Representations
of a
Physical Universe

Alex Atanasov & Aaron Hillman

May 29, 2017

Todo list

Preface

Dedicated to $\otimes$*,*
we'll decide at the end

Contents

Acknowledgements

The authors would like to thank:

∇^2 John C. Dell

$\oint\!\!\!\oint$ Jonathan A. Osborne

$\mathcal{U}_q(\mathfrak{sl}_2)$ Igor B. Frenkel

$\mathcal{L}_X Y$ Vincent E. Moncrief

$\{-,-\}$ Francesco Iachello

$\int \frac{d^4p}{(2\pi)^4}$ Walter D. Goldberger

$\langle \mathcal{O}\mathcal{O}\mathcal{O}\mathcal{O} \rangle$ David Poland

$\circ - \circ$ You Qi

$\frac{j(j+1)}{2}$ James C. Ellenbogen

$\omega_i \mathrm{d}q^i$ Ted A. Jacobson

For the valuable conversations and lectures that they offered, which led up to the development of this text.

Part 0

Prerequisite Material

Part 1

A Language for Nature

Chapter 1

Harmonics: Fourier Analysis

This chapter develops the ideas of Fourier analysis, a framework whose reach now extends across the science and engineering disciplines. There is no arguing the fact that it is one of the most powerful tools that mathematics has ever developed, not just for its wide practical application, and not just for its use in formulating physical theories, but for going further and influencing the way that mathematics itself is thought about and done.

Another mathematical idea that has had similar impact across the scientific fields is linear algebra: the in-depth study of the phenomenon of linearity. Aside from allowing us to go between points in physical space and tuples of numbers, linear algebra gives another powerful tool that, as of yet, has not been touched upon in this book: the eigenvector. Eigenvectors give a natural way of handling an object, once it is represented by a linear operator on a vector space. When that object is the stress tensor of a material, they describe the principal axes, when it is a high dimensional dataset, they allow for the principal-component analysis, and for markov chains and stochastic processes, they give information about long-term dynamics.

In this section, we will not just reflect on the uses and power of linear algebra and Fourier analysis, but we will go further and show that they are one and the same: Fourier analysis, viewed through the right lense, is plain linear algebra with eigenvalues. This will set the stage on which the next chapter will give us the real performance: generalizing the ideas of Fourier analysis to get something much more

unified and powerful, which it can be said to constitute a cornerstone of modern mathematics: representation theory.

The chapter is organized in stages that will form stepping stones between elementary linear algebra and full-blown analysis. We begin by performing Fourier analysis on N points connected in a discrete ring. Then, taking the limit as $N \to \infty$, we obtain the Fourier series on the circle S^1, i.e. the interval $[0, 1]$ with periodic boundary conditions. Generalizing this to arbitrary intervals $[-k, k]$, we take another limit as $k \to \infty$ and obtain the Fourier transform on the line. From here, we study higher-dimensional generalizations of these ideas: to 2-D lattices of points and continuous n-D Euclidean space. We conclude by translating these ideas to more general arrangements of N connected points: Fourier analysis on graphs, leading to spectral graph theory.

1.1 Discrete, Bounded: Eigenvalues

We begin with the definition of a **finite graph**

Definition 1.1. A finite graph G is a finite set of **vertices** $V = \{v_1, \ldots, v_n\}$ together with a set of **edges** $E = \{e_1, \ldots, e_m\}$ such that each edge is an unordered pair of vertices (v_i, v_j) indicating a connection between two vertices.

Occasionally, we will call the vertices "points" on the graph.

For this section, we consider the special case of N vertices connected in a ring by having each vertex joined to two nearest neighbors as below. Denote this graph by G.

Illustrate N points connected in a ring

We have considered functions on a manifold M, that associate a real or complex number to each point $p \in M$. Here we will consider the simpler, more finite case of functions f that associate to each *vertex* $v \in G$ a (real or complex) number $f(v)$ called the value of f at v. The ideas of functions on graphs are in many ways analogous to the notions of functions on manifolds, as we shall see.

For this ring with N points, a function can be described entirely

in terms of N numbers: the values it takes at each point.

$$f \approx \begin{pmatrix} f(v_1) \\ \vdots \\ f(v_N) \end{pmatrix}$$

Here we write '$\approx$' instead of '$=$' because of the same point that was made in chapter 1. This collection of N numbers certainly *represents* the function f, but it is *not* f. That is, we can identify the space of real-valued functions on these N-points with $\mathbb{R}^N$, and the space of complex-valued functions with $\mathbb{C}^N$, but this choice requires that we pick a basis.

But what basis have we picked on the function space? Our basis $\{b_i\}_{i=1}^N$ consists of functions $b_i(v_j)$ that are zero everywhere except for when $i = j$, where $b_i(v_i) = 1$. Then any function f can clearly be written as

$$f = \sum_i f(v_i) b_i \tag{1.1}$$

it is easy to see that evaluating this at a vertex v_j makes every b_i vanish except for $i = j$, giving total value $f(v_j)$ as required. This set of basis functions seems to be the obvious one. We will call it the **position basis**, taking vernacular from physics.

Although the space of real-valued functions on G is isomorphic *as a vector space* to $\mathbb{R}^N$, it has much more structure than just an ordinary vector space. We know that we can multiply two functions f and g to obtain a new one $h = fg$ simply by multiplying values pointwise $h(v_i) = f(v_i)g(v_i)$. Since the functions on G are a vector space, we thus have a way to multiply two vectors in the space to get another.

Moreover, we can define an inner product on this space of functions

Definition 1.2 (Inner Product of Functions on a G)**.** Let f, g be to functions defined on G. We define an inner product $(\cdot, \cdot)$ on this space by

$$(f, g) = \sum_{i=1}^{N} f(v_i) g(v_i) \tag{1.2}$$

1.2 Continuous, Bounded: Fourier Series

1.3 Continuous, Unbounded: Fourier Transforms

1.4 Fourier Analysis on Euclidean Spaces

1.5 Fourier Analysis on Graphs: Spectral Graph Theory

1.6 Vector Spaces of Functions

We will work with functions just over $\mathbb{R}$ at first, the setting that we are most used to. It is no great surprise that functions $f : \mathbb{R} \to \mathbb{R}$ form a vector space, since we can add and scale any two functions $af + bg$ and get another function as their linear combination. We can therefore try to apply ideas from linear algebra to this vector space. Note first that it is infinite dimensional.

There is an important subspace worth looking at: $C^\infty(\mathbb{R})$ of smooth functions. Since finite linear combinations of smooth functions are still smooth, $C^\infty(\mathbb{R})$ is a subspace. We only allow for finite linear combinations of functions for now, as otherwise we need to figure out what we mean by infinite sequences and series of functions to "converge" to a limit in this vector space.

Example 1.3. The set of polynomials up to degree n on $\mathbb{R}$, $P^n(\mathbb{R})$, is a vector space. The set of polynomials up to degree n restricted to an interval $[a, b]$, $P^n[a, b]$, is also a vector space.

Note we need to include all the polynomials of degree $\leqslant n$ as well, not just degree n. Otherwise something like $(x^5 + x^3) - x^5$ would not be in the space.

In particular, now, let's consider the interval $[0, 1]$ to be of interest. We care about the space of *all functions* on $[0, 1]$. Again, this is still infinite dimensional, but say we restricted *further* and only cared about the function at the points $0, 1/4, 1/2, 3/4$ and 1. These are 5

points of interest. A function restricted to these 5 points can be represented just as a vector of 5 numbers: $[f(0)\ f(1/4)\ f(1/2)\ f(3/4)\ f(1)]$.

The linear combinations of functions on $[0,1]$ when restricted to these 5 points are then just linear combinations of these vectors in $\mathbb{R}^5$. In numerics, this would be a *low resolution* way to view the function, equivalent to something like the graphic.

5 box functions interpolating f

If we increased the number of points we sampled at, say $f(k/n)$ as k runs from 0 to n, then we represent functions by vectors in $\mathbb{R}^{n+1}$. As n gets big our resolution gets better and better. In general, though, infinitely many different functions will get sent to the same representation as a vector in $\mathbb{R}^{n+1}$ (so this discrete approximation from the set of all functions on $[0,1]$ to just $n+1$ evenly-spaced points has a huge kernel).

This type of idea is what lies behind Riemann summation for defining the integral operator. In fact, the discrete approximation of the integral operator on functions, when viewed as vectors in $\mathbb{R}^{n+1}$

Example 1.4 (Riemann Sums)**.**

$$\int_{approx} f dx = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\frac{1}{n} = \begin{pmatrix}\frac{1}{n} & \dots & \frac{1}{n}\end{pmatrix}\begin{pmatrix} f(0) \\ \vdots \\ f(1)\end{pmatrix}$$

The Riemann sum in this discrete space is just the linear functional $\begin{pmatrix}\frac{1}{n} & \dots & \frac{1}{n}\end{pmatrix}$ acting on the vector representing the function to give the desired value. If the full function f is Riemann-integrable, then as $n \to \infty$ this will approach the true value $\int_0^1 f dx$.

This approximation in $\mathbb{R}^{n+1}$ is not just any vector space. It has an notion of *length* for a vector

$$||f_n||^2 = \frac{1}{n}\sum_{i=0}^{n+1} f(k/n)^2 \tag{1.3}$$

As the sum of the squares of the function values at every point. The factor of $1/n$ out front is so that in the limit of $n \to \infty$ this gives a **norm**, a notion of size, on the space of functions whose squares are Riemann-integrable (such functions are called **square-integrable**)

on $[0, 1]$:

$$||f||^2 = \int_0^1 f(x)^2 dx \tag{1.4}$$

On $[0, 1]$, the norm of a function is its root mean square value. In fact this notion of length also gives a natural notion of inner product (both on the discrete space and the full space):

$$\langle f|g\rangle = \int_0^1 f(x)g(x)dx \tag{1.5}$$

Looking at it this way, it is *exactly* the same as how we defined a dot product: multiply the corresponding components together and sum that up. The difference is that now "components" have become "functional values" and the sum has become an integral.

Naturally, we can define an inner product on any interval $[a, b]$ between square integrable functions on that interval:

$$\langle f|g\rangle = \int_a^b f(x)g(x)dx \tag{1.6}$$

It is similarly possible to define an inner product for square-integrable functions on the *whole* real line:

$$\langle f|g\rangle = \lim_{L\to\infty} \int_{-L}^{L} f(x)g(x)dx = \int_{\mathbb{R}} f(x)g(x)dx \tag{1.7}$$

As in Chapter **??**, an inner product on a real vector space is as follows.

Definition 1.5 (Real Inner Product)**.** For a real vector space V, an inner product $\langle -|-\rangle$ on V is a bilinear map into $\mathbb{R}$ so that:

1. (Linear) $\langle a\mathbf{x} + b\mathbf{y}|\mathbf{z}\rangle = a\langle \mathbf{x}|\mathbf{z}\rangle + b\langle \mathbf{y}|\mathbf{z}\rangle$
2. (Symmetric) $\langle \mathbf{x}|\mathbf{y}\rangle = \langle \mathbf{y}|\mathbf{x}\rangle$
3. (Positive-definite) $\langle \mathbf{x}|\mathbf{x}\rangle \geqslant 0$ and equality holds iff $\mathbf{x} = 0$. A vector space that possesses an inner product is called an ***inner product space***.

Inner product spaces are specific examples of **Metric Spaces**.

Definition 1.6 (Metric Space)**.** A set X is a metric space if there is a non-negative function $d : X \times X \to \mathbb{R}$ that satisfies

1. (Symmetry) $d(x, y) = d(y, x)$
2. (Positive Definite) $d(x, y)$=0 iff $x = y$
3. (Triangle Inequality) $d(x, z) \leqslant d(x, y) + d(y, z)$

Just as on the real line and on $\mathbb{R}^n$, having a notion of distance allowed us to define closed and open sets, the same is true on metric spaces. In particular, inner product spaces have a natural topology defined by open and closed balls defining neighborhoods, and so as vector spaces that are also topological spaces, they are called **Topological Vector Spaces**.

Returning back to the inner product space of functions on $\mathbb{R}$ this inner product is far from useless. If both $|f| = 1$ and $|g| = 1$ have length then the inner product, in some well-defined sense, measures how *correlated* f is with g. If the integral of f against g is zero then there's no way to predict the behavior of g based on the behavior of f. If it is positive, then a positive f means a positive g *more often than not*, while if the integral is negative then a positive f means a *negative* g more often than not. Such a quantity in general:

$$\frac{\langle f|g\rangle}{||f||\ ||g||}$$

is geometrically what the cosine of the angle between two vectors was in $\mathbb{R}^{n+1}$. Functions that are "in the same direction" would have a positive cosine, while functions "pointing in opposite directions" would have a negative one. Uncorrelated, orthogonal functions would have a cosine of zero.

Exercise on correlation of functions and the inner product

So the vector space of square-integrable functions on $[0, 1]$ is an inner product space. It is denoted[1] by $L^2[0, 1]$. Similarly, for a general

[1] **HERE SAY SOMETHING ABOUT LEBESGUE VS RIEMANN INTEGRATION**

Riemann vs Lebesgue in a footnote

interval $[a, b]$ we denote the square integrable functions as $L^2[a, b]$, and for functions square-integrable on the *entire real line*, we write $L^2(\mathbb{R})$

The space $L^2[0, 1]$ is more than just an inner product space. It is a **Hilbert space**, an object that is of central importance in quantum mechanics, signal processing, and mathematical analysis. A Hilbert space is a vector space that allows us to take limits of vectors, too, not just finite sums, so that we can do analysis on it.

If we have a sequence of vectors x_n on our Hilbert space that *accumulates*, in the sense that for any ε we can pick an N large enough so that all the x_i after x_N are within ε of x_N

$$|x_N - x_i| < \varepsilon \tag{1.8}$$

Ex: Show this is the same as forming a Cauchy sequence

Figure: All vectors being within a neighborhood of x_N, accumulating.

Then, if we were working in $\mathbb{R}^n$, this would mean that the sequence should converge within some closed ball $B_\varepsilon(x_N)$. In fact, convergent sequences are *exactly* the ones that accumulate like this.

We want a Hilbert space H, infinite-dimensional or otherwise, to have this property of being **complete** in the inner product. That is, whenever a sequence of vectors accumulate, that sequence will in fact **converge** to a **limit** vector *still in* H.

Definition 1.7 (Real Hilbert Space)**.** A real Hilbert space H is a real inner product space that is **complete** with respect to its inner product.

Example 1.8. The rationals $\mathbb{Q}$ are not a complete metric space. The reals are (they are called the **completion** of the rationals).

Example 1.9. The space of continuous functions is not complete. Consider the sequence of functions:

$$f_n(x) = \begin{cases} -1, & \text{if } x \leqslant -1/n \\ nx, & \text{if } -1/n < x < 1/n \\ 1, & \text{otherwise} \end{cases} \tag{1.9}$$

While each f_n is continuous, the limit is the sign function of x, which is *not* continuous. Similar arguments show that the space of smooth

functions is also not complete, by considering $f_n(x) = e^{-n\pi x^2}$ approaching the zero function with a jump discontinuity at zero.

Ex: Being complete is with respect to a metric, not a topology (reals are homeomorphic to incomplete $[0, 1]$)

Theorem 1.10. *$L^2[a, b]$ is complete, as is $L^2(\mathbb{R})$.*

Proof. Any introductory text in real analysis and measure theory proves this result after developing the necessary background in proper Lebesgue integration, c.f.

Provide a measure theory reference for L^2 being complete, like Stein, but probably not Stein tb$\hbar$.

□

This is an extraordinary property, which places L^2 spaces above all others in their importance to classical analysis. This property will allow us to talk about sequences on these spaces, and hence do analysis on these spaces by talking about sequences of functions converging to a limit point "in the L^2 norm".

The two properties of having an inner product as well as having this notion of convergence in a norm are beautifully useful in formulating the original ideas of Fourier.

1.7 Trigonometric Series

Let us investigate the phenomenon of a wave in a 1-dimensional medium, like a string vibrating in one dimension. At a point x_0, the string has some displacement $u(x_0)$. Its neighbors have displacements $u(x_0 + h)$ and $u(x_0 - h)$.

Draw A String for the Wave equation

Because of **Hooke's Law** $\mathbf{F} = k\mathbf{x}$, the rightward neighbor will exert a force $\mathbf{F}$ on the string at x_0 proportional to the difference in displacement $u(x_0 + h) - u(x_0)$. Similarly, the leftward neighbor will exert a force proportional to $u(x_0 - h) - u(x_0)$. The total force at x_0 then is some scalar multiple times the sum of both these contributions:

$$\mathbf{F}_{x_0} = k[u(x_0 + h) - u(x_0) + u(x_0 - h) - u(x_0)]. \tag{1.10}$$

The mass constant k between successive neighbors can be related to the total spring constant K by $k = KL/h$. This is a simple exercise.

Make an exercise for the total spring constant

.

We also know from **Newton's Second Law** that the force is equal to the mass times the acceleration of the displacement of the string at x_0

$$\mathbf{F}_{x_0} = m\frac{\partial^2 u}{\partial t^2}. \tag{1.11}$$

The mass of the particle at x_0, because it is only a small section of the string, becomes infinitesimal as $h \to 0$ and can be expressed as ρh where ρ is the mass density of the string. Together, then:

$$\rho h\frac{\partial^2 u}{\partial t^2} = \frac{KL}{h}[u(x_0 + h) + u(x_0 - h) - 2u(x_0)] \tag{1.12}$$

This gives rise to the equation:

$$\frac{\partial^2 u}{\partial t^2} = \frac{KL}{\rho}\ \frac{u(x_0 + h) + u(x_0 - h) - 2u(x_0)}{h^2} \tag{1.13}$$

The constant KL/ρ on the right hand side has units of length squared per second squared, and can be interpreted as a velocity squared c^2. The more interesting quantity on the right hand side is what succeeds that. Note that it can be re-written as:

$$\frac{1}{h}\left[\frac{u(x_0 + h) - u(x_0)}{h} - \frac{u(x_0) - u(x_0 - h)}{h}\right]$$

As $h \to 0$ the inside approaches a difference of the derivatives $\frac{\partial u}{\partial x}$ at neighboring points $x_0 + h/2, x_0 - h/2$, so that the whole term becomes the *second derivative* $\partial^2 u/\partial x^2$. The equation for the wave on the string is then:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \tag{1.14}$$

This is the **wave equation** of central importance across the subfields of physics.

The wave equation in one dimension is important for studying simple vibrations of strings, metal beams, and waves of current in

trasmission lines and cavity resonators. However, the higher dimensional generalizations of the wave equation find even more use, especially in studying the propagation of wave energy through physical space. It is worth deriving this as well. The waves propagate through a medium with speed c depending on the properties of the medium.

One way is just to note that now we have $u(x, y)$, and as before we will have that $u(x \pm h, y \pm h)$ will contribute a term that now looks like

$$\frac{u(x+h,y)+u(x-h,y)+u(x,y+h)+u(x,y-h)-4u(x,y)}{h^2}$$

$$\begin{aligned}=&\frac{u(x+h,y)+u(x-h,y)-2u(x,y)}{h^2}\\&+\frac{u(x,y+h)+u(x,y-h)-2u(x,y)}{h^2}\end{aligned}$$

$$\to \frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}$$

so that the wave equation in 2 dimensions is

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2} \tag{1.15}$$

This quantity, $\frac{\partial^2 u}{\partial x^2}$ in one dimension and $\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}$ in two, and more generally

$$\sum_{k=1}^{n}\frac{\partial^2 u}{\partial x_i^2}$$

in n dimensions, is called the **Laplacian** of u. It is the divergence of the gradient, and so is often denoted $\nabla^2 u$, or otherwise just $\triangle u$.

Now for the 2-dimensional argument, you could ask why we did not take into account diagonal neighbors like $u(x + h, y + h)$. One argument could be that they are not "directly connected" to (x, y), but a better answer is that we *can* account for the diagonal neighbors (and associated $\sqrt{2}$s pop up because their distance away is not exactly h), and in the limit as $h \to 0$, the finite differences obtained would still converge to the same Laplacian.

Concept 1.11 (Laplacian, Intuitively)**.** There are many different finite-difference approximations for the Laplacian, all proportional to the idea:

$$\triangle u \propto \frac{\text{sum over } n \text{ neighbors of } u(\text{neighbors}) - n * u(x,y)}{\text{Area}}$$

In this way, if the average value of u on a small circle around x is greater than the value of u at x, the Laplacian is positive, and vice versa gives that the Laplacian is negative.

It is natural for the average behavior of the neighbors at point to "drive" the behavior of the point in the next instant. For waves, the mean value of the neighbors of that point drives that point to *accelerate* in that direction.

Similarly, for heat flow, a similar derivation using Newton's law of *cooling*, instead of motion, gives that this Laplacian: the imbalance between the point and its environment, drives the point to simply *move* to stabilize that imbalance:

Make a heat equation derivation exercise, from Newton's law of COOLING this time

Concept 1.12 (Wave and Heat Equations)**.** The motion of waves is governed by the equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \triangle u$$

for some constant speed c depending on the properties of the medium.

Further, the flow of heat is governed by the equation

$$\tau\frac{\partial u}{\partial t} = \triangle u$$

for some constant τ depending on the properties of the medium.

Fourier arrived at the idea of trigonometric series while studying heat flow. We will look at the example of the vibrating string, pinned down at its edges.

We now begin with a string defined on the interval $[0, 1]$. u satisfies the wave equation, as well as the **boundary conditions** $u(0) = u(1) = 0$. If we have an initial shape of the string $u_0(x)$ as well as an initial velocity $\partial u/\partial t$ at $t = 0$, given by $v_0(x)$, then we'd like to see what dynamics emerge.

1.8 Spectral Theory on Graphs

We begin with the definition of a **finite graph**

Definition 1.13. A finite graph G is a finite set of vertices $V = \{v_1, \ldots, v_n\}$ together with a set of edges $E = \{e_1, \ldots, e_m\}$ such that each edge is an unordered pair of vertices (v_i, v_j) indicating that they are connected.

Draw a figure of a graph

All of this information can be captured in terms of an n by n matrix called the **adjacency matrix**.

Definition 1.14 (Adjacency Matrix)**.** The adjacency matrix A for a graph G is a matrix whose i, jth entry is 1 if (v_i, v_j) is an edge, and 0 otherwise.

The adjacency matrix acting on a vector with 0s everywhere except for the ith slot will give a vector with 0s everywhere except for the slots of the **neighbors** of v_i. That is, all vertices v_j such that (v_i, v_j) is an edge. We denote the set of neighbor vertices to v_i by $N(v_i)$.

Note that this matrix is symmetric, and consists of only zeroes and ones. Similarly we define the **degree matrix** as

Definition 1.15 (Degree Matrix)**.** The degree matrix D for a graph is a diagonal matrix whose (i, i)th entry is the *number of neighbors* of v_i, that is $|N(v_i)|$

Graphs appear throughout modern mathematics and computer science, to the point that we expect that the reader is familiar with them at least at a rudimentary level. The reason we put a focus on graphs in a section that is about Fourier analysis is the following:

Concept 1.16 (Discrete Approximation)**.** In almost all practical matters of engineering and applied math, manifolds are approximated by graphs.

The reason for why we care about this is that it makes everything finite-dimensional, turning all analysis into algebra and allowing us to not worry about notions of convergence and Hilbert spaces. Because of the immense use of such finite methods in practical engineering,

physics, and computer science, we have this topic as a section of its own, not just a mere exercise of example. We will illustrate its applicability to science through a series of examples.

Example 1.17 (Signal Processing)**.** A signal $f(t)$ on a time interval $t \in [0,1]$ has the interval approximated as a graph G_n of vertices $v_0, \ldots, v_n$. v_k represents the point $k/n \in [0,1]$.

v_0 is connected only to v_1, v_n is connected only to v_{n-1} and otherwise v_i is connected to its immediate neighbors $v_{i\pm1}$. The signal is then a function on the set of $n+1$ vertices.

Consider the derivation of the wave and heat equations in the preceding chapter.

Example 1.18 (Image Processing)**.** A $2D$ image represented by $f(x,y)$ on the interval $[0,1] \times [0,1] = [0,1]^2$ is approximated as a graph G_n of vertices $v_{i,j}$ for $0 \leqslant i,j \leqslant n$. The vertex $v_{i,j}$ represents the point $(i/n, j/n) \in [0,1]^2$.

As before, each vertex v_ij on the interior is connected to its four nearest neighbors $v_{i\pm1,j\pm1}$, vertices on the corners are only connected to two such neighbors, and otherwise vertices on the edges are connected to the three such neighbors. Images are then functions on this graph of $(n+1)$ vertices.

In these examples, we can form a **discrete derivative** operator.

Example 1.19 (Discrete Derivative)**.** In the first example, we can form a discrete derivative of a function on these $(n+1)$ vertices as:

$$\Delta f(v_i) = \begin{cases} \frac{f(v_{i+1})-f(v_i)}{(1/n)}, & \text{if } i = 0 \\ \frac{f(v_i)-f(v_{i-1})}{(1/n)}, & \text{if } i = n \\ \frac{f(v_{i+1})-f(v_{i-1})}{2(1/n)}, & \text{otherwise} \end{cases} \tag{1.16}$$

This derivative is called a **central difference**. Of course we could have also picked just a **forward difference**

$$\Delta f(v_i) = \begin{cases} \frac{f(v_{i+1})-f(v_i)}{(1/n)}, & \text{if } i < n \\ \frac{f(v_i)-f(v_{i-1})}{(1/n)}, & \text{if } i = n \end{cases} \tag{1.17}$$

or the other way, as a **backward difference**. All of these operators are linear and representable as $(n+1)$ by $(n+1)$ matrices acting on

the set of functions on the graph, equivalent to[2] $\mathbb{C}^{n+1}$. Their kernel, just like the regular derivative operator, is one-dimensional: the space of constant functions.

Similarly, in the second example we can form discrete derivatives corresponding to the "x" and "y" directions in the same way as before, operating on either the first or the second index of the $v_{i,j}$, respectively.

Much like picking a derivative $\partial/\partial q^i$ is coordinate covariant, so is picking a derivative on a graph not a natural thing. We need to pick a sort of "flow" on the graph along which our derivative vector field is pointing. On a graph given at random, its not at all obvious if there is any natural way to define "derivative". This is further evidence that the coordinate-dependent notion of "derivative" is not of central importance in physics.

Random scrambled graph "where's your derivative now, fam?"

On the other hand, there *is* an operator independent of coordinate system that we *can* form: the **graph Laplacian**. In the previous section, for 1-dimension we defined the Laplacian to be the sum of the function values at the two neighboring points $x_{i\pm1}$ minus 2 times the function value at the point x_0 itself, all divided by h, the infinitesimal length associated to the point x_0. For a 2-dimensional plane, we defined the Laplacian to be the sum of the function values at the 4 neighboring points minus 4 times the value of the function at (x_0, y_0), all divided by h^2, the infinitesimal *area* associated to the point (x_0, y_0).

We now generalize this to a graph. A given point may have k neighbors. We take the sum of the function f at all k of these neighbors and subtract k times the function value at f. As a matrix, this can be seen as:

$$\sum_{v_i \in N(v_0)} [f(v_i) - f(v_0)] = \sum_{v_i \in N(v_0)} f(v_i) - D(v_0) f(v_0) \tag{1.18}$$

$$= (A - D) f \tag{1.19}$$

This last equality recasts it in terms of the linear operators of the adjacency and degree matrix acting on the vector f.

[2]Note we work over complex vector spaces for the same reason as in the previous section: we can find all the eigenvalues of an operator when working over $\mathbb{C}$.

This thing should be proportional (up to the factor of that infinitesimal area associated to each vertex, assuming each vertex has equal area) to the Laplacian as we've defined it before. For this reason, in graph theory we don't care about an overall constant that multiplies our Laplacian, and so there are multiple conventions. Note that the operators A and D are *invariants*, associated to the graph, and so the Laplacian is invariant as well. On any graph, no matter how scrambled-looking it is, there is a good definition for a Laplacian.

Definition 1.20 (Laplacian Matrix of a Graph)**.** For a graph $G = (V, E)$ represented by adjacency matrix A giving degree matrix D, the Laplacian can be written as

$$\triangle = A - D \tag{1.20}$$

Because we care about the matrix itself and not the constant multiplying it, different parts of literature define it differently as:

- (reversed) $\triangle = D - A$
- (normalized) $\triangle = 1 - D^{-1}A$
- (symmetric normalized) $\triangle = 1 - D^{-1/2}AD^{-1/2}$

We will use the symmetric normalized, as it is common in mathematical literature.

Applications to Molecular Orbital Theory

Regardless of what Laplacian definition we use, the matrix is symmetric. By the spectral theorem, this guarantees us that the eigenvalues are all real and that the eigenvectors span the space of functions on the graph, and are in fact *orthogonal*. This is as powerful a result as we could have ever hoped for, because now, if we write the heat equation as

$$-\triangle f = \frac{\partial f}{\partial t}. \tag{1.21}$$

We can expand f in an eigenbasis h_i of eigenvalues λ_i determined from the matrix for $\triangle$:

$$f = \sum_{i=1}^{n} c_i h_i \tag{1.22}$$

where c_i is obtained from orthogonality of the eigenbasis, $c_i = \langle h_i | f \rangle$ so that:

$$-\triangle f = \sum_{i=1}^{n} c_i(-\lambda_i)h_i = \sum_{i=1}^{n} c_i \lambda_i h_i = \frac{\partial f}{\partial t} = \sum_{i=1}^{n} c_i \dot{h}_i. \tag{1.23}$$

As a vector equation, this implies that components must agree:

$$-\lambda_i h_i = \dot{h}_i \Rightarrow h_i(t) = h_i(0)e^{-\lambda t} \tag{1.24}$$

Thus, we get the full solution to the heat equation:

$$f(t) = \sum_{i=0}^{n} c_i e^{-\lambda t} h_i \tag{1.25}$$

In shorthand then, the diffusion of heat operator can be written as the exponential:

$$e^{-t\triangle}$$

And for this reason we can call the Laplacian the *infinitesimal generator* of diffusion.

One important subset of graphs that is worth focusing on is the set of all **k-regular** graphs. These are graphs such that each vertex v has exactly k neighbors: $\forall v \ |N(v)| = k$.

Observation 1.21. *The vector* $(1, \ldots, 1)$ *is an eigenvector of* Δ *with eigenvalue* 0*.*

The matrix A acting on $(1, 0 \ldots, 0)$ gives the vector of 1s on all neighbors of v_1 and 0s elsewhere. It's easy to see then that A acting on $(1, \ldots, 1)$ will give a vector with $|N(v_i)|$ on its ith entry. But this is exactly the same as the action of D on $(1, \ldots, 1)$, so $A - D$ gives zero on this vector.

The equivalent on manifolds is that the constant function has vanishing Laplacian.

1.9 Spectral Theory on $\mathbb{R}$

Chapter 2

Beyond Harmonics: Representation Theory

2.1 The Representations of Finite Groups

2.2 Character Theory

2.3 The Representations of Topological Groups

2.4 Representations Induced

Part 2

Physics

Chapter 3

Symmetries of the sphere: $\mathrm{SU}(2)$ and friends

Chapter 4

Classical Mechanics and Symplectic Geometry

Chapter 5

Einstein's Gravity

Part 3

Advanced Topics

Chapter 6

An Introduction to Quantization

Chapter 7

Classification of Simple Lie Algebras over $\mathbb{C}$

Conclusion

Appendices

Chapter A

Eigenvalues and the Jordan Normal Form

Should this just be an appendix section?

IDK how to make chapter -> appendix which is a big problem here

Eigenvalues of a linear tranformation and their associated eigenvectors hold great value, The eigenvectors determine subspaces of V where the action of the transformation becomes as *simple as possible*: literally scalar multiplication. If we can decompose V, by using eigenvectors, into a series of spaces along which the transformation is simply scaling, we can **diagonalize** the matrix, and all the information and dynamics that it represents become extraordinarily simple to work with. The difficult-to-work-with process of matrix multiplication collapses down to multiplication by scalars.

When we studied a linear transformation $T : V \to V$ on a real vector space V of dimension n. T takes a vector $\mathbf{v}_i$ and maps it to another vector $T\mathbf{v}_i \in V$. If we pick a specific basis $B\{\mathbf{e}_i\}$ for our space, then T can be represented by a **matrix**, A^i_j whose ijth entries in the basis B can be explicitly found:

$$A^i_j = \begin{pmatrix} T(\mathbf{e}_1) & \dots & T(\mathbf{e}_n) \end{pmatrix}_B = \begin{pmatrix} T(\mathbf{e}_1)^1 & \dots & T(\mathbf{e}_n)^1 \\ \vdots & \dots & \vdots \\ T(\mathbf{e}_1)^n & \dots & T(\mathbf{e}_n)^n \end{pmatrix}_B \tag{A.1}$$

Note this same transformation can be viewed through different bases

$\mathbf{e}'_i$, producing different matrix representations depending on the basis. Therefore we should realize, just like tuples of numbers only represented physical vectors once a coordinate system was chosen, the same is true for how matrices only *represent* the linear operators we care about. In particular, since we can write matrix multiplication in a basis as

$$(Tv)^i = A^i_j v^j \tag{A.2}$$

Now if we change basis from B to B' and we have a matrix Q^j_i that takes the components v^i of a vector in the B basis and gives us the B' basis $(v')^i = Q^j_i v^j$ then we can transform A^i_j according to:

$$(A^i_j)_{B'} = Q^i_k (A^k_l)_B (Q^{-1})^l_j \tag{A.3}$$

Accordingly, $A_{B'} = QA_BQ^{-1}$. This should make perfect sense: if we're given a vector's components v'^i in the B' basis, first we act by Q^{-1} to get back into the B basis, then we act by A, since we know how T acts in this basis, and then we map that vector back into the B' basis with Q.

One operator can be represented by many different matrices. Here, we will attempt to *classify* all the possible linear operators on a space. This would be a great thing, since linear operators can be so diverse, ranging from stretching to shearing to rotation, and much more.

We start with the **trivial** linear operator, 0. Nothing can be more basic than the action of sending everything to zero. The simplest linear transformation that isn't just 0 is multiplication by a scalar, λ. This is basic multiplication, and as a matrix is represented just as

$$\lambda = \lambda I = \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \end{pmatrix} = \lambda \delta^i_j$$

Note that if we change basis, the matrix representation for scalar multiplication *doesn't* change: regardless of how you look at it, multiplication by λ is multiplication by λ. The set of linear operators that are just simple scalar multiplication form a subspace of all linear operators, and are called **simple**. Simple linear operators are literally just numbers, acting on the space by scaling everything the same way. Note the trivial 0 operator is an example of a simple operator.

Of course, the set of linear operators that are simple is a very small subset of the set of all linear operators. A linear operator on V does not in general act simply on V, but the next best thing we could hope for is that T becomes simple when its action is restricted to a *subspace* of V.

What would this mean, mathematically? We want to find vectors $\mathbf{v}$ so that

$$T\mathbf{v} = \lambda\mathbf{v} \tag{A.4}$$

Note that such a set of $\mathbf{v}$ is *also* a vector space.

Proposition A.1. *The set of vectors where $T\mathbf{v} = \lambda\mathbf{v}$, for some specific λ, is a vector space, denoted V_λ.*

Proof. If $\mathbf{v}, \mathbf{w}$ have

$$T\mathbf{v} = \lambda\mathbf{v}, T\mathbf{w} = \lambda\mathbf{w}$$

then

$$T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w}) = a\lambda\mathbf{v} + b\lambda\mathbf{w} = \lambda(a\mathbf{v} + b\mathbf{w})$$

□

We then have that V_λ is exactly the subspace where T acts simply as λ. When $\lambda = 0$ this is exactly what we referred to as $\ker T$. How can we find this subspace? The trick all lies in noting that the number λ is *itself* just a linear transformation. Then finding when Equation (A.4) is satisfied is the same as finding when:

$$(T - \lambda)\mathbf{v} = 0 \tag{A.5}$$

So this is no different from calculating a kernel of the shifted operator $T - \lambda$. If we know how to find the kernel, the subspace where $T = 0$, we also know how to find the subspace where $T = \lambda$ by finding the kernel where $T - \lambda = 0$. It's a really simple idea.

But it's worth asking the question "does $T - \lambda$ even *have* a nontrivial kernel?". That is, for which λ is V_λ nontrivial? We know a quick way to answer this question. $T - \lambda$ will have a nontrivial kernel exactly when:

$$\det(T - \lambda) = 0 \tag{A.6}$$

In a specific basis, T can be represented by a matrix T_{ij} and λ as a operator is always the matrix $\lambda I = \lambda\delta_{ij}$ so that we can directly

compute this in a basis by computing the matrix determinant $\det(T_{ij} - \lambda\delta_{ij})$. It is not hard to see that this is a polynomial

$$\begin{pmatrix} T_1^1 - \lambda & \dots & T_1^n \\ \vdots & & \vdots \\ T_1^n & \dots & T_n^n - \lambda \end{pmatrix} \tag{A.7}$$

It's not difficult to see that the resulting determinant becomes a *polynomial of degree n* in λ. It is called the **characteristic polynomial** $p_T(\lambda)$ of the transformation T.

The determinant of a transformation is itself independent of the coordinate system we use, so the characteristic polynomial is an invariant, not dependent on the basis that gives us our T_{ij}. The roots of this polynomial are special numbers, characteristic of the transformation. These "characteristic values" or, from German, **eigenvalues**, of T are thus also invariants independent of coordinate system. They are exactly the values for which V_λ is nontrivial, so that T acts simply as one of these λ_i on each V_{λ_i}.

But over the reals, there are transformations that don't even have eigenvalues:

Example A.2. The transformation representing counterclockwise rotation of the basis, which can be represented by the real matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has no real eigenvalues.

We see this immediately since its characteristic polynomial is λ^2+1 which has no real roots. It does, however, have complex roots $\lambda = \pm i$ and we shouldn't ignore this fact. If we want to find a vector subspace on which this transformation acts simply as $\pm i$, we need to *extend* the number system that we are working over, from the reals to the complex numbers. In fact, working over the complex numbers is completely *ideal* if we want $p_T(\lambda)$ to have roots, giving us eigenvalues. The reason is because of this foundational theorem:

Theorem A.3 (Fundamental Theorem of Algebra)**.** *Every non-constant polynomial with coefficients in* $\mathbb{C}$ *has a root in* $\mathbb{C}$*.*

Proof. (Optional, for those familia with complex analysis). The easiest way to do this is by use of complex analysis. Any introductory text

in complex analysis will give the tools necessary to form this proof. If the reader is familiar with Cauchy's integral formula (**which can be derived in one of the exercises in the preceding chapter**)

Make an exercise in the preceding chapter to derive all of complex analysis' integral formulae

, then he can understand the following:

For a polynomial p of a complex variable z, assume $p(z)$ did not have a root in $\mathbb{C}$. Then $f(z) := 1/p(z)$ is bounded on $\mathbb{C}$ so that $|f(z)| \leqslant B$, and is also holomorphic in fact entire on the complex plane. This means it admits a Taylor series about $z = 0$

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \tag{A.8}$$

with a_k obtained from the Cauchy integral formula by:

$$a_k = \frac{f^{(k)}(0)}{k!} = \frac{1}{2\pi i} \oint_{\partial D_R(0)} \frac{f(\zeta)}{\zeta^{k+1}} d\zeta \tag{A.9}$$

We are integrating over the boundary of the disk of radius R enclosing zero. Since this can be any disk, we let R becomes large so that $|\zeta| = |Re^{i\theta}| = |R|$ also becomes large, and so

$$\begin{aligned} |a_k| &= \frac{1}{2\pi} \left| \oint_{\partial D_R(0)} \frac{f(\zeta)}{\zeta^{k+1}} d\zeta \right| \\ &\leqslant \frac{1}{2\pi} \oint_{D_R(0)} \frac{|f(\zeta)|}{R^{k+1}} |d\zeta| \\ &\leqslant \frac{1}{2\pi} (2\pi R) \max_{\partial D_R(0)} \frac{f}{R^{k+1}} \\ &\leqslant \frac{B}{R^k} \end{aligned}$$

since R can be anything, we let it go to infinity and obtain $|a_k| \leqslant 0$ for all $k \geqslant 1$. This gives us that $f(z)$ is constant, and also a powerful theorem in general:

Theorem A.4 (Liouville)**.** *Any bounded entire function on $\mathbb{C}$ is constant.*

In particular $1/p(z)$ is a constant, so $p(z)$ must be a constant as well. If a polynomial does not have roots in $\mathbb{C}$, it must be a constant function. □

Since every polynomial p has a root r, we can write it as $p(z) = (z - r_1)p'(z)$, with p' a polynomial of degree $n - 1$. Applying the fundamental theorem to p', now, and continuing inductively immediately shows that.

Corollary A.5. *Every non-constant polynomial with coefficients in $\mathbb{C}$ has exactly n roots, counted with multiplicity, in $\mathbb{C}$.*

That means that a polynomial over $\mathbb{C}$, $p(x) = \sum_{i=0}^{n} a_i x^i$ splits entirely into a product of linear factors $c(x - r_1) \dots (x - r_n)$ over $\mathbb{C}$. In particular, this means that our characteristic polynomial will have n roots, counted with multiplicity. Although the complex numbers are less intuitive to be introduced to than the real numbers were, they have actually made things *easier* and much more interesting. So instead of disregarding these numbers because we think they are are unphysical, let us accept them, and later try to physically interpret what they mean!

The key word of Corollary A.5 is "with multiplicity". Our polynomial will always have exactly n roots, but they need not all be distinct. Let's say that k of them are distinct $\lambda_1 \dots \lambda_k$, each with multiplicity m_i (so $\sum_{i=1}^{k} m_i = n$). We can write the polynomial as

$$p_A(t) = (-1)^n \prod_{i=1}^{k} (t - \lambda_i)^{m_i} \tag{A.10}$$

First let's assume no root is repeated for this polynomial. Then we have $\lambda_1 \dots \lambda_n$ distinct. Each λ_i gives rise to a corresponding eigenvector $\mathbf{v}_i$.

Lemma A.6. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. Let $\mathbf{v}_1$ correspond to λ_1, $\mathbf{v}_2$ to λ_2. If they are linearly dependent then

$$\exists a, b \neq 0 \text{ s.t. } a\mathbf{v}_1 + b\mathbf{v}_2 = 0 \tag{A.11}$$

But then

$$T(0) = T(a\mathbf{v}_1 + b\mathbf{v}_2) = a\lambda_1\mathbf{v}_1 + b\mathbf{v}_2 = 0 \tag{A.12}$$

Subtracting from this λ_1 times the above equation, we get that $b(\lambda_2 - \lambda_1)\mathbf{v}_2 = 0$, and since $b \neq 0, \mathbf{v}_1 \neq 0$, we get $\lambda_1 = \lambda_2$. □

So when all the λ_i are distinct, this gives rise to n pairwise linearly independent eigenvectors $\mathbf{v}_i$ of eigenvalue λ_i. But this forms a *basis* for our vector space! In this basis, we get that

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \tag{A.13}$$

This means on *each* 1-dimensional subspace $V_{\lambda_i} = \text{span}\{\mathbf{v}_i\}$, T acts simply, and the entire space V can be written as a *direct sum* (c.f. Section **??**) of the V_{λ_i}

$$V = \bigoplus_{i=1}^{n} V_{\lambda_i} \tag{A.14}$$

When the space V can be decomposed as a direct sum of subspaces V_i on which the operator T acts simply, T is called **semisimple**. There is another word you may have heard in place of semisimple in the past, which means the same thing: **diagonalizable**.

Definition A.7 (Semisimplicity)**.** A linear operator $T \in \text{End}(V)$ is called *semisimple* or *diagonalizable* if there exists a decomposition

$$V = \bigoplus_{i} V_i \tag{A.15}$$

such that T acts simply, as λ_i on each V_i. We write that the operator T itself can be written as a direct sum of the simple linear operators:

$$T = \bigoplus_{i} T_i \tag{A.16}$$

Where T_i is the restriction of T to V_i, $T|_{V_i}$.

Graphic of restriction of an operator to subspaces, and now its just a constant scalar

Note Equation (A.13) was a special case of semisimplicity, when all V_i had dimension 1. In general, a semisimple matrix when viewed

in a basis $\mathbf{v}_j$ so that each $\mathbf{v}_j$ lies in one of the V_i will look just like Equation (A.13), *allowing* for repeated λ_i.

But from our previous analysis, this implies that for a linear operator whose characteristic polynomial *doesn't* have repeated roots (i.e all roots are **simple**) that there the eigenvectors form a basis for the space, in which the operator looks diagonal.

Theorem A.8 (Semisimplicity for T Lacking Repeated Eigenvalues)**.** *If the characteristic polynomial of a linear operator T does not have repeated roots, then T is semisimple, and V decomposes as.*

$$V = \bigoplus_{i=1}^{n} \ker(T - \lambda_i) \tag{A.17}$$

Proof. To reiterate the idea, this follows immediately from Lemma A.6, which means the eigenvectors form a basis for V. So V can be decomposed as a direct sum of the eigenvector spaces, on each of which T acts simply by definition, so T is semisimple. □

Now note that almost all polynomials do not have repeated roots. An intuitive sketch of this is that if you pick a polynomial at random, the roots will have some random distribution on $\mathbb{C}$. The chance that two roots will happen to be at the exact same point and form a double root is infinitesimal: the same as the chance of picking a random real number and getting an integer. So a linear operator picked at random, we'd expect, would not have any double roots. That is *almost all linear operators are semisimple.* There's a way to view almost any operator so that the operator looks diagonal.

The converse of Theorem A.8 is not true. The characteristic polynomial can have repeated roots and the operator can still be semisimple. For example simple operators λ are trivially semisimple and their characteristic polynomials are $(t-\lambda)^n$. Are all operators just semisimple, diagonal when viewed in the right eigenbasis?

The classic example of an operator that is not semisimple is the one represented by the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The characteristic polynomial for this is $p_A(t) = t^2$, the same as the characteristic polynomial for the zero matrix, however the eigenspace

$V_0 = \ker A$ is only one-dimensional, and the characteristic polynomial tells us there are no other eigenspaces. So although the eigenvalue 0 has **algebraic multiplicity** 2, we say the associated eigenspace only has **geometric multiplicity** 1, since $\ker T - \lambda = 1$.

This operator is on the other side of the coin from semisimple: it is **nilpotent**. A nilpotent operator T is one so that $T^k = 0$ for some number k. The minimum such k for which $T^k = 0$ is called the **nilpotence degree** of T.

Theorem A.9 (Primary Decomposition of a Linear Operator)**.** *For a general $T \in \mathrm{End}(V)$, with its distinct eigenvalue roots to its characteristic polynomial labelled by $\lambda_1, \ldots, \lambda_k$, and each λ_i with multiplicity denoted m_i, we can decompose V as a direct sum:*

$$V = \bigoplus_{i=1}^{k} \ker\left[(T - \lambda_i)^{m_i}\right] \tag{A.18}$$

There is in fact a connection between this statement and how an integer n can be decomposed into a product of power of "simple" integers: primes, as $n = p_1^{\alpha_1} \ldots p_n^{\alpha_n}$.

Chapter B

Category Theory

We begin by going back to elementary linear algebra. Linear transformations between vector spaces $T : V \to W$ satisfy

$$T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w}) \tag{B.1}$$

That is, they respect linear combinations of vectors.

In more abstract language, transformations that respect algebraic structure (like linear transformations respect vector space structure) generally called **homomorphisms**. We denote the set of linear transformations from V to W by $\mathrm{Hom}(V, W)$. Linear transformations from V to itself, $\mathrm{Hom}(V, V)$ are called **endomorphisms** and are denoted by $\mathrm{End}(V) := \mathrm{Hom}(V, V)$. A particularly special endomorphism on V is the *identity* map 1_V that acts trivially on V, sending every vector to itself.

Each linear transformation in $T \in \mathrm{Hom}(V, W)$ has two important subspaces associated with it: its kernel and its image. We say the kernel, $\ker T$ is the subspace of V that T sends to zero. The image $\mathrm{im}\, T$ is the subspace of W that T maps into.

Now certain linear transformations have the property that they map every vector $\mathbf{v} \in V$ to a *unique* vector in W. Section **??** has shown that this is equivalent to $\ker T = 0$. Such linear transformations are called one-to-one, and they are also called **injective**.

On the other side, linear transformations can also have the property that they map *to* every vector $w \in W$. That is, $\mathrm{im}\, T = W$. Such linear transformations are called onto, or **surjective**.

When a linear transformation is both injective and surjective, it is **bijective**, meaning there is an inverse $T^{-1} : V \to W$ so that

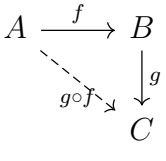
$T \circ T^{-1} = 1_V$ and $T^{-1} \circ T = 1_W$. These are called **bijections** and also **isomorphisms** between V and W. Endomorphisms from V to itself that are isomorphisms are called **automorphisms** and are denoted by $\mathrm{Aut}(V)$ or $\mathrm{GL}(V)$.

This is the inspiration for the much more general setting: Category Theory.

A **category** is simply something that consists of **objects** that are linked together by arrows called **morphisms**, also just called **maps**. A morphism f takes the **source object** A to a **target object** B. Such morphisms can be drawn using **diagrams** as

$$A \xrightarrow{f} B$$

We denote the class of morphisms from A to B by $\hom(A, B)$. A morphism from A to itself is called an **endomorphism** and the class of endomorphisms is denoted by $\mathrm{end}(A) = \hom(A, A)$. If f is a morphism from object A to B and g is a morphism from object B to C then we can **compose** g with f to form another morphism $g \circ f$ according to the following diagram.



We say "the diagram commutes" to mean that we can go either direction in the diagram, along f then g or just along $g \circ f$ and get the same result.

Further, there is a special morphism to every object A, the **identity morphism** denoted by 1_A such that for any morphism $f : B \to A$ into A, $1_A \circ f = f$ and for any morphism $g : A \to B$ out of A, we have $g \circ 1_A = g$. Note this implies that the identity is unique, as if there were another $1'_A$, then by these properties, $1'_A = 1_A \circ 1'_A = 1_A$.

The class of all vector spaces forms a category, denoted **Vect**, where morphisms are linear maps. This was our motivating example. Further along the same line of thought, groups and rings are categories, whose morphisms are exactly the algebraic homomorphisms associated with those algebraic structures. Abelian groups also form a

category, that is a **subcategory** of the category of groups. A subcategory S of C is a category whose objects and morphisms are objects and morphisms in C, respectively.

Even more simply, the class[1] of all sets forms a category, **Set**, where morphisms are exactly functions between sets.

More interestingly, the class of all *topological spaces* forms a category, with morphisms being the *continuous functions*. A sub-category of this is the category of manifolds **Man**. If we instead want our morphisms to be smooth maps, we can form the subcategory of smooth manifolds, denoted $\mathbf{Man}^{\infty}$.

[1]We keep using the word class instead of set exactly because we don't want to run into the paradox of saying "the set of all sets". The word "class" refers to a mathematical construction designed to avoid this, which is not focused on here.

Notation

References

Index