

P231: Methods of Theoretical Physics

Flip Tanedo

flip.tanedo@ucr.edu

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

Abstract

Notes for P231 based on the 2019 course.

1 What this course is about

Physics 231: Methods of Theoretical Physics is a course for first-year physics and astronomy graduate students. It is a ‘crash course’ in mathematical methods necessary for graduate courses in electrodynamics, quantum mechanics, and statistical mechanics. The style of the course is a boot camp rather than a rigorous theorem–proof mathematics lecture. Where possible, the emphasis is on physical intuition rather than mathematical precision.

1.1 Green’s functions

The primary goal of this course is to use Green’s functions to solve linear differential equations. The general form of such a differential equation is

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

where \mathcal{O} is some differential operator that encodes some kind of physical dynamics, $s(x)$ is the source of those dynamics, and $f(x)$ is a response that we would like to determine. Colloquially, the Green’s function is the operator \mathcal{O}^{-1} . But what the heck does that even mean?

We will approach this problem by analogy to linear algebra, where a linear transformation A acting on a vector space can give equations like:

$$A\mathbf{v} = \mathbf{w} , \tag{1.2}$$

whose solution is

$$\mathbf{v} = A^{-1}\mathbf{w} . \tag{1.3}$$

We will connect the notion of a linear differential operator to a matrix in infinite dimensional space to give a working definition of \mathcal{O}^{-1} . We will then pull out a bag of tricks from complex analysis to formally solve $\mathcal{O}^{-1}s(x)$ given \mathcal{O} and $s(x)$.

1.2 Mathematical Niceness

In my personal approach to mathematics in physics, I have found it useful to have the notion of a **nice** mathematical situation. This is not a formal idea, and it is one of the many things mathematicians find ridiculous about me. But as a physicist, the concept of mathematical *niceness* is helpful.

The idea is simply that the physical systems that we tend to care about are all *nice*. What that means is that they rarely populate the degenerate ‘special exceptions’ that mathematicians worry about. Those exceptions may be the difference between a mathematical theorem being true or false for general cases. We rarely care about general cases—we care about the cases that nature seems to populate.

As such, we rarely have to worry about whether a function is smooth, or even the precise details of a perturbation expansion’s radius of convergence. We rarely deal with operators that are not Hermitian. We rarely worry about the existence of orthonormal bases—the spaces in physics are all *nice*. We rarely have to worry about these degenerate cases, and so most of our energy goes toward understanding the mathematical properties of the *nice* cases that describe nature.

Every once in a while, we *do* have to worry about the exceptional cases. Those scenarios are the most interesting of all. That’s when the mathematical formalism that we use to describe physics grabs us by the collar and says, *listen to me—something important is happening and it probably has to do with nature!* Usually this is what happens when a calculation tells us that a physical result is infinite.

That being said, in this course, we will focus on *nice* functions and *nice* operators and *nice* boundary conditions, etc. For the most part, this is what we need to make progress on our physical models and it’s worth spending our time learning to work with *nice* limits. Leave the degenerate cases to the mathematicians for now. Eventually, though, you may 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’re missing the point.

1.3 Physics vs. Mathematics, I

Let’s make one point clear:

$$\text{Physics} \neq \text{Mathematics} . \tag{1.4}$$

There are many ways that this is true. Physics is intimately tied to experimental results and empirical science. Physicists will Taylor expand to their hearts’ content—sometimes even when the expansion is not formally justified. Physicists use explicit coordinates, mathematicians abhor this.

Physicists seek to uncover a truth about *this* universe. We use mathematical formalism to describe the universe. Sometimes the formalism that we need even spurs on developments in formal mathematics. However, let us be clear that our descriptions of the universe are *mathematical models*. Our models have limits of validity: where they are tested, where they break down.

Sometimes our models will break down in a way that is manifest: the classical potential of an electron orbiting a proton appears to have a divergence when the electron is arbitrarily close to the proton. In fact, quantum mechanically the problem gets worse: if the electron can literally overlap with the proton in some ‘wavefunctiony’ sense, then there’s some piece of the electron that is hitting the singularity. The solution is that there is additional physics that our description failed to account for.¹

¹If one has not thought about this problem, please do so. What is the new physics that resolves the Coulomb singularity?

1.4 Physics vs. Mathematics, II

In this class we will write lots of equations. We will use binary relations like $=$ or \neq . Sometimes to make a point we'll write \cong or \equiv or \doteq to mean something like 'definition' or 'tautologically equivalent to' or some other variant of 'even more equal than equal.'

As physicists, though, the most important binary relation is none of those things. Usually what we really care about is in \sim .² This tells how something *scales*. If I double a quantity on the right-hand side, how does the quantity on the left-hand side scale? Does it depend linearly? Quadratically? Non-linearly? The answer encodes something important about the underlying physics of the system and lies at the heart of the *imagine a cow is a sphere* joke.

By the way, implicit in this is the idea that in this class, we will not care about stray factors of 2. As my adviser used to say, if you're worried about a factor of 2, then your additional homework is to figure out that factor of 2.

1.5 Physics vs. Mathematics, III

There is another way in which physics is different from mathematics, and it is far more prosaic. Quantities in physics have units. We don't just deal with numbers, we deal with kilograms, electron volts, meters. It turns out that dimensional analysis is a big part of what we do as physicists.

2 Dimensional Analysis

You'll be surprised how far you can go in physics by thinking deeply about dimensional analysis. Here we'll only get you started. To go one step further, you may read more about the Buckingham Pi theorem³ or dive into neat applications⁴⁵.

2.1 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*:

$$(3 \text{ apples}) \times \left(\frac{\$1}{\text{apple}} \right) = \$3 . \quad (2.1)$$

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

$$1 \text{ apple} = \$1 \quad \Rightarrow \quad 1 = \frac{\$1}{1 \text{ apple}} . \quad (2.2)$$

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

All that matters is that the conversion is constant. Indeed, the constants of nature make very good 'exchange rates.' For example, in high-energy physics we like to use **natural units**. This is the curious statement that

$$\hbar = c = 1 . \quad (2.3)$$

²I use this the same way as \propto , which is completely different from 'approximately.' \approx .

³<https://aapt.scitation.org/doi/10.1119/1.1987069>

⁴<https://aapt.scitation.org/doi/full/10.1119/1.3535586?ver=pdfcov>

⁵<http://inspirehep.net/record/153032?ln=en>

At face value, this doesn't 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),

$$c = 3 \times 10^{10} \text{ cm/s} . \quad (2.4)$$

If $c = 1$, then this means

$$1 \text{ s} = 3 \times 10^{10} \text{ cm} . \quad (2.5)$$

This, in turn, connects a unit of time to a unit of distance. By measuring time, the constant c automatically gives us 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.

2.2 Quantifying the 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:

$$[Q] = L^a M^b T^c . \quad (2.6)$$

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.

As a simple example, we may want to figure out the units of force. We remember that $\mathbf{F} = m\mathbf{a}$, so

$$[\mathbf{F}] = [m][\mathbf{a}] = M \times LT^{-2} = L^1 M^1 T^{-2} . \quad (2.7)$$

Life is even easier in natural units, where $c = 1$ 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 typically write $[Q]$ to mean the mass-dimension of a quantity. To revert back to conventional units, one simply multiplies by appropriate factors of $1 = c$ and $1 = \hbar$.

2.3 Usage: Sanity Check

The simplest use of dimensional analysis is to check your work. It simply does not make sense to ever write

$$1 + (3 \text{ cm}) . \quad (2.8)$$

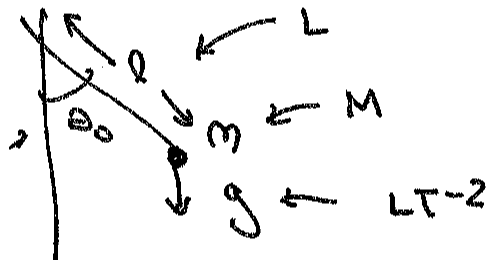
This does not make sense. You cannot sum together terms of different dimension. Similarly, $\sin(3 \text{ cm})$ does not make sense. What about $e^{5 \text{ cm}}$? This doesn't make sense because

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots \quad (2.9)$$

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

2.4 Solving problems

Here's 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 this:



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 :

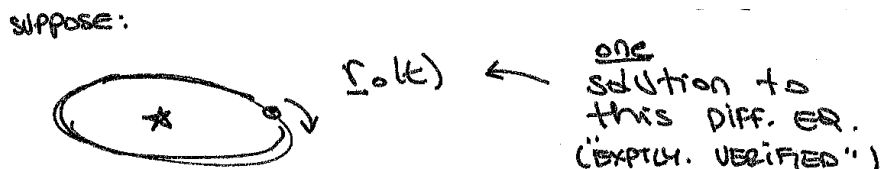
$$\tau \sim g^{-1/2} \ell^{1/2} . \quad (2.10)$$

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

$$\tau = f(\theta_0) \sqrt{\frac{\ell}{g}} . \quad (2.11)$$

2.5 Scaling

A large part of physics has to do with scaling relations. Here's a somewhat contrived example of how this works. Suppose you have some static, central potential $U(\mathbf{r})$. Maybe it's some planet orbiting a star.



The force law gives:

$$m\ddot{\mathbf{r}} = -\frac{\partial U}{\partial \mathbf{r}} . \quad (2.12)$$

Suppose we are given a solution, $\mathbf{r}_0(t)$. Perhaps this is a trajectory that is experimentally verified. Dimensional analysis gives a way to scale this solution into other solutions. For example, let us scale time by defining a new variable t' :

$$t \equiv \alpha t' . \quad (2.13)$$

If 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 $\sim T^{-2}$, 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:

$$m \left(\frac{d}{dt} \right)^2 \mathbf{r}_0(t) = m\alpha^{-2} \left(\frac{d}{dt'} \right)^2 \mathbf{r}_0(\alpha t') . \quad (2.14)$$

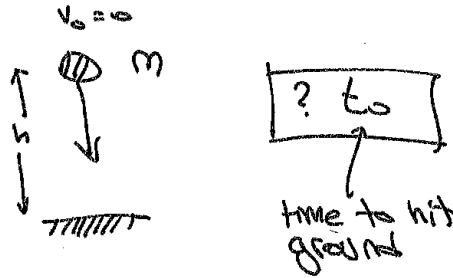
This begs us to define a new mass $m' = m\alpha^{-2}$. We thus have

$$m' \ddot{\mathbf{r}}(\alpha t') = -\frac{\partial U}{\partial \mathbf{r}} . \quad (2.15)$$

What this tells us is that $\mathbf{r}_1(t') = \mathbf{r}_0(\alpha t')$ is a solution in the same potential that traces the same trajectory but at α times the speed and with mass m' . For example, if $\alpha = 2$, then $\mathbf{r}_1(t)$ traces the same trajectory at double the velocity with one fourth of the mass.

2.6 Error Estimates

This section is based on a lovely *American Journal of Physics* article by Craig Bohren.⁶ Let's go back to another high school physics problem.



Suppose you drop a mass m from height h that is initially at rest. How long before this hits the ground? You can integrate the force equation to get

$$t_0 = \sqrt{\frac{2h}{g}} . \quad (2.16)$$

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 what the actual answer is.

How good of an approximation is it?

One way to do this is to do the next-to-leading order ('NLO') calculation, taking into account a more realistic (and hence more complicated) model and then compare to t_0 . But this is stupid. Why do we need to do a *hard* calculation to justify doing an *easy* one? If we're going to do the hard calculation anyway, what's the point of ever doing the easy one?

⁶<https://doi.org/10.1119/1.1574042>

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

$$\epsilon = \frac{t_1 - t_0}{t_0} . \quad (2.17)$$

This is a dimensionless quantity that determines how far off t_0 is from a more realistic calculation, t_1 . Ideally we don't actually have to do work to get t_1 .

Let's assume that we're not completely nuts and that we're in a regime where the error is small⁷. Then the error is a function of some dimensionless parameters, ξ , in the system. We define these ξ so that as $\xi \rightarrow 0$, $\epsilon(\xi) \rightarrow 0$. In other words, the approximation gets better as the ξ are made smaller. By Taylor expansion:

$$\epsilon(\xi) = \epsilon(0) + f'(0)\xi + \mathcal{O}(\xi^2) . \quad (2.18)$$

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

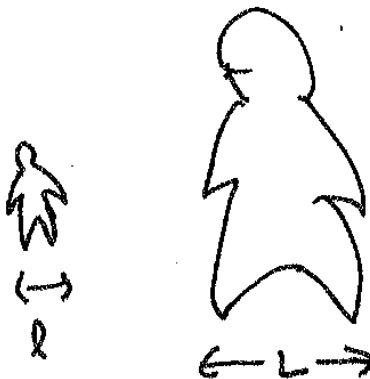
Here's how it works in practice. One effect that we miss in our toy calculation of t_0 is that the earth is round with radius R . This means that assuming a constant g is an approximation. We have two choices for a dimensionless parameter ξ :

$$\xi = \frac{h}{R} \quad \text{or} \quad \xi = \frac{R}{h} . \quad (2.19)$$

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

2.7 Bonus: Allometry

There's 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 :



If both people exercised at the same rate, which one loses more absolute weight? By how much? Let's assume that weight loss is primarily from the conversion of organic molecules into carbon dioxide.

⁷Note 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?'

3 Linear Algebra Review

The usual joke is that we all know linear algebra from kindergarten, at which point students roll their eyes. As physicists, linear algebra is part of our DNA from the vector calculus of our first electrodynamics course to quantum mechanics. Why, then should we patronize ourselves with yet another review of linear algebra?

Our goal is to understand Green's functions as *matrix inverses*. They are the inverse of matrices that represent the differential operator of the differential equation that we'd like to solve. These matrices act on a space of functions. Our goal is to use linear algebra to build an intuition for Green's functions. We start with reminders of facts from your mathematical 'childhood,' but throughout it all we would like to keep the following in mind:

$$\text{function space} = \infty\text{-dimensional vector space} . \quad (3.1)$$

3.1 The basics

A **linear transformation** A acts on a vector \mathbf{v} as $A\mathbf{v}$. This transformation satisfies

$$A(\alpha\mathbf{v} + \beta\mathbf{w}) = \alpha A\mathbf{v} + \beta A\mathbf{w} . \quad (3.2)$$

Here α and β are numbers. This is conventionally matrix multiplication. The result is also a vector. One way that we like to think about vectors is as columns of elements:

$$\begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^N \end{pmatrix} , \quad (3.3)$$

where N is the **dimension** of the vector space. Our notation is that v^i refers to the i^{th} component of \mathbf{v} . Sometimes—as physicists—we refer to v^i as the vector itself, which is a slight abuse of notation that occasionally causes confusion.

Matrix multiplication may actually have been something you learned in grade school. Here's how it works in two dimensions. A transformation that takes vectors into vectors takes the following form:

$$A = \begin{pmatrix} A^1_1 & A^1_2 \\ A^2_1 & A^2_2 \end{pmatrix} . \quad (3.4)$$

We're being fancy and using upper and lower indices. The significance is not pertinent at the moment, but practitioners of special relativity should nod with familiarity. If you're squeamish about the indices, don't worry: the elements of A have two indices, the first one is written a little higher than the second one. This notation is neither mathematics nor physics, it's a convention that we use for future convenience.

The action of A on \mathbf{v} is:

$$A\mathbf{v} = \begin{pmatrix} A^1_1 & A^1_2 \\ A^2_1 & A^2_2 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} A^1_1 v^1 + A^1_2 v^2 \\ A^2_1 v^1 + A^2_2 v^2 \end{pmatrix} . \quad (3.5)$$

Look at this carefully. The components of the new vector $(A\mathbf{v})^i$ are sums. In each term, the second/lower index of an A element multiplies the component of \mathbf{v} with the same index. The first/upper index of A tells you whether that term should be in $(A\mathbf{v})^1$ or $(A\mathbf{v})^2$.

A generic component of $(A\mathbf{v})$ is

$$(A\mathbf{v})^i = \sum_j A^i_j v^j = A^i_j v^j \quad (\text{Einstein convention}) . \quad (3.6)$$

On the right-hand side we use Einstein notation: *we implicitly sum over repeated upper/lower indices*. We will use this notation from now on. If you are at all in doubt about this, please work out the 2×2 case carefully and compare to the succinct notation above.

If A and B are linear transformations, then $A + B$ is a linear transformation. The components of $A + B$ are simply the piecewise sum of the corresponding components of A and B :

$$(A + B)^i_j = A^i_j + B^i_j . \quad (3.7)$$

3.2 Linear Transformations and Vector Space

Let's be a little more pedantic. We should *not* think of a vector \mathbf{v} as some 'column of numbers.' The power of this formalism is that a vector space is abstract. The layer of abstraction is encoded in the basis vectors, which we write as $\mathbf{e}_{(i)}$. For a space of dimension N , there are N such vectors indexed by the subscript. Let us more formally write the vector \mathbf{v} as

$$\mathbf{v} = v^1 \mathbf{e}_{(1)} + v^2 \mathbf{e}_{(2)} + \cdots = v^i \mathbf{e}_{(i)} . \quad (3.8)$$

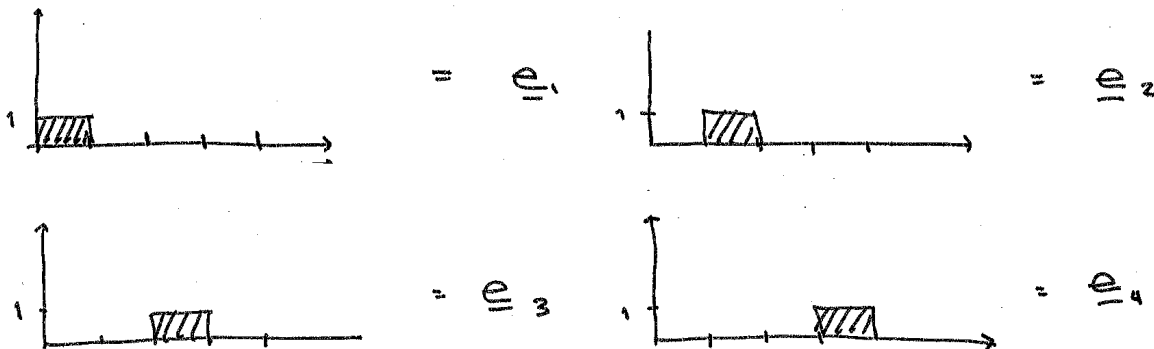
These basis vectors may be unit vectors in space. They may be represented by unit column vectors, e.g.

$$\mathbf{e}_{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{e}_{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \cdots . \quad (3.9)$$

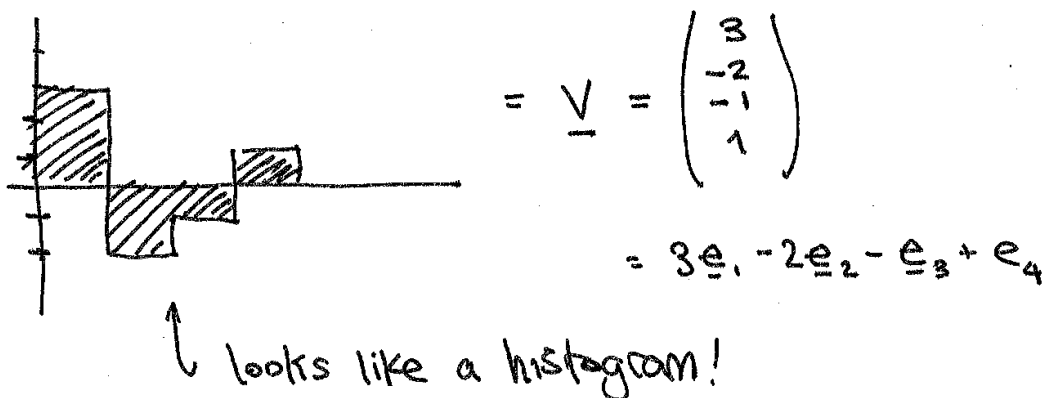
But these may be more general objects. For example, you can specify a color of light by specifying the red/green/blue content. We could have $\mathbf{e}_{(1)}$ be a unit amount of red light, $\mathbf{e}_{(2)}$ be a unit amount of green light, and $\mathbf{e}_{(3)}$ be a unit amount of blue light. Then a 3-vector \mathbf{v} would correspond to light of a particular color. This color space is a vector space.

3.3 A funny vector space: histogram space

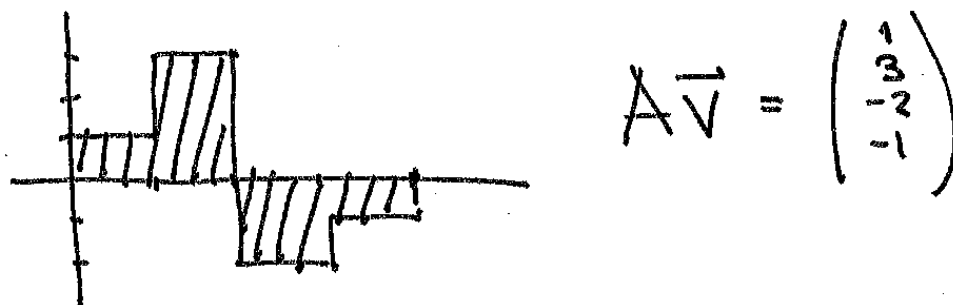
Here's a funny vector space that we're going to use as a pedagogical crutch. Imagine histogram-space. The basis vectors are:



This is a basis for a histogram over unit bins from $x = 0$ to $x = 4$. A vector in this space is, for example:

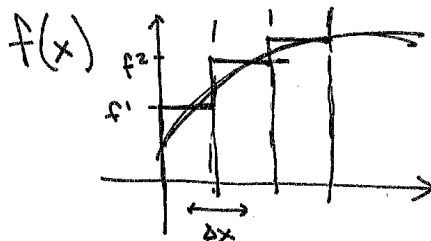


We can perform a linear transformation A on \underline{v} which outputs another vector. Let's say it's this:



Sanity check: from this, can you derive what A is? Answer: no.

The power of this admittedly strange formalism is that we can think of these histograms as approximations of continuous functions:



Thus a vector in this approximate (discretized) *function* space is

$$\underline{f} = \begin{pmatrix} f^1 \\ f^2 \\ \vdots \\ f^N \end{pmatrix}. \quad (3.10)$$

3.4 Derivative Operators

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

$$\mathbf{f}' = \frac{1}{\Delta x} \begin{pmatrix} f^2 - f^1 \\ f^3 - f^2 \\ \vdots \\ f^{i+1} - f^i \\ \vdots \end{pmatrix} . \quad (3.11)$$

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

$$\mathbf{f}' = D\mathbf{f} . \quad (3.12)$$

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

$$(f')^N = \frac{1}{\Delta x} (f^{N+1} - f^N) \quad (3.13)$$

but now we have no idea what f^{N+1} is. That was never a component in our vector space. There is no $\mathbf{e}_{(N+1)}$ basis vector.

This demonstrates an important lesson that we'll need when we move more formally to function spaces: *boundary conditions are part of the definition of the function space.*

In the present case, let's assume Dirichlet boundary conditions. A convenient way to impose this is to define what happens to all functions outside the domain of the function space:

$$f^{i>N} = f^{i<1} = 0 . \quad (3.14)$$

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

$$(f')^N = \frac{1}{\Delta x} (f^{N+1} - f^N) = \frac{-f^N}{\Delta x} . \quad (3.15)$$

Alternatively, we could have also imposed periodic boundary conditions:

$$f^i = f^{i+kN} \quad k \in \mathbb{Z} . \quad (3.16)$$

This would then give

$$(f')^N = \frac{1}{\Delta x} (f^{N+1} - f^N) = \frac{1}{\Delta x} (f^1 - f^N) . \quad (3.17)$$

One could have also defined a backward derivative where $(f')^i \sim f^i - f^{i-1}$. The second derivative may be defined symmetrically:

$$(f'')^i = \frac{(f^{i+1} - f^i) - (f^i - f^{i-1})}{\Delta x^2} . \quad (3.18)$$

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

3.5 Derivatives in other function space bases

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

$$\mathbf{e}_{(0)} = 1 \qquad \mathbf{e}_{(1)} = x \qquad \mathbf{e}_{(2)} = x^2 . \qquad (3.19)$$

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

$$f(x) = ax^2 + bx + c \qquad \Rightarrow \qquad \mathbf{f} = \begin{pmatrix} c \\ b \\ a \end{pmatrix} . \qquad (3.20)$$

The derivative operator has an easy representation in this space:

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} . \qquad (3.21)$$

We can see that

$$D\mathbf{f} = \begin{pmatrix} b \\ 2a \\ 0 \end{pmatrix} \qquad D^2\mathbf{f} = \begin{pmatrix} 2a \\ 0 \\ 0 \end{pmatrix} \qquad D^3\mathbf{f} = 0 . \qquad (3.22)$$

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

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

3.6 Locality

3.7 Row Vectors and all that

In high school we didn't distinguish between row vectors and column vectors. They both seemed to convey the same information. The row vectors were just 'tipped over.' Perhaps you noticed that they follow the rules of 'matrix multiplication' act on a vector:

$$(w_1 \ w_2 \ \cdots) \begin{pmatrix} v^1 \\ v^2 \\ \vdots \end{pmatrix} = w_1 v^1 + w_2 v^2 + \cdots . \qquad (3.23)$$

In fact, this is like \mathbf{w}^T is a function that acts linearly on \mathbf{v} :

$$\mathbf{w}^T(\mathbf{v}) = w_1 v^1 + w_2 v^2 + \cdots . \qquad (3.24)$$

Indeed, let us be a bit more formal about this. This layer of formalism is uncharacteristic of our approach in this course, but this underpins so much of the mathematical structure of our physical theories that it is worth getting right from the beginning.

Let V be a vector space. It contains vectors, \mathbf{v} . Sometimes these are called contravariant vectors or kets. They have basis vectors $\mathbf{e}_{(i)}$.

Now introduce a related but *completely distinct* vector space called V^* . This is the space of **dual vectors** to V . A **dual vector** is what you may know as a **row vector**, a **ket**, a **covariant vector**, or a [differential] **one-form**. These are all words for the *same idea*. An element of V^* , say (\mathbf{w}^T) is a linear function that takes vectors and spits out numbers:

$$\mathbf{w}^T \in V^* \Rightarrow \mathbf{w}^T : V \rightarrow \mathbb{R} . \quad (3.25)$$

Don't think about \mathbf{w}^T as some kind of operation on a vector $\mathbf{w} \in V$; at least not yet. For now the ' T ' is just part of the name of \mathbf{w}^T . The two spaces V and V^* are totally different. We haven't said anything about how to turn elements of V into elements of V^* or vice versa. It should be clear that there is a sense of 'duality' here: the vectors V are also linear functions that take a dual vector and spit out a number.

Let us call the basis of dual vectors $\tilde{\mathbf{e}}^{(i)}$. This notation is cumbersome, so we'll change to something different soon. The upper index is deliberate. The defining property of $\tilde{\mathbf{e}}^{(i)}$ is:

$$\tilde{\mathbf{e}}^{(i)} (\mathbf{e}_{(j)}) \equiv \delta_j^i . \quad (3.26)$$

One may check that this gives

$$(w_i \tilde{\mathbf{e}}^{(i)}) (v^j \mathbf{e}_{(j)}) = w_i v^j \delta_j^i = w_i v^i = w_1 v^1 + w_2 v^2 + \cdots . \quad (3.27)$$

3.8 Orthonormal Bases

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

There are many things to be said about this, none of them are particularly edifying without a full discussion. With no apologies, I'll make the following [perhaps perplexing] remarks:

1. There is no such thing as a 'position vector.' Positions refer to some base space, whereas vectors (like differential operators) act on the tangent space at a point of that base space.
2. A given tangent space is 'nice' and has a nice orthonormal basis.
3. That basis may not be the same for neighboring tangent spaces (perhaps due to coordinates, perhaps due to intrinsic curvature).

In this course these nuances will not come up. In the rest of your life you'll still have to deal with curvilinear coordinates. But suffice it to say that our study of function space will be nice an orthonormal. Of course, we haven't yet given a definition of 'normality.'

3.9 Bra-Ket Notation

There is neither any physics nor mathematics contained in a choice of notation. However, a convenient notation does simplify our lives. Let us introduce bra-ket notation. In this notation, we denote vectors by kets:

$$|v\rangle = v^i |i\rangle , \quad (3.28)$$

where $|i\rangle = \mathbf{e}_{(i)}$ is the basis of vectors that span the vector space V . There is nothing new or different about this object, $\mathbf{v} = |v\rangle$.

We denote dual vectors (row-vectors, one-forms) as bras:

$$\langle w| = w_i \langle i| , \quad (3.29)$$

where $\langle i| = \widetilde{\mathbf{e}}^{(i)}$. The orthonormality of this basis is encoded in

$$\langle i|j\rangle = \delta_j^i . \quad (3.30)$$

In bra-ket notation a linear transformation A has a basis

$$A = A^i{}_j |i\rangle \langle j| . \quad (3.31)$$

The notation $|i\rangle \langle j|$ is shorthand for $|i\rangle \otimes \langle j|$. If the \otimes doesn't mean anything to you, that's fine. It doesn't mean much to me either. Matrix multiplication proceeds as before:

$$A\mathbf{v} = A|v\rangle = A^i{}_j |i\rangle \langle j| v^k |k\rangle = A^i{}_j v^k |i\rangle \langle j|k\rangle = A^i{}_j v^k |i\rangle \delta_k^j = A^i{}_j v^j |i\rangle . \quad (3.32)$$

Observe that the power of the notation is clear: the object with the index v^i is just a number. It commutes with everything. All of the vector-ness is carried in the basis objects: the bras, kets, and ket-bras. Those do not commute. But they have a well defined way in which kets act on bras (or vice versa).⁸

3.10 Eigenvectors are nice

Give a sufficiently *nice* linear transformation, A , there is a particularly convenient basis: the eigenvectors of A . These are kets $|\lambda\rangle$ such that

$$A|\lambda\rangle = \lambda|\lambda\rangle . \quad (3.33)$$

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

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

$$|v\rangle = v^i |\lambda_{(i)}\rangle , \quad (3.34)$$

Then the action of A on this vector is easy:

$$A|v\rangle = \sum_i \lambda_{(i)} v^i |\lambda_{(i)}\rangle . \quad (3.35)$$

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

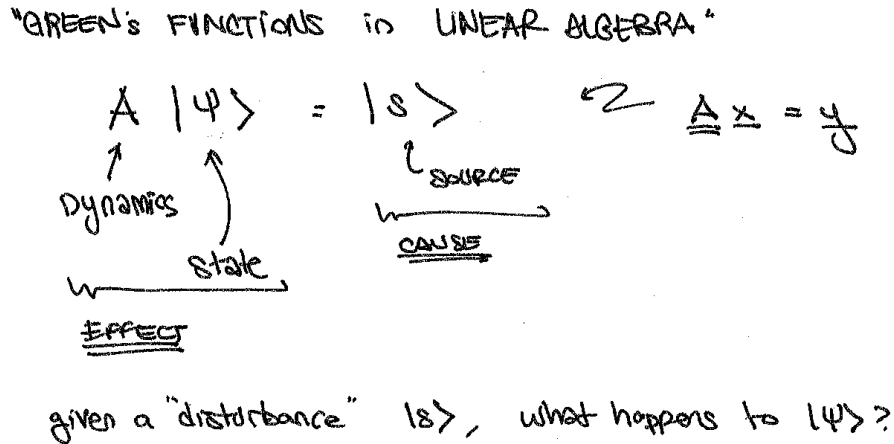
$$A^{-1}|v\rangle = \sum_i \lambda_{(i)}^{-1} v^i |\lambda_{(i)}\rangle . \quad (3.36)$$

⁸This is where the \oplus notation is handy. It keeps track of which kets/bras might hit which other bras/kets. This falls under the name of multi-linear algebra.

3.11 The Green's Function Problem

Going back to the big picture: recall that we want to solve differential equations of the form $\mathcal{O}f(x) = s(x)$. If we had a sense of the *eigenfunctions* of \mathcal{O} , then we could expand $s(x)$ in a basis of those eigenfunctions and then apply \mathcal{O}^{-1} to both sides.

The analog is this:



The operator A encodes the *physics* of the system, the underlying dynamics. This is presumably local: it is a near-diagonal matrix coming from one or two powers of derivatives. The ket $|s\rangle$ is the source. This is the thing that *causes* the dynamics. The ket $|\psi\rangle$ is some state that we would like to determine.

3.12 Metrics

Thus far we have introduced vector spaces. The dual vector space is a set of linear functions that act on elements of a vector space; these are bras/row-vectors/one-forms. Let us now introduce a new piece of machinery: a **metric**. This is also known as an **inner product** or a **dot product**. A space with a metric is called a metric space. We only state this fact to emphasize that we are *adding this structure by hand*. Vector spaces don't come with metrics—someone makes up a metric and slaps it onto the vector space.

The **metric** is a function that takes two vectors and spits out a number. It is linear in each argument. In other words, a metric g is:

$$g : V \times V \rightarrow \mathbb{R} . \quad (3.37)$$

Occasionally one may want a metric defined such that the output is a complex number. We thus have:

$$g(\alpha \mathbf{v} + \beta \mathbf{w}, \delta \mathbf{x} + \gamma \mathbf{y}) = \alpha \delta g(\mathbf{v}, \mathbf{x}) + \alpha \gamma g(\mathbf{v}, \mathbf{y}) + \beta \delta g(\mathbf{w}, \mathbf{x}) + \beta \gamma g(\mathbf{w}, \mathbf{y}) . \quad (3.38)$$

One more special assumption about the metric is that it is **symmetric**:

$$g(\mathbf{v}, \mathbf{w}) = g(\mathbf{w}, \mathbf{v}) . \quad (3.39)$$

In indices one may write

$$g = g_{ij} |i\rangle \otimes \langle j| \quad (3.40)$$

so that

$$g(\mathbf{v}, \mathbf{w}) = g_{ij} v^i w^j . \quad (3.41)$$

Here we see the usefulness of the \otimes notation. It tells us that the bras and kets resolve as follows:

$$g_{ij} \langle i| \otimes \langle j| (v^k |k\rangle) (w^\ell |\ell\rangle) = g_{ij} v^k w^\ell \langle i|k\rangle \langle j|\ell\rangle = g_{ij} v^k w^\ell \delta_k^i \delta_\ell^j = g_{ij} v^i w^j . \quad (3.42)$$

For ordinary Euclidean space in flat coordinates, the metric is simply the unit matrix: $g_{ij} = \text{diag}(1, \dots, 1)$. In Minkowski space there's a relative minus sign between space and time. In curvilinear coordinates things get ugly.

Here's the neat thing about metrics. We can take a metric and pre-load it with a vector. Given a metric g and a vector \mathbf{v} , we may define a function

$$g(\mathbf{v}, \quad) . \quad (3.43)$$

This simply means that the function $f(\mathbf{w}) = g(\mathbf{v}, \mathbf{w})$. Observe that $f(\mathbf{w})$ is a linear function that takes elements of V and returns a number. In other words, this is a *dual vector* (row-vector, one-form). Observe what having a machine like the metric has done for us: it has allowed us to convert vectors into dual vectors:

$$g(\mathbf{v}, \quad) = g_{ij} v^i \langle j| . \quad (3.44)$$

Similarly, one may define an inverse metric g^{-1} such that $g^{-1}g = \mathbb{1}$. In a slight abuse of notation, the inverse metric is written with two upper indices: g^{ij} . Note that we do not write the $^{-1}$. The inverse metric will *raise* the index on a lower-index object, while the metric *lowers* the index of an upper-index object.⁹

3.13 Hermitian Conjugate

Now that we can go between column and row vectors, thanks to the metric and its inverse, it is worth thinking a bit about what we meant by the 'transpose' operator. The transpose precisely turned a column vector \mathbf{v} into an associated row vector \mathbf{v}^T . The generalization of this idea is the Hermitian conjugate, † .

Acknowledgements

P.T. thanks *your name here* for useful comments and discussions.

⁹Of course: what's really happening is that the metric has a basis $\langle i| \otimes \langle j|$ while the inverse metric has a basis $|i\rangle \otimes |j\rangle$.