

UNIVERSITY OF CALIFORNIA, RIVERSIDE

PHYSICS 231, FALL 2022

Flip Tanedo

Physics 231: Methods of Theoretical Physics

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

Lecture Notes

September 21, 2022



Flip Tanedo, *Physics 231: Methods of Theoretical Physics*, A mathematical methods course for first-year graduate students in physics & astronomy, September 21, 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 21, 2022

Contents

Abstract 5

- 1 Mathematical Methods 9
 - 1.1 Green's functions and this course 9
 - 1.2 This is not what I expected 10
 - 1.3 The non-mathematical idea of mathematical niceness 11
 - 1.4 The unbearable arrogance of physicists 12
 - 1.5 Footnotes 14
 - 1.6 One last piece of advice 14
- 2 Physics versus Mathematics 17
 - 2.1 The most important binary relation 17
 - 2.2 Units 18
- 3 Dimensional Analysis 19
 - 3.1 Converting Units 19
 - 3.2 Quantifying units 20
 - 3.3 Dimensional analysis at work 21

I LINEAR ALGEBRA

- 4 Finite-Dimensional Linear Algebra 33
 - 4.1 Yet another review of linear algebra 33
 - 4.2 What is Linear? 33
 - 4.3 Vectors and Vector Spaces 35
 - 4.4 Notation for Vectors 36
 - 4.5 Basis Vectors 37
 - 4.6 Nice basis vectors 39
 - 4.7 The metric, inner product, or dot product 41
 - 4.8 Inner product as a projection 45
 - 4.9 Linear functions of vectors 47
 - 4.10 Summation convention 53
 - 4.11 Linearity and Indices 54
 - 4.12 Duality 56
 - 4.13 Metric and Indices 60
 - 4.14 Linear Transformations 61

- 4.15 Indices and Symmetries 62
- 4.16 Multilinear Notation 62
- 4.17 Examples of Finite-Dimensional Vector Spaces 63
- 4.18 Determinants 63

5 Vector fields 65

6 Function Space 67

APPENDIX

A Conventions 71

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

¹In other words, this is *your* Ph.D, craft it appropriately.

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:

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

In this equation, \mathcal{O} is a *differential operator* that encodes some kind of physical dynamics², $s(x)$ 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:

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

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} .

²A **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, $\mathcal{O} = (d/dx)^2 + 3x(d/dx)$.

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

$$A\boldsymbol{v} = \boldsymbol{w}, \quad (1.3)$$

whose solution is

$$\boldsymbol{v} = A^{-1}\boldsymbol{w}. \quad (1.4)$$

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 infinities 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 undergraduate 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.

¹Carl 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 Wronskians 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

²3Blue1Brown. *How to lie using visual proofs*. 2022. URL: <https://youtu.be/VYQV1VoWoPY>.

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:*

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

¹This 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.

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

³Full 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, $r \rightarrow 0$, 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

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:

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

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

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

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

²The 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:

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

This is a truth in many different respects¹:

- ◊ 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.

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

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

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

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 \equiv or \doteq to mean something like 'definition' or 'tautologically equivalent to' or some other variant of *even more equal than equal*.

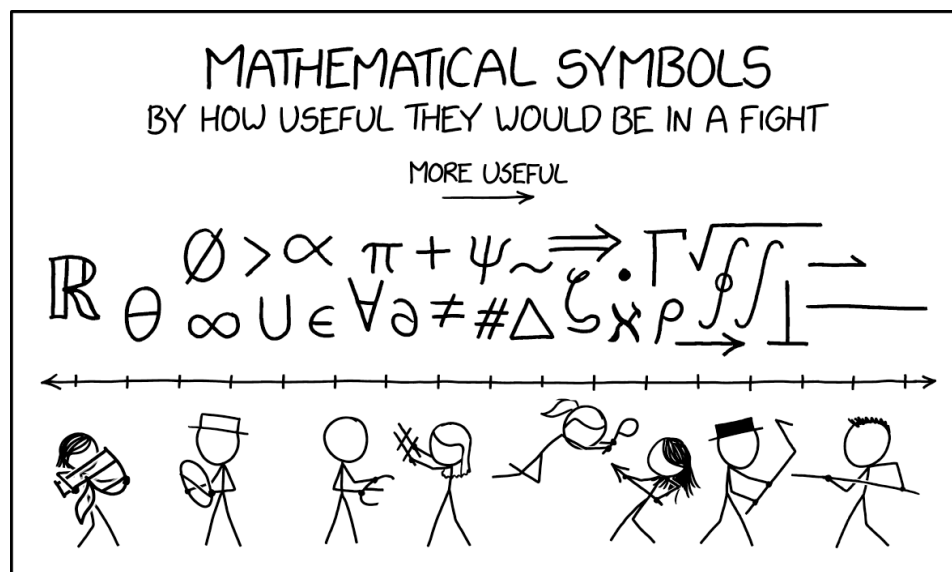
As physicists the most important binary relation is none of those things². What we usually care about is \sim .³ The symbol \sim tells us how how something *scales*. If I double a quantity on the right-hand side, how does the quantity on

²I thank Yuval Grossman teaching me this.

³I 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 \sim 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

¹That 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 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) and 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*:

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

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

$$1 \text{ apple} = \$1 \quad \Rightarrow \quad 1 = \frac{\$1}{1 \text{ apple}} . \quad (3.2)$$

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**:

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

At face value, this does not make sense. \hbar has units of action, c is a speed, and 1 is dimensionless. 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),

$$c = 3 \times 10^{10} \text{ cm/s} . \quad (3.4)$$

If $c = 1$, then

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

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:

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

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*

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

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:*

$$[F] = [c^{-3} \hbar F] = M^2. \quad (3.8)$$

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:

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

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

$$e^x = 1 + x + \frac{1}{2!}x^2 + \dots \quad (3.10)$$

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

Exercise 3.3.1. 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?

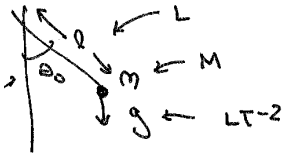


Figure 3.1.
Sketch of a simple pendulum.

Solving problems

Here is a common problem in introductory physics. Assume you have a pendulum 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 acceleration, $[g] = LT^{-2}$. 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 :

$$\tau \sim g^{-1/2} \ell^{1/2}. \quad (3.11)$$

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:

$$\tau = f(\theta_0) \sqrt{\frac{\ell}{g}}. \quad (3.12)$$

[Flip: To do: include problems from Robinett AJP article on dimensional analysis, doi: 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 *Mathematical Methods of Classical Mechanics*.¹ Suppose you have some static, central potential $U(\mathbf{r})$. Maybe it's some planet orbiting a star.

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

SUPPOSE:

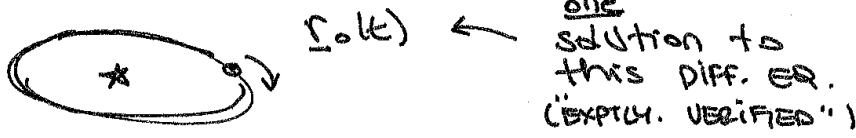


Figure 3.2.

A orbital trajectory, $r_0(t)$.

The force law gives:

$$m\ddot{\mathbf{r}} = -\frac{\partial U}{\partial \mathbf{r}}. \quad (3.13)$$

Suppose we are given a solution, $\mathbf{r}_0(t)$. Perhaps this is a trajectory that is experimentally 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' :

$$t \equiv \alpha t'. \quad (3.14)$$

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:

$$m \left(\frac{d}{dt} \right)^2 \mathbf{r}_0(t) = m\alpha^{-2} \left(\frac{d}{dt'} \right)^2 \mathbf{r}_0(\alpha t'). \quad (3.15)$$

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

$$m' \left(\frac{d}{dt'} \right)^2 \mathbf{r}_0(\alpha t') = -\frac{\partial U}{\partial \mathbf{r}_0}. \quad (3.16)$$

What this tells us is that we may define a new trajectory, $\mathbf{r}_1(t') \equiv \mathbf{r}_0(\alpha t')$, 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:

$$m' \left(\frac{d}{dt} \right)^2 \mathbf{r}_1(t) = -\frac{\partial U}{\partial \mathbf{r}_1}, \quad (3.17)$$

which is indeed¹ (3.13) with a new mass m' and a trajectory $\mathbf{r}_1(t') \equiv \mathbf{r}_0(\alpha t')$. For ¹We were able to swap \mathbf{r}_0 with \mathbf{r}_1 simply because U only depends on the position.

example, if $\alpha = 2$, then $\mathbf{r}_1(t)$ traces the same trajectory at double the velocity with one fourth of the mass.

Exercise 3.3.2. *I missed something in the example above. In order for a planet of mass m' to have trajectory $\mathbf{r}_1(t')$, what is the mass of the star compared to the original mass M_\star ?*¹

¹Thanks to Eric Zhang (2021) for pointing this out.

Example 3.3.1. *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. Anyone 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

¹⁰Frederick P. Brooks. *The mythical 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.

You can integrate the force equation to get

$$t_0 = \sqrt{\frac{2h}{g}}. \quad (3.18)$$

¹¹Craig 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>.

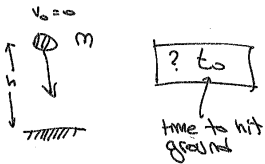


Figure 3.3. Dropping a ball of mass m .

This is the *exact* answer *within our model* of the system. The model made several 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 know is: *how good of an approximation is it?*

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 appreciate 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

$$\varepsilon = \frac{t_1 - t_0}{t_0} . \quad (3.19)$$

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 small¹. Then the error is a function of some dimensionless parameters, ξ , in the system. We define these ξ so that as $\xi \rightarrow 0$, $\varepsilon(\xi) \rightarrow 0$. In other words, the approximation gets better as the ξ are made smaller. By Taylor expansion:

$$\varepsilon(\xi) = \varepsilon(0) + \varepsilon'(0)\xi + \mathcal{O}(\xi^2) . \quad (3.20)$$

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.²

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 :

$$\xi = \frac{h}{R} \quad \text{or} \quad \xi = \frac{R}{h} . \quad (3.21)$$

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)

¹Note the error has to be dimensionless in order for us to be able to call it ‘small,’ otherwise it begs the question of ‘small with respect to what?’

²There 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.

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 3.3.3. 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:

$$\ddot{h} = \frac{-g}{\left(1 + \frac{h}{R}\right)^2} . \quad (3.22)$$

Taylor expand to reduce this to an equation of the form

$$\frac{d^2 q}{ds^2} = -1 + 2q , \quad (3.23)$$

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

$$q(s) = c_1 e^{\sqrt{2}s} + c_2 e^{-\sqrt{2}s} + \frac{1}{2} . \quad (3.24)$$

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

$$\frac{2q(s) - 1}{2q(0) - 1} = \cosh(\sqrt{2}s) . \quad (3.25)$$

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

$$s_1 = \frac{2q_0}{1 - 2q_0} . \quad (3.26)$$

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 $\mathcal{O}(h/R)$.

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

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 3.3.4. *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.*

Exercise 3.3.5. *David Hu won his first IgNobel prize for determining that mammals take about 21 seconds to urinate, largely independently of their size¹. Can you use dimensional analysis to argue why this would be the case? It may be helpful to refer to the paper¹². 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.*

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:

$$P + \rho gh = \frac{1}{2}\rho v^2, \quad (3.27)$$

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

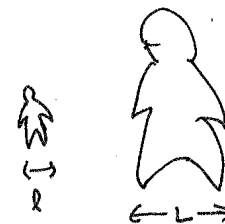


Figure 3.4.
Two mathematically similar people.

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

¹²Patricia 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>.

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 3.3.6. 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 individual 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 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 Physics¹⁴ and the references therein.

Exercise 3.3.7. 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 3.3.8. 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:

Argue from dimensional analysis that $v_{\text{spr}}/L \sim b\rho/\sigma$. (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 3.3.9. 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 Thomas McMahon’s article “The Mechanical Design of Trees” in Scientific American volume 233 (1975)³. McMahon was the first to propose a physical explanation

¹³Ezra Klein. Transcript: Ezra Klein interviews Céline Goudner. Aug. 2021. URL: <https://www.nytimes.com/2021/08/06/podcasts/transcript-ezra-klein-interviews-celine-goudner.html>.

¹Early in the COVID-19 pandemic, many physicists became armchair modelers of epidemics. Some of this was driven by hubris about our mathematical intuition. Many of the physicists lost interest when their models aligned poorly with what actually happened.

¹⁴Nicole 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>.

²https://starwars.fandom.com/wiki/All_Terrain_Armored_Transport

³<https://www.jstor.org/stable/24949846>

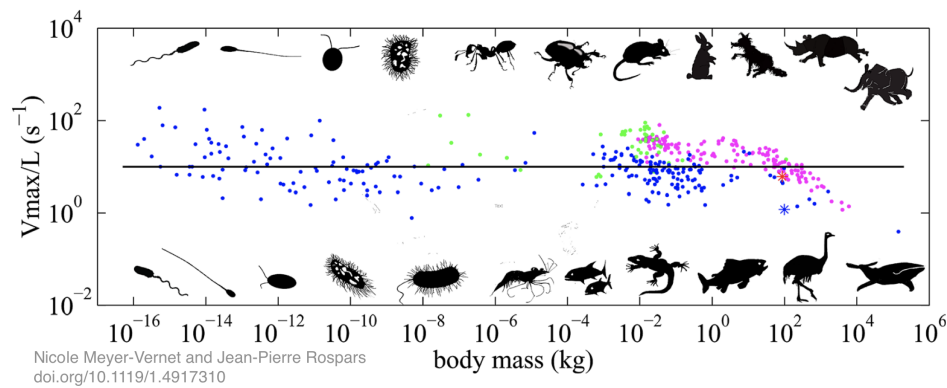


Figure 3.5.
Image from Meyer-Verneta
and Rospars. [14]

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

¹⁵Robert 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>.

PART I

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:

$$\text{differential operator} = \infty\text{-dimensional matrix} . \quad (4.1)$$

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:

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

$$\text{function space} = \infty\text{-dimensional vector space} . \quad (4.2)$$

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 things², 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.

²For 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 $f : \mathbb{R} \rightarrow \mathbb{R}$.

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 $f : V \rightarrow W$ is **linear** if and only if¹

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

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 numbers²: 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.

¹One can also write “ \Leftrightarrow ” to mean ‘if and only if.’

²More generally, these are elements of a **field**: sets of objects with addition and multiplication defined.

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,

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

This map is not linear because

$$R(\theta_1 + \theta_2) \neq R(\theta_1) + R(\theta_2), \quad (4.7)$$

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?

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.

¹Mathematicians chuckle at freshman physics textbooks.

¹⁶Mikio 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 \mathbf{v} and \mathbf{w} be any two vectors in the vector space V ; we write this as $\mathbf{v}, \mathbf{w} \in V$. Further, let α, β be numbers. We have the following rules:

- ◊ The sum $\alpha\mathbf{v} + \beta\mathbf{w}$ is a vector in V .
- ◊ The zero vector $\mathbf{0}$ is a vector in V such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$.

¹Please 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, $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$, and associative, $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$.

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: $\mathbf{0} + \mathbf{x} = \mathbf{x}$. 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 $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ 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 colors¹⁷ or the overlapping sciences of physics and neurological perception of color¹⁸.*

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

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

²As a student I was once very because a textbook introduced the notation $\vec{\partial}$. 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} .

¹See 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 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 extension² in physics is the notation v^i . Technically, v^i refers to the i^{th} 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.

²https://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 $2\boldsymbol{v}$ and $3\boldsymbol{v}$, 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

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

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

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

$$\hat{\mathbf{e}}_{(1)} \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \hat{\mathbf{e}}_{(2)} \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \quad (4.9)$$

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 \mathbf{v} as a linear combination,

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

We could have defined a different basis, for example

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

In this primed basis the vector \mathbf{v} would have different components,

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

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.’

¹Please humor the unusual notation with a lower index in parenthesis. We explain this unusual choice in Section 4.13.

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,

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

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

Exercise 4.5.2. In the example above, write \mathbf{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 \mathbf{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 \mathbf{v} are unique.

¹Then 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)

4.6 Nice basis vectors

You have likely been trained to *assume* that a basis is *nice*, following the notion of assumed *nice*ness² 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.

²Section 1.3.

Linear independence

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

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

where the sum is over all basis vectors $\mathbf{e}_{(j)}$ except the i^{th} basis vector. It should be obvious³ that a proposed basis with a linearly dependent basis vector

³In the sense of Section 1.4.

1. Does not uniquely define the components of some vectors \mathbf{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

¹From 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 :*

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

These vectors are obviously linearly independent. Assuming the usual Euclidean inner product, they 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 \mathbf{v} be the vector that is pointing in the $\hat{\mathbf{x}}$ direction of \mathbb{R}^2 . What are the components of \mathbf{v} with respect to the basis in (4.15)? Similarly, let \mathbf{w} be the vector that is pointing in the $\hat{\mathbf{y}}$ direction of \mathbb{R}^2 . What are the components of \mathbf{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 4.7.2.

4.7 The metric, inner product, or dot product

The **metric** (inner/dot product) on a real vector space V is a bilinear¹ map from $V \times V \rightarrow \mathbb{R}$.² 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 \mathbf{v} with respect to \mathbf{w} or vice versa: their ‘overlap’ should be the same. We use the angle bracket notation where $\langle \mathbf{v}, \mathbf{w} \rangle$ is the inner product of two vectors \mathbf{v} and \mathbf{w} .³ Summarizing the above properties in equations:

$$\langle \alpha \mathbf{v} + \beta \mathbf{w}, \mathbf{u} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle + \beta \langle \mathbf{w}, \mathbf{u} \rangle \quad (4.16)$$

$$\langle \mathbf{v}, \alpha \mathbf{w} + \beta \mathbf{u} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{u} \rangle \quad (4.17)$$

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle. \quad (4.18)$$

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

$$|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \quad (4.19)$$

Because we want lengths to make sense, we make a further requirement that any sensible metric is *positive definite*,⁴

$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \quad (4.20)$$

¹This simply means linear in each argument.

²We also care about *complex* vector spaces, in which case the metric is a map from $V \times V \rightarrow \mathbb{C}$.

³This is deliberately suggestive of the ket notation for vectors, Sec. 4.4.

⁴Mathematicians use the adjective **Riemannian** to mean ‘positive definite’ when applied to metrics. Physicists often assume that their metrics are Riemannian.

for any vector $\mathbf{v} \in V$. Both the angle bracket notation $\langle \mathbf{v}, \mathbf{w} \rangle$ and the dot product notation $\mathbf{v} \cdot \mathbf{w}$ 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 $V \times V$:

$$g(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle. \quad (4.21)$$

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

Example 4.7.1. *The Euclidean metric in Cartesian coordinates is the usual dot product from childhood. In \mathbb{R}^2 , the dot product of two vectors is $\mathbf{v} \cdot \mathbf{w} = v^1 w^1 + v^2 w^2$. The generalization to \mathbb{R}^N is*

$$\langle \mathbf{v}, \mathbf{w} \rangle = g(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^N v^i w^i, \quad (4.22)$$

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

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

$$|\mathbf{v}| = \sqrt{\mathbf{v}^T \mathbf{v}}.$$

Observe that there is no inner product in this expression: the ‘row vector’ \mathbf{v}^T naturally acts on the column vector \mathbf{v} by matrix multiplication.² This suggests that the inner/dot product is related to

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w}. \quad (4.23)$$

To be fully general, we write

$$\mathbf{v} \cdot \mathbf{w} \equiv \mathbf{v}^T \mathbf{g} \mathbf{w} \quad \mathbf{g} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.24)$$

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

¹Some suggest that the g stands for ‘gravity’ <https://hsm.stackexchange.com/questions/3435/>, see also https://www.reddit.com/r/math/comments/p2i6qu/why_is_a_riemannian_metric_tensor_denoted_by_g/.

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

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

$$ds^2 = dx^2 + dy^2 . \quad (4.25)$$

This generalizes to the form

$$ds^2 = \sum_{i,j=1}^2 g_{ij} dx^i dx^j , \quad (4.26)$$

where g_{ij} are the components of \mathbf{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 are 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.

¹Terms 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,

²We deliberately postpone defining what we mean by tensor.

$$g(\mathbf{v}, \mathbf{w}) \equiv \sum_{i,j} g_{ij} v^i w^j . \quad (4.27)$$

For the case of \mathbb{R}^2 in Cartesian coordinates, (4.27) matches (4.22) when we take the components of g_{ij} to be

$$g_{ij} = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} , \quad (4.28)$$

where we define a cousin³ of the Kronecker δ . If you are not familiar with summing over indices, please check this explicitly: (4.27) has a double sum over indices i and j , whereas (4.22) 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 (4.27) is precisely the same as the one that shows up in (4.26).

³Formally 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 4.7.1. 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 (4.24). Why can any bilinear function of two vectors in \mathbb{R}^2 be written with respect to the four

components of \mathbf{g} ? What are the restrictions on the components of \mathbf{g} given that metric is symmetric and positive definite?

Example 4.7.2. *The two-dimensional Minkowski space metric in the standard basis is*

$$\langle \mathbf{v}, \mathbf{w} \rangle = v^0 w^0 - v^1 w^1, \quad (4.29)$$

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

$$\mathbf{v} = \begin{pmatrix} v^0 \\ v^1 \end{pmatrix}, \quad (4.30)$$

with v^0 being the time-like component. The generalization to $(d + 1)$ -dimensional Minkowski space¹ is simply

$$\langle \mathbf{v}, \mathbf{w} \rangle = v^0 w^0 - \vec{\mathbf{v}} \cdot \vec{\mathbf{w}}, \quad (4.31)$$

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.

Exercise 4.7.2. *The **Gram–Schmidt** procedure. Given the two vectors*

$$\mathbf{v} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad \mathbf{w} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad (4.32)$$

derive an orthonormal basis on \mathbb{R}^2 using the canonical² 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 4.7.3. *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?*

¹This 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 $\mathbb{R}^{(2,d)}$.

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

Exercise 4.7.4. Write out the Euclidean metric for polar coordinates in \mathbb{R}^2 , (r, θ) . These are related to ordinary Cartesian coordinates by:

$$(x, y) = (r \cos \theta, r \sin \theta) . \quad (4.33)$$

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

4.8 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, $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$, it does not matter which vector is being projected onto the other. This projection gives a definition of the angle, θ , between two vectors:

$$\langle \mathbf{v}, \mathbf{w} \rangle \equiv |\mathbf{v}| |\mathbf{w}| \cos \theta . \quad (4.34)$$

Writing everything explicitly in terms of the inner product:

$$\cos \theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle}} . \quad (4.35)$$

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 (4.35) generalizes to *any* inner product.¹

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

$$v^i = \langle \hat{\mathbf{e}}_{(i)}, \mathbf{v} \rangle . \quad (4.36)$$

Exercise 4.8.1. Explain what goes wrong when the basis is not orthonormal. Give an explicit example in \mathbb{R}^2 . Define a non-orthonormal basis and a vector \mathbf{v} with components v^i with respect to the basis. Then take the inner products with the basis vectors to show that $v^i \neq \langle \hat{\mathbf{e}}_{(i)}, \mathbf{v} \rangle$.

¹One 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.

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

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

$$\langle \hat{\mathbf{e}}_{(i)}, \hat{\mathbf{e}}_{(j)} \rangle = \delta_{ij} . \quad (4.37)$$

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

$$g_{ij} = \langle \hat{\mathbf{e}}_{(i)}, \hat{\mathbf{e}}_{(j)} \rangle . \quad (4.38)$$

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

$$\langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \sum_i v^i \hat{\mathbf{e}}_{(i)}, \sum_j w^j \hat{\mathbf{e}}_{(j)} \right\rangle = \sum_{i,j} v^i w^j \langle \hat{\mathbf{e}}_{(i)}, \hat{\mathbf{e}}_{(j)} \rangle = \sum_{i,j} g_{ij} v^i w^j , \quad (4.39)$$

where the last equality identifies the metric components g_{ij} in (4.27). 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 4.8.2. Show that a basis is linearly independent if and only if

$$\det \langle \hat{\mathbf{e}}_{(i)}, \hat{\mathbf{e}}_{(j)} \rangle \neq 0 . \quad (4.40)$$

¹This phrase is code for “you can prove it, but we do not prove it here.”

³The absolute value is important here because the space-time metric has a relative sign between timelike and spacelike components, (4.31).

Because the metric is symmetric, it turns out² 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

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 choose 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

4.9 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,

$$\mathbf{v} = \begin{pmatrix} v^1 \\ \vdots \\ v^N \end{pmatrix}, \quad (4.41)$$

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

$$\tilde{\mathbf{w}} = (w_1, w_2, \dots, w_N), \quad (4.42)$$

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 $\tilde{\mathbf{w}}$ and \mathbf{v} :

1. Rotate the first object by 90° in the clockwise direction. This means that $\tilde{\mathbf{w}}$ 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: $w_1 v^1 + w_2 v^2 + \dots$.

¹An 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.

²In 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.

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

$$\tilde{\mathbf{w}}\mathbf{v} = \sum_i w_i v^i, \quad (4.43)$$

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 $\tilde{\mathbf{w}}$ and on \mathbf{v} . This means if we replace $\tilde{\mathbf{w}}$ with $2\tilde{\mathbf{w}}$, then the number on the right-hand side doubles, and similarly with \mathbf{v} . In fact, we can be even more explicit about the role of $\tilde{\mathbf{w}}$ as a function that takes in vectors by writing this action as $\tilde{\mathbf{w}}[\mathbf{v}]$.

Exercise 4.9.1. 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 identical¹ 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.

¹Perhaps a more mathematical word is ‘isomorphic.’

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 $\tilde{\mathbf{e}}^{(i)}$ so that a row vector $\tilde{\mathbf{w}}$ may be expanded as

$$\tilde{\mathbf{w}} = \sum_i w_i \tilde{\mathbf{e}}^{(i)}, \quad (4.44)$$

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 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 basis²

We underline the following point:

²If 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.

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

Given an orthonormal basis $\{\hat{e}_{(i)}\}$ 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.¹ These are the basis linear functions that satisfy:

¹Is it clear that these are all the same things? We can then just call them dual vectors or bras.

$$\tilde{e}^{(i)}[\hat{e}_{(j)}] = \delta_j^i = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (4.45)$$

We read this as a basis dual vector $\tilde{e}^{(i)}$ acting² on a basis vector $\hat{e}_{(j)}$. This canonical choice of basis realizes the rule (4.43) with the coefficients w_i identified as the components of \tilde{w} .

²I always thought about this as Pac-Man eating the ghosts.

δ_j^i 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.

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, $V \times V \rightarrow \mathbb{R}$. A row vector takes in a single vector and spits out a number, $V \rightarrow \mathbb{R}$.³ 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.

³This 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 4.9.2. *If you do not already know the answer, then pause and take a moment to think how you would create a linear function from $V \rightarrow \mathbb{R}$ 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.*⁴

⁴*Inception* (2010), Christopher Nolan.

The idea is that we can *pre-load* the metric with a vector. Pick any vector, v . Now insert v into the first argument of the inner product, $\langle v, _ \rangle$. 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, $\langle \mathbf{v}, _ \rangle$ is a linear function from $V \rightarrow \mathbb{R}$. This is precisely a dual vector. In fact, let us give this object a name:

$$\tilde{\mathbf{v}} = \langle \mathbf{v}, _ \rangle. \quad (4.46)$$

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 (4.45)¹, we can find the components of a dual vector by feeding it basis vectors:

$$\tilde{\mathbf{v}} [\hat{\mathbf{e}}_{(i)}] = \sum_j v_j \hat{\mathbf{e}}^{(j)} [\hat{\mathbf{e}}_{(i)}] = v_j. \quad (4.47)$$

Applying our definition of $\tilde{\mathbf{v}}$ as a pre-loaded metric (4.46),

$$\tilde{\mathbf{v}} [\hat{\mathbf{e}}_{(i)}] = \sum_j v^j \langle \hat{\mathbf{e}}_{(j)}, \hat{\mathbf{e}}_{(i)} \rangle = v^i. \quad (4.48)$$

From this we find that $v_j = g_{ji} v^i$, where we have used the tensor notation of the metric (4.27). 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:

$$\mathbf{v}^T = \tilde{\mathbf{v}} = \langle \mathbf{v}, _ \rangle. \quad (4.49)$$

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, $\mathbf{v}^\dagger = \tilde{\mathbf{v}}$. If you are not sure whether a space is real or complex, you can say Hermitian conjugate to cover both cases and sound fancy.²

²This 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.³ 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,

³We 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.

and the action on the components is clear: $v_i = \sum_j g_{ij} v^j$. In other words, the metric g_{ij} is a machine that can lower indices.

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, $\mathbf{v} = |v\rangle$. A dual vectors are written as a bra, $\tilde{\mathbf{w}} = \langle w|$. The basis bras and kets are written as follows:

$$|v\rangle = v^i |i\rangle \quad \Leftrightarrow \quad \mathbf{v} = v^i \hat{\mathbf{e}}_{(i)} \quad (4.50)$$

$$\langle w| = w_j \langle j| \quad \Leftrightarrow \quad \tilde{\mathbf{w}} = w_j \tilde{\mathbf{e}}^{(j)} . \quad (4.51)$$

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 transpose¹. Given a ket $|v\rangle$, there is a natural understanding what the corresponding bra is with respect to the inner product:

$$\langle v| = \langle v|_{\perp} , \quad (4.52)$$

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

$$\langle v| [|w\rangle] = \langle v|w\rangle = \langle v, w\rangle . \quad (4.53)$$

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

¹We should really say ‘Hermitian conjugate’ because the bra/ket notation is most commonly used for Hilbert spaces, which are complex vector spaces.

The form notation

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

$$\mathbf{v} = v^i \hat{\mathbf{e}}_{(i)} \leftrightarrow v^i \frac{\partial}{\partial x^i} \quad \tilde{\mathbf{w}} = w_i \tilde{\mathbf{e}}^{(i)} \leftrightarrow w_i dx^i. \quad (4.54)$$

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 4.9.3. *Confirm that the set of partial derivatives $\partial_x, \partial_y, \partial_z$ 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

$$dx^i \left[\frac{\partial}{\partial x^j} \right] \equiv \delta_j^i, \quad (4.55)$$

which you already know from (4.45). 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

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

4.10 *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

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:

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

$$\tilde{\boldsymbol{w}}\boldsymbol{v} = w_i v^i = w_1 v^1 + w_2 v^2 + \dots, \quad (4.56)$$

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

²Nevermind the indignities of using a whiteboard.

Exercise 4.10.1. *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.*

³A 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.

$$v^i w^i = w^1 v^1 + w^2 v^2 + \dots \quad (4.57)$$

$$g_{ij} v^i w^j u^j = g_{11} v^1 w^1 u^1 + g_{12} v^1 w^2 u^2 + \dots + g_{21} v^2 w^1 u^1 + \dots \quad (4.58)$$

$$g_{ii} v^i w^i = g_{11} v^1 w^1 + g_{22} v^2 w^2 + \dots \quad (4.59)$$

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.

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

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

³There 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.

Example 4.10.1. *Linear transformations are linear functions that take vectors to vectors, $V \rightarrow V$. These are represented by matrices, see Sec. 4.14 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 $A^i_i = A^1_1 + A^2_2 + \dots$. 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.*

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?

4.11 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.⁴ 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.

⁴You 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 4.11.1. 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 4.11.1. Suppose you have a linear function in one dimension, $f(\mathbf{x})$. For simplicity, let \mathbf{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

$$f(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) . \quad (4.60)$$

The only way this is possible is if

$$f(\mathbf{x}) = a\mathbf{x} , \quad (4.61)$$

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

Example 4.11.2. Same as the previous example, but now suppose that \mathbf{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.

$$f \left[\alpha \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \beta \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \right] = \alpha f \left[\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \right] + \beta f \left[\begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \right] . \quad (4.62)$$

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

$$f \left[\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \right] = ax^1 + (\text{independent of } x^1) , \quad (4.63)$$

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

$$f \left[\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \right] = bx^2 + (\text{independent of } x^2) , \quad (4.64)$$

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

$$f(\mathbf{x}) = f_1 x^1 + f_2 x^2 = \begin{pmatrix} f_1 & f_2 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} , \quad (4.65)$$

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

$$f(\mathbf{x}) = (f_1 \langle 1| + f_2 \langle 2|) (x^1 |1\rangle + x^2 |2\rangle) \quad (4.66)$$

$$= f_1 x^1 \langle 1|1\rangle + f_2 x^1 \langle 2|1\rangle + f_1 x^2 \langle 1|2\rangle + f_2 x^2 \langle 2|2\rangle \quad (4.67)$$

$$= f_i x^i , \quad (4.68)$$

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

4.12 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 4.12.1. 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. 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^* : $V = (V^*)^*$.

¹It has all the same mathematical properties and structures.

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} . For numbers α and β , $(\alpha f + \beta g)$ is also a linear function whose output on $x \in X$ is simply

²Astute readers notice that this picture requires a metric.

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x). \quad (4.69)$$

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 4.12.2. *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.³ 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 linear⁴, then $|v\rangle$ is obviously a linear function on dual vectors.

³Earlier 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.

⁴Check this if it is not obvious!

Example 4.12.1. Suppose you have a vector \mathbf{v} in \mathbb{R}^2 . I happen to have a couple of basis covectors, $\tilde{\mathbf{e}}^{(1)}$ and $\tilde{\mathbf{e}}^{(2)}$. These are not necessarily the canonical basis; i.e. $\tilde{\mathbf{e}}^{(1)} \neq (1 \ 0)$. We know that my basis covectors can ‘eat’ your vector and produce a number. When we do this experiment, we get the following numbers:

$$\tilde{\mathbf{e}}^{(1)}[\mathbf{v}] = a \qquad \tilde{\mathbf{e}}^{(2)}[\mathbf{v}] = b \qquad (4.70)$$

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

$$\mathbf{v}[\tilde{\mathbf{w}}] \equiv w_1 \tilde{\mathbf{e}}^{(1)}[\mathbf{v}] + w_2 \tilde{\mathbf{e}}^{(2)}[\mathbf{v}] = a w_1 + b w_2. \qquad (4.71)$$

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 transpose¹ of a vector, that is, how to convert an element of V into an element of V^* . This was the observation (4.49) that we can ‘preload’ the inner product with a vector and the resulting object is a covector.

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*.²

Let us confirm this in words. The metric takes two vectors and spits out a number. Equivalently, it takes one vector \mathbf{v} and it spits out a covector, (4.46), $\tilde{\mathbf{v}}$. With this latter interpretation, the *inverse metric* takes the covector $\tilde{\mathbf{v}}$ and returns the original vector \mathbf{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^* , (4.45). 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.³ The metric is

¹Hermitian conjugate.

²That 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.

³This is probably among the top ten ways to “say you are a physicist without saying that you are a physicist.”

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 :

$$(g^{-1})^{ik} g_{kj} = \delta_j^i, \quad (4.72)$$

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.¹

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 simplify the notation by simply dropping the inverse symbol:

$$(g^{-1})^{ij} \equiv g^{ij}. \quad (4.73)$$

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

$$g^{ik} g_{kj} = \delta_j^i. \quad (4.74)$$

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 4.12.2. *As a quick sanity check, is there any ambiguity when writing the inverse metric as g^{ij} coming from the interpretation that $(g^{-1})^{ij} = g^{ij} = g^{ik} g_{kt} g^{tj}$? In the last equality, we interpreted the inverse metric as an object that raises indices. When using the defining relation $g^{ik} g_{kj} = \delta_j^i$ then the relation in question becomes a tautology.²*

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

¹To 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.

²That’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.

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.

4.13 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

$$(g_{ij} \langle i| \otimes \langle j|) (v^k |k\rangle) = v^k g_{ij} \langle i| \langle j|k\rangle = v^k g_{ij} \langle i| \delta_k^j = v^j g_{ij} \langle i| \equiv \tilde{v}_i \langle i|. \quad (4.75)$$

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 \mathbf{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 $|k\rangle$.

In general 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:

$$\mathbf{T} \equiv T^{ijk}_{\ell mn} |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell| \otimes \langle m| \otimes \langle n| . \quad (4.76)$$

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

$$\mathbf{T}\mathbf{v} = T^{ijk}_{\ell mn} v^p |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell| \otimes \langle m|p\rangle \otimes \langle n| \quad (4.77)$$

$$= T^{ijk}_{\ell mn} v^m |i\rangle \otimes |j\rangle \otimes |k\rangle \otimes \langle \ell| \otimes \langle n| . \quad (4.78)$$

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

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

4.14 Linear Transformations

[Flip: I am here]

Example 4.14.1. *Transformation of fields. Mixture of active and passive.*

Insert: idea that linearity of a transformation means a matrix representation.

4.15 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 reason why physicists love indices: they are a natural language for capitalizing on the symmetries of a system.¹

¹The 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.

4.16 Multilinear Notation

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. Because these maps can take many vectors/covectors as arguments, 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.²

²There are various phobias about insects with many legs. Perhaps for similar reasons.

Example 4.16.1. *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 $f(\mathbf{v}, \mathbf{w})$ where we understand that f itself is a vector. We are assuming that we all agree on some basis and there’s no need to explicitly write $\mathbf{v} = v^i \hat{\mathbf{e}}_{(i)}$. Multilinearity means*

$$f(\alpha \mathbf{v} + \beta \mathbf{u}, \mathbf{w}) = \alpha f(\mathbf{v}, \mathbf{w}) + \beta f(\mathbf{u}, \mathbf{w}) \quad (4.79)$$

$$f(\mathbf{v}, \alpha \mathbf{w} + \beta \mathbf{x}) = \alpha f(\mathbf{v}, \mathbf{w}) + \beta f(\mathbf{v}, \mathbf{x}), \quad (4.80)$$

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

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

$$f^i(\alpha \mathbf{v} + \beta \mathbf{u}, \mathbf{w}) = \alpha f^i(\mathbf{v}, \mathbf{w}) + \beta f^i(\mathbf{u}, \mathbf{w}) \quad (4.81)$$

$$f^i(\mathbf{v}, \alpha \mathbf{w} + \beta \mathbf{x}) = \alpha f^i(\mathbf{v}, \mathbf{w}) + \beta f^i(\mathbf{v}, \mathbf{x}), \quad (4.82)$$

where the relations are separately true for each index i .

4.17 Examples of Finite-Dimensional Vector Spaces

Euclidean space

Minkowski space

Exercise 4.17.1. *The twin paradox in a closed universe.*

Spin space

The algebra of special unitary matrices

Functions

4.18 Determinants

5 *Vector fields*

6 *Function Space*

APPENDIX

A *Conventions*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis.

Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.[20]

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

- allometry, 27
- basis, 38
- bra, 47
- canonical, 44
- contraction, 53
- contravariant vector, 47
- covariant vector, 47
- covector, 47
- differential one forms, 52
- differential operator, 9
- dimension, 20, 39
- dimensional analysis, 19
- dot product, 41
- dual vector, 47
- dual vector space, 48
- dummy index, 54
- Dynkin diagrams, 45
- error, 25
- function, 33
- Green's function, 9
- Hermitian conjugate, 50
- inner product, 41
- isomorphic, 57
- ket, 47
- Kronecker delta, 49
- linear, 34
- linear combination, 36
- linear independence, 39
- metric, 41
- metric space, 41
- model, 27
- natural units, 20
- nice, 11
- norm, 41
- obvious, 12
- one-form, 47
- orthogonal, 40
- perturbation theory, 11
- Riemannian, 41
- row vector, 47
- tensor product, 60
- transpose, 50
- vector, 35
- vector space, 35

This document was typeset using \LaTeX and Sylvain Kern's `tufte-style-thesis` class.
The package is available on GitHub:
<https://github.com/sylvain-kern/tufte-style-thesis/>.
Feel free to contribute!