply it to infinite dimensional function spaces.

[1] But by causality and locality these are the ones we care about in physics.

[2] For example, the Hilbert Hotel puzzle.

4.2 *What is Linear?*

A function, $f(x)$, takes some kind of input x and produces some kind of output. When the inputs and outputs are real numbers, $x, f(x) \in \mathbb{R}$, then we can plot this relation between inputs and outputs on the $\mathbb{R}^2$ by the mapping $(x, f(x))$.

The curve $f(x)$ is the set of all points $(x, y) \in \mathbb{R}^2$ such that $y = f(x)$. Said in yet another way, $y = f(x)$ is a constraint on $\mathbb{R}^2$ that slices it into a one-dimensional subspace. To emphasize that f takes in real numbers and spits out real numbers, we can write $f : \mathbb{R} \to \mathbb{R}$.

Before we end up getting too pedantic about what a curve is, recall that even children can tell you that the equation $f(x) = mx + b$ defines a line in the two dimensional plane. In this relation, m is the slope and b is the intercept of the line with the y-axis. We must be more restrictive: we set $b = 0$ and impose that a linear relation between two numbers x and $f(x)$ takes the form $f(x) = mx$.

The reason for our apparent pedantry is to generalize the definition of linearity to extend beyond the picture of curves on $\mathbb{R}^2$. A relation $f(x)$ is linear if the following are true:

$$f(\alpha x) = \alpha f(x) \tag{4.3}$$

$$f(x + y) = f(x) + f(y) \, . \tag{4.4}$$

We assume that x and y are two objects of the same type, and α is simply some number.

Exercise 4.2.1. *Confirm that $f(x) = mx$ is a linear function between real numbers for any value of $m \in \mathbb{R}$.*

We can state this all more formally. Suppose there is some collection of objects that we call V. Let x and y be two such objects: $x, y \in V$. Further, let α and β be two numbers: $\alpha, \beta \in \mathbb{R}$. Then a function $f : V \to W$ is **linear** if and only if[1]

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{4.5}$$

Let us dissect this a bit. We stated that x and y have to be the same type of object, members of the class V. This echoes our discussion of dimensional analysis: if x are apples and y are Pokemon, then $2.5x + 7y$ is nonsensical and so any function of such an object is nonsensical. We defined α and β to be numbers[2]: these just count how many objects in V are being fed into our function. For now we assume that these are real numbers, but we will soon generalize to complex numbers.

We wrote $f : V \rightarrow W$, which means that the output of the function $f(x)$ is an object of class W. In our toy example $f(x) = mx$, $V = W = \mathbb{R}$. In general, W and V could be two totally different classes of objects. W could be real numbers, something with units, a mathematical object with more structure, or any other class of objects. No matter what V and W are, the relation (4.5) tells us when a function is linear.

Example 4.2.1. *Let f be a function that maps angles θ to rotation matrices,*

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} . \tag{4.6}$$

This map is not linear because

$$R(\theta_1 + \theta_2) \neq R(\theta_1) + R(\theta_2) , \tag{4.7}$$

as you can check for the trivial case of $\theta_1 = \theta_2 = 0$.

Exercise 4.2.2. *Let f be a function that maps driving speed to the probability of being pulled over by the police or highway patrol. Explain why f cannot be linear.*

We can diagnose the linearity of a function, even when that function inputs and outputs objects that are more general than simply numbers. Thus far, we have been rather coy what it means to be an "object in class V." This leads us to the notion of a vector space.

4.3   *Vectors and Vector Spaces*

For our purposes in this course, we consider functions of **vectors**. Not all objects of interest are vectors, for example, coordinates are decidedly not vectors. The term *position vector* is thus a faux pas.[1] On the contrary, infinitesimal differences of coordinates are vectors called **tangent vectors**. So what are vectors and vector spaces?

[1]Mathematicians chuckle at freshman physics textbooks.

We focus on an imprecise, but physically intuitive working definition. For those who prefer a more mathematical discussion that is still tied to physical intuition, see *Geometry, Topology, and Physics* by Nakahara.[16] A **vector** is an element of a vector space. That is, a **vector space** is the 'complete' collection of a given type of vector.

[16]Mikio Nakahara. *Geometry, topology and physics*. Graduate student series in physics. Bristol: Hilger, 1990.

We have already assumed that you can add vectors and rescale them by numbers; that is, we can take **linear combinations** of vectors. The assumption is implicit in our definition of linearity, (4.5). Let v and w be any two vectors in the vector space V; we write this as $v, w \in V$. Further, let α, β be numbers. We have the following rules:

⬦ The sum $\alpha v + \beta w$ is a vector in V.

⬦ The zero vector $\mathbf{0}$ is a vector in V such that $\mathbf{0} + v = v$.

[1]Please refer to your favorite linear algebra book.

Formally there are other rules, but you probably already assumed them.[1] These include the fact that vector addition is commutative, $v + w = w + v$, and associative, $(v + w) + u = v + (w + u)$.

Example 4.3.1. *The existence of a meaningful zero vector is one example why 'position space' or coordinate space is not a vector space. The zero vector $\mathbf{0}$ is the one that leaves other vectors unchanged upon addition: $\mathbf{0} + x = x$. Clearly this depends on the coordinate system and violates our intuition that physics should be independent of the choice of coordinates. Further, the notion of 'adding positions' is physically nonsensical. On the contrary, differences of positions $\Delta x = x_1 - x_2$ are* meaningful. *For example, the force between two non-relativistic point particles depends the vector that is the difference between the particle positions.*

Exercise 4.3.1. Color Space. *Colors for digital media have many representations. One popular representation is* RGB *where the intensity of red, green, and blue components are specified. For example, $(25, 174, 40)$ is a pleasant shade of green. This triplet of numbers certainly looks like a vector: it is an ordered collection of numbers. Is the space of* RGB *colors a vector space?* Answer: no. Make a list of reasons why. *Like all meaningful questions, this rabbit hole runs deep; for example, see articles on the geometry of colors[17] or the overlapping sciences of physics and neurological perception of color[18].*

[17]Joseph W Weinberg. "The geometry of colors". In: *General relativity and gravitation* 7.1 (1976), pp. 135–169.

[18]Alexander Logvinenko and Vladimir Levin. *Foundations of Colour Science: From Colorimetry to Perception.* Wiley, 2022. ISBN: 978-1-119-88591-7.

[2]As a student I was once very because a textbook introduced the notation $\vec{\vec{\partial}}$. Only much later would I come to appreciate the perspicacity of that symbol.

4.4   *Notation for Vectors*

Physicists use a few different notations for vectors depending on the context. In these notes we also use whichever notation is most convenient for the task and we are not ashamed to change notation as needed. Students should be nimble to be able to make use of different notations.[2] It is absolutely critical that even though we can refer to a mathematical object with different notation,

the underlying idea is the same. This insight is the key to never getting lost in the forest of special functions (Bessel, Legendre, spherical harmonics, etc.) that appear in physics.

Bold/over-arrow notation

Thus far we refer to vectors with a boldfaced and italicized symbol,[1] $\boldsymbol{v}$. For the author this is a comfortable and familiar notation from school. On the board in a classroom, we sometimes write the vector with an underline $\underline{v}$ since boldfaced can be difficult to write freehand. A related notation uses arrows, $\vec{v}$.

[1] See ISO 80000-2:2019 from the International Organization for Standardization.

Ket notation

Another popular notation in physics is the bra–ket notation. Vectors are kets, $|v\rangle$. We we review below, there are also 'row vectors' that we call bras, $\langle w|$. This is all a tongue-in-cheek joke because you can combine a bra and a ket to form a *braket*, $\langle w | v \rangle$, which we identify as the inner product (dot product) of two vectors $\langle w, v \rangle = \boldsymbol{w} \cdot \boldsymbol{v}$.

Index notation

A common metaphorical extension[2] in physics is the notation v^i. Technically, v^i refers to the ith component of the vector $\boldsymbol{v}$.[3] However, since physicists often write their equations component-wise, we often slip into the shorthand of using v^i to mean either a given component or the entire vector. The appropriate meaning is usually clear from context.

[2] `https://en.wikipedia.org/wiki/Metaphorical_extension`

[3] i^{th} component with respect to what? See the following section on basis vectors.

Unadorned notation

We identify functions as vectors in an infinite dimensional space. In this case, it is often sufficient to avoid any additional adornment, so we write a function as f. The analog of a component of a vector, v^i, is the function at a point, $f(x)$.

4.5 *Basis Vectors*

The number of vectors in a vector space is formally infinite. If $\boldsymbol{v}$ is a vector, then so is $2v$ and $3v$, not to mention $0.999\boldsymbol{v}$, ad nauseum. Fortunately we can pick a standard set of reference vectors and define any other vector with respect

[1]Please humor the unusual
notation with a lower index in
parenthesis. We explain this
unusual choice in Section ??.

to those reference vectors. We call those reference vectors **basis vectors** $\hat{e}_{(i)}$ and the entire set a **basis** for the vector space.[1] We may express any vector v in the vector space as a linear combination of basis vectors:

$$v = \sum_i v^i \hat{e}_{(i)} \, . \tag{4.8}$$

Example 4.5.1. *In two dimensional Euclidean space,* $\mathbb{R}^2$ *we have a canonical basis*

$$\hat{e}_{(1)} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad \hat{e}_{(2)} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} \, . \tag{4.9}$$

Canonical *is just a fancy way to say* the obvious choice. *The first basis vector points in the x direction, the cond points in the y direction. We can thus represent a vector v as a linear combination,*

$$v = \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = v^1 \hat{e}_{(1)} + v^2 \hat{e}_{(2)} \, . \tag{4.10}$$

We could have defined a different basis, for example

$$\hat{e}'_{(1)} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad\qquad \hat{e}'_{(2)} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \, . \tag{4.11}$$

In this primed basis the vector v would have different components,

$$v = v'^1 \hat{e}'_{(1)} + v'^2 \hat{e}'_{(2)} \, , \tag{4.12}$$

where clearly $v'^i \neq v^i$.

Exercise 4.5.1. *Suppose $v^1 = 3$ and $v^2 = -4.5$ in the unprimed basis from the example above. Find the corresponding primed components v'^1 and v'^2. You do not have to do this in any fancy systematic way.* Suggestion: *just draw the vectors on $\mathbb{R}^2$. You do not often get to do this, especially with more abstract vectors. But when you can do something the simple way, you should do it that way and then think about how to generalize 'the simple way.'*

The number of basis vectors required to describe any vector is called the **dimension** of the vector space. The dimension of $\mathbb{R}^2$ is two. If you specify fewer basis vectors than the dimension of the vector space, then there are vectors that you cannot describe. If you specify more basis vectors than the dimension of the vector space, then there is not a unique way to specify the vector components.[1]

Example 4.5.2. *In $\mathbb{R}^2$, suppose we specified* three *basis vectors,*

$$\hat{e}_{(1)} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{e}_{(2)} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \hat{e}_{(3)} \equiv \frac{-1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \qquad (4.13)$$

The vector $v = \hat{e}_{(1)} + \hat{e}_{(2)}$ can equivalently be written as $v = -\sqrt{2}e_{(3)}$.

Exercise 4.5.2. *In the example above, write v with respect to the three basis vectors in a way that has not yet been specified. Repeat this exercise until it is obvious that there are an infinite number of ways of writing v with respect to the three basis vectors. Contrast this to the case where we restrict to any pair of the basis vectors, in which case the components of v are unique.*

4.6   *Nice basis vectors*

You have likely been trained to *assume* that a basis is *nice*, following the notion of assumed *niceness*[2] in physics. As we move towards abstract vector spaces, it is worth explicitly stating these assumptions so that we can make a note of where they may break and what other mathematical structure we need to define them. The two assumptions are linear independence and orthonormality.

Linear independence

Two vectors v and w are **linearly independent** if they are not proportional to each other: $v \neq \alpha w$ for any number α. The basis vectors for any reasonable basis are linearly independent—any given basis vector $e_{(i)}$ cannot be written as a linear combination of the other basis vectors:

$$\hat{e}_{(i)} \neq \sum_{i \neq j} \hat{e}_{(j)}, \qquad (4.14)$$

where the sum is over all basis vectors $e_{(j)}$ except the ith basis vector. It should be obvious[3] that a proposed basis with a linearly dependent basis vector

[1] *Then shalt thou count to three, no more, no less. Three shall be the number thou shalt count, and the number of the counting shall be three. Four shalt thou not count, neither count thou two, excepting that thou then proceed to three. Five is right out. (Monty Python & The Holy Grail, 1975)*

[2] Section 1.3.

[3] In the sense of Section 1.4.

1. Does not uniquely define the components of some vectors $\boldsymbol{v}$ in the vector space.

2. Either has more basis vectors than the dimension of the space, or there are vectors in the space that cannot be described by the basis.

Thus every basis vector in any reasonable basis is linearly independent from the other basis vectors.

Orthonormality

Okay, this is actually two different conditions: orthogonal and normal. The basis vectors of a nice basis are *orthogonal* to every other basis vector and *normalized* to have unit length. **Orthogonal** means that the vectors are perpendicular.

Eh... then what do we mean by *perpendicular*? We certainly have some notion of two directions being perpendicular from everyday life: north is perpendicular to west because if we move five steps north we are stationary in the east–west direction. But how do we define this mathematically? In fact, while we are at it: how do we define 'unit length' with respect to these vectors?[1]

[1]From a physics perspective: in what units?

Example 4.6.1. *Orthogonality and linear independence are related but are not the same. Orthogonal vectors are linearly independent, but linear independence does not imply orthogonality. For example, consider the basis of $\mathbb{R}^2$:*

$$\hat{\boldsymbol{e}}_{(1)} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad \hat{\boldsymbol{e}}_{(2)} \equiv \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \qquad\qquad (4.15)$$

These vectors are obviously linearly independent. Assuming the usual Euclidean inner product, the are also not *orthogonal. The observant student will notice that linear independence does not require one to define an inner product, whereas orthogonality is only defined with respect to some inner product.*

Exercise 4.6.1. *Let $\boldsymbol{v}$ be the vector that is pointing in the $\hat{\boldsymbol{x}}$ direction of $\mathbb{R}^2$. What are the components of $\boldsymbol{v}$ with respect to the basis in (4.15)? Similarly, let $\boldsymbol{w}$ be the vector that is pointing in the $\hat{\boldsymbol{y}}$ direction of $\mathbb{R}^2$. What are the components of $\boldsymbol{w}$ with respect to the basis in (4.15)?*

Unlike linear independence, orthonormality is not a strictly necessary condition for having a basis—though it is hard to imagine a scenario where one would *not* use an orthonormal basis. However, the notions of orthogonality and

normalization depend on *additional* mathematical structure that we have to impose/assume/invent for our vector space. Formally, we say that we promote the vector space to a **metric space**. The additional structure that we define is a machine that takes two vectors and tells us something about the 'distance' between them. We call this machine the **metric**, **inner product**, or **dot product**; each phrase refers to the same thing. Once we have a metric, the Gram–Schmidt procedure assures us that we can construct an orthonormal basis from a linearly independent basis, see Exercise 4.8.2.

At this point we should take a deep breath and state explicitly that we've been assuming an orthonormal basis. In this course we will continue to use an orthonormal basis. You may object to this and say that you used to believe in orthonormal bases until you were forced to write down the gradient (or worse, the Laplacian) in spherical coordinates.

Exercise 4.6.2. *Are polar coordinates an orthonormal basis? You can generalize this to spherical or cylindrical coordinates. Take a moment to think about this. Do vectors in polar coordinates even make sense as vectors? Do we have a sensible addition rule? These questions are unfair because they make the assumption that the vector space described by these coordinates is "the same" as the coordinate space. This seems perfectly innocuous until you have learned ab it differential geometry or general relativity. If this piques your interest, I encourage you to read more about this.*[1]

There are many things to be said about non-Cartesian ("curvilinear") coordinate systems and orthonormality. None of them are particularly edifying without a full discussion. With no apologies, I make the following [perhaps perplexing] remarks, illustrated in a figure below:

1. There is no such thing as a 'position vector.' Positions refer to some base space, whereas vectors (like differential operators) act on the tangent space at a point of that base space.

2. A given tangent space is 'nice' and has a nice orthonormal basis.

3. That basis may not be the same for neighboring tangent spaces (perhaps due to coordinates, perhaps due to intrinsic curvature).

In this course these nuances will not come up. In the rest of your life you will still have to deal with curvilinear coordinates.[2] Suffice it to say that our study of function space will be nice an orthonormal.

[1] A good reference that presents both the physical intuition and mathematical formalism is Sean Carroll's *Spacetime and Geometry*; Carroll's appendices are an excellent crash course in differential geometry.

[2] There is an excellent discussion of non-coordinate bases and curvilinear coordinates in Bernard Schutz's *A First Course in General Relativity*. See Chapter 5.6 (in the first edition), "Noncoordinate bases."
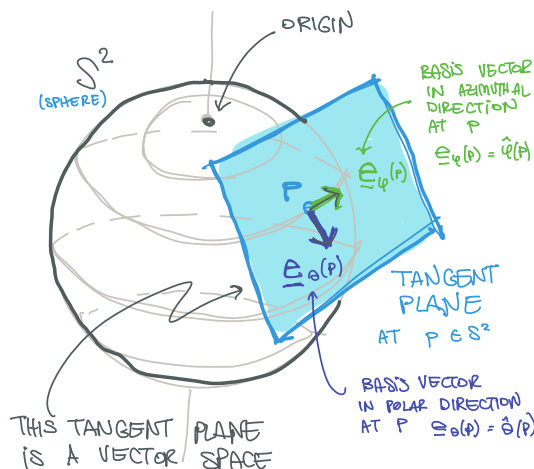
4.7   *Examples of Finite-Dimensional Vector Spaces*

Space and spacetime vectors

There are some obvious examples of vector spaces. The one you are most used to is $\mathbb{R}^3$, the 'ordinary' (Cartesian) three-dimensional space. The components of $\boldsymbol{v} \in \mathbb{R}^3$ are simply three real numbers v^i multiplying the 'obvious' basis vectors:

$$v^i \hat{\boldsymbol{e}}_{(i)} = v^1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v^2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v^3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \tag{4.16}$$

We can generalized to $\mathbb{R}^n$ with n basis vectors. From special relativity you may be familiar with four-vectors with components (p^0, p^1, p^2, p^3). We index starting with a zero for historical conventions, but it serves to indicate that the *timelike* component p^0 is special. At this level, it looks like spacetime is $\mathbb{R}^4$. An indeed, at this level that would be accurate. We will see in the next section that when we extend from a vector space to a *metric space*, there is an important distinction and spacetime is in fact $\mathbb{R}^{1,3}$.

This is a good place to remind ourselves that there is no such thing as a 'position vector.' What people really mean by a position vector is the vector between some arbitrarily chosen origin and a given point on space.

There are more abstract versions of vector spaces.

Matrices

The space of 2 × 2 Hermitian[1] matrices is a vector space with basis elements

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{4.17}$$

Convince yourself that you can form any Hermitian 2 matrix out of these four basis matrices. You may recognize the familiar Pauli matrices σ^i from quantum mechanics.

Quantum states

While we have mentioned quantum mechanics, the space of quantum mechanical states is a kind of vector space. This is why the bra-ket notation that we usually learn in quantum mechanics is equally applicable to linear algebra. In fact, the big *aha!* moment of the first time you learned quantum mechanics should have been when you realized that quantum mechanics is simply physicists doing linear algebra. As a bonus, quantum mechanics has primed us to already be comfortable with *complex* vector spaces.

There are a few caveats. In quantum mechanics two states that differ by a phase are equivalent. This means that a state is not just a given vector $|\psi\rangle$ in the complex vector space, but a *ray* where $|\psi\rangle$ is identified[2] with all rephasings, $|\psi\rangle \cong e^{i\theta}|\psi\rangle$. This identification simply means that two vectors that only differ by a phase correspond to the same quantum state: for example, they correspond to the same observable (if the states are directly observable).

Functions

This is a powerful idea hidden behind a trivial idea. Functions are vectors in a vector space that we call *function space*. I am pretty sure that mathematicians do not call it that, but we use this terminology anyway. This should be *obviously* true. Consider two functions f and g. Maybe $f(x) = x^3$ and $g(x) = x + 2$, it does not matter. Clearly you can take a linear combination of the two and the result is a function:

$$\alpha f(x) + \beta g(x) = \alpha x^3 + \beta x + 2\beta . \tag{4.18}$$

This is a totally valid function that we could call $(\alpha f + \beta g)(x)$.

Exercise 4.7.1. *What are some possible bases for function space?*

A judicious choice of basis for function space can make our lives much easier. This is a central theme of this course and justifies the beastiary of special functions that you meet in graduate school. Let us ignore the subtleties of defining a function space for now—we get to this soon enough. The following example gives a taste of what it means to work with functions as vectors.

Exercise 4.7.2. *Here is a cute two-dimensional function space that gives us a shortcut to calculate a particular* indefinite *integral. Consider a two dimensional vector space spanned by the functions*

$$|f_1\rangle = f_1(x) = e^{ax} \cos bx \qquad\qquad |f_2\rangle = f_2(x) = e^{ax} \sin bx \,, \qquad (4.19)$$

where a and b are constants. Forget orthonormality or boundary conditions for this problem. The derivative d/dx is a linear operator that acts on this space. Write down the derivative as a 2×2 matrix in the above basis, D.

Invert D in the usual way that you learned to invert 2×2 matrices during your childhood[1]. Call this matrix D^{-1}.

Now stop and think: the inverse of a derivative is an indefinite integral[2]. Thus acting with D^{-1} on the vector $|f_1\rangle$ should be understood as an integral of $f_1(x)$. Show that, indeed,

$$D^{-1}|f_1\rangle = \int dx \, e^{ax} \cos bx \,. \qquad (4.20)$$

Feel free to use Mathematica *to do the indefinite integral on the right-hand side. Pat yourself on the back if you can do it without a computer.*

More exotic examples

We saw in Example 4.3.1 that color space is almost—but not quite—a vector space. There are, however, plenty of unusual vector spaces.

One fun example is the space of Fibbonacci sequences. These are sequences of numbers $\{x_1, x_2, x_3, \cdots\}$ such that $x_n = x_{n-1} + x_{n-2}$. You generate a Fibbonacci sequence by picking some x_1 and x_2 and determining all subsequent values using the defining rule.

[1] Stuck? Here's a life pro tip: http://bfy.tw/KG2Z

[2] Ignore the constant term.

Exercise 4.7.3. *Show that the Fibbonacci sequences form a vector space. What is the dimension of this vector space? Hint: the dimension is* not infinite.

Another example that shows up in mathematical physics is something called a **root space**. This is a space spanned by a finite number of lattice basis vectors. The space is defined so that vectors are *integer* multiples of the basis vectors. Surprisingly, root space has an intimate connection to the study of continuous symmetries—the representation theory of Lie groups.

Along these lines, you could consider the vector space of n-bits, $(\mathbb{Z}_2)^n$. The bits are either zero or one and you impose modular arithmetic. You can find several other great examples by judicious Googling.[1]

[1] For example, I grabbed some of these examples from `https://math.stackexchange.com/q/5233/1032899`.

4.8 *The metric, inner product, or dot product*

The **metric** (inner/dot product product) on a real vector space V is a bilinear[2] map from $V \times V \to \mathbb{R}$.[3] This means that it is a machine that takes two vectors and returns out a number. We use the metric to measure one vector with respect to another. We further impose that the metric is symmetric: it should not matter whether we measure v with respect to w or vice versa: their 'overlap' should be the same. We use the angle bracket notation where $\langle v, w \rangle$ is the inner product of two vectors v and w.[4] Summarizing the above properties in equations:

[2] This simply means linear in each argument.

[3] We also care about *complex* vector spaces, in which case the metric is a map from $V \times V \to \mathbb{C}$.

[4] This is deliberately suggestive of the ket notation for vectors, Sec. 4.4.

$$\langle \alpha v + \beta w, u \rangle = \alpha \langle v, u \rangle + \beta \langle w, u \rangle \tag{4.21}$$

$$\langle v, \alpha w + \beta u \rangle = \alpha \langle v, w \rangle + \beta \langle v, u \rangle \tag{4.22}$$

$$\langle v, w \rangle = \langle w, v \rangle. \tag{4.23}$$

The **norm** of a vector is simply its length with respect to the metric. We write the norm of a vector v as $|v|$, or sometimes as $\|v\|$ when we really want to be fancy. The norm is simply the square root of the inner product of the vector with itself:

$$|v| = \sqrt{\langle v, v \rangle}. \tag{4.24}$$

Because we want lengths to make sense, we make a further requirement that any sensible metric is *positive definite*,[5]

[5] Mathematicians use the adjective **Riemannian** to mean 'positive definite' when applied to metrics. Physicists often assume that their metrics are Riemannian, with the exception of spacetime, which is semi-Riemannian... because time is funny.

$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0 \tag{4.25}$$

for any vector $\boldsymbol{v} \in V$. Both the angle bracket notation $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ and the dot product notation $\boldsymbol{v} \cdot \boldsymbol{w}$ make it clear that the metric is an operation between two vectors. There is another notation that highlights the idea of the metric as a map/function on $V \times V$:

$$g(\boldsymbol{v}, \boldsymbol{w}) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle . \tag{4.26}$$

The convention of calling the metric g is common in general relativity.[1]

Example 4.8.1. *The Euclidean metric in Cartesian coordinates is the usual dot product from childhood. In $\mathbb{R}^2$, the dot product of two vectors is $\boldsymbol{v} \cdot \boldsymbol{w} = v^1 w^1 + v^2 w^2$. The generalization to $\mathbb{R}^N$ is*

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = g(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{w} = \sum_{i=1}^{N} v^i w^i , \tag{4.27}$$

where v^i and w^i are the ith Cartesian components of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ respectively. There are a few different ways of writing this out.

Let $\boldsymbol{v}^{\mathrm{T}}$ be the transpose of $\boldsymbol{v}$. We do not want to dwell on what this means (see below), but suffice it to say that if $\boldsymbol{v}$ is what a child would call a column vector, then $\boldsymbol{v}^{\mathrm{T}}$ is what the child would call a row vector. The child may then say that the length of $\boldsymbol{v}$ is simply

$$|\boldsymbol{v}| = \sqrt{\boldsymbol{v}^{\mathrm{T}} \boldsymbol{v}} .$$

Observe that there is no inner product in this expression: the 'row vector' $\boldsymbol{v}^{\mathrm{T}}$ naturally acts on the column vector $\boldsymbol{v}$ by matrix multiplication.[2] This suggests that the inner/dot product is related to

[2]As we discuss below, the row vector is a linear function from $V \to \mathbb{R}$.

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^{\mathrm{T}} \boldsymbol{w} . \tag{4.28}$$

To be fully general, we write

$$\boldsymbol{v} \cdot \boldsymbol{w} \equiv \boldsymbol{v}^{\mathrm{T}} \mathbf{g} \boldsymbol{w} \qquad\qquad \mathbf{g} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{4.29}$$

Other metrics on $\mathbb{R}^2$ may be defined the components of the matrix $\mathbf{g}$.

Yet another definition of the metric is with respect to the differential line element. For the Euclidean metric in $\mathbb{R}^2$,

$$ds^2 = dx^2 + dy^2 \ . \tag{4.30}$$

This generalizes to the form

$$ds^2 = \sum_{i,j=1}^{2} g_{ij} dx^i dx^j \ , \tag{4.31}$$

*where g_{ij} are the components of **g**. Here the differentials are linear functions that take in one of the vector arguments of the metric and spit out a number.[1] This notation is familiar if you have studied some general relativity, but is likely completely mysterious for those with no exposure to differential geometry. Do not worry if you fare not [yet] familiar with the notation: take it as an indication that there are elegant connections between calculus and linear algebra that are waiting for you to discover them.*

[1]Terms like $dx^2 = dx \otimes dx$ where the first dx is a linear function of the first argument, and the second dx is a linear function of the second argument.

The above exercise motivates writing the inner product in *tensor* notation[2],

$$g(\boldsymbol{v}, \boldsymbol{w}) \equiv \sum_{i,j} g_{ij} v^i w^j \ . \tag{4.32}$$

[2]We deliberately postpone defining what we mean by tensor.

For the case of $\mathbb{R}^2$ in Cartesian coordinates, (4.32) matches (4.27) when we take the components of g_{ij} to be

$$g_{ij} = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} , \tag{4.33}$$

where we define a cousin[3] of the Kronecker δ. If you are not familiar with summing over indices, please check this explicitly: (4.32) has a double sum over indices i and j, whereas (4.27) only has a single sum. The δ_{ij} thus collapses one of the sums by forcing one of the indices to be exactly the same as the other index. The g_{ij} in (4.32) is precisely the same as the one that shows up in (4.31).

[3]Formally the Kronecker δ should have one upper and one lower index, as we discuss below. Our abuse of notation, however, is mostly harmless. Yes, that was a Douglas Adams reference.

Exercise 4.8.1. *Given the restriction that a metric is a bilinear function of two vectors, explain why a general $\mathbb{R}^2$ metric may be written in the form (4.29). Why can any bilinear function of two vectors in $\mathbb{R}^2$ be written with respect to the four*

components of **g**? *What are the restrictions on the components of* **g** *given that metric is symmetric and positive definite?*

Example 4.8.2. *The two-dimensional Minkowski space metric in the standard basis is*

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v^0 w^0 - v^1 w^1 \,, \tag{4.34}$$

where we use the standard relativistic notation that the components of a vector are labelled as

$$\boldsymbol{v} = \begin{pmatrix} v^0 \\ v^1 \end{pmatrix} \,, \tag{4.35}$$

with v^0 being the time-like component. The generalization to $(d + 1)$-dimensional Minkowski space[1] is simply

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v^0 w^0 - \vec{\mathbf{v}} \cdot \vec{\mathbf{w}} \,, \tag{4.36}$$

where $\vec{\mathbf{v}}$ are the d-dimensional spatial components and $\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}$ represents the ordinary Euclidean inner product on space.

[1] This is called $\mathbb{R}^{(1,d)}$ or $\mathbb{R}^{(d,1)}$. Occasionally we invoke spaces with more than one time dimension. Some maximally symmetric spacetimes with negative curvature—think of a Pringles potato chip—are described as surfaces in $\mathbb{R}^{(2,d)}$.

Exercise 4.8.2. *The **Gram–Schmidt** procedure. Given the two vectors*

$$\boldsymbol{v} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad\qquad \boldsymbol{w} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \,, \tag{4.37}$$

[2] This is a fancy word that we take to mean 'the usual choice.' In this case we mean the Euclidean metric.

derive an orthonormal basis on $\mathbb{R}^2$ using the canonical[2] metric. You should not need instructions for how to do this, even if you do not remember the Gram–Schmidt procedure: there are only two vectors and only a small number of operations you can do with them.

Exercise 4.8.3. *The **Gram–Schmidt** procedure, generalized. Based on the previous exercise, give instructions for how to systematically derive an orthonormal basis from a set of d linearly independent vectors in a d-dimensional metric space. Does it matter what order you process the original d vectors in this procedure?*

Exercise 4.8.4. *Write out the Euclidean metric for polar coordinates in* $\mathbb{R}^2$, (r, θ). *These are related to ordinary Cartesian coordinates by:*

$$(x, y) = (r \cos \theta, r \sin \theta) .\qquad(4.38)$$

The answer is not $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_r w_r + v_\theta w_\theta$. *You know the Euclidean metric in Cartesian coordinates and you know the change of basis, so this is straightforward.*

4.9 Inner product as a projection

[**Flip**: insert figure]

The inner product has a natural geometric interpretation as the projection of one vector onto the other. By the symmetry of the metric, $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$, it does not matter which vector is being projected onto the other. This projection gives a definition of the angle, θ, between two vectors:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle \equiv |\boldsymbol{v}||\boldsymbol{w}| \cos \theta .\qquad(4.39)$$

Writing everything explicitly in terms of the inner product:

$$\cos \theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle \langle \boldsymbol{w}, \boldsymbol{w} \rangle}} .\qquad(4.40)$$

This definition of the angle obviously matches what we expect for vectors in $\mathbb{R}^2$. In fact, with a little bit of thought, it obviously matches what we expect for vectors in $\mathbb{R}^d$ with $d > 1$: The two vectors live in a two-dimensional subspace of $\mathbb{R}^d$ and so the angle reduces to the $\mathbb{R}^2$ case. What is nice about this definition is that (4.40) generalizes to *any* inner product.[1]

Given an orthonormal basis $\{ \hat{\boldsymbol{e}}_{(i)} \}$,[2] the components of a vector $\boldsymbol{v}$ with respect to this basis are simply the inner products:

$$v^i = \langle \hat{\boldsymbol{e}}_{(i)}, \boldsymbol{v} \rangle .\qquad(4.41)$$

Exercise 4.9.1. *Explain what goes wrong when the basis is not orthonormal. Give an explicit example in* $\mathbb{R}^2$. *Define a non-orthonomal basis and a vector* $\boldsymbol{v}$ *with components* v^i *with respect to the basis. Then take the inner products with the basis vectors to show that* $v^i \neq \langle \hat{\boldsymbol{e}}_{(i)}, \boldsymbol{v} \rangle$.

[1]One place this shows up is the classification of continuous symmetry groups. **Dynkin diagrams** indicate the 'angles' between different directions in the space of symmetry transformations.

[2]The notation $\{\cdots\}$ refers to a set of objects. In this case we mean the set of basis vectors: $\hat{\boldsymbol{e}}_{(1)}, \hat{\boldsymbol{e}}_{(2)}, \cdots$.

Returning to our notion of a 'nice basis' in Sec. 4.6, the orthonormality of a set of vectors is only defined with respect to a metric. A basis is orthonormal if

$$\langle \hat{\boldsymbol{e}}_{(i)}, \hat{\boldsymbol{e}}_{(j)} \rangle = \delta_{ij} \ . \tag{4.42}$$

Conversely, given an orthonormal set of basis vectors the components of the metric g_{ij} are,[1]

[1]This is sometimes callsed the **Gram matrix**

$$g_{ij} = \langle \hat{\boldsymbol{e}}_{(i)}, \hat{\boldsymbol{e}}_{(j)} \rangle \ . \tag{4.43}$$

This identification is obvious upon invoking the linearity of the inner prodcut and the expansion of any vector as a linear combination of basis vectors:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \left\langle \sum_i v^i \hat{\boldsymbol{e}}_{(i)}, \sum_j w^j \hat{\boldsymbol{e}}_{(j)}, \right\rangle = \sum_{i,j} v^i w^j \left\langle \hat{\boldsymbol{e}}_{(i)}, \hat{\boldsymbol{e}}_{(j)}, \right\rangle = \sum_{i,j} g_{ij} v^i w^j \ , \tag{4.44}$$

where the last equality identifies the metric components g_{ij} in (4.32). The metric, which gives us a sense of angle and length, encodes nothing more and nothing less than the inner products of the basis vectors.

Exercise 4.9.2. *Show that a basis is linearly independent if and only if*

$$\det \langle \hat{\boldsymbol{e}}_{(i)}, \hat{\boldsymbol{e}}_{(j)} \rangle \neq 0 \ . \tag{4.45}$$

[2]This phrase is code for "you can prove it, but we do not prove it here."

Because the metric is symmetric, it turns out[2] that one can always choose a basis where the metric is diagonal. This is tantamount to saying that there is an orthogonal basis. One can further impose that $|\det g_{ij}| = 1$ for an orthonormal basis.[3]

[3]The absolute value is important here because the spacetime metric has a relative sign between timelike and spacelike components, (4.36).

You may suspect that all this implies that there is always a *correct* basis—or a class of correct bases—where the metric is always nice and diagonal. For many cases this is true and this is the origin of funny prefactors that show up in some of the standard mathematical expressions we use. However, there are exceptions to this. In general relativity, the statement that one can always diagonalize the metric is a *local* statement. Physically this is equivalent to saying that at every point in spacetime one can chose a local inertial frame that is 'free falling.' That is to say that even in a curved spacetime, at a given point you can choose coordinates where everything is Minkowski. However, the coordinates in which everything is 'flat' at point p do not generally match the coordinates in which

everything is 'flat' at point $q \neq p$. The mismatch between these coordinate systems is at the root of the mathematical structure of general relativity: it is the physical origin of the notions of connection (covariant derivative), parallel transport, and moving frames.[1]

Example 4.9.1. *We saw that Fibbonacci sequences form a vector space. Give an example of a reasonable metric on this vector space?. Answer: a good choice is the two-dimensional Euclidean metric applied to the first two elements. When the first two elements of two Fibbonacci sequences match, then the entire sequences are identical. In fact, with this definition, the space of Fibbonacci sequences is identical (isomorphic) to $\mathbb{R}^2$.*

4.10   *Linear functions of vectors*

Recalling our definition of linearity in (4.5), we now focus on the class of linear functions that take in a vector and spit out a number. To say that mathematically, we consider linear functions from $V \rightarrow \mathbb{R}$. This class of functions turns out to be significant enough to have a name. In fact, it is so significant that it has several names—and we do not always make it clear that we mean the same thing: **row vector**, **bra**, **covariant vector**, **one-form**, **dual vector**, **covector**.[2] All of these refer to the same idea of being a linear function on vectors that returns a number. By the way, there is a corresponding plurality of names for what we have thus far been calling a vector: column vector, ket, and **contravariant vector** all mean 'vector.'

The prototypical example here is a row vector. If we write a vector as a column of N numbers,

$$\boldsymbol{v} = \begin{pmatrix} v^1 \\ \vdots \\ v^N \end{pmatrix}, \tag{4.46}$$

then a row vector $\tilde{\boldsymbol{w}}$ is—no surprise here—a row of N numbers:

$$\tilde{\boldsymbol{w}} = \begin{pmatrix} w_1, w_2, \cdots, w_N \end{pmatrix}, \tag{4.47}$$

[1] An excellent introduction that is grounded in physical intuition is Carroll's *Spacetime and Geometry*. The appendices are a crash course in a graduate differential geometry course.

[2] In japanese the prefix *ko-* means 'little.' That has nothing to do with the 'co' in 'covector,' but I recently learned about *ko* in the context of the San-X characters Rillakuma and Korillakuma. The characters are delightful, but do not confuse yourself by thinking that vectors and covectors are related in the same way that Rillakuma and Korillakuma are related. It is really not even clear to me how the latter two are related.

where we have very deliberately written the components with a lower index. We recall that the elementary school rule for matrix multiplication between two objects $\tilde{w}$ and v:

1. Rotate the first object by 90° in the clockwise direction. This means that $\tilde{w}$ is rotated into a column of numbers.

2. Multiply the elements of the two objects that are the same height. In this case, we multiply w_1 by v^1, w_2 by v^2, and so forth.

3. Sum together these multiplications: $w_1 v^1 + w_2 v^2 + \cdots$.

[**Flip**: insert figure] In this way, the row vector $\tilde{w}$ acts on the column vector v as

$$\tilde{w} v = \sum_i w_i v^i \, , \qquad\qquad (4.48)$$

for some numbers w_i. The right-hand side here is clearly a number. Furthermore, it should be clear that the right hand side depends linearly on both $\tilde{w}$ *and* on v. This means if we replace $\tilde{w}$ with $2\tilde{w}$, then the number on the right-hand side doubles, and similarly with v. In fact, we can be even more explicit about the role of $\tilde{w}$ as a function that takes in vectors by writing this action as $\tilde{w}[v]$.

Exercise 4.10.1. *Show that the space linear functions that map $V \rightarrow \mathbb{R}$ is also a vector space with the same dimension as V. We call this the **dual vector space**, V^*. Two thoughts should occur to you:*

1. *The word 'dual' seems to imply that the linear functions on V are somehow identical[1] to the vectors in V.*

2. *The prior point is 'obvious' from the elementary school perspective that row vectors are simply column vectors that have been written horizontally.*

Basis row vectors

We can define a basis of row vectors/one-forms/bras/covariant vectors analogously to the basis of vectors. In the boldface notation, we could write these as $\tilde{e}^{(i)}$ so that a row vector $\tilde{w}$ may be expanded as

$$\tilde{w} = \sum_i w_i \tilde{e}^{(i)} \, , \qquad\qquad (4.49)$$

where we observe the careful placement of indices. The w_i are simply the components of the row vector with respect to the specified basis. These components

[1]Perhaps a more mathematical word is 'isomorphic.'

are simply numbers. All of the *row-vector-ness* of a row vector is carried in the basis row-vectors. Everything that makes a given row vector unique is specified in the components w_i that specify a particular linear combination of basis row-vectors.

In principle one can choose any set of basis row vectors satisfying linear independence. Obviously we prefer to have an orthonormal basis[1]

We underline the following point:

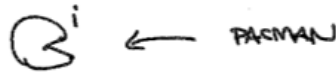> If you know the what the basis objects do, then you know everything about the space.

[1]If you are paying close attention you notice that we have not yet defined an inner product for the row vectors. Such an inner product is necessary to define 'orthonormality' on this space of row vectors.

Given an orthonormal basis $\left\{ \hat{\boldsymbol{e}}_{(i)} \right\}$ for the vector space V, there is a natural choice for a nice basis for the dual vector space V^* of row vectors/dual vectors/bras/covariant vectors/one-forms.[2] These are the basis linear functions that satisfy:

[2]Is it clear that these are all the same things? We can then just call them dual vectors or bras.

$$\tilde{\boldsymbol{e}}^{(i)}\left[\hat{\boldsymbol{e}}_{(j)}\right] = \delta^i_j \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise .} \end{cases} \tag{4.50}$$

We read this as a basis dual vector $\tilde{\boldsymbol{e}}^{(i)}$ acting[3] on a basis vector $\hat{\boldsymbol{e}}_{(j)}$. This canonical choice of basis realizes the rule (4.48) with the coefficients w_i identified as the components of $\tilde{w}$.

δ^i_j is the Kronecker δ with one upper and one lower index. When summing over the Kronecker δ with respect to either index, it enforces that the term only contributes when the two indices match. It replaces the sum over i with a fixed value $i = j$, and vice versa.

It is perhaps useful to use a slightly different notation based on Pac-Man. Rather than writing $\tilde{\boldsymbol{e}}^{(i)}$, lets write the basis dual vectors as

[3]I always thought about this as Pac-Man eating the ghosts. *Pac-Man* is a video game invented by Toru Iwatani for Bandai Namco in 1980. The game development started nearly the same time that—in a different part of the world—Michael Green and John Schwartz began a fruitful collaboration that would lead to the first string revolution four years later. When Green and Schwartz discovered the remarkable cancellation of anomalies in their theory, *Pac-Man* had just been released on the Nintendo Entertainment System.

In this notation, the action of a basis dual vector on a basis vector is simply Pac-Man eating the basis vectors:

$$\mathbb{C}^j \, \underline{e}_{(i)} \; = \; \delta^j_{\;i}$$

$$\mathbb{C}^1 \, \underline{e}_{(2)} \; = \; 0$$

$$\mathbb{C}^5 \, \underline{e}_{(5)} \; = \; 1$$

So we can write (4.48) as

$$\underline{w}^T \, \underline{v} \;\; = \; \left(w_1 \, \mathbb{C}^1 + w_2 \, \mathbb{C}^2 + \cdots \right)\left(v^1 \underline{e}_1 + \cdots \right)$$

$$\uparrow$$
$$\text{(this is } \underline{w} \cdot \underline{v} \text{)} \qquad \text{no dot!}$$

$$= \; w_1 v^1 \, \overset{=1}{\overbrace{\mathbb{C}^1 \underline{e}_1}} \; + \; w_1 v^2 \, \overset{=0}{\overbrace{\mathbb{C}^1 \underline{e}_2}} \; + \cdots$$

$$+ \, w_2 v^1 \underset{=0}{\underbrace{\mathbb{C}^2 \underline{e}_1}} \; + \; w_2 v^2 \underset{=1}{\underbrace{\mathbb{C}^2 \underline{e}_2}} \; + \cdots$$

$$= \; w_1 v^1 + w_2 v^2 + \cdots$$

$$\text{So we get} \quad \underline{w}^T \, \underline{v} \; = \; \sum_i w_i v^i$$
$$\uparrow \qquad \uparrow$$
$$\text{``}w_i\text{''} \quad \text{``}v^j\text{''}$$

Creating dual vectors with a metric

Let us be very clear: these dual vectors (row vectors, one-forms, bras, covariant vectors) *do not require a metric.* The metric takes in two vectors and spits out a number, $V \times V \to \mathbb{R}$. A row vector takes in a single vector and spits out a number, $V \to \mathbb{R}$.[1] These are two very different classes of objects.

However, if our vector space happens to be a metric space—that is, if we happen to have a metric lying around that we care to attach to V—then we can use the metric to create dual vectors out of vectors.

[1]This is a good place to remind ourselves that eventually we replace $\mathbb{R}$ with the complex numbers, $\mathbb{C}$. One may further generalize to different *fields*—generalizations of 'numbers'.

Exercise 4.10.2. *If you do not already know the answer, then pause and take a moment to think how you would create a linear function from $V \to \mathbb{R}$ given a*

bilinear function from $V \times V \rightarrow \mathbb{R}$. This is a good way to check if you are parsing this information actively or if you have fallen asleep.[1]

[1] *Inception* (2010), Christopher Nolan.

The idea is that we can *pre-load* the metric with a vector. Pick any vector, $\boldsymbol{v}$. Now insert $\boldsymbol{v}$ into the first argument of the inner product, $\langle \boldsymbol{v}, \_ \rangle$. This object now takes in a vector to fill in the second argument of the inner product and uses the machinery of the inner product to spit out a number. By the linearity of the inner product, $\langle \boldsymbol{v}, \_ \rangle$ is a linear function from $V \rightarrow \mathbb{R}$. This is precisely a dual vector. In fact, let us give this object a name:

$$\tilde{\boldsymbol{v}} = \langle \boldsymbol{v}, \_ \rangle . \qquad (4.51)$$

Ta-da! We have successfully constructed a dual vector out of a vector. What are the components of this dual vector? Assuming the canonical basis for the dual space (4.50)[2], we can find the components of a dual vector by feeding it basis vectors:

[2] This is by construction: we only have the basis of V and create a basis for V^* out of it.

$$\tilde{\boldsymbol{v}} \left[\hat{\boldsymbol{e}}_{(i)} \right] = \sum_j v_j \tilde{\boldsymbol{e}}^{(j)} \left[\hat{\boldsymbol{e}}_{(i)} \right] = v_j . \qquad (4.52)$$

Applying our definition of $\tilde{\boldsymbol{v}}$ as a pre-loaded metric (4.51),

$$\tilde{\boldsymbol{v}} \left[\hat{\boldsymbol{e}}_{(i)} \right] = \sum_j v^j \langle \hat{\boldsymbol{e}}_{(j)}, \hat{\boldsymbol{e}}_{(i)} \rangle = v^i . \qquad (4.53)$$

From this we find that $v_j = g_{ji} v^i$, where we have used the tensor notation of the metricm (4.32). For the Euclidean metric where $g_{ji} = \delta_{ji}$ we have $v_i = v^i$: that is, the row vector has the same components as the column vector from which it was created. We now recognize that this operation of 'pre-loading a metric' with a vector to create a dual vector is simply what we had been calling the **transpose** of a vector:

$$\boldsymbol{v}^{\mathrm{T}} = \tilde{\boldsymbol{v}} = \langle \boldsymbol{v}, \_ \rangle . \qquad (4.54)$$

There is a generalization of the transpose to complex vector spaces called the **Hermitian conjugate**, which is the combined transpose and complex conjugation. The symbol is the dagger, $\boldsymbol{v}^{\dagger} = \tilde{\boldsymbol{v}}$. If you are not sure whether a space is real or complex, you can say Hermitian conjugate to cover both cases and sound fancy.[3]

[3] This reminds me of people who say *compote* instead of *jam*.

We recognize that if the metric is not simply $g_{ij} = \delta_{ij}$, then the components of the dual vector are different from the components of the vector.[1] We can appreciate the utility of the upper and lower indices: something with an upper index is a component of a vector, while something with a lower index is a component of a dual vector. We can use the metric to convert from vectors to dual vectors, and the action on the components is clear: $v_i = \sum_i g_{ij} v^j$. In other words, the metric g_{ij} is a machine that can lower indices.

[1]We know that we can diagonalize the metric with an auspicious choice of basis. However, there are plenty of non-trivial diagonal metrics. Consider Minkowski space or polar coordinates.

Please appreciate that the phrase 'lower the index' is really shorthand for *convert the components of a vector into the components of the associated dual vector by taking the transpose/pre-loading the metric with that vector.* It is kind of a mouthful, so we say 'lower the index' as a shortcut. Sometimes we even *think* about the metric as acting on the indices of the components. But this too is a mental shortcut for understanding that we are moving between elements of a vector space and linear functions on those vectors.

Bra and Ket notation

The utility of the bra/ket notation in quantum mechanics is now evident. A vector is written as a ket, $\boldsymbol{v} = |v\rangle$. A dual vectors are written as a bra, $\tilde{\boldsymbol{w}} = \langle w|$. The basis bras and kets are written as follows:

$$|v\rangle = v^i |i\rangle \qquad \Longleftrightarrow \qquad \boldsymbol{v} = v^i \hat{\boldsymbol{e}}_{(i)} \qquad (4.55)$$

$$\langle w| = w_j \langle j| \qquad \Longleftrightarrow \qquad \tilde{\boldsymbol{w}} = w_j \tilde{\boldsymbol{e}}^{(j)} . \qquad (4.56)$$

One downside of the notation is that you have to mentally remember that the basis bra $\langle j|$ is secretly a lower index while the basis ket $|i\rangle$ is secretly an upper index. On the other hand, the notation has the feature of connecting a vector to its transpose[2]. Given a ket $|v\rangle$, there is a natural understanding what the corresponding bra is with respect to the inner product:

[2]We should really say 'Hermitian conjugate' because the bra/ket notation is most commonly used for Hilbert spaces, which are complex vector spaces.

$$\langle v| = \langle v|\_\rangle , \qquad (4.57)$$

where it is understood that the action of $\langle v|$ on $|w\rangle$ is

$$\langle v| [|w\rangle] = \langle v|w\rangle = \langle v, w\rangle . \qquad (4.58)$$

We emphasize that $\langle v|w \rangle$ is a bra acting on a ket, wheras $\langle v|w \rangle$ is the inner product of two kets.

The form notation

A separate notation for vectors and dual vectors is inspired by differential geometry:

$$\boldsymbol{v} = v^i \hat{\boldsymbol{e}}_{(i)} \leftrightarrow v^i \frac{\partial}{\partial x_i} \qquad\qquad \tilde{\boldsymbol{w}} = w_i \tilde{\boldsymbol{e}}^{(i)} \leftrightarrow w_i dx^i \,. \qquad (4.59)$$

The notation is perplexing the first time you see it. Everything looks familiar from calculus, but is now being used in a completely unusual context. Somehow:

⋄ Partial derivatives are the basis vectors. How odd! Partial derivatives are supposed to act on functions. Evidently the elements of the vector space are linear combinations of partial derivatives.

⋄ The infinitesimal elements, for example dx or dy, are dual vectors. This is even more strange. Based on our intuition from freshman physics, there is no sense in which an "infinitesimal displacement" acts on a partial derivative to spit out a number! Though it is perhaps fair to say that the linear combination of infinitesimal displacements is still an infinitesimal displacement—whatever that means.

Exercise 4.10.3. *Confirm that the set of partial derivatives ∂_x, ∂_y, ∂_z form a three-dimensional vector space. Here $\partial_x = \partial/\partial x$ and so forth. The vector space is the space of first order partial derivative operators acting on a function. This is trivial: do linear combinations of these operators live in the space? If you are caught up on "but what are these operators acting on?" then you are focused on the wrong thing.*

This all implies further structure that we have not specified. The further structure is the idea of a bundle over a manifold. We give a hint of this in Chap. 5.

Let us be clear. We are *defining* the object dx as a basis for linear functions acting on the vector space of partial derivatives. In this context, we call these

row vectors **differential one forms**. The action of a differential one-form on a partial derivative is

$$dx^i \left[\frac{\partial}{\partial x^j} \right] \equiv \delta^i_j \, , \qquad (4.60)$$

which you already know from (4.50). The utility of this notation is illuminated in the places where differential geometry intersects with physics. We present the notation here for those who anticipate meeting this formalism in the near future: by emphasizing that this curiosity is nothing more than vectors and dual vectors, they may then focus on the ways in which this story plays out on differential manifolds.[1]

[1]Manifold is another fancy word for what physicists usually call space or spacetime. It is like saying *pommes frites* instead of *french fries*.

4.11   *Summation convention*

From now on we invoke the summation convention that is standard in special and general relativity:

> Whenever you see exactly one upper index that is the same as exactly one lower index, there is an implied sum over that index.

This is the notational equivalent of tipping at a restaurant in the United States: you are never explicitly told to tip, but it is implied that a tip for the server should added to the final bill.[19]

[19]K. Segrave. *Tipping: An American Social History of Gratuities*. McFarland, Incorporated, Publishers, 2009. ISBN: 9780786442461.

The summing over two indices is called **contraction** of these indices. This comes from the mental shorthand of replacing (dual) vectors with their components. If we think of $\boldsymbol{v}$ as v^i and $\tilde{\boldsymbol{w}}$ as w_i, then we can think of the action of $\tilde{\boldsymbol{w}}$ on $\boldsymbol{v}$ as a contraction of the indices of their components:

$$\tilde{\boldsymbol{w}}\boldsymbol{v} = w_i v^i = w_1 v^1 + w_2 v^2 + \cdots \, , \qquad (4.61)$$

where we are using the summation convention on the right hand side. The summation convention comes about from the observation that there are only certain ways in which objects in linear algebra can combine. It is incredibly useful and saves several pen/pencil/chalk/keystrokes,[2] but leads to the common mental shorthand that an object *is* its components.[3]

[2]Nevermind the indignities of using a whiteboard.

[3]A vector $\boldsymbol{v}$ is fully specified by its components v^i once you agree on a basis. However, the vector is not its components. The components tell you the specific linear combination of basis vectors.

Exercise 4.11.1. *Which of the following is an appropriate use of the summation convention? Anything that is not appropriate is nonsense and should ideally never appear in your work.*

$$v^i w^i = w^1 v^1 + w^2 v^2 + \cdots \tag{4.62}$$

$$g_{ij} v^i w^i u^j = g_{11} v^1 w^1 u^1 + g_{12} v^1 w^1 u^2 + \cdots + g_{21} v^2 w^2 u^1 + \cdots \tag{4.63}$$

$$g_{ii} v^i w^i = g_{11} v^1 w^1 + g_{22} v^2 w^2 + \cdots \tag{4.64}$$

Answer: these are all inappropriate.

A few useful guidelines:

1. The summation convention is a shorthand. You can start by writing (explicitly or just mentally) the full expression with a big clunky summation symbol. Then simply erase the summation.

2. Conversely, if you have an expression where a contraction does not make sense even when writing out the summation symbol explicitly, then the expression is probably wrong.

3. Once you use an index, say *i*, in a contraction, *never* use that index again in the same expression. This leads to confusion. We say that the contracted index is a **dummy index**. This means that it is no longer a real index because it has been summed over.[1]

4. If the same index appears more than once, go directly to jail.[2]

5. If you contract two indices of the same reflect on the questionable decisions brought you to this point.[3]

6. When there is any possibility of ambiguity, explicitly say whether or not a repeated index is meant to be summed over.

Example 4.11.1. *Linear transformations are linear functions that take vectors to vectors, $V \to V$. These are represented by matrices, see Sec. 4.15 The components of a matrix are written with an upper index (the first index) and a lower index (the second index): $A^i{}_k$. The trace of the matrix is $A^i{}_i = A^1{}_1 + A^2{}_2 + \cdots$. When there is no ambiguity, sometimes we write A to mean the trace of A. This makes sense since the trace of a matrix has no indices: it is simply a number. However, this is often tons of ambiguity since there are other objects formed out of a matrix that are simply numbers, like the determinant.*

[1] The phrase dummy index reminds me of the dummy plug system in *Neon Genesis Evangelion*.

[2] Do not pass go. Do not collect $200. This is a reference to *Monopoly*.

[3] There are some times when it is convenient to break this rule and allow contraction of same-height indices when (1) the metric is $g_{ij} = \delta_{ij}$ and (2) there is little chance of confusion.

Bonus: how would you write the determinant of a matrix in terms of indices? Answer: we have not yet defined the necessary machinery to write out the determinant. What sort of machinery is missing?

4.12   Linearity and Indices

You should now appreciate that there is something to this whole index notation business that is intrinsically tied to the idea of linear maps.[1] One way of understanding this is that the point of writing indices is to treat objects as linear maps with respect to a vector space.

[1]You may also be wondering why we only care about linear maps. Let us remind ourselves that the essence of calculus is to squeeze the most out of a linear approximation to our functions.

Exercise 4.12.1. *Consider a linear function $f(x)$ where x is a real number. In grade school you may have learned that linear functions take the form $f(x) = ax + b$. Indeed, this equation plots to a line in the (x, f) plane. Show that with our definition, a linear function $f(x)$ cannot have a constant term, $b = 0$. HINT: $b \neq 0$ means there is some $x_0 \neq 0$ such that $f(x) = 0$. That means $f(x) + f(x_0) = f(x)$. Where does this mess things up?*

Example 4.12.1. *Suppose you have a linear function in one dimension, $f(\boldsymbol{x})$. For simplicity, let $\boldsymbol{x}$ be a vector in the simple vector space, $\mathbb{R}$. We normally think of the real numbers as, well, numbers. It should be clear that they are also the most boring vectors: sums of real numbers are real numbers, you can rescale real numbers by numbers to get a real number.*

Linearity means that

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y}) . \tag{4.65}$$

The only way this is possible is if

$$f(\boldsymbol{x}) = a\boldsymbol{x} , \tag{4.66}$$

where a is some number that completely specifies the linear function f. In fact, a is the single component *of f.*

Example 4.12.2. *Same a the previous example, but now suppose that x is an element of the two dimensional vector space $\mathbb{R}^2$. Let us write out the linearity condition in the column vector notation.*

$$f\left[\alpha\begin{pmatrix}x^1\\x^2\end{pmatrix}+\beta\begin{pmatrix}y^1\\y^2\end{pmatrix}\right]=\alpha f\left[\begin{pmatrix}x^1\\x^2\end{pmatrix}\right]+\beta f\left[\begin{pmatrix}y^1\\y^2\end{pmatrix}\right]. \tag{4.67}$$

Consider the case $x^2=y^2=0$. Restricting to this subspace reduces us back to the previous example, from which we deduce that

$$f\left[\begin{pmatrix}x^1\\x^2\end{pmatrix}\right]=ax^1+\left(\text{independent of }x^1\right), \tag{4.68}$$

for some constant a. This is the only way to satisfy linearity. By a similar argument for $x^1=y^1=0$, we see that

$$f\left[\begin{pmatrix}x^1\\x^2\end{pmatrix}\right]=bx^2+\left(\text{independent of }x^2\right), \tag{4.69}$$

*for some constant b. We know that there cannot be an overall constant term that is independent of both x^1 and x^2 since this would violate the linearity condition, see exercise above. In fact, we can change notation a bit and write $a\equiv f_1$ and $b\equiv f_2$, making the **components** of the linear function f clear. We can then write the linear map as*

$$f(\boldsymbol{x})=f_1x^1+f_2x^2=\begin{pmatrix}f_1 & f_2\end{pmatrix}\begin{pmatrix}x^1\\x^2\end{pmatrix}, \tag{4.70}$$

where in the last equality we have returned to the usual grade school row-vector/column-vector notation. If we insert the basis vectors and covectors, this is simply

$$f(\boldsymbol{x})=(f_1\langle1|+f_2\langle2|)\left(x^1|1\rangle+x^2|2\rangle\right) \tag{4.71}$$

$$=f_1x^1\langle1|1\rangle+f_2x^1\langle2|1\rangle f_1x^2\langle1|2\rangle+f_2x^2\langle2|2\rangle \tag{4.72}$$

$$=f_ix^i, \tag{4.73}$$

which of course returns the same linear combination of components due to $\langle i|j\rangle=\delta^i_j$.

4.13   *Duality*

Thus far the relation between vectors and dual vectors seem a bit one sided. Vectors are the 'objects' that seem to have some intrinsic meaning on their own. Dual vectors, on the other hand, appear to exist only in relation to vectors: they are linear functions on the space of vectors. As the word *dual* implies, however, both the vector space V and its dual space V^* are on equal footing. One could swap $V \leftrightarrow V^*$ and all of our equations would still be valid.

Exercise 4.13.1. *Before moving on, ask yourself if the qualitative idea of this duality is obvious or not obvious. Even if you have not been exposed to this idea before, it may be clear where we are going. Hint: it may be useful to think about spaces equipped with a metric first, since the metric is itself a [bi-]linear map.*

This duality is manifest in the following two statements:

[1]It has all the same mathematical properties and structures.

1. The space of dual vectors V^* is **isomorphic**[1] to the space of vectors V. We write this as $V \cong V^*$.

2. The elements of V are linear functions that map elements V^* to numbers. In other words, V is the dual space of V^*: $V = (V^*)^*$.

The first statement is obvious if we think about dual vectors as row vectors. Were those not simply column vectors that had been 'tipped over' using the metric?[2] More generally, it should be obvious upon reflection that a *linear combination of linear functions is, itself, a linear function.* Suppose f and g are linear functions from some space X to $\mathbb{R}$. T for numbers α and β, $(\alpha f + \beta g)$ is bealso a linear function whose output on $x \in X$ is simply

[2]Astute readers notice that this picture requires a metric.

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) \,. \tag{4.74}$$

This linearity does not depend on the nature of the domain, X. Thus far, we have shown that V^* is itself a bona fide vector space. We have, of course, defined V^* to be the space of linear functions on $X = V$, and this is sufficient to argue that not only is V^* a vector space, but it is isomorphic to the vector space of its arguments, V.

Exercise 4.13.2. *Argue why the linear vector space V^* is isomorphic to V. We are less interested in rigor and more interested in the idea that there is a plausible one-to-one map between elements of the vector space and elements of the dual vector*

space. It is sufficient to argue based on linearity that for each basis element in V there is a natural choice of basis element in V^* for $\langle i \,|\, j \rangle = \delta_i^j$.

The second condition to show this duality is this idea that not only do dual vectors 'eat' vectors (and poop numbers), but the relation works the other way around as well: vectors can 'eat' dual vectors.[1] This is obvious from bra and ket notation: when we write $\langle w \,|\, v \rangle$, it is not clear which one is 'acting' on the other nor does it matter. If you happen to have a vector $|v\rangle$, then the one thing you can do if I give you a dual vector $\langle w|$ is to produce a number $\langle w \,|\, v \rangle$. Because this action is linear[2], then $|v\rangle$ is obviously a linear function on dual vectors.

[1]Earlier we wrote that dual basis vectors 'eat' vectors the way that Pac-Man eats ghosts. Duality is the idea that ghosts also eat Pac-Man.

[2]Check this if it is not obvious!

Example 4.13.1. *Suppose you have a vector $\boldsymbol{v}$ in $\mathbb{R}^2$. I happen to have a couple of basis covectors, $\tilde{\boldsymbol{e}}^{(1)}$ and $\tilde{\boldsymbol{e}}^{(2)}$. These are not necessarily the canonical basis; i.e. $\tilde{\boldsymbol{e}}^{(1)} \neq \begin{pmatrix} 1 & 0 \end{pmatrix}$. We know that my basis covectors can 'eat' your vector and produce a number. When we do this experiment, we get the following numbers:*

$$\tilde{\boldsymbol{e}}^{(1)}\left[\boldsymbol{v}\right] = a \qquad\qquad \tilde{\boldsymbol{e}}^{(2)}\left[\boldsymbol{v}\right] = b \qquad\qquad (4.75)$$

This means I can write $\boldsymbol{v}$ as a linear function of dual vectors $\tilde{\boldsymbol{w}} = w_1\tilde{\boldsymbol{e}}^{(1)} + w_2\tilde{\boldsymbol{e}}^{(2)}$ as follows:

$$\boldsymbol{v}\left[\tilde{\boldsymbol{w}}\right] \equiv w_1\tilde{\boldsymbol{e}}^{(1)}\left[\boldsymbol{v}\right] + w_2\tilde{\boldsymbol{e}}^{(2)}\left[\boldsymbol{v}\right] = aw_1 + bw_2 \,. \qquad\qquad (4.76)$$

This is really something of a trivial statement. The key point is that we do not have to think about which of V or V^* is the 'proper' vector space and which one is the dual space of linear functions. They are each proper vector spaces and they are each dual spaces of linear functions on the other.

We implied earlier that all of this becomes a bit more clear when we have a metric. In fact, half of the isomorphism is obvious with the metric because we know that the metric allows us to define the transpose[3] of a vector, that is, how to convert an element of V into an element of V^*. This was the observation (4.54) that we can 'preload' the inner product with a vector and the resulting object is a covector.

[3]Hermitian conjugate.

What is left for us to argue is whether there is a 'dual metric' which we can preload with a covector so that the resulting object is a vector: a linear function

on the space of covectors. There is a natural object for this that does not require any additional structure: the *inverse metric*.[1]

[1]That the metric should be invertible comes along with the idea of 'niceness.' However, there are plenty of examples where a metric may not be invertible at a given point. This happens, for example, at $r = 0$ when using spherical coordinates.

Let us confirm this in words. The metric takes two vectors and spits out a number. Equivalently, it takes one vector v and it spits out a covector, (4.51), $\tilde{v}$. With this latter interpretation, the *inverse metric* takes the covector $\tilde{v}$ and returns the original vector v. This is exactly the machine that we want to highlight the dual relation between V and V^*.

Given a basis for V there is a natural basis for V^*, (4.50). This is the basis where $\langle i | j \rangle = \delta_i^j$. Clearly one could have equivalently started with a basis for V^* and written the canonical basis for $(V^*)^* = V$. Once we agree on the basis for the space and its dual, we are free to work with just the indices.[2] The metric is an object with two lower indices, g_{ij}. The inverse metric is a matrix that, when acting on the metric, gives unity. In this case, unity is the Kronecker δ:

[2]This is probably among the top ten ways to "say you are a physicist without saying that you are a physicist."

$$\left(g^{-1}\right)^{ik} g_{kj} = \delta_j^i \,, \tag{4.77}$$

where we have used the summation convention to leave the sum over k implicit. The order of indices on δ_j^i clearly does not matter. We have intuited from the index structure that the components of the inverse metric must have two upper indices. This is because the summation convention tells us that we can only sum an upper and a lower index.[3]

[3]To be clear: the mathematical structure says that the only sensible contractions are between covector indices and vector indices. The summation convention does two things: (1) it gives a mnemonic to make sure that only these mathematically sensible contractions occur, and (2) it saves us from writing $\sum$ all over the place. It is the first point that is 'deep' in this convention.

There is one more standard notational simplification that we impose. Because the metric on V has two lower indices and its inverse metric (the metric on V^*) has two upper indices, there is really no way to confuse them as long as we are writing out indices explicitly. So let us simply the notation by simply droping the inverse symbol:

$$\left(g^{-1}\right)^{ij} \equiv g^{ij} \,. \tag{4.78}$$

This looks a bit unusual, but the result is that we have

$$g^{ik} g_{kj} = \delta_j^i \,. \tag{4.79}$$

The metric lowers indices, while the inverse metric raises indices. This is obvious: you can think about the inverse metric as the metric for V^* and then swap 'raised/upper index' with 'lower index' in our discussion of the metric.

Example 4.13.2. *As a quick sanity check, is there any ambiguity when writing the inverse metric as* g^{ij} *coming from the interpretation that* $(g^{-1})^{ij} = g^{ij} = g^{ik}g_{k\ell}g^{\ell j}$? *In the last equality, we interpreted the inverse metric as an object that raises indices. When using the defining relation* $g^{ik}g_{kj} = \delta^i_j$ *then the relation in question becomes a tautology.*[1]

> **Remark:** The word *dual* shows up often in theoretical physics and in different contexts. The dual space V^* of a vector space is one manifestation of this. More generally, a dual description of something is a transformation of that thing that is completely equivalent but perhaps illuminates a different aspect of the thing. A concrete example is *electromagnetic duality* in classical electrodynamics without sources. In this case, there is a duality where we swap the electric and magnetic fields with one another. Like many of the most fascinating dualities in physics, this is a strong–weak duality where the electric coupling of the dual theory goes like the inverse of the electric coupling in the original theory. Because these couplings are often expansion parameters in perturbation theory, strong–weak dualities give a way to apply perturbation theory to systems that do not appear to have perturbative limits at first glance. As the name implies, applying the duality transformation twice returns you to the original description. Two of the most fascinating class of dualities at this time are the holographic principle (AdS/CFT correspondence) and the suite of electromagnetic dualities in supersymmetric gauge theories (Seiberg dualities). The dual of a linear space is a much humbler idea in comparison.

[1] That's how mathematicians say 'more obvious than obvious.' This, in turn, is a phrase I learned from Tony Zee and that I greatly enjoy using to pedagogical and sarcastic effect.

4.14 *Metric and Indices*

Formally, the metric is an object that 'has on indices' in the same way a vector 'has no indices' because $\boldsymbol{v} = v^i |i\rangle$ in bra-ket notation. Of course, physicists love to write v^i to mean "the vector $\boldsymbol{v}$" because the indices are convenient so long as the meaning of the basis vectors $|i\rangle$ is unambiguous to everyone. The analog for an object like the metric is $\mathbf{g} = g_{ij} \langle i| \otimes \langle j|$. Here we meet the **tensor product**, $\otimes$. What $\langle i| \otimes \langle j|$ means is that $\langle i|$ and $\langle j|$ are *not* acting on one another. They are both waiting for separate kets to come along to spit out numbers. This simply

returns us to the observation that the metric can be thought of a machine that either:

1. Takes one vector and spits out a covector in a linear way.

2. Takes two vectors and spits out a number in a linear way.

Let us see how this works in each case to demonstrate how to 'read' the tensor product. The tensor $\mathbf{g}$ acting on a vector $\boldsymbol{v}$ gives

$$\left(g_{ij} \langle i | \otimes \langle j | \right) \left(v^k | k \rangle \right) = v^k g_{ij} \langle i | \langle j | k \rangle = v^k g_{ij} \langle i | \delta_k^j = v^j g_{ij} \langle i | \equiv \tilde{v}_i \langle i | . \quad (4.80)$$

We have defined $\tilde{v}_i \equiv v^j g_{ij}$ to be excessively pedantic. We are free to drop the tilde and write $v_i \equiv v^j g_{ij}$ because there are no other objects that are named v with one lower index. Here it is implicit that the second bra in $\mathbf{g}$, $\langle j |$, acts on the ket in $\boldsymbol{v}$, $| k \rangle$. In fact, because $g_{ij} = g_{ji}$ by assumption for the metric, it does not matter whether it is the first bra or the second bra in the tensor product $\mathbf{g} = g_{ij} \langle i | \otimes \langle j |$ that acts on the ket $| k \rangle$.

At this point, you would be justified in feeling that we have really made a big deal out of converting vectors into covectors and vice versa. Is this not just some glorified transpose? Why introduce all this mathematical structure? We return to this thread in Section 4.18. First, let us broaden our set of objects that can exist and act on our (co-)vector spaces.

4.15   *Linear Transformations*

We have come far enough that you may be wondering, *if this is linear algebra, where are all the matrices?* From here on out, we downplay the idea of a 'matrix' because it can lead to some ambiguities: not every mathematical object that can be expressed as a two-dimensional array behaves in the same way.
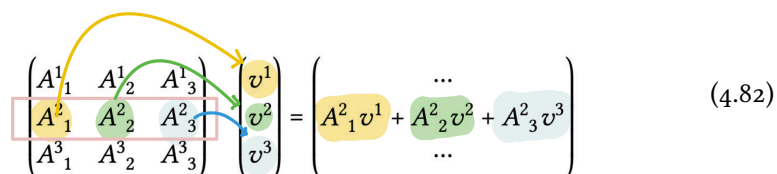
Example 4.15.1. *The metric $\mathbf{g}$ is an object with two indices, g_{ij}. In an N-dimensional vector space, it is often written as an $N \times N$ array,*

$$\mathbf{g} = \begin{pmatrix} g_{11} & g_{12} & \cdots \\ g_{21} & \ddots & \cdots \\ \vdots & \vdots & g_{NN} \end{pmatrix} . \quad (4.81)$$

If we accept some sloppiness, we could even call this a matrix. However, this is not the same kind of object that you normally think about as a matrix.

When you first learned linear algebra[1], you learned that matrices act on vectors using the usual matrix multiplication rule. That rule looked something like this:

$$\begin{pmatrix} A^1{}_1 & A^1{}_2 & A^1{}_3 \\ A^2{}_1 & A^2{}_2 & A^2{}_3 \\ A^3{}_1 & A^3{}_2 & A^3{}_3 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} = \begin{pmatrix} \cdots \\ A^2{}_1 v^1 + A^2{}_2 v^2 + A^2{}_3 v^3 \\ \cdots \end{pmatrix} \tag{4.82}$$

The mechanical picture for this rule is best shown in real time. A rough summary is as follows. If matrix A acts on vector v, the value of the ith component of Av is calculated by:

1. Take the ith *row* of A.

2. Rotate it clockwise by one quarter of a turn. Now it is aligned with the vector v. Mentally place them right next to each other.

3. Multiply the components that are horizontally next to each other.

4. Sum together each product. This is $(Av)^i$.

Exercise 4.15.1. *Write this procedure in index notation. Answer:* $(Av)^i = A^i{}_j v^j$. *Please take time to make sure that this is obvious. If it is not: work out the implied sum explicitly. You can even use the explicit case above, where $i = 2$ and the indices run from $i, j \in \{1, 2, 3\}$.*

What is critical to emphasize at this stage is not yet the mechanical definition of the matrix multiplication procedure, but the following facts:

1. When a matrix acts on a vector, the result is a vector. This is in contrast to a number or a covector or anything else.

2. When a matrix acts on twice a vector, the result is twice the previous result. That is: $A(2v) = 2Av$. The same is true for any rescaling by a number.

3. When a matrix acts on the sum of two vectors, the result is the sum of the matrix acting on each vector independently: $A(v + w) = Av + Aw$.

What this tells us is that the object that we were calling a 'matrix' colloquially—and that we may have been [erroneously] equating with a 2D array of numbers—is really a linear map from $V \rightarrow V$. Since this linear map takes objects (vectors) in V and turns them into other objects in V, we call this a **linear transformation**. So from now on we will be deliberate and talk about linear transformations from V to V.

A quick remark: perhaps in "Algebra II" you worked with non-square matrices. A $n \times m$ matrix is a linear transformation that maps an m-dimensional vector into an n-dimensional vector. Clearly these are different vector spaces. In this course we are not interested in non-square matrices—there is an *reason* we want to stick to the same vector space. And anyway, they will be infinite dimensional.

Exercise 4.15.2. *The "Algebra II" notion of a matrix is a square array of numbers that turns vectors into other vectors. From the definition of this multiplication, it should be clear that the transformation is linear. Show that* any *linear transformation on a vector $\boldsymbol{v}$ can be represented as matrix multiplication. As a hint, you can start with the case of a one-dimensional vector where $\boldsymbol{v}$ is just a number. Remember that our definition of linear is slightly different from the $y = mx + b$ notion in basic algebra.*

The above exercise makes it clear that the Algebra II idea of a matrix is exactly the same as a linear transformation from $V \rightarrow V$. The latter phase is one that we can generalize to tensors.

Let us add mathematical structure to our picture. Earlier we met the tensor product of two covectors, $\langle i | \otimes \langle j |$. This is an object that linearly 'eats' a vector using $\langle j |$ and then spits out a covector with basis $\langle i |$. Another way of thinking about this is that the $\langle j |$ acts on the vector and spits out a number. That number contributes to the coefficient of $\langle i |$. This tensor structure is a map from $V \rightarrow V^*$. What does the tensor structure of a linear transformation $V \rightarrow V$ look like?

Following the same line of thought, it should be no surprise that a linear transformation from $V \rightarrow V$ can be expanded in terms of a basis of "ket–bras" $|i\rangle \otimes \langle j |$. Because these linear transformations are so common, we typically drop the tensor product symbol and write

$$|i\rangle \otimes \langle j | \rightarrow |i\rangle \langle j | \ . \tag{4.83}$$

Thus the object that we used to call a 'matrix' A is expanded in components $A^i_{\ j}$ as

$$A = A^i_{\ j} \, |i\rangle \langle j| \ . \tag{4.84}$$

Note the index structure carefully. The basis coefficient (or component) $A^i_{\ j}$ has one upper index and one lower index. The upper index is first—it appears to the left—and indicates that the first object in the basis tensor product is a basis vector $|i\rangle$. The second index is lowered and indicates that the second object in the basis tensor is a basis covector, $\langle j|$.

The index structure of the coefficients now tells us everything there is to know about the basis. As a physicist, we may write[1] "$A = A^i_{\ j}$." By this we mean that we can *fill in* the basis $|i\rangle\langle j|$ just by looking at the indices. If we take this very glib shorthand, then we can imagine a linear transformation A acting on a vector $\boldsymbol{v}$ by simply looking at their indices:

$$(A\boldsymbol{v})^i = A^i_{\ j} \boldsymbol{v}^j \ . \tag{4.85}$$

We can tell ourselves the story that the "lower index" of $A^i_{\ j}$ acts on the "upper index" of $\boldsymbol{v}^j$ by contraction so that the ith component of $A\boldsymbol{v}$ is the sum $A^i_{\ j}\boldsymbol{v}^j$. Of course, what is actually going on is that the basis objects are doing all of the work "under the hood" of enforcing the linear transformation:

$$\left(A^i_{\ j} \, |i\rangle \langle j| \right) \left(\boldsymbol{v}^k \, |k\rangle \right) = A^i_{\ j} \boldsymbol{v}^k \, |i\rangle \langle j|k\rangle \quad = A^i_{\ j} \boldsymbol{v}^k \delta^j_k \, |i\rangle = \left(A^i_{\ j} \boldsymbol{v}^j\right) |i\rangle \ . \tag{4.86}$$

Exercise 4.15.3. *Okay, one last chance. If it is not completely obvious to you that* $\left(A^i_{\ j}\boldsymbol{v}^j\right)$ *is indeed the* i^{th} *component of* $A\boldsymbol{v}$, *then work this out explicitly for the simple case of a* 2×2 *matrix.*

Exercise 4.15.4. *Show that the space of linear transformations from* $V \rightarrow V$ *(the space of "matrices") is a vector space. The set of ket–bras (nobody really calls them that)* $|i\rangle\langle j|$ *are a basis for this space. When* $V = \mathbb{R}^n$, *we call the space* $\mathrm{GL}(N, \mathbb{R})$, *the* general linear group.

Exercise 4.15.5. *Show that the a valid definition for a metric on the vector space* $\mathrm{GL}(N, \mathbb{R})$ *is*

$$\langle A, B \rangle = \mathrm{Tr}\left(A^{\mathrm{T}} B\right) \equiv \left(A^{\mathrm{T}}\right)^i_{\ j} B^j_{\ i} \ , \tag{4.87}$$

the trace of $A^{\mathrm{T}}B$ where A^{T} is the transpose of A: $\left(A^{\mathrm{T}}\right)^i_{\ j} \equiv g_{ik}A^k_{\ \ell}g^{\ell j}$. We motivate this definition below.

Exercise 4.15.6. *Consider* $\mathrm{GL}(2, \mathbb{R})$. *A standard basis is* $|I\rangle$ *with* $I \in \{1, 2, 3, 4\}$ *and defined with respect to the* $|i\rangle\langle j|$ *as:*

$$|1\rangle = |1\rangle\langle 1| \qquad |2\rangle = |1\rangle\langle 2| \qquad |3\rangle = |2\rangle\langle 1| \qquad |4\rangle = |2\rangle\langle 2| . \qquad (4.88)$$

Suppose $|A\rangle$ *is an element of* $\mathrm{GL}(2, \mathbb{R})$ *with components* A^I *in this standard basis. Now suppose we have a different basis written with respect to diagonal, symmetric, and antisymmetric matrices:*

$$\left|1'\right\rangle = |1\rangle\langle 1| \qquad\qquad\qquad \left|2'\right\rangle = \frac{1}{2}\left(|1\rangle\langle 2| + |2\rangle\langle 1|\right) \qquad (4.89)$$

$$\left|3'\right\rangle = \frac{1}{2}\left(|1\rangle\langle 2| - |2\rangle\langle 1|\right) \qquad \left|4'\right\rangle = |2\rangle\langle 2| . \qquad (4.90)$$

What are the components of $|A\rangle$ *in the primed basis,* $(A')^I$?

What else can we do with a linear transformation A from $V \rightarrow V$? We already know that if we feed A a vector, we get another vector. We could also 'pre-load' A with a covector in the same way that we 'pre-loaded' the metric. Suppose $\tilde{w}$ is such a covector. Then

$$\tilde{w}A = \left(\tilde{w}_i \langle i|\right)\left(A^j_{\ k} |j\rangle\langle k|\right) = \tilde{w}_i A^i_{\ k} \langle k| \qquad (4.91)$$

is also a covector. This is obvious because it takes the form (coefficient)$^i \langle i|$ where the $\langle i|$ are the basis covectors. This means that we can think of the linear transformations from $V \rightarrow V$ as, equivalently, linear transformations from $V^* \rightarrow V^*$. Alternatively, if we feed A a covector and a vector, then we can form a *bilinear* function from $V \times V^* \rightarrow \mathbb{R}$ through

$$A(\tilde{w}, v) = \tilde{w}Av . \qquad (4.92)$$

Here 'bilinear' means that $A(\tilde{w}, v)$ is linear in each of its two arguments.

Thus we can interpret the linear transformation A in three ways:

1. A linear map from $V \rightarrow V$

2. A linear map from $V^* \rightarrow V^*$

3. A bilinear map from $V \times V^* \rightarrow \mathbb{R}$.

Each of these interpretations can be understood from the index structure of the components $A^i_{\ j}$.

1. We can contract the lower j index with the upper index of a vector's components, $A^i_{\ j} v^j$. After the contraction, there is only an upper index left so we know that the object is a vector.

2. We can contract the upper i index with the lower index of a covector's components, $A^i_{\ j} \tilde{w}_i = \tilde{w}_i A^i_{\ j}$. After the contraction, there is only a lower index left so we know that the object is a covector.

3. We can contract each of the indices with the corresponding indices of a vector a covector. After the contraction, there are no indices left over and the object is simply a number.

We see that the array of numbers ("matrix") A and its index structure tell us everything about the linear transformation, no matter *which* linear transformation we are using. From now on, we shall refer to linear transformations by their coordinates, e.g. $A^i_{\ j}$ rather than as a function $f(v)$.

Transformation = action on all basis vectors

[**Flip**: Need to revise this]

If you know the action of a linear transformation on all basis vectors, then you know the transformation. When written as a matrix, a linear transformation is simply stating what that linear trasformation does to your basis vectors. If a matrix B has elments

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = B^i_{\ j} |i\rangle\langle j| , \tag{4.93}$$

then this simply means that acting on basis vectors $e_{(1)} = |1\rangle$ and $e_{(2)} = |2\rangle$ gives

$$B|1\rangle = a|1\rangle + c|2\rangle \tag{4.94}$$
$$B|2\rangle = b|1\rangle + d|2\rangle . \tag{4.95}$$

In column vector notation:

$$B\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} \qquad\qquad B\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix} . \tag{4.96}$$

So knowing the action on basis vectors is the same as knowing the transformation itself.

Given a linear operator A, the inverse operator A^{-1} is defined by

$$A^{-1}A = \mathbb{1}_{N \times N} . \tag{4.97}$$

For an N-dimensional vector space, this represents N^2 different equations: one for each element. The inverse operator, by the way, is also linear. Suppose I told you the action of the inverse transformation A^{-1} on your basis vectors, vis-a-vis (4.95):

$$A^{-1}|1\rangle = x|1\rangle + y|2\rangle \tag{4.98}$$
$$A^{-1}|2\rangle = z|1\rangle + w|2\rangle . \tag{4.99}$$

Then you know exactly how A^{-1} acts on a general vector $|s\rangle = s^1|1\rangle + s^2|2\rangle$:

$$A^{-1}|s\rangle = A^{-1}\left(s^1|1\rangle + s^2|2\rangle \right) \tag{4.100}$$
$$= s^1 A^{-1}|1\rangle + s^2 A^{-1}|2\rangle \tag{4.101}$$
$$= (s^1 x + s^2 z)|1\rangle + (s^1 y + s^2 w)|2\rangle . \tag{4.102}$$

You can now keep this in mind when we say we want to solve $A|\psi\rangle = |s\rangle$. If we knew the action of A^{-1} on some basis of the space, then the problem is simple:

$$|\psi\rangle = \psi^i|i\rangle = \left(A^{-1} \right)^i{}_j |i\rangle\langle j| s^k|k\rangle \tag{4.103}$$
$$= \left(A^{-1} \right)^i{}_j s^k |i\rangle\langle j||k\rangle \tag{4.104}$$
$$= \left(A^{-1} \right)^i{}_j s^j |i\rangle . \tag{4.105}$$

We can write this as an equation for each component:

$$\psi^i = \sum_j \left(A^{-1} \right)^i{}_j s^j . \tag{4.106}$$

We've restored the explicit sum over j as a convenient reminder. The quantity $\left(A^{-1}\right)^i_{\ j}$ is what we would like to identify with a Green's function; see Section ?? for a hint of that. From what we've learned in this subsection: knowing the components $\left(A^{-1}\right)^i_{\ j}$ is exactly the same as *knowing the action of the inverse operator on basis vectors.*

4.16 *Adjoint*

Recall the definition of the transpose/Hermitian conjugate in (4.54). The 'dagger' operation † converts a vector into a covector using the inner product. For real vector spaces this is simply the transpose, for complex vector spaces it is the transpose and complex conjugation. We now define an operation on linear transformations A that borrows the same notation: the adjoint. The adjoint of a linear transformation A is denoted $A^\dagger$ and is defined by the inner product relation:

$$\langle A^\dagger w, v \rangle = \langle w, Av \rangle. \tag{4.107}$$

This should look familiar from the bra-ket notation in quantum mechanics where we have $\langle w|Av \rangle = \langle A^\dagger w|v \rangle$. The bra $\langle w|$ is the Hermitian conjugate (dual) of some ket $\langle w| = |w\rangle^\dagger$, which is in turn defined with respect to the inner product. Following these definitions, one sees that this is equivalent to the adjoint definition (4.107). This justifies the 'overloading' of the † operator: if $u = Aw$, then

$$\langle u| = (A|w\rangle)^\dagger = (|w\rangle)^\dagger A^\dagger = \langle w|A^\dagger. \tag{4.108}$$

When A is a matrix, then we can think about † as literally meaning "transpose and complex conjugate." When V is a real vector space—so that linear transformations from $V \rightarrow V$ are also real—then † is equivalent to transpose.

Example 4.16.1. *Show that*

$$(AB)^\dagger = B^\dagger A^\dagger. \tag{4.109}$$

With foresight about naming variables, start with

$$\langle B^{\dagger} \boldsymbol{w}, \boldsymbol{v} \rangle = \langle \boldsymbol{w}, B \boldsymbol{v} \rangle \, . \tag{4.110}$$

Now assume that $\boldsymbol{w} \equiv A^{\dagger}\boldsymbol{u}$. Plugging this in and using the definition of the adjoint gives:

$$\langle B^{\dagger} A^{\dagger} \boldsymbol{u}, \boldsymbol{v} \rangle = \langle A^{\dagger}\boldsymbol{u}, B\boldsymbol{v} \rangle = \langle \boldsymbol{u}, (AB)\boldsymbol{v} \rangle = \langle (AB)^{\dagger}\boldsymbol{u}, \boldsymbol{v} \rangle \, . \tag{4.111}$$

This proves the proposed relation. You should recognize this from the real case where $(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}}$.

What does the adjoint of a matrix look like in terms of indices? In fact, one could ask the following question. What is the index structure of $A^{\dagger}$? We could even restrict for simplicity to real spaces so that we really want to know: what are the positions of the indices of A^{T}? We know that A^{T} is still a linear map from $V \to V$. This means it has one upper and one lower index. We can then ask how these indices are oriented:

$$A^{i}_{\ j} = \left(A^{\mathrm{T}}\right)^{j}_{\ i} \qquad \text{or} \qquad A^{i}_{\ j} = \left(A^{\mathrm{T}}\right)^{\ i}_{j} \, . \tag{4.112}$$

Which one is it? The fact that we write $A^{\mathrm{T}}\boldsymbol{w}$ in the defining relation for the adjoint means that A^{T} is a linear transformation from $V \to V$, which tells us that it is the same class of object as A and should thus have the same index structure. Write this index structure out explicitly in the definition of the adjoint, (4.107), written with the inner product notation replaced by the metric:

$$\left(A^{\mathrm{T}}\right)^{i}_{\ k} w^{k} v^{j} g_{ij} = w^{i} A^{j}_{\ k} v^{k} g_{ij} \, . \tag{4.113}$$

We are about to do something slick, but first we re-label the indices for future convenience. Relabeling dummy indices does not change anything.

$$\left(A^{\mathrm{T}}\right)^{i}_{\ k} w^{k} v^{j} g_{ij} = w^{k} A^{i}_{\ j} v^{j} g_{ki} \, . \tag{4.114}$$

On the right-hand side we chose indices so that we have $A^{i}_{\ j}$ as a factor. Conveniently, on left-hand side the dummy indices of the w and v coefficients match.

Because the above equality is true for *any* kets $\boldsymbol{v}$ and $\boldsymbol{w}$, then we can simply cancel $w^k v^j$ on both sides:

$$\left(A^{\mathrm{T}}\right)^i_{\ k} g_{ij} = A^i_{\ j} g_{ki} . \tag{4.115}$$

Exercise 4.16.1. *Explain why step from* (4.114) *to* (4.115) *is valid.*

Now we may contract both sides with $g^{j\ell}$. Using the fact that this is the inverse metric so that $g^{j\ell} g_{ij} = \delta^\ell_i$, we have

$$\left(A^{\mathrm{T}}\right)^\ell_{\ k} = g_{ki} A^i_{\ j} g^{j\ell}. \tag{4.116}$$

This gives an index definition of A^{T}.

Exercise 4.16.2. *Using similar manipulations, show that*

$$A^i_{\ j} = g_{j\ell} \left(A^{\mathrm{T}}\right)^\ell_{\ k} g^{ki} = \left(A^{\mathrm{T}}\right)^{\ j}_i . \tag{4.117}$$

We see that we end up with an unusual index structure on the right-hand side. At first glance we may be worried that this is not the index structure of a linear transformation from $V \rightarrow V$ like we are used to. However, we also now have the rule that indices are raised and lowered uniquely by the metric. Thus the expression $\left(A^{\mathrm{T}}\right)^{\ j}_i$ is completely determined from $\left(A^{\mathrm{T}}\right)^i_{\ j}$ through the implicit action of the metric.

Exercise 4.16.3. *Show that* $\left(A^\dagger\right)^\dagger = A$.

You may recall from linear algebra that there was something *nice* about symmetric matrices. It is the same thing that is *nice* about Hermitian operators in quantum mechanics. Both are **self-adjoint**, meaning $A^\dagger = A$. Self-adjoint operators are nice because they have real eigenvalues, even if the vector spaces they act on are not necessarily real.[1]

[1] You may care about this if A is supposed to correspond to a physical observable.

4.17 *Multilinear Transformations*

Linear algebra should feel very familiar because most of us have played with matrices in various forms throughout our mathematical education. If you are like me, then at some point in your education you hear the word '**tensor**' and

you feel like it is some exotic new class of object with a slew of unknown mathematical structure. This is not the case. We now make a point to understand what a tensor is and to realize that it is simply an extension of linear maps between vector spaces.

From the previous section we know that given a vector space V, perhaps $\mathbb{R}^n$, we can consider the space of linear transformations from $V \to V$.[1] This space formalizes what we called *matrices*. It gave us three maps: $V \to V$, $V^* \to V^*$, and $V \times V^* \to \mathbb{R}$. In the latter case, we called the transformation *bilinear* because it is linear in each argument.

The generalization that we now make is a linear transformation of the form:[2]

$$V \times \cdots \times V \times V^* \times \cdots \times V^* \to V\mathbb{R} \qquad (4.118)$$

That is: a linear map that takes some number of vectors and some number of covectors and returns a number. Because the map is linear in each argument, we say that they are *multilinear*. Maybe the reason why this is 'scary' the first time we see it is that the components of tensors, in general, have many indices.[3] From our experience with "matrices," we know that the multilinear map above can be interpreted in many ways. The same object may, for example, map some number of vectors and covectors into some other number of vectors and covectors.

Example 4.17.1. *Consider a multilinear map that takes two vectors and spits out a third vector. This is a tensor. As stated, we can think of this as a function $f(\boldsymbol{v}, \boldsymbol{w})$ where we understand that f itself is a vector.[4] We are assuming that we all agree on some basis and there's no need to explicitly write $\boldsymbol{v} = v^i \hat{\boldsymbol{e}}_{(i)}$. Multilinearity means*

$$f(\alpha\boldsymbol{v} + \beta\boldsymbol{u}, \boldsymbol{w}) = \alpha f(\boldsymbol{v}, \boldsymbol{w}) + \beta f(\boldsymbol{u}, \boldsymbol{w}) \qquad (4.119)$$

$$f(\boldsymbol{v}, \alpha\boldsymbol{w} + \beta\boldsymbol{x}) = \alpha f(\boldsymbol{v}, \boldsymbol{w}) + \beta f(\boldsymbol{v}, \boldsymbol{x}), \qquad (4.120)$$

where α and β are numbers. The function f is simply linear in each of its two arguments: it is multi*linear. (There's really nothing deep here, eh?) Then we recognize*

[1] If you did Exercise 4.15.4, then you know that these transformations are themselves a vector space. Then you could consider linear transformations on the space of linear transformations...

[2] And eventually we also include complex functions, where we replace $\mathbb{R} \to \mathbb{C}$.

[3] There are various phobias about insects with many legs. Perhaps for similar reasons.

[4] As I squint at my screen I realize that it may not be clear that f is boldfaced. You can mentally put a $\vec{f}$ everywhere.

that f itself has an index. So the above multilinearity conditions may be written in component form:

$$f^i(\alpha \boldsymbol{v} + \beta \boldsymbol{u}, \boldsymbol{w}) = \alpha f(\boldsymbol{v}, \boldsymbol{w}) + \beta f(\boldsymbol{u}, \boldsymbol{w}) \qquad (4.121)$$

$$f^i(\boldsymbol{v}, \alpha \boldsymbol{w} + \beta \boldsymbol{x}) = \alpha f(\boldsymbol{v}, \boldsymbol{w}) + \beta f(\boldsymbol{v}, \boldsymbol{x}) \,, \qquad (4.122)$$

where the relations are separately true for each index i.

Because you did Exercise (4.15.2)—*you did do this exercise, right?*—you now should intuit that you can represent the information in a multilinear transformation as some array of numbers. This is the same as saying that you can represent a linear transformation as a $N \times N$ array that we call a matrix.[1]

> [1]Even though I keep saying we need to stop calling it that.

Exercise 4.17.1. *Consider the multilinear map f in Example 4.17.1. Suppose the vector space is 2-dimensional, for simplicity. Write out the 2^3 components of the tensor corresponding to f. As extra credit[2], convince yourself that you can write this as a $2 \times 2 \times 2$ cube of numbers.*

> [2]Whenever I mention 'extra credit' please know that I am imagining the skit "That's Numberwang!" from the British comedy show *That Mitchell And Webb Look.*

Exercise 4.17.2. *Convince yourself[3] that the function f in Example 4.17.1 is* equivalently *all of the following:*

- ⋄ *A linear map from $V \times V \to V$.*

- ⋄ *A linear map from $V \to V^* \times V$. Note that the output of this map is a linear transformation from $V \to V$.*

- ⋄ *A linear map from $V^* \to V^* \times V^*$.*

- ⋄ *A linear map from $V^* \times V \to V^*$.*

- ⋄ *A linear map from $V \times V \times V^* \to \mathbb{R}$.*

Are there other linear maps one can construct from this?

> [3]This is my way of saying: you do not have to do a full mathematical proof, but think about this until it is absolutely clear that it is true.

Tensors are precisely these multilinear maps. This is in exactly the same that vectors and covectors are linear maps on each other. It is also in exactly the same way that linear transformations from $V \to V$ are also linear transformations from $V^* \to V^*$ and bilinear maps from $V \times V^* \to \mathbb{R}$. We could read all of this off of the index structure of the linear transformation. In the same way, the index structure of the multilinear transformations tell us all of these things.

4.18   *Tensor Indices*

Multilinear maps (tensors) are linear combinations of a tensor product of basis vectors and covectors. For every vector argument that a multilinear map takes in, there is a basis covector (a basis linear map from $V \rightarrow \mathbb{R}$). For every covector argument that a multilinear map takes in, there is a basis vector (a basis linear map from $V^* \rightarrow \mathbb{R}$). Alternatively, you do not have to think of the basis covector as taking in a vector as an argument, but rather outputting a covector as an output. All of this is clear from the index notation: a lower index can either contract with an upper index (eating a vector) or it can be a leftover index after other contracts (pooping a covector).

When you act with a tensor product on some object, one must specify how the bras and kets in the product are supposed to 'hit' the object. For example, consider the following tensor product:

$$\mathsf{T} \equiv T^{ijk}_{\ell mn} |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell | \otimes \langle m | \otimes \langle n | \,. \tag{4.123}$$

I can act with T on a vector $\boldsymbol{v} = v^i |i\rangle$. To do this, one of the basis bras in T needs to act on the ket in $\boldsymbol{v}$. We need to specify which one.[1] I could tell you in words that T acts on $\boldsymbol{v}$ through its fourth argument—the one corresponding to the $\langle k |$ basis bra. That is cumbersome, but gives the following:

$$\mathsf{T}\boldsymbol{v} = T^{ijk}_{\ell mn} v^p |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell | \otimes \langle m|p\rangle \otimes \langle n | \tag{4.124}$$

$$= T^{ijk}_{\ell mn} v^m |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell | \otimes \langle n | \,. \tag{4.125}$$

We have used $\langle m|p\rangle = \delta^m_p$. Now we notice one of the benefits of "lazy physicist" index notation: suppose we just looked at the *components* of $\mathsf{T}\boldsymbol{v}$. These are the coefficients $T^{ijk}_{\ell mn} v^m$. Simply by looking at these components, we know that

1. The object $\mathsf{T}\boldsymbol{v}$ has three vector indices and two dual vector indices. The upper index is contracted with a lower index is not a 'free' index, it is a 'dummy' index that has been summed over.

2. We know that the basis vectors must be $|i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell | \otimes \langle n |$ simply because they must contract the indices of the components.

[1] In practice we rarely actually have to to this. This is because we usually symmetrize and antisymmetrize the upper and lower indices so that it does not matter which index is acting on the object. In my life I am fortunate to rarely have to work with objects with large numbers of indices with mixed symmetrizations and antisymmetrizations. Then there are the folks who work on tensor networks...

3. We know *which* index in T contracts with the index in $\boldsymbol{v}$ because that is the dummy index contraction. We can see from $T^{ijk}_{\ell mn}v^m$ that the fifth index of T contracts with the upper index of $\boldsymbol{v} = v^m|m\rangle$.

In Section 4.19 we present one more powerful feature of index notation: our ability to read off how a given tensor changes when we either (a) transform our coordinates or (b) transform the physical object that the tensor describes.

Exercise 4.18.1. *This is the tensor version of Exercise 4.15.2. Convince yourself that any multilinear map that takes arguments from a vector space V and its dual V^* may be represented as a tensor. It helps to reduce this problem to multiple instances of linear maps as in Exercise 4.15.2.*

4.19 *Indices and Symmetries*

Earlier we made the point that writing objects in terms of indices makes quite a lot of sense when those objects are linear maps between vector spaces. There is a second reasons why physicists love indices: they are a natural language for capitalizing on the symmetries of a system.[1]

4.20 *Eigenvalues and Eigenvectors*

Given a sufficiently *nice* linear transformation, A^i_{j}, there is a particularly convenient basis: the **eigenvectors** of A. These are vectors $|\lambda\rangle$ such that

$$A|\lambda\rangle = \lambda|\lambda\rangle \,. \tag{4.126}$$

In other words, A acts on the eigenvector by rescaling. The rescaling coefficient is the eigenvalues. For *nice* transformations (see Section 1.3), there is a complete set of such vectors to span the vector space.

Exercise 4.20.1. *If you are particularly keen, figure out (or look up) the mathematical conditions for a transformation A to have a set of eigenvectors that form a basis of the vector space. As an added bonus, check that these eigenvectors are orthogonal.*

If you write a general vector $|v\rangle$ in terms of this eigenbasis,

$$|v\rangle = v^i|\lambda_{(i)}\rangle \,, \tag{4.127}$$

[1] The mathematical study of symmetries is called group theory. Sometimes people distinguish this from representation theory, which is loosely the study of how objects transform with respect to symmetries—in contrast to the underlying structure of a class of symmetries. This lies outside the scope of our course, but you are likely to have had some training in representation theory when you invoked Clebsch–Gordan coefficients in quantum mechanics.

Then the action of A on this vector is easy:

$$A|v\rangle = \sum_i \lambda_{(i)} v^i |\lambda_{(i)}\rangle \,. \tag{4.128}$$

In fact, assuming that all of the eigenvalues are non-zero, even the matrix inverse is easy:

$$A^{-1}|v\rangle = \sum_i \lambda_{(i)}^{-1} v^i |\lambda_{(i)}\rangle \,. \tag{4.129}$$

[1]This is a general truth that "inverse problems" are very difficult. This is well known in applied mathematics, but the era of 'big science' has made it clear that many of the frontiers in physics and astronomy are inverse problems.

Hey, that is kind of a big deal. Taking the inverse of a matrix is a much harder problem than acting with the matrix on a vector.[1] The first time you see this should have brought a deep joy to your life: if you can decompose a matrix (linear transformation) into its eigenvectors and eigenvalues, then taking the inverse transformation is simple.

4.21   *What about traces and determinants?*

4.22   *Determinants*

5   *Vector fields*

6 Function Space

6.1 Histogram Space

A funny vector space: histogram space

Here's a funny vector space that we're going to use as a pedagogical crutch. Imagine histogram-space. The basis vectors are:
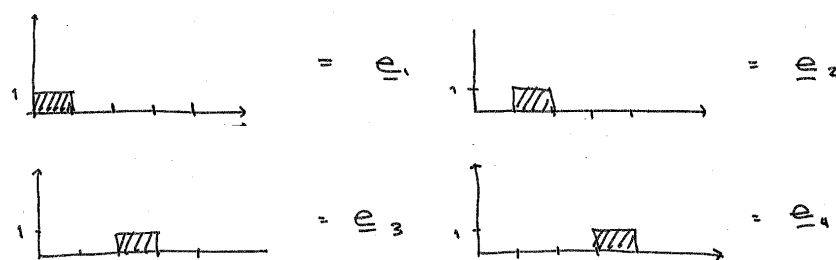
This is a basis for a histogram over unit bins from $x = 0$ to $x = 4$. A vector in this space is, for example:
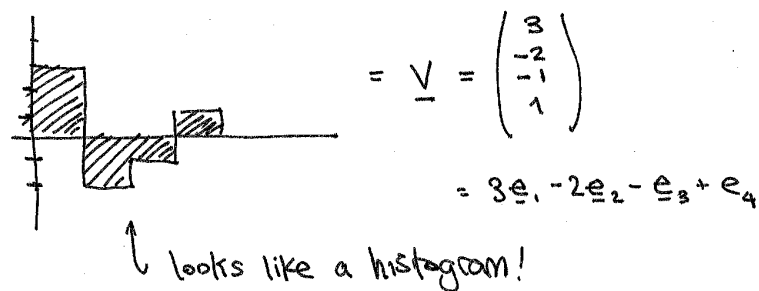

We can perform a linear transformation A on $\boldsymbol{v}$ which outputs another vector. Let's say it's this:

$$A\vec{v} = \begin{pmatrix} 1 \\ 3 \\ -2 \\ -1 \end{pmatrix}$$
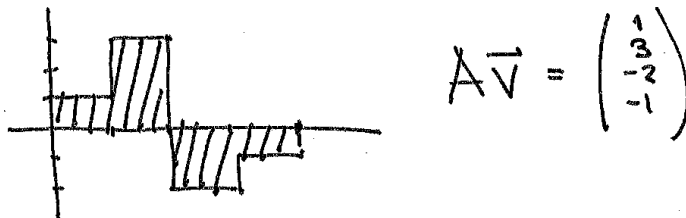
Exercise 6.1.1. *From the image above, can you derive what A is?*

The answer to the above exercise is *no*. Please make sure you convince yourself why: there are many different transformations that convert to old histogram into the new histogram. If you're not convinced: the matrix A is 4×4 and thus has 16 entries that we need to define. The matrix equation $A\boldsymbol{v} = \boldsymbol{w}$ for known vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ encodes only four equations.

The power of this admittedly strange formalism is that we can think of these histograms as approximations of continuous functions:
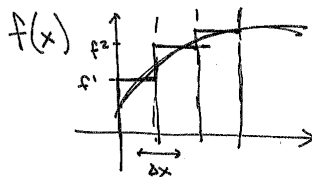
Thus a vector in this approximate (discretized) *function* space is

$$f = \begin{pmatrix} f^1 \\ f^2 \\ \vdots \\ f^N \end{pmatrix}. \tag{6.1}$$

In what way are polar coordinates an orthonormal basis?

Derivative Operators

Our discretized function space allows us to define a [forward] derivative[1]:

[1]One could have also defined a backward derivative where $(f')^i \sim f^i - f^{i-1}$. Note that you *cannot* try to make this symmetric by defining a 'centered' derivative like $(f')^i \sim f^{i+1/2} - f^{i-1/2}$ because there's no such thing as a fractional index. If you tried to write $(f')^i \sim f^{i+1} - f^{i-1}$ you're making a worse approximation. If you're like me, the fact

$$f' = \frac{1}{\Delta x} \begin{pmatrix} f^2 - f^1 \\ f^3 - f^2 \\ \vdots \\ f^{i+1} - f^i \\ \vdots \end{pmatrix} . \tag{6.2}$$

This is familiar if you've ever had to manually program a derivative into a computer program. Note that the right-hand side looks like a linear transformation of f. In other words, we expect to be able to write a matrix D so that

$$f' = Df . \tag{6.3}$$

One problem is apparent: what happens at the 'bottom' of the vector? What is the last component of the derivative, f'^N? Formally, this is

$$(f')^N = \frac{1}{\Delta x}(f^{N+1} - f^N) \tag{6.4}$$

but now we have no idea what f^{N+1} is. That was never a component in our vector space. There is no $e_{(N+1)}$ basis vector. This demonstrates and important lesson that we'll need when we move more formally to function spaces:

> **Boundary conditions are part of the definition of the function space**.

That was so important that I put the whole damn sentence in boldface and set it in the middle of the line. The significance of boundary conditions may be a bit surprising—but think of this as part of the definition of which functions we allow into our function space.

Example 6.1.1. *When one first learns about Fourier series with Dirichlet boundary conditions, one finds that the Fourier expansion* only *contains sines. The solution to the wave equation in such a system is some function that is zero at each endpoint. So the function space relevant to the system is composed only of functions that are zero at each endpoint.*

For now let's assume **Dirichlet boundary conditions**. A convenient way to impose this is to define what happens to all functions outside the domain of the function space:

$$f^{i>N} = f^{i<1} = 0 \ . \tag{6.5}$$

This solves the problem of the derivative on the last component:

$$(f')^N = \frac{1}{\Delta x}(f^{N+1} - f^N) = \frac{-f^N}{\Delta x} \ . \tag{6.6}$$

Alternatively, we could have also imposed **periodic boundary conditions**:

$$f^i = f^{i+kN} \qquad\qquad k \in \mathbb{Z} \ . \tag{6.7}$$

This would then give

$$(f')^N = \frac{1}{\Delta x}(f^{N+1} - f^N) = \frac{1}{\Delta x}(f^1 - f^N) \ . \tag{6.8}$$

Periodic boundary conditions amount to wrapping the x-axis into a circle. Older folks sometimes call this *Asteroids* boundary conditions. I'd also accept *Star Control* boundary conditions. Periodic boundary conditions show up *all* the time in physics. Sometimes they show up in obvious places, like the Brillouin zone of a crystal lattice. Other times they show up in not-so-obvious places like the boundary conditions of the known universe. In addition to being crucial for a well-defined function space, the boundary conditions of a system establish its topology[1].

Exercise 6.1.2. *We don't know anything about the universe outside the Hubble radius. Why do you think it would be reasonable in a physical model to assume that it has periodic boundary conditions? Hint: what would happen to the x-momentum of an asteroid in the classic arcade game* Asteroids *if the game did not have periodic boundary conditions?*

The second derivative may be defined symmetrically:

$$(f'')^i = \frac{(f^{i+1} - f^i) - (f^i - f^{i-1})}{\Delta x^2} \ . \tag{6.9}$$

[1] I cannot over-emphasize the importance of topology in contemporary physics. Most of the physics you will learn in your first year graduate courses are intrinsically *local* because the laws of physics are causal. Topological quantities are *global*, they are integrals over an entire space. Because winding numbers (and their higher-dimensional cousins) are quantized, they are robust against perturbations. The number of holes in a donut is one, whether or not it's been slightly squished in the box. By the way, the best donuts in Southern California are from *Sidecar Doughnuts* in Costa Mesa. Get the Basil Eggs Benedict donut before 11am; you can thank me later.

You may pontificate about the reason why the first derivative does have a symmetric discretization while the second derivative does.

Optional: why histogram space is dumb

The notion of building up continuous functions $f(x)$ from the histogram basis above turns out to be problematic from the point of view of mathematical rigor. You can read more about the problems in Appel, *Mathematics for Physicists* Chapter 9.1 (Insufficiency of Vector Spaces), or Hassani, *Mathematical Physics*, Chapter 7 (Hilbert Spaces). The problem in an infinite dimensional vector space is that one can take a convergent sequence of vectors whose limiting vector is not part of the vector space. This has to do with the definition of a norm of a vector with an infinite number of terms. An infinite dimensional vector space that *does* contain all of its limits (it is complete) is called a **Hilbert space**.

Derivatives in other function space bases

There are other ways to write a discrete basis of functions. Here is a natural one for functions that are up to second-order polynomials:

$$\boldsymbol{e}_{(0)} = 1 \qquad\qquad \boldsymbol{e}_{(1)} = x \qquad\qquad \boldsymbol{e}_{(2)} = x^2 \,. \qquad (6.10)$$

Let's sidestep questions about orthonormality for the moment. Clearly linear combinations of these basis functions can produce any quadratic function:

$$f(x) = ax^2 + bx + c \qquad\qquad \Longrightarrow \qquad\qquad \boldsymbol{f} = \begin{pmatrix} c \\ b \\ a \end{pmatrix}. \qquad (6.11)$$

The derivative operator has an easy representation in this space:

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}. \qquad (6.12)$$

We can see that

$$Df = \begin{pmatrix} b \\ 2a \\ 0 \end{pmatrix} \qquad\qquad D^2f = \begin{pmatrix} 2a \\ 0 \\ 0 \end{pmatrix} \qquad\qquad D^3f = 0 \,. \qquad (6.13)$$

The last line is, of course, the realization that the third-derivative of a quadratic function vanishes. Feel free to attach mathy words to this like *kernel*.

There are other bases that we may use for function space. A particularly nice one that we will use over and over is the Fourier basis, which we usually refer to as *momentum space*. The basis vectors are things like sines, cosines, or oscillating exponentials. These do not vanish for any power of D.

Locality

Notice that in the histogram basis, the derivative matrix D is sparse: it is zero everywhere away from the diagonal. The only non-zero elements on the ith row are around the $(i \pm 1)^{\text{th}}$ column. Higher powers of D sample further away, but the non-zero elements are always clustered near the diagonal.

This is simply a notion of **locality**. Remember the Taylor expansion:

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \cdots \,. \qquad (6.14)$$

If we think about the histogram as a discretization of a continuous function, then it is clear what the higher derivatives are doing. Given a function $f(x) = f$, one might like to know about the function around some point x_0 corresponding to some index i. That is: $f^i = f(x_0)$. If you'd like to learn more about the function around that point, one can express the derivative at x_0. Thus Df says something about the slope, D^2f says something about the curvature, and so on. Because each successive power of D samples terms further away from f^i, you can tell that these higher order terms are learning about the function further and further away from x_0.

Now think about the types of differential equations that you've encountered in physics. They often include one or two derivatives. You hardly ever see three, four, or more derivatives[1]. There's a reason for this: at the scales that we can access experimentally, nature appears to be local. Our mathematical models of nature typically have locality built in[2]. Physics at one spacetime point should

---

[1] With some thought, it may also be clear why spatial derivatives typically appear squared.

[2] A recent counterexample: `https://www.quantamagazine.org/physicists-discover-geometry-underlying-particle-physics-20130917/`

not depend on spacetime points that are far away.

This may be familiar from the idea of causality—the idea that *A causes B* therefore *A* must have happened *before B*. One of the key results in special relativity is that causality can be tricky if two events do not occur at the same spacetime point. More carefully, *A* can only cause *B* if there is a timelike separation of the appropriate sign. If we want to build causal theories of nature, then the dynamics at x_0 should not rely on what is happening at x_1, a finite distance away.[1]

[1] This is different from saying that information cannot propagate from x_0 to x_1; such propagation could come from some causal excitation of the electromagnetic field traveling every infinitesimal distance between the two positions. This is reminiscent of the classical Zeno's paradox.

APPENDIX

A *Conventions*

[20] [19]

Bibliography

[1]   Carl Bender. *Mathematical Physics - Lecture 1*. Nov. 2011. URL: https://pirsa.org/11110040.

[2]   3Blue1Brown. *How to lie using visual proofs*. 2022. URL: https://youtu.be/VYQVlVoWoPY.

[3]   John Baez. *Struggles with the Continuum (Part 6)*. 2016. URL: https://johncarlosbaez.wordpress.com/2016/09/21/struggles-with-the-continuum-part-6/.

[4]   J. Holt. *When Einstein Walked with Gödel: Excursions to the Edge of Thought*. Farrar, Straus and Giroux, 2018. ISBN: 9780374717841.

[5]   Randall Munroe. *Mathematical Symbol Fight*. 2016. URL: https://xkcd.com/2343/.

[6]   John H. Evans. "Dimensional Analysis and the Buckingham Pi Theorem". In: *American Journal of Physics* 40.12 (1972), pp. 1815–1822. DOI: 10.1119/1.1987069. URL: https://doi.org/10.1119/1.1987069.

[7]   R. W. Robinett. "Dimensional analysis as the other language of physics". In: *American Journal of Physics* 83.4 (2015), pp. 353–361. DOI: 10.1119/1.4902882. URL: https://doi.org/10.1119/1.4902882.

[8]   Fredrick Olness and Randall Scalise. "Regularization, renormalization, and dimensional analysis: Dimensional regularization meets freshman E&M". In: *American Journal of Physics* 79.3 (2011), pp. 306–312. DOI: 10.1119/1.3535586. URL: https://doi.org/10.1119/1.3535586.

[9]   Paul M. Stevenson. "Dimensional Analysis in Field Theory: An Elementary Introduction to Broken Scale Invariance and the Renormalization Group Equations". In: *Annals Phys.* 132 (1981), p. 383. DOI: 10.1016/0003-4916(81)90072-5.

[10]   Frederick P. Brooks. *The mythical man-month – Essays on Software-Engineering*. Addison-Wesley, 1975.

[11]   Craig F. Bohren. "Dimensional analysis, falling bodies, and the fine art of not solving differential equations". In: *American Journal of Physics* 72.4 (2004), pp. 534–537. DOI: 10.1119/1.1574042. URL: https://doi.org/10.1119/1.1574042.

[12] Patricia J. Yang et al. "Duration of urination does not change with body size". In: *Proceedings of the National Academy of Sciences* 111.33 (2014), pp. 11932–11937. DOI: 10.1073/pnas.1402289111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1402289111.

[13] Ezra Klein. *Transcript: Ezra Klein interviews Céline Gounder*. Aug. 2021. URL: https://www.nytimes.com/2021/08/06/podcasts/transcript-ezra-klein-interviews-celine-gounder.html.

[14] Nicole Meyer-Vernet and Jean-Pierre Rospars. "How fast do living organisms move: Maximum speeds from bacteria to elephants and whales". In: *American Journal of Physics* 83.8 (2015), pp. 719–722. DOI: 10.1119/1.4917310. URL: https://doi.org/10.1119/1.4917310.

[15] Robert D. Howe and Richard E. Kronauer. "Thomas McMahon: A Dedication in Memoriam". In: *Annual Review of Biomedical Engineering* 3.1 (2001). PMID: 11712277, pp. xv–xxxix. DOI: 10.1146/annurev.bioeng.3.1.0. URL: https://doi.org/10.1146/annurev.bioeng.3.1.0.

[16] Mikio Nakahara. *Geometry, topology and physics*. Graduate student series in physics. Bristol: Hilger, 1990.

[17] Joseph W Weinberg. "The geometry of colors". In: *General relativity and gravitation* 7.1 (1976), pp. 135–169.

[18] Alexander Logvinenko and Vladimir Levin. *Foundations of Colour Science: From Colorimetry to Perception*. Wiley, 2022. ISBN: 978-1-119-88591-7.

[19] K. Segrave. *Tipping: An American Social History of Gratuities*. McFarland, Incorporated, Publishers, 2009. ISBN: 9780786442461.

[20] Robert Bringhurst. *The Elements of Typographic Style*. Hartley & Marks Publishers, 2002. ISBN: 978-0881791327.

Index