IV

Linear Algebra

Part IV: Contents

9 Vector spaces

This is a pretty light chapter. The point of it is to define what a vector space and a basis are. These are intuitive concepts that you may already know.

§9.1 The definitions of a ring and field

Prototypical example for this section: $\mathbb{Z}$, $\mathbb{R}$, *and* $\mathbb{C}$ *are rings; the latter two are fields.*

I'll very informally define a ring/field here, in case you skipped the earlier chapter.

- A **ring** is a structure with a *commutative* addition and multiplication, as well as subtraction, like $\mathbb{Z}$. It also has an additive identity 0 and multiplicative identity 1.
- If the multiplication is invertible like in $\mathbb{R}$ or $\mathbb{C}$, (meaning $\frac{1}{x}$ makes sense for any $x \neq 0$), then the ring is called a **field**.

In fact, if you replace "field" by "$\mathbb{R}$" everywhere in what follows, you probably won't lose much. It's customary to use the letter R for rings, and k or K for fields.

Finally, in case you skipped the chapter on groups, I should also mention:

- An **additive abelian group** is a structure with a commutative addition, as well as subtraction, plus an additive identity 0. It doesn't have to have multiplication. A good example is $\mathbb{R}^3$ (with addition componentwise).

§9.2 Modules and vector spaces

Prototypical example for this section: Polynomials of degree at most n.

You intuitively know already that $\mathbb{R}^n$ is a "vector space": its elements can be added together, and there's some scaling by real numbers. Let's develop this more generally.

Fix a commutative ring R. Then informally,

An R-module is any structure where you can add two elements and scale by elements of R.

Moreover, a **vector space** is just a module whose commutative ring is actually a field. I'll give you the full definition in a moment, but first, examples...

Example 9.2.1 (Quadratic polynomials, aka my favorite example)

My favorite example of an $\mathbb{R}$-vector space is the set of polynomials of degree at most two, namely

$$\left\{ ax^2 + bx + c \mid a, b, c \in \mathbb{R} \right\}.$$

Indeed, you can add any two quadratics, and multiply by constants. You can't multiply two quadratics to get a quadratic, but that's irrelevant – in a vector space there need not be a notion of multiplying two vectors together.

In a sense we'll define later, this vector space has dimension 3 (as expected!).

Example 9.2.2 (All polynomials)

The set of *all* polynomials with real coefficients is an $\mathbb{R}$-vector space, because you can *add any two polynomials* and *scale by constants*.

Example 9.2.3 (Euclidean space)

(a) The complex numbers
$$\{a + bi \mid a, b \in \mathbb{R}\}$$
form a real vector space. As we'll see later, it has "dimension 2".

(b) The real numbers $\mathbb{R}$ form a real vector space of dimension 1.

(c) The set of 3D vectors
$$\{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$
forms a real vector space, because you can add any two triples component-wise. Again, we'll later explain why it has "dimension 3".

Example 9.2.4 (More examples of vector spaces)

(a) The set
$$\mathbb{Q}[\sqrt{2}] = \left\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\right\}$$
has a structure of a $\mathbb{Q}$-vector space in the obvious fashion: one can add any two elements, and scale by rational numbers. (It is not an $\mathbb{R}$-vector space — why?)

(b) The set
$$\{(x, y, z) \mid x + y + z = 0 \text{ and } x, y, z \in \mathbb{R}\}$$
is a 2-dimensional real vector space.

(c) The set of all functions $f \colon \mathbb{R} \to \mathbb{R}$ is also a real vector space (since the notions $f + g$ and $c \cdot f$ both make sense for $c \in \mathbb{R}$).

Now let me write the actual rules for how this multiplication behaves.

Definition 9.2.5. Let R be a commutative ring. An R-**module** starts with an additive abelian group $M = (M, +)$ whose identity is denoted $0 = 0_M$. We additionally specify a left multiplication by elements of R. This multiplication must satisfy the following properties for $r, r_1, r_2 \in R$ and $m, m_1, m_2 \in M$:

(i) $r_1 \cdot (r_2 \cdot m) = (r_1 r_2) \cdot m$.

(ii) Multiplication is distributive, meaning
$$(r_1 + r_2) \cdot m = r_1 \cdot m + r_2 \cdot m \text{ and } r \cdot (m_1 + m_2) = r \cdot m_1 + r \cdot m_2.$$

(iii) $1_R \cdot m = m$.

(iv) $0_R \cdot m = 0_M$. (This is actually extraneous; one can deduce it from the first three.)

If R is a field we say M is an R-**vector space**; its elements are called **vectors** and the members of R are called **scalars**.

Abuse of Notation 9.2.6. In the above, we're using the same symbol $+$ for the addition of M and the addition of R. Sorry about that, but it's kind of hard to avoid, and the point of the axioms is that these additions should be related. I'll try to remember to put $r \cdot m$ for the multiplication of the module and $r_1 r_2$ for the multiplication of R.

Question 9.2.7. In Example 9.2.1, I was careful to say "degree at most 2" instead of "degree 2". What's the reason for this? In other words, why is

$$\left\{ax^2 + bx + c \mid a, b, c \in \mathbb{R}, a \neq 0\right\}$$

not an $\mathbb{R}$-vector space?

A couple less intuitive but somewhat important examples...

Example 9.2.8 (Abelian groups are $\mathbb{Z}$-modules)
(Skip this example if you're not comfortable with groups.)

(a) The example of real polynomials

$$\left\{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\right\}$$

is also a $\mathbb{Z}$-module! Indeed, we can add any two such polynomials, and we can scale them by integers.

(b) The set of integers modulo 100, say $\mathbb{Z}/100\mathbb{Z}$, is a $\mathbb{Z}$-module as well. Can you see how?

(c) In fact, *any* abelian group $G = (G, +)$ is a $\mathbb{Z}$-module. The multiplication can be defined by

$$n \cdot g = \underbrace{g + \cdots + g}_{n \text{ times}} \qquad (-n) \cdot g = n \cdot (-g)$$

for $n \geq 0$. (Here $-g$ is the additive inverse of g.)

Example 9.2.9 (Every ring is its own module)

(a) $\mathbb{R}$ can be thought of as an $\mathbb{R}$-vector space over itself. Can you see why?

(b) By the same reasoning, we see that *any* commutative ring R can be thought of as an R-module over itself.

§9.3 Direct sums

Prototypical example for this section: $\{ax^2 + bx + c\} = \mathbb{R} \oplus x\mathbb{R} \oplus x^2\mathbb{R}$, *and* $\mathbb{R}^3$ *is the sum of its axes.*

Let's return to Example 9.2.1, and consider

$$V = \left\{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\right\}.$$

Even though I haven't told you what a dimension is, you can probably see that this vector space "should have" dimension 3. We'll get to that in a moment.

The other thing you may have noticed is that somehow the x^2, x and 1 terms don't "talk to each other". They're totally unrelated. In other words, we can consider the three sets

$$\begin{aligned} x^2\mathbb{R} &:= \left\{ax^2 \mid a \in \mathbb{R}\right\} \\ x\mathbb{R} &:= \{bx \mid b \in \mathbb{R}\} \\ \mathbb{R} &:= \{c \mid c \in \mathbb{R}\}. \end{aligned}$$

In an obvious way, each of these can be thought of as a "copy" of $\mathbb{R}$.

Then V quite literally consists of the "sums of these sets". Specifically, every element of V can be written *uniquely* as the sum of one element from each of these sets. This motivates us to write

$$V = x^2\mathbb{R} \oplus x\mathbb{R} \oplus \mathbb{R}.$$

The notion which captures this formally is the **direct sum**.

Definition 9.3.1. Let M be an R-module. Let M_1 and M_2 be subsets of M which are themselves R-modules. Then we write $M = M_1 \oplus M_2$ and say M is a **direct sum** of M_1 and M_2 if every element from M can be written uniquely as the sum of an element from M_1 and M_2.

Example 9.3.2 (Euclidean plane)
Take the vector space $\mathbb{R}^2 = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}\}$. We can consider it as a direct sum of its x-axis and y-axis:

$$X = \{(x, 0) \mid x \in \mathbb{R}\} \text{ and } Y = \{(0, y) \mid y \in \mathbb{R}\}.$$

Then $\mathbb{R}^2 = X \oplus Y$.

This gives us a "top-down" way to break down modules into some disconnected components.

By applying this idea in reverse, we can also construct new vector spaces as follows. In a very unfortunate accident, the two names and notations for technically distinct things are exactly the same.

Definition 9.3.3. Let M and N be R-modules. We define the **direct sum** $M \oplus N$ to be the R-module whose elements are pairs $(m, n) \in M \times N$. The operations are given by

$$(m_1, n_1) + (m_2, n_2) = (m_1 + m_2, n_1 + n_2).$$

and

$$r \cdot (m, n) = (r \cdot m, r \cdot n).$$

For example, while we technically wrote $\mathbb{R}^2 = X \oplus Y$, since each of X and Y is a copy of $\mathbb{R}$, we might as well have written $\mathbb{R}^2 \cong \mathbb{R} \oplus \mathbb{R}$.

Abuse of Notation 9.3.4. The above illustrates an abuse of notation in the way we write a direct sum. The symbol $\oplus$ has two meanings.

- If V is a *given* space and W_1 and W_2 are subspaces, then $V = W_1 \oplus W_2$ means that "V *splits* as a direct sum $W_1 \oplus W_2$" in the way we defined above.
- If W_1 and W_2 are two *unrelated* spaces, then $W_1 \oplus W_2$ is *defined* as the vector space whose *elements* are pairs $(w_1, w_2) \in W_1 \times W_2$.

You can see that these definitions "kind of" coincide.

In this way, you can see that V should be isomorphic to $\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$; we had $V = x^2\mathbb{R} \oplus x\mathbb{R} \oplus \mathbb{R}$, but the 1, x, x^2 don't really talk to each other and each of the summands is really just a copy of $\mathbb{R}$ at heart.

Definition 9.3.5. We can also define, for every positive integer n, the module

$$M^{\oplus n} := \underbrace{M \oplus M \oplus \cdots \oplus M}_{n \text{ times}}.$$

§9.4 Linear independence, spans, and basis

Prototypical example for this section: $\{1, x, x^2\}$ *is a basis of* $\{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\}$.

The idea of a basis, the topic of this section, gives us another way to capture the notion that

$$V = \left\{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\right\}$$

is sums of copies of $\{1, x, x^2\}$. This section should be very intuitive, if technical. If you can't see why the theorems here "should" be true, you're doing it wrong.

Let M be an R-module now. We define three very classical notions that you likely are already familiar with. If not, fall upon your notion of Euclidean space or V above.

Definition 9.4.1. A **linear combination** of some vectors $v_1, \ldots, v_n$ is a sum of the form $r_1v_1 + \cdots + r_nv_n$, where $r_1, \ldots, r_n \in R$. The linear combination is called **trivial** if $r_1 = r_2 = \cdots = r_n = 0_R$, and **nontrivial** otherwise.

Definition 9.4.2. Consider a finite set of vectors $v_1, \ldots, v_n$ in a module M.

- It is called **linearly independent** if there is no nontrivial linear combination with value 0_M. (Observe that $0_M = 0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n$ is always true – the assertion is that there is no other way to express 0_M in this form.)
- It is called a **generating set** if every $v \in M$ can be written as a linear combination of the $\{v_i\}$. If M is a vector space we say it is **spanning** instead.
- It is called a **basis** (plural **bases**) if every $v \in M$ can be written *uniquely* as a linear combination of the $\{v_i\}$.

The same definitions apply for an infinite set, with the proviso that all sums must be finite.

So by definition, $\{1, x, x^2\}$ is a basis for V. It's not the only one: $\{2, x, x^2\}$ and $\{x + 4, x - 2, x^2 + x\}$ are other examples of bases, though not as natural. However, the set $S = \{3 + x^2, x + 1, 5 + 2x + x^2\}$ is not a basis; it fails for two reasons:

- Note that $0 = (3 + x^2) + 2(x + 1) - (5 + 2x + x^2)$. So the set S is not linearly independent.
- It's not possible to write x^2 as a sum of elements of S. So S fails to be spanning.

With these new terms, we can say a basis is a linearly independent and spanning set.

Example 9.4.3 (More examples of bases)

(a) Regard $\mathbb{Q}[\sqrt{2}] = \left\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\right\}$ as a $\mathbb{Q}$-vector space. Then $\{1, \sqrt{2}\}$ is a basis.

(b) If V is the set of all real polynomials, there is an infinite basis $\{1, x, x^2, \dots\}$. The condition that we only use finitely many terms just says that the polynomials must have finite degree (which is good).

(c) Let $V = \{(x, y, z) \mid x + y + z = 0 \text{ and } x, y, z \in \mathbb{R}\}$. Then we expect there to be a basis of size 2, but unlike previous examples there is no immediately "obvious" choice. Some working examples include:

- $(1, -1, 0)$ and $(1, 0, -1)$,
- $(0, 1, -1)$ and $(1, 0, -1)$,
- $(5, 3, -8)$ and $(2, -1, -1)$.

Exercise 9.4.4. Show that a set of vectors is a basis if and only if it is linearly independent and spanning. (Think about the polynomial example if you get stuck.)

Now we state a few results which assert that bases in vector spaces behave as nicely as possible.

Theorem 9.4.5 (Maximality and minimality of bases)

Let V be a vector space over some field k and take $e_1, \dots, e_n \in V$. The following are equivalent:

(a) The e_i form a basis.

(b) The e_i are spanning, but no proper subset is spanning.

(c) The e_i are linearly independent, but adding any other element of V makes them not linearly independent.

Remark 9.4.6 — If we replace V by a general module M over a commutative ring R, then (a) $\implies$ (b) and (a) $\implies$ (c) but not conversely.

Proof. Straightforward, do it yourself if you like. The key point to notice is that you need to divide by scalars for the converse direction, hence V is required to be a vector space instead of just a module for the implications (b) $\implies$ (a) and (c) $\implies$ (a). □

Theorem 9.4.7 (Dimension theorem for vector spaces)

If a vector space V has a finite basis, then every other basis has the same number of elements.

Proof. We prove something stronger: Assume $v_1, \dots, v_n$ is a spanning set while $w_1, \dots, w_m$ is linearly independent. We claim that $n \geq m$.

Question 9.4.8. Show that this claim is enough to imply the theorem.

Let $A_0 = \{v_1, \dots, v_n\}$ be the spanning set. Throw in w_1: by the spanning condition, $w_1 = c_1 v_1 + \cdots + c_n v_n$. There's some nonzero coefficient, say c_n. Thus

$$v_n = \frac{1}{c_n} w_1 - \frac{c_1}{c_n} v_1 - \frac{c_2}{c_n} v_2 - \dots.$$

Thus $A_1 = \{v_1, \dots, v_{n-1}, w_1\}$ is spanning. Now do the same thing, throwing in w_2, and deleting some element of the v_i as before to get A_2; the condition that the w_i are linearly independent ensures that some v_i coefficient must always not be zero. Since we can eventually get to A_m, we have $n \geq m$. □

Remark 9.4.9 (Generalizations)

- The theorem is true for an infinite basis as well if we interpret "the number of elements" as "cardinality". This is confusing on a first read through, so we won't elaborate.
- In fact, this is true for modules over any commutative ring. Interestingly, the proof for the general case proceeds by reducing to the case of a vector space.

The dimension theorem, true to its name, lets us define the **dimension** of a vector space as the size of any finite basis, if one exists. When it does exist we say V is **finite-dimensional**. So for example,

$$V = \left\{ ax^2 + bx + c \mid a, b, c \in \mathbb{R} \right\}$$

has dimension three, because $\{1, x, x^2\}$ is a basis. That's not the only basis: we could as well have written

$$\left\{ a(x^2 - 4x) + b(x + 2) + c \mid a, b, c \in \mathbb{R} \right\}$$

and gotten the exact same vector space. But the beauty of the theorem is that no matter how we try to contrive the generating set, we always will get exactly three elements. That's why it makes sense to say V has dimension three.

On the other hand, the set of all polynomials $\mathbb{R}[x]$ is *infinite-dimensional* (which should be intuitively clear).

A basis $e_1, \dots, e_n$ of V is really cool because it means that to specify $v \in V$, I only have to specify $a_1, \dots, a_n \in k$, and then let $v = a_1 e_1 + \cdots + a_n e_n$. You can even think of v as $(a_1, \dots, a_n)$. To put it another way, if V is a k-vector space we always have

$$V = e_1 k \oplus e_2 k \oplus \cdots \oplus e_n k.$$

§9.5 Linear maps

Prototypical example for this section: Evaluation of $\{ax^2 + bx + c\}$ *at* $x = 3$.

We've seen homomorphisms and continuous maps. Now we're about to see linear maps, the structure preserving maps between vector spaces. Can you guess the definition?

Definition 9.5.1. Let V and W be vector spaces over the same field k. A **linear map** is a map $T \colon V \to W$ such that:

(i) We have $T(v_1 + v_2) = T(v_1) + T(v_2)$ for any $v_1, v_2 \in V$.[1]

(ii) For any $a \in k$ and $v \in V$, $T(a \cdot v) = a \cdot T(v)$.

If this map is a bijection (equivalently, if it has an inverse), it is an **isomorphism**. We then say V and W are **isomorphic** vector spaces and write $V \cong W$.

Example 9.5.2 (Examples of linear maps)

(a) For any vector spaces V and W there is a trivial linear map sending everything to $0_W \in W$.

(b) For any vector space V, there is the identity isomorphism $\mathrm{id} \colon V \to V$.

(c) The map $\mathbb{R}^3 \to \mathbb{R}$ by $(a, b, c) \mapsto 4a + 2b + c$ is a linear map.

(d) Let V be the set of real polynomials of degree at most 2. The map $\mathbb{R}^3 \to V$ by $(a, b, c) \mapsto ax^2 + bx + c$ is an *isomorphism*.

(e) Let V be the set of real polynomials of degree at most 2. The map $V \to \mathbb{R}$ by $ax^2 + bx + c \mapsto 9a + 3b + c$ is a linear map, which can be described as "evaluation at 3".

(f) Let W be the set of functions $\mathbb{R} \to \mathbb{R}$. The evaluation map $W \to \mathbb{R}$ by $f \mapsto f(0)$ is a linear map.

(g) There is a map of $\mathbb{Q}$-vector spaces $\mathbb{Q}[\sqrt{2}] \to \mathbb{Q}[\sqrt{2}]$ called "multiply by $\sqrt{2}$"; this map sends $a + b\sqrt{2} \mapsto 2b + a\sqrt{2}$. This map is an isomorphism, because it has an inverse "multiply by $1/\sqrt{2}$".

In the expression $T(a \cdot v) = a \cdot T(v)$, note that the first $\cdot$ is the multiplication of V and the second $\cdot$ is the multiplication of W. Note that this notion of isomorphism really only cares about the size of the basis:

Proposition 9.5.3 (n-dimensional vector spaces are isomorphic)
If V is an n-dimensional vector space, then $V \cong k^{\oplus n}$.

Question 9.5.4. Let $e_1, \ldots, e_n$ be a basis for V. What is the isomorphism? (Your first guess is probably right.)

Remark 9.5.5 — You could technically say that all finite-dimensional vector spaces are just $k^{\oplus n}$ and that no other space is worth caring about. But this seems kind of rude. Spaces often are more than just triples: $ax^2 + bx + c$ is a polynomial, and so it has some "essence" to it that you'd lose if you compressed it into (a, b, c). Moreover, a lot of spaces, like the set of vectors (x, y, z) with $x + y + z = 0$, do not have an obvious choice of basis. Thus to cast such a space into $k^{\oplus n}$ would require you to make arbitrary decisions.

[1] In group language, T is a homomorphism $(V, +) \to (W, +)$.

§9.6 What is a matrix?

Now I get to tell you what a matrix is! This is fun, because now I can finally explain to you how to *derive* the recipes for matrix multiplication, rather than being told.

This section is so important, and also revelatory for so many students, that I'm actually going to do it twice. The first time, I'm going to work in an extremely special case, namely $V = W = \mathbb{R}^2$, using lots of numbers. (This is how I explained this concept when I taught it to first-year undergraduate students that didn't have proof experience.) Then the second time, we'll do it in modern language without all the numbers.

§9.6.i Extended example with $\mathbb{R}^2$, suitable for the general public

Throughout this section, I'll work specifically with $\mathbb{R}^2$, whose elements I will write as $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ rather than (x, y) (you'll see why when I talk about matrix multiplication).

Pop quiz:

- **Question 1**: Suppose that you're given a linear map $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} \pi \\ 9 \end{smallmatrix}\right]$ and $T\left(\left[\begin{smallmatrix} 100 \\ 100 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 0 \\ 12 \end{smallmatrix}\right]$. What are $T\left(\left[\begin{smallmatrix} 103 \\ 104 \end{smallmatrix}\right]\right)$ and $T\left(\left[\begin{smallmatrix} 203 \\ 204 \end{smallmatrix}\right]\right)$?

  Answer 1: just add them.

$$T\left(\begin{bmatrix} 103 \\ 104 \end{bmatrix}\right) = \begin{bmatrix} \pi \\ 9 \end{bmatrix} + \begin{bmatrix} 0 \\ 12 \end{bmatrix} = \begin{bmatrix} \pi \\ 21 \end{bmatrix}$$
$$T\left(\begin{bmatrix} 203 \\ 204 \end{bmatrix}\right) = \begin{bmatrix} \pi \\ 9 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 12 \end{bmatrix} = \begin{bmatrix} \pi \\ 33 \end{bmatrix}.$$

- **Question 2**: Suppose that you're given a linear map $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$ and $T\left(\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right]$. What is $T\left(\left[\begin{smallmatrix} 50 \\ 70 \end{smallmatrix}\right]\right)$?

  Answer 2:

$$T\left(\begin{bmatrix} 50 \\ 70 \end{bmatrix}\right) = 50\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 70\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 190 \\ 430 \end{bmatrix}.$$

So what this example illustrates is that the requirements on a linear map $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ are so strong that if you just know $T\left(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]\right)$ and $T\left(\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\right)$ then you can *compute* the values of T at any other point. That's true for any two basis vectors (i.e., Question 1 could have been asked for inputs much nastier than the cherry-picked $\left[\begin{smallmatrix} 103 \\ 104 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 203 \\ 204 \end{smallmatrix}\right]$, and it would still be solvable), but of course $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ is an especially convenient choice.

Now we can give the following definition:

Definition 9.6.1. For a linear transform $T \colon \mathbb{R}^2 \to \mathbb{R}^2$, its *matrix* is an encoding of T obtained by gluing the column vectors

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \qquad \text{and} \qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$$

together to get a 2×2 array of numbers.

For example,

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \text{ and } T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \iff T \text{ encoded as } \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Now, what happens if you apply the matrix multiplication rule from high school to the column vector $\left[\begin{smallmatrix}50\\70\end{smallmatrix}\right]$? Well, you get that

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}50\\70\end{bmatrix} = \begin{bmatrix}1\cdot 50+2\cdot 70\\3\cdot 50+4\cdot 70\end{bmatrix} = \begin{bmatrix}190\\430\end{bmatrix}$$

...and you can see we're actually just doing the second pop quiz question again. So:

If $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is encoded as a 2×2 matrix M, then multiplication of M with a (column) vector $v \in \mathbb{R}^2$ is defined to coincide with $T(v)$.

Remark 9.6.2 (The identity matrix deserves its name) — This also gives a more natural reason why the 2×2 identity matrix is $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$ rather than the explanation high school gives (namely, "well, try multiplying by it and notice you get the same thing"). If id is the identity function, then $\mathrm{id}\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, so that's the first column of the matrix; similarly $\mathrm{id}\left(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ is the second column.

Now, what happens if we bring two maps S and T into the game, and compose them? We can do the same game with $S\circ T$.

- **Question 3**: Suppose that you're given a linear map $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}1\\3\end{smallmatrix}\right]$ and $T\left(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}2\\4\end{smallmatrix}\right]$. Then you're given a second linear map $S\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $S\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}5\\7\end{smallmatrix}\right]$ and $S\left(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right) = \left[\begin{smallmatrix}6\\8\end{smallmatrix}\right]$. What are $S\left(T\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]\right)\right)$ and $S\left(T\left(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right)\right)$?

  Answer 3:

$$S\left(T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\right) = S\left(\begin{bmatrix}1\\3\end{bmatrix}\right) = 1\begin{bmatrix}5\\7\end{bmatrix} + 3\begin{bmatrix}6\\8\end{bmatrix} = \begin{bmatrix}23\\31\end{bmatrix}.$$
$$S\left(T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)\right) = S\left(\begin{bmatrix}2\\4\end{bmatrix}\right) = 2\begin{bmatrix}5\\7\end{bmatrix} + 4\begin{bmatrix}6\\8\end{bmatrix} = \begin{bmatrix}34\\46\end{bmatrix}.$$

Since $S\circ T$ is itself a linear map, we now know its matrix encoding:

$$S\circ T = \begin{bmatrix}23 & 34\\31 & 46\end{bmatrix}.$$

Now, you might have learned some matrix multiplication rule in school as a definition. If you execute that definition on the matrices for S and T, you should get

$$\underbrace{\begin{bmatrix}5 & 6\\7 & 8\end{bmatrix}}_{\text{encoding of } S}\ \underbrace{\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}}_{\text{encoding of } T} = \begin{bmatrix}5\cdot 1+6\cdot 3 & 5\cdot 2+6\cdot 4\\7\cdot 1+8\cdot 3 & 7\cdot 2+8\cdot 4\end{bmatrix} = \begin{bmatrix}23 & 34\\31 & 46\end{bmatrix}$$

It's the encoding for $S\circ T$ — indeed, you can see why, because if you trace through the work in Answer 3, it's actually the same arithmetic being carried out.

This shows why our Napkin definition of matrix as the *encoding* of a linear function is better than what many of you have seen. In high school, the recipe for matrix multiplication is provided as an unnatural definition, e.g., in cute pictures like Figure 9.1. However, for us, the recipe in Figure 9.1 is a *theorem*: we can *derive* how to get the encoding of $S\circ T$ given the encodings of S and T.

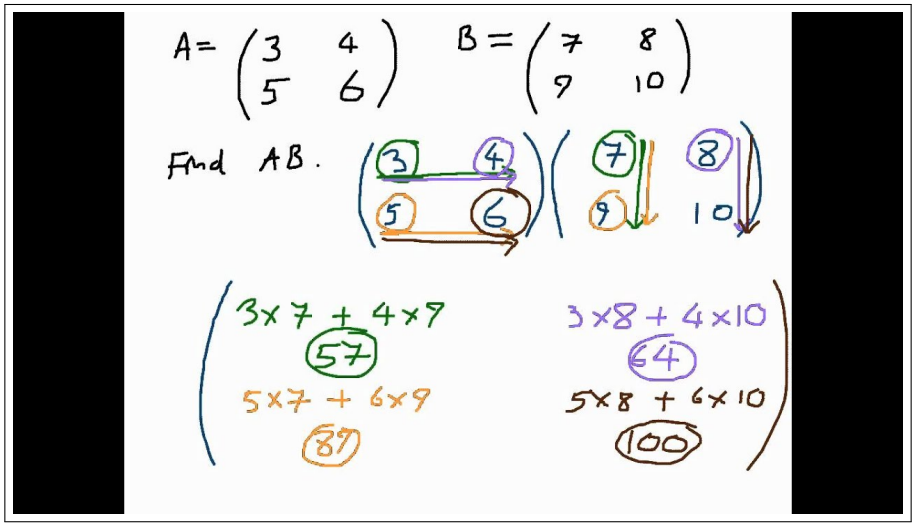


Figure 9.1: Matrix multiplication as taught in American high school: "here's a recipe, trust me bro". Image from [Ma12].

§9.6.ii General discussion, back to Napkin levels of abstraction

Let's go back to modern language, where we work with finite-dimensional spaces over any field, and any basis of the spaces (rather than a fixed basis like in the previous section).

Pick a finite-dimensional vector space V with *some* basis $e_1, \dots, e_m$ and a vector space W with basis $w_1, \dots, w_n$. Suppose I have a map $T\colon V \to W$ and I want to tell you what T is. It would be awfully inconsiderate of me to try and tell you what $T(v)$ is at every point v. But we saw I only have to tell you what $T(e_1)$, …, $T(e_m)$ are, because from there you can work out $T(a_1e_1 + \cdots + a_me_m)$ for yourself:

$$T(a_1e_1 + \cdots + a_me_m) = a_1T(e_1) + \cdots + a_mT(e_m).$$

Since the e_i are a basis, that tells you all you need to know about T.

Example 9.6.3 (Extending linear maps)
Let $V = \{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\}$. Then $T(ax^2+bx+c) = aT(x^2)+bT(x)+cT(1)$.

Now I can even be more concrete. I could tell you what $T(e_1)$ is, but seeing as I have a basis of W, I can actually just tell you what $T(e_1)$ is in terms of this basis. Specifically, there are unique $a_{11}, a_{21}, \dots, a_{n1} \in k$ such that

$$T(e_1) = a_{11}w_1 + a_{21}w_2 + \cdots + a_{n1}w_n.$$

So rather than telling you the value of $T(e_1)$ in some abstract space W, I could just tell you what $a_{11}, a_{21}, \dots, a_{n1}$ were. Then I'd repeat this for $T(e_2)$, $T(e_3)$, all the way up to $T(e_m)$, and that would tell you everything you need to know about T.

That's where the matrix T comes from! It's a concise way of writing down all mn numbers I need to tell you.

To be explicit, the matrix for T is defined as the array

$$T = \underbrace{\begin{bmatrix} | & | & & | \\ T(e_1) & T(e_2) & \dots & T(e_m) \\ | & | & & | \end{bmatrix}}_{m \text{ columns}} \Big\} n \text{ rows}$$
$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

To drive this point home,

A matrix is the laziest possible way to specify a linear map from V to W.

Example 9.6.4 (An example of a matrix)

Here is a concrete example in terms of a basis. Let $V = \mathbb{R}^3$ with basis e_1, e_2, e_3 and let $W = \mathbb{R}^2$ with basis w_1, w_2. If I have $T \colon V \to W$ then uniquely determined by three values, for example:

$$\begin{aligned} T(e_1) &= 4w_1 + 7w_2 \\ T(e_2) &= 2w_1 + 3w_2 \\ T(e_3) &= w_1 \end{aligned}$$

The columns then correspond to $T(e_1)$, $T(e_2)$, $T(e_3)$:

$$T = \begin{bmatrix} 4 & 2 & 1 \\ 7 & 3 & 0 \end{bmatrix}$$

Example 9.6.5 (An example of a matrix after choosing a basis)

We again let $V = \{ax^2 + bx + c\}$ be the vector space of polynomials of degree at most 2. We fix the basis 1, x, x^2 for it.
Consider the "evaluation at 3" map, a map $V \to \mathbb{R}$. We pick 1 as the basis element of the RHS; then we can write it as a 1×3 matrix

$$\begin{bmatrix} 1 & 3 & 9 \end{bmatrix}$$

with the columns corresponding to $T(1)$, $T(x)$, $T(x^2)$.

From here you can actually work out for yourself what it means to multiply two matrices. Suppose we have picked a basis for three spaces U, V, W. Given maps $T \colon U \to V$ and $S \colon V \to W$, we can consider their composition $S \circ T$, i.e.

$$U \xrightarrow{T} V \xrightarrow{S} W.$$

Matrix multiplication is defined exactly so that the matrix ST is the same thing we get from interpreting the composed function $S \circ T$ as a matrix, as we saw last section.

In particular, since function composition is associative, it follows that matrix multiplication is as well.

This means you can define concepts like the determinant or the trace of a matrix both in terms of an "intrinsic" map $T\colon V \to W$ and in terms of the entries of the matrix. Since the map T itself doesn't refer to any basis, the abstract definition will imply that the numerical definition doesn't depend on the choice of a basis.

§9.7 Subspaces and picking convenient bases

Prototypical example for this section: Any two linearly independent vectors in $\mathbb{R}^3$.

Definition 9.7.1. Let M be a left R-module. A **submodule** N of M is a module N such that every element of N is also an element of M. If M is a vector space then N is called a **subspace**.

Example 9.7.2 (Kernels)

The **kernel** of a map $T\colon V \to W$ (written $\ker T$) is the set of $v \in V$ such that $T(v) = 0_W$. It is a subspace of V, since it's closed under addition and scaling (why?).

Example 9.7.3 (Spans)

Let V be a vector space and $v_1, \dots, v_m$ be any vectors of V. The **span** of these vectors is defined as the set

$$\{a_1 v_1 + \cdots + a_m v_m \mid a_1, \dots, a_m \in k\}.$$

Note that it is a subspace of V as well!

Question 9.7.4. Why is 0_V an element of each of the above examples? In general, why must any subspace contain 0_V?

Subspaces behave nicely with respect to bases.

Theorem 9.7.5 (Basis completion)

Let V be an n-dimensional space, and V' a subspace of V. Then

(a) V' is also finite-dimensional.

(b) If $e_1, \dots, e_m$ is a basis of V', then there exist $e_{m+1}, \dots, e_n$ in V such that $e_1, \dots, e_n$ is a basis of V.

Proof. Omitted, since it is intuitive and the proof is not that enlightening. (However, we will use this result repeatedly later on, so do take the time to internalize it now.) □

A very common use case is picking a convenient basis for a map T.

Theorem 9.7.6 (Picking a basis for linear maps)

Let $T\colon V \to W$ be a map of finite-dimensional vector spaces, with $n = \dim V$, $m = \dim W$. Then there exists a basis $v_1, \dots, v_n$ of V and a basis $w_1, \dots, w_m$ of W, as well as a nonnegative integer k, such that

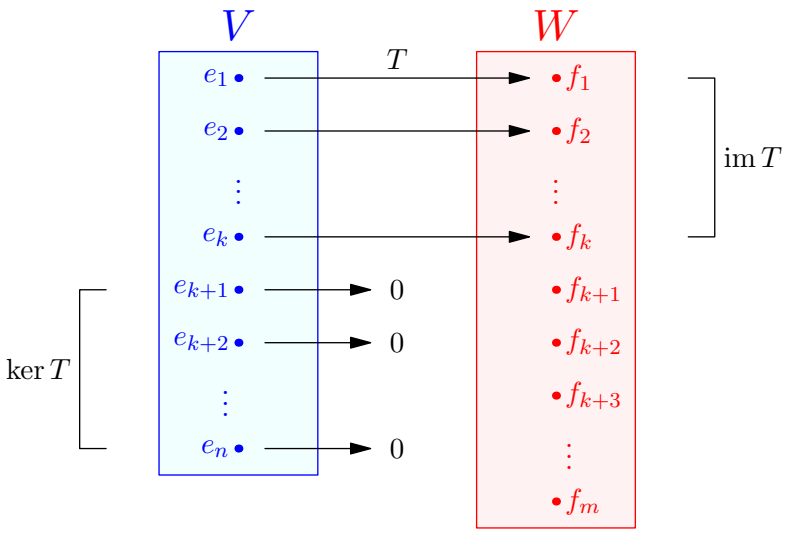
$$T(v_i) = \begin{cases} w_i & \text{if } i \le k \\ 0_W & \text{if } i > k. \end{cases}$$

Moreover $\dim \ker T = n - k$ and $\dim T^{\text{img}}(V) = k$.

Sketch of Proof. You might like to try this one yourself before reading on: it's a repeated application of Theorem 9.7.5.

Let $\ker T$ have dimension $n-k$. We can pick $v_{k+1}, \dots, v_n$ a basis of $\ker T$. Then extend it to a basis $v_1, \dots, v_n$ of V. The map T is injective over the span of $v_1, \dots, v_k$ (since only 0_V is in the kernel) so its images in W are linearly independent. Setting $w_i = T(v_i)$ for each i, we get some linearly independent set in W. Then extend it again to a basis of W. □

This theorem is super important, not only because of applications but also because it will give you the right picture in your head of how a linear map is supposed to look. I'll even draw a cartoon of it to make sure you remember:



In particular, for $T\colon V \to W$, one can write $V = \ker T \oplus V'$, so that T annihilates its kernel while sending V' to an isomorphic copy in W.

A corollary of this (which you should have expected anyways) is the so called rank-nullity theorem, which is the analog of the first isomorphism theorem.

Theorem 9.7.7 (Rank-nullity theorem)

Let V and W be finite-dimensional vector spaces. If $T\colon V \to W$, then

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Question 9.7.8. Conclude the rank-nullity theorem from Theorem 9.7.6.

§9.8 A cute application: Lagrange interpolation

Here's a cute application[2] of linear algebra to a theorem from high school.

Theorem 9.8.1 (Lagrange interpolation)

Let $x_1, \dots, x_{n+1}$ be distinct real numbers and $y_1, \dots, y_{n+1}$ any real numbers. Then there exists a *unique* polynomial P of degree at most n such that

$$P(x_i) = y_i$$

for every i.

When $n = 1$ for example, this loosely says there is a unique line joining two points.

Proof. The idea is to consider the vector space V of polynomials with degree at most n, as well as the vector space $W = \mathbb{R}^{n+1}$.

> **Question 9.8.2.** Check that $\dim V = n + 1 = \dim W$. This is easiest to do if you pick a basis for V, but you can then immediately forget about the basis once you finish this exercise.

Then consider the linear map $T \colon V \to W$ given by

$$P \mapsto (P(x_1), \dots, P(x_{n+1})) .$$

This is indeed a linear map because, well, $T(P+Q) = T(P) + T(Q)$ and $T(cP) = cT(P)$. It also happens to be injective: if $P \in \ker T$, then $P(x_1) = \cdots = P(x_{n+1}) = 0$, but $\deg P \le n$ and so P can only be the zero polynomial.

So T is an injective map between vector spaces of the same dimension. Thus it is actually a bijection, which is exactly what we wanted. □

§9.9 Pedagogical digression: Arrays of numbers are evil

(This whole section is Evan yapping about how to *teach* linear algebra, so it can be safely skipped.)

As I'll stress repeatedly, a matrix represents a *linear map between two vector spaces*. Writing it in the form of an $m \times n$ matrix is merely a very convenient way to see the map concretely. But it obfuscates the fact that this map is, well, a map, not an array of numbers.

If you took high school precalculus, you'll see everything done in terms of matrices. To any typical high school student, a matrix is an array of numbers. No one is sure what exactly these numbers represent, but they're told how to magically multiply these arrays to get more arrays. They're told that the matrix

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

[2] Source: Communicated to me by Joe Harris at the first Harvard-MIT Undergraduate Math Symposium.

is an "identity matrix", because when you multiply by another matrix it doesn't change. Then they're told that the determinant is some magical combination of these numbers formed by this weird multiplication rule. No one knows what this determinant does, other than the fact that $\det(AB) = \det A \det B$, and something about areas and row operations and Cramer's rule.

Then you go into linear algebra in college, and you do more magic with these arrays of numbers. You're told that two matrices T_1 and T_2 are similar if

$$T_2 = ST_1S^{-1}$$

for some invertible matrix S. You're told that the trace of a matrix $\operatorname{Tr} T$ is the sum of the diagonal entries. Somehow this doesn't change if you look at a similar matrix, but you're not sure why. Then you define the characteristic polynomial as

$$p_T(X) = \det(XI - T).$$

Somehow this also doesn't change if you take a similar matrix, but now you really don't know why. And then you have the Cayley-Hamilton theorem in all its black magic: $p_T(T)$ is the zero map. Out of curiosity you Google the proof, and you find some ad-hoc procedure which still leaves you with no idea why it's true.

This is terrible. What's so special about $T_2 = ST_1S^{-1}$? Only if you know that the matrices are linear maps does this make sense: T_2 is just T_1 rewritten with a different choice of basis.

I really want to push the opposite view. Linear algebra is the study of *linear maps*, but it is taught as the study of *arrays of numbers*, and no one knows what these numbers mean. And for a good reason: the numbers are meaningless. They are a highly convenient way of encoding the matrix, but they are not the main objects of study, any more than the dates of events are the main objects of study in history.

The other huge downside is that people get the impression that the only (real) vector space in existence is $\mathbb{R}^{\oplus n}$. As explained in Remark 9.5.5, while you *can* work this way if you're a soulless robot, it's very unnatural for humans to do so.

When I took Math 55a as a freshman at Harvard, I got the exact opposite treatment: we did all of linear algebra without writing down a single matrix. During all this time I was quite confused. What's wrong with a basis? I didn't appreciate until later that this approach was the morally correct way to treat the subject: it made it clear what was happening.

Throughout the Napkin, I've tried to strike a balance between these two approaches, using matrices when appropriate to illustrate the maps and to simplify proofs, but ultimately writing theorems and definitions in their *morally correct* form. I hope that this has both the advantage of giving the "right" definitions while being concrete enough to be digested. But I would like to say for the record that, if I had to pick between the high school approach and the 55a approach, I would pick 55a in a heartbeat.

§9.10 A word on general modules

Prototypical example for this section: $\mathbb{Z}[\sqrt{2}]$ *is a* $\mathbb{Z}$*-module of rank two.*

I focused mostly on vector spaces (aka modules over a field) in this chapter for simplicity, so I want to make a few remarks about modules over a general commutative ring R before concluding.

Firstly, recall that for general modules, we say "generating set" instead of "spanning set". Shrug.

The main issue with rings is that our key theorem Theorem 9.4.5 fails in spectacular ways. For example, consider $\mathbb{Z}$ as a $\mathbb{Z}$-module over itself. Then $\{2\}$ is linearly independent, but it cannot be extended to a basis. Similarly, $\{2, 3\}$ is spanning, but one cannot cut it down to a basis. You can see why defining dimension is going to be difficult.

Nonetheless, there are still analogs of some of the definitions above.

Definition 9.10.1. An R-module M is called **finitely generated** if it has a finite generating set.

Definition 9.10.2. An R-module M is called **free** if it has a basis. As said before, the analogue of the dimension theorem holds, and we use the word **rank** to denote the size of the basis. As before, there's an isomorphism $M \cong R^{\oplus n}$ where n is the rank.

Example 9.10.3 (An example of a $\mathbb{Z}$-module)
The $\mathbb{Z}$-module

$$\mathbb{Z}[\sqrt{2}] = \left\{ a + b\sqrt{2} \mid a, b \in \mathbb{Z} \right\}$$

has a basis $\{1, \sqrt{2}\}$, so we say it is a free $\mathbb{Z}$-module of rank 2.

Abuse of Notation 9.10.4 (Notation for groups)**.** Recall that an abelian group can be viewed a $\mathbb{Z}$-module (and in fact vice-versa!), so we can (and will) apply these words to abelian groups. We'll use the notation $G \oplus H$ for two abelian groups G and H for their Cartesian product, emphasizing the fact that G and H are abelian. This will happen when we study algebraic number theory and homology groups.

§9.11 A few harder problems to think about

General hint: Theorem 9.7.6 will be your best friend for many of these problems.

Problem 9A†. Let V and W be finite-dimensional vector spaces with nonzero dimension, and consider linear maps $T \colon V \to W$. Complete the following table by writing "sometimes", "always", or "never" for each entry.

	T injective	T surjective	T isomorphism
If $\dim V > \dim W \ldots$			
If $\dim V = \dim W \ldots$			
If $\dim V < \dim W \ldots$			

Problem 9B† (Equal dimension vector spaces are usually isomorphisms)**.** Let V and W be finite-dimensional vector spaces with $\dim V = \dim W$. Prove that for a map $T \colon V \to W$, the following are equivalent:

- T is injective,
- T is surjective,
- T is bijective.

Problem 9C. Let's say a *magic square* is a 3×3 matrix of real numbers where the sum of all diagonals, columns, and rows is equal, such as $\left[\begin{smallmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{smallmatrix}\right]$. Find the dimension of the set of magic squares, as a real vector space under addition.

Problem 9D (Multiplication by $\sqrt{5}$)**.** Let $V = \mathbb{Q}[\sqrt{5}] = \left\{a + b\sqrt{5}\right\}$ be a two-dimensional $\mathbb{Q}$-vector space, and fix the basis $\{1, \sqrt{5}\}$ for it. Write down the 2×2 matrix with rational coefficients that corresponds to multiplication by $\sqrt{5}$.

Problem 9E (Multivariable Lagrange interpolation)**.** Let $S \subset \mathbb{Z}^2$ be a set of n lattice points. Prove that there exists a nonzero two-variable polynomial p with real coefficients, of degree at most $\sqrt{2n}$, such that $p(x, y) = 0$ for every $(x, y) \in S$.

Problem 9F (Putnam 2003)**.** Do there exist polynomials $a(x)$, $b(x)$, $c(y)$, $d(y)$ such that

$$1 + xy + (xy)^2 = a(x)c(y) + b(x)d(y)$$

holds identically?

Problem 9G (TSTST 2014)**.** Let $P(x)$ and $Q(x)$ be arbitrary polynomials with real coefficients, and let d be the degree of $P(x)$. Assume that $P(x)$ is not the zero polynomial. Prove that there exist polynomials $A(x)$ and $B(x)$ such that

(i) Both A and B have degree at most $d/2$,

(ii) At most one of A and B is the zero polynomial,

(iii) P divides $A + Q \cdot B$.

Problem 9H* (Idempotents are projection maps)**.** Let $P \colon V \to V$ be a linear map, where V is a vector space (not necessarily finite-dimensional). Suppose P is **idempotent**, meaning $P(P(v)) = P(v)$ for each $v \in V$, or equivalently P is the identity on its image. Prove that

$$V = \ker P \oplus \operatorname{im} P.$$

Thus we can think of P as *projection* onto the subspace $\operatorname{im} P$.

Problem 9I*. Let V be a finite dimensional vector space. Let $T \colon V \to V$ be a linear map, and let $T^n \colon V \to V$ denote T applied n times. Prove that there exists an integer N such that

$$V = \ker T^N \oplus \operatorname{im} T^N.$$

10 Eigen-things

This chapter will develop the theory of eigenvalues and eigenvectors, the so-called "Jordan canonical form". (Later on we will use it to define the characteristic polynomial.)

§10.1 Why you should care

We know that a square matrix T is really just a linear map from V to V. What's the simplest type of linear map? It would just be multiplication by some scalar λ, which would have associated matrix (in any basis!)

$$T = \begin{bmatrix} \lambda & 0 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda \end{bmatrix}.$$

That's perhaps *too* simple, though. If we had a fixed basis $e_1, \dots, e_n$ then another very "simple" operation would just be scaling each basis element e_i by λ_i, i.e. a **diagonal matrix** of the form

$$T = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

These maps are more general. Indeed, you can, for example, compute T^{100} in a heartbeat: the map sends $e_i \to \lambda_i^{100} e_i$. (Try doing that with an arbitrary $n \times n$ matrix.)

Of course, most linear maps are probably not that nice. Or are they?

Example 10.1.1 (Getting lucky)

Let V be some two-dimensional vector space with e_1 and e_2 as basis elements. Let's consider a map $T \colon V \to V$ by $e_1 \mapsto 2e_1$ and $e_2 \mapsto e_1 + 3e_2$, which you can even write concretely as

$$T = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \quad \text{in basis } e_1, e_2.$$

This doesn't look anywhere as nice until we realize we can rewrite it as

$$\begin{aligned} e_1 &\mapsto 2e_1 \\ e_1 + e_2 &\mapsto 3(e_1 + e_2). \end{aligned}$$

So suppose we change to the basis e_1 and $e_1 + e_2$. Thus in the new basis,

$$T = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{in basis } e_1, e_1 + e_2.$$

So our completely random-looking map, under a suitable change of basis, looks like the very nice maps we described before!

In this chapter, we will be *making* our luck, and we will see that our better understanding of matrices gives us the right way to think about this.

§10.2 Warning on assumptions

Most theorems in this chapter only work for

- finite-dimensional vector spaces V,
- over a field k which is *algebraically closed*.

On the other hand, the definitions work fine without these assumptions.

§10.3 Eigenvectors and eigenvalues

Let k be a field and V a vector space over it. In the above example, we saw that there were two very nice vectors, e_1 and $e_1 + e_2$, for which V did something very simple. Naturally, these vectors have a name.

Definition 10.3.1. Let $T \colon V \to V$ and $v \in V$ a *nonzero* vector. We say that v is an **eigenvector** if $T(v) = \lambda v$ for some $\lambda \in k$ (possibly zero, but remember $v \neq 0$). The value λ is called an **eigenvalue** of T.

We will sometimes abbreviate "v is an eigenvector with eigenvalue λ" to just "v is a λ-eigenvector".

Of course, no mention to a basis anywhere.

Example 10.3.2 (An example of an eigenvector and eigenvalue)

Consider the example earlier with $T = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

(a) Note that e_1 and $e_1 + e_2$ are 2-eigenvectors and 3-eigenvectors.

(b) Of course, $5e_1$ is also an 2-eigenvector.

(c) And, $7e_1 + 7e_2$ is also a 3-eigenvector.

So you can quickly see the following observation.

Question 10.3.3. Show that the λ-eigenvectors, together with $\{0\}$ form a subspace.

Definition 10.3.4. For any λ, we define the λ-**eigenspace** as the set of λ-eigenvectors together with 0.

This lets us state succinctly that "2 is an eigenvalue of T with one-dimensional eigenspace spanned by e_1".

Unfortunately, it's not exactly true that eigenvalues always exist.

Example 10.3.5 (Eigenvalues need not exist)

Let $V = \mathbb{R}^2$ and let T be the map which rotates a vector by 90° around the origin.

Then $T(v)$ is not a multiple of v for any $v \in V$, other than the trivial $v = 0$.

However, it is true if we replace k with an algebraically closed field.[1]

Theorem 10.3.6 (Eigenvalues always exist over algebraically closed fields)
Suppose k is an *algebraically closed* field. Let V be a finite dimensional k-vector space. Then if $T\colon V \to V$ is a linear map, there exists an eigenvalue $\lambda \in k$.

Proof. (From [Ax97]) The idea behind this proof is to consider "polynomials" in T. For example, $2T^2 - 4T + 5$ would be shorthand for $2T(T(v)) - 4T(v) + 5v$. In this way we can consider "polynomials" $P(T)$; this lets us tie in the "algebraically closed" condition. These polynomials behave nicely:

Question 10.3.7. Show that $P(T) + Q(T) = (P+Q)(T)$ and $P(T) \circ Q(T) = (P \cdot Q)(T)$.

Let $n = \dim V < \infty$ and fix any nonzero vector $v \in V$, and consider vectors v, $T(v)$, $\dots$, $T^n(v)$. There are $n+1$ of them, so they can't be linearly independent for dimension reasons; thus there is a nonzero polynomial P such that $P(T)$ is zero when applied to v. WLOG suppose P is a monic polynomial, and thus $P(z) = (z - r_1)\dots(z - r_m)$ say. Then we get

$$0 = (T - r_1\mathrm{id}) \circ (T - r_2\mathrm{id}) \circ \cdots \circ (T - r_m\mathrm{id})(v)$$

(where id is the identity matrix). This means at least one of $T - r_i\mathrm{id}$ is not injective, i.e. has a nontrivial kernel, which is the same as an eigenvector. □

So in general we like to consider algebraically closed fields. This is not a big loss: any real matrix can be interpreted as a complex matrix whose entries just happen to be real, for example.

§10.4 The Jordan form

So that you know exactly where I'm going, here's the main theorem.

Definition 10.4.1. A **Jordan block** is an $n \times n$ matrix of the following shape:

$$\begin{bmatrix} \lambda & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}.$$

In other words, it has λ on the diagonal, and 1 above it. We allow $n = 1$, so $\begin{bmatrix}\lambda\end{bmatrix}$ is a Jordan block.

[1] A field is **algebraically closed** if all its polynomials have roots, the archetypal example being $\mathbb{C}$.

Theorem 10.4.2 (Jordan canonical form)

Let $T\colon V \to V$ be a linear map of finite-dimensional vector spaces over an algebraically closed field k. Then we can choose a basis of V such that the matrix T is "block-diagonal" with each block being a Jordan block.

Such a matrix is said to be in **Jordan form**. This form is unique up to rearranging the order of the blocks.

As an example, this means the matrix should look something like:

$$\begin{bmatrix}
\lambda_1 & 1 & & & & & & & \\
0 & \lambda_1 & & & & & & & \\
& & \lambda_2 & & & & & & \\
& & & \lambda_3 & 1 & 0 & & & \\
& & & 0 & \lambda_3 & 1 & & & \\
& & & 0 & 0 & \lambda_3 & & & \\
& & & & & & \ddots & & \\
& & & & & & & \lambda_m & 1 \\
& & & & & & & 0 & \lambda_m
\end{bmatrix}$$

Question 10.4.3. Check that diagonal matrices are the special case when each block is 1×1.

What does this mean? Basically, it means *our dream is almost true*. What happens is that V can get broken down as a direct sum

$$V = J_1 \oplus J_2 \oplus \cdots \oplus J_m$$

and T acts on each of these subspaces independently. These subspaces correspond to the blocks in the matrix above. In the simplest case, $\dim J_i = 1$, so J_i has a basis element e for which $T(e) = \lambda_i e$; in other words, we just have a simple eigenvalue. But on occasion, the situation is not quite so simple, and we have a block of size greater than 1; this leads to 1's just above the diagonals.

I'll explain later how to interpret the 1's, when I make up the word *descending staircase*. For now, you should note that even if $\dim J_i \geq 2$, we still have a basis element which is an eigenvector with eigenvalue λ_i.

Example 10.4.4 (A concrete example of Jordan form)

Let $T\colon k^6 \to k^6$ and suppose T is given by the matrix

$$T = \begin{bmatrix}
5 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 7 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 3
\end{bmatrix}.$$

Reading the matrix, we can compute all the eigenvectors and eigenvalues: for any

constants $a, b \in k$ we have

$$\begin{aligned} T(a \cdot e_1) &= 5a \cdot e_1 \\ T(a \cdot e_2) &= 2a \cdot e_2 \\ T(a \cdot e_4) &= 7a \cdot e_4 \\ T(a \cdot e_5 + b \cdot e_6) &= 3\left[a \cdot e_5 + b \cdot e_6\right]. \end{aligned}$$

The element e_3 on the other hand, is not an eigenvector since $T(e_3) = e_2 + 2e_3$.

§10.5 Nilpotent maps

Bear with me for a moment. First, define:

Definition 10.5.1. A map $T \colon V \to V$ is **nilpotent** if T^m is the zero map for some integer m. (Here T^m means "T applied m times".)

What's an example of a nilpotent map?

Example 10.5.2 (The "descending staircase")
Let $V = k^{\oplus 3}$ have basis e_1, e_2, e_3. Then the map T which sends

$$e_3 \mapsto e_2 \mapsto e_1 \mapsto 0$$

is nilpotent, since $T(e_1) = T^2(e_2) = T^3(e_3) = 0$, and hence $T^3(v) = 0$ for all $v \in V$.

The 3×3 descending staircase has matrix representation

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

You'll notice this is a Jordan block.

Exercise 10.5.3. Show that the descending staircase above has 0 as its only eigenvalue.

That's a pretty nice example. As another example, we can have multiple such staircases.

Example 10.5.4 (Double staircase)
Let $V = k^{\oplus 5}$ have basis e_1, e_2, e_3, e_4, e_5. Then the map

$$e_3 \mapsto e_2 \mapsto e_1 \mapsto 0 \text{ and } e_5 \mapsto e_4 \mapsto 0$$

is nilpotent.

Picture, with some zeros omitted for emphasis:

$$T = \begin{bmatrix} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ & & & 0 & 1 \\ & & & 0 & 0 \end{bmatrix}$$

You can see this isn't really that different from the previous example; it's just the same idea repeated multiple times. And in fact we now claim that *all* nilpotent maps have essentially that form.

Theorem 10.5.5 (Nilpotent Jordan)

Let V be a finite-dimensional vector space over an algebraically closed field k. Let $T\colon V \to V$ be a nilpotent map. Then we can write $V = \bigoplus_{i=1}^{m} V_i$ where each V_i has a basis of the form v_i, $T(v_i)$, $\dots$, $T^{\dim V_i - 1}(v_i)$ for some $v_i \in V_i$, and such that $T^{\dim V_i}(v_i) = 0$.

Hence:

Every nilpotent map can be viewed as independent staircases.

Each chain v_i, $T(v_i)$, $T(T(v_i))$, $\dots$ is just one staircase. The proof is given later, but first let me point out where this is going.

Here's the punch line. Let's take the double staircase again. Expressing it as a matrix gives, say

$$S = \begin{bmatrix} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ & & & 0 & 1 \\ & & & 0 & 0 \end{bmatrix}.$$

Then we can compute

$$S + \lambda\text{id} = \begin{bmatrix} \lambda & 1 & 0 & & \\ 0 & \lambda & 1 & & \\ 0 & 0 & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{bmatrix}.$$

It's a bunch of λ Jordan blocks! This gives us a plan to proceed: we need to break V into a bunch of subspaces such that $T - \lambda\text{id}$ is nilpotent over each subspace. Then Nilpotent Jordan will finish the job.

§10.6 Reducing to the nilpotent case

Definition 10.6.1. Let $T\colon V \to V$. A subspace $W \subseteq V$ is called T-**invariant** if $T(w) \in W$ for any $w \in W$. In this way, T can be thought of as a map $W \to W$.

In this way, the Jordan form is a decomposition of V into invariant subspaces.

Now I'm going to be cheap, and define:

Definition 10.6.2. A map $T\colon V \to V$ is called **indecomposable** if it's impossible to write $V = W_1 \oplus W_2$ where both W_1 and W_2 are nontrivial T-invariant spaces.

Picture of a *decomposable* map:

$$\left[\begin{array}{cc|ccc} \multicolumn{2}{c|}{\multirow{2}{*}{W_1}} & 0 & 0 & 0 \\ & & 0 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{3}{c}{\multirow{3}{*}{W_2}} \\ 0 & 0 & \multicolumn{3}{c}{} \\ 0 & 0 & \multicolumn{3}{c}{} \end{array}\right]$$

As you might expect, we can break a space apart into "indecomposable" parts.

Proposition 10.6.3 (Invariant subspace decomposition)
Let V be a finite-dimensional vector space. Given any map $T\colon V \to V$, we can write
$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_m$$
where each V_i is T-invariant, and for any i the map $T\colon V_i \to V_i$ is indecomposable.

Proof. Same as the proof that every integer is the product of primes. If V is not decomposable, we are done. Otherwise, by definition write $V = W_1 \oplus W_2$ and then repeat on each of W_1 and W_2. □

Incredibly, with just that we're almost done! Consider a decomposition as above, so that $T\colon V_1 \to V_1$ is an indecomposable map. Then T has an eigenvalue λ_1, so let $S = T - \lambda_1 \mathrm{id}$; hence $\ker S \neq \{0\}$.

Question 10.6.4. Show that V_1 is also S-invariant, so we can consider $S\colon V_1 \to V_1$.

By Problem 9I*, we have
$$V_1 = \ker S^N \oplus \operatorname{im} S^N$$
for some N. But we assumed T was indecomposable, so this can only happen if $\operatorname{im} S^N = \{0\}$ and $\ker S^N = V_1$ (since $\ker S^N$ contains our eigenvector). Hence S is nilpotent, so it's a collection of staircases. In fact, since T is indecomposable, there is only one staircase. Hence V_1 is a Jordan block, as desired.

§10.7 (Optional) Proof of nilpotent Jordan

The proof is just induction on $\dim V$. Assume $\dim V \geq 1$, and let $W = T^{\text{img}}(V)$ be the image of V. Since T is nilpotent, we must have $W \subsetneq V$. Moreover, if $W = \{0\}$ (i.e. T is the zero map) then we're already done. So assume $\{0\} \subsetneq W \subsetneq V$.

By the inductive hypothesis, we can select a good basis of W:
$$\begin{aligned}\mathcal{B}' = \Big\{ & T(v_1), T(T(v_1)), T(T(T(v_1))), \dots \\ & T(v_2), T(T(v_2)), T(T(T(v_2))), \dots \\ & \dots, \\ & T(v_\ell), T(T(v_\ell)), T(T(T(v_\ell))), \dots \Big\}\end{aligned}$$
for some $T(v_i) \in W$ (here we have taken advantage of the fact that each element of W is itself of the form $T(v)$ for some v).

Also, note that there are exactly ℓ elements of $\mathcal{B}'$ which are in $\ker T$ (namely the last element of each of the ℓ staircases). We can thus complete it to a basis $v_{\ell+1}, \dots, v_m$ (where $m = \dim \ker T$). (In other words, the last element of each staircase plus the $m - \ell$ new ones are a basis for $\ker T$.)

Now consider

$$\begin{aligned}\mathcal{B} = \Big\{ & v_1, T(v_1), T(T(v_1)), T(T(T(v_1))), \dots \\ & v_2, T(v_2), T(T(v_2)), T(T(T(v_2))), \dots \\ & \dots, \\ & v_\ell, T(v_\ell), T(T(v_\ell)), T(T(T(v_\ell))), \dots \\ & v_{\ell+1}, v_{\ell+2}, \dots, v_m \Big\}.\end{aligned}$$

Question 10.7.1. Check that there are exactly $\ell + \dim W + (\dim \ker T - \ell) = \dim V$ elements.

Exercise 10.7.2. Show that all the elements are linearly independent. (Assume for contradiction there is some linear dependence, then take T of both sides.)

Hence $\mathcal{B}$ is a basis of the desired form.

§10.8 Algebraic and geometric multiplicity

Prototypical example for this section: The matrix T below.

This is some convenient notation: let's consider the matrix in Jordan form

$$T = \begin{bmatrix} 7 & 1 & & & & & \\ 0 & 7 & & & & & \\ & & 9 & & & & \\ & & & 7 & 1 & 0 \\ & & & 0 & 7 & 1 \\ & & & 0 & 0 & 7 \end{bmatrix}.$$

We focus on the eigenvalue 7, which appears multiple times, so it is certainly "repeated". However, there are two different senses in which you could say it is repeated.

- *Algebraic*: You could say it is repeated five times, because it appears five times on the diagonal.
- *Geometric*: You could say it really only appears two times: because there are only two eigen*vectors* with eigenvalue 7, namely e_1 and e_4.

  Indeed, the vector e_2 for example has $T(e_2) = 7e_2 + e_1$, so it's not really an eigenvector! If you apply $T - 7\text{id}$ to e_2 twice though, you do get zero.

Question 10.8.1. In this example, how many times do you need to apply $T - 7\text{id}$ to e_6 to get zero?

Both these notions are valid, so we will name both. To preserve generality, we first state the "intrinsic" definition.

Definition 10.8.2. Let $T \colon V \to V$ be a linear map and λ a scalar.

- The **geometric multiplicity** of λ is the dimension $\dim V_\lambda$ of the λ-eigenspace.

- Define the **generalized eigenspace** V^λ to be the subspace of V for which $(T - \lambda\text{id})^n(v) = 0$ for some $n \geq 1$. The **algebraic multiplicity** of λ is the dimension $\dim V^\lambda$.

(Silly edge case: we allow "multiplicity zero" if λ is not an eigenvalue at all.)

However in practice you should just count the Jordan blocks.

Example 10.8.3 (An example of eigenspaces via Jordan form)
Retain the matrix T mentioned earlier and let $\lambda = 7$.

- The eigenspace V_λ has basis e_1 and e_4, so the geometric multiplicity is 2.
- The generalized eigenspace V^λ has basis e_1, e_2, e_4, e_5, e_6 so the algebraic multiplicity is 5.

To be completely explicit, here is how you think of these in practice:

Proposition 10.8.4 (Geometric and algebraic multiplicity vs Jordan blocks)
Assume $T \colon V \to V$ is a linear map of finite-dimensional vector spaces, written in Jordan form. Let λ be a scalar. Then

- The geometric multiplicity of λ is the number of Jordan blocks with eigenvalue λ; the eigenspace has one basis element per Jordan block.
- The algebraic multiplicity of λ is the sum of the dimensions of the Jordan blocks with eigenvalue λ; the eigenspace is the direct sum of the subspaces corresponding to those blocks.

Proof. Definition 10.8.2 was essentially chosen to be a basis-free rephrasing of this proposition. □

Question 10.8.5. Show that the geometric multiplicity is always less than or equal to the algebraic multiplicity.

This actually gives us a tentative definition:

- The trace is the sum of the eigenvalues, counted with algebraic multiplicity.
- The determinant is the product of the eigenvalues, counted with algebraic multiplicity.

This definition is okay, but it has the disadvantage of requiring the ground field to be algebraically closed. It is also not the definition that is easiest to work with computationally. The next two chapters will give us a better definition.

§10.9 A few harder problems to think about

Problem 10A (Sum of algebraic multiplicities)**.** Given a 2018-dimensional complex vector space V and a map $T \colon V \to V$, what is the sum of the algebraic multiplicities of all eigenvalues of T?

Problem 10B (The word "diagonalizable")**.** A linear map $T\colon V \to V$ (where $\dim V$ is finite) is said to be **diagonalizable** if it has a basis $e_1, \dots, e_n$ such that each e_i is an eigenvector.

(a) Explain the name "diagonalizable".

(b) Suppose we are working over an algebraically closed field. Then show that T is diagonalizable if and only if for any λ, the geometric multiplicity of λ equals the algebraic multiplicity of λ.

Problem 10C (Switcharoo)**.** Let V be the $\mathbb{C}$-vector space with basis e_1 and e_2. The map $T\colon V \to V$ sends $T(e_1) = e_2$ and $T(e_2) = e_1$. Determine the eigenspaces of T.

Problem 10D. Suppose $T\colon \mathbb{C}^{\oplus 2} \to \mathbb{C}^{\oplus 2}$ is a linear map of $\mathbb{C}$-vector spaces such that $T^{2011} = \mathrm{id}$. Must T be diagonalizable?

Problem 10E (Writing a polynomial backwards)**.** Define the complex vector space V of polynomials with degree at most 2, say $V = \{ax^2 + bx + c \mid a, b, c \in \mathbb{C}\}$. Define $T\colon V \to V$ by

$$T(ax^2 + bx + c) = cx^2 + bx + a.$$

Determine the eigenspaces of T.

Problem 10F (Differentiation of polynomials)**.** Let $V = \mathbb{R}[x]$ be the infinite-dimensional real vector space of all polynomials with real coefficients. Note that $\frac{d}{dx}\colon V \to V$ is a linear map (for example it sends x^3 to $3x^2$). Which real numbers are eigenvalues of this map?

Problem 10G (Differentiation of functions)**.** Let V be the infinite-dimensional real vector space of all infinitely differentiable functions $\mathbb{R} \to \mathbb{R}$. Note that $\frac{d}{dx}\colon V \to V$ is a linear map (for example it sends $\cos x$ to $-\sin x$). Which real numbers are eigenvalues of this map?

11 Dual space and trace

You may have learned in high school that given a matrix

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

the trace is the sum along the diagonals $a+d$ and the determinant is $ad-bc$. But we know that a matrix is somehow just encoding a linear map using a choice of basis. Why would these random formulas somehow not depend on the choice of a basis?

In this chapter, we are going to give an intrinsic definition of $\operatorname{Tr} T$, where $T\colon V \to V$ and $\dim V < \infty$. This will give a coordinate-free definition which will in particular imply the trace $a+d$ doesn't change if we take a different basis.

In doing so, we will introduce two new constructions: the *tensor product* $V \otimes W$ (which is a sort of product of two spaces, with dimension $\dim V \cdot \dim W$) and the *dual space* $V^\vee$, which is the set of linear maps $V \to k$ (a k-vector space). Later on, when we upgrade from a vector space V to an inner product space, we will see that the dual space $V^\vee$ gives a nice interpretation of the "transpose" of a matrix. You'll already see some of that come through here.

The trace is only defined for finite-dimensional vector spaces, so if you want you can restrict your attention to finite-dimensional vector spaces for this chapter. (On the other hand we do not need the ground field to be algebraically closed.)

The next chapter will then do the same for the determinant.

§11.1 Tensor product

Prototypical example for this section: $\mathbb{R}[x] \otimes \mathbb{R}[y] = \mathbb{R}[x,y]$.

We know that $\dim(V \oplus W) = \dim V + \dim W$, even though as sets $V \oplus W$ looks like $V \times W$. What if we wanted a real "product" of spaces, with multiplication of dimensions?

For example, let's pull out my favorite example of a real vector space, namely

$$V = \left\{ ax^2 + bx + c \mid a,b,c \in \mathbb{R} \right\}.$$

Here's another space, a little smaller:

$$W = \left\{ dy + e \mid d,e \in \mathbb{R} \right\}.$$

If we take the direct sum, then we would get some rather unnatural vector space of dimension five (whose elements can be thought of as pairs $(ax^2+bx+c, dy+e)$). But suppose we want a vector space whose elements are *products* of polynomials in V and W; it would contain elements like $4x^2y + 5xy + y + 3$. In particular, the basis would be

$$\left\{ x^2y, x^2, xy, x, y, 1 \right\}$$

and thus have dimension six.

For this we resort to the *tensor product*. It does exactly this, except that the "multiplication" is done by a scary[1] symbol $\otimes$: think of it as a "wall" that separates the elements between the two vector spaces. For example, the above example might be written as

$$4x^2 \otimes y + 5x \otimes y + 1 \otimes y + 3 \otimes 1.$$

[1]Seriously, $\otimes$ looks *terrifying* to non-mathematicians, and even to many math undergraduates.

(This should be read as $(4x^2 \otimes y) + (5x \otimes y) + \dots$; addition comes after $\otimes$.) Of course there should be no distinction between writing $4x^2 \otimes y$ and $x^2 \otimes 4y$ or even $2x^2 \otimes 2y$. While we want to keep the x and y separate, the scalars should be free to float around.

Of course, there's no need to do everything in terms of just the monomials. We are free to write

$$(x+1) \otimes (y+1).$$

If you like, you can expand this as

$$x \otimes y + 1 \otimes y + x \otimes 1 + 1 \otimes 1.$$

Same thing. The point is that we can take any two of our polynomials and artificially "tensor" them together.

The definition of the tensor product does exactly this, and nothing else.[2]

Definition 11.1.1. Let V and W be vector spaces over the same field k. The **tensor product** $V \otimes_k W$ is the abelian group generated by elements of the form $v \otimes w$, subject to relations

$$\begin{aligned}(v_1 + v_2) \otimes w &= v_1 \otimes w + v_2 \otimes w \\ v \otimes (w_1 + w_2) &= v \otimes w_1 + v \otimes w_2 \\ (c \cdot v) \otimes w &= v \otimes (c \cdot w).\end{aligned}$$

As a vector space, its action is given by $c \cdot (v \otimes w) = (c \cdot v) \otimes w = v \otimes (c \cdot w)$.

Here's another way to phrase the same idea. We define a **pure tensor** as an element of the form $v \otimes w$ for $v \in V$ and $w \in W$. But we let the $\otimes$ wall be "permeable" in the sense that

$$(c \cdot v) \otimes w = v \otimes (c \cdot w) = c \cdot (v \otimes w)$$

and we let multiplication and addition distribute as we expect. Then $V \otimes W$ consists of sums of pure tensors.

Example 11.1.2 (Infinite-dimensional example of tensor product: two-variable polynomials)

Although it's not relevant to this chapter, this definition works equally well with infinite-dimensional vector spaces. The best example might be

$$\mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{R}[y] = \mathbb{R}[x,y].$$

That is, the tensor product of polynomials in x with real polynomials in y turns out to just be two-variable polynomials $\mathbb{R}[x,y]$.

Remark 11.1.3 (Warning on sums of pure tensors) — Remember the elements of $V \otimes_k W$ really are *sums* of these pure tensors! If you liked the previous example, this fact has a nice interpretation — not every polynomial in $\mathbb{R}[x,y] = \mathbb{R}[x] \otimes_{\mathbb{R}} \mathbb{R}[y]$ factors as a polynomial in x times a polynomial in y (i.e. as pure tensors $f(x) \otimes g(y)$). But they all can be written as sums of pure tensors $x^a \otimes y^b$.

[2] I'll only define this for vector spaces for simplicity. The definition for modules over a commutative ring R is exactly the same.

As the example we gave suggested, the basis of $V \otimes_k W$ is literally the "product" of the bases of V and W. In particular, this fulfills our desire that $\dim(V \otimes_k W) = \dim V \cdot \dim W$.

Proposition 11.1.4 (Basis of $V \otimes W$)
Let V and W be finite-dimensional k-vector spaces. If $e_1, \dots, e_m$ is a basis of V and $f_1, \dots, f_n$ is a basis of W, then the basis of $V \otimes_k W$ is precisely $e_i \otimes f_j$, where $i = 1, \dots, m$ and $j = 1, \dots, n$.

Proof. Omitted; it's easy at least to see that this basis is spanning. □

Example 11.1.5 (Explicit computation)
Let V have basis e_1, e_2 and W have basis f_1, f_2. Let $v = 3e_1 + 4e_2 \in V$ and $w = 5f_1 + 6f_2 \in W$. Let's write $v \otimes w$ in this basis for $V \otimes_k W$:

$$\begin{aligned} v \otimes w &= (3e_1 + 4e_2) \otimes (5f_1 + 6f_2) \\ &= (3e_1) \otimes (5f_1) + (4e_2) \otimes (5f_1) + (3e_1) \otimes (6f_2) + (4e_2) \otimes (6f_2) \\ &= 15(e_1 \otimes f_1) + 20(e_2 \otimes f_1) + 18(e_1 \otimes f_2) + 24(e_2 \otimes f_2). \end{aligned}$$

So you can see why tensor products are a nice "product" to consider if we're really interested in $V \times W$ in a way that's more intimate than just a direct sum.

Abuse of Notation 11.1.6. Moving forward, we'll almost always abbreviate $\otimes_k$ to just $\otimes$, since k is usually clear.

Remark 11.1.7 — Observe that to define a linear map $V \otimes W \to X$, I only have to say what happens to each pure tensor $v \otimes w$, since the pure tensors *generate* $V \otimes W$. But again, keep in mind that $V \otimes W$ consists of *sums* of these pure tensors! In other words, $V \otimes W$ is generated by pure tensors.

Remark 11.1.8 — Much like the Cartesian product $A \times B$ of sets, you can tensor together any two vector spaces V and W over the same field k; the relationship between V and W is completely irrelevant. One can think of the $\otimes$ as a "wall" through which one can pass scalars in k, but otherwise keeps the elements of V and W separated. Thus, $\otimes$ is **content-agnostic**.
This also means that even if V and W have some relation to each other, the tensor product doesn't remember this. So for example $v \otimes 1 \neq 1 \otimes v$, just like $(g, 1_G) \neq (1_G, g)$ in the group $G \times G$.

§11.2 Dual space

Prototypical example for this section: Rotate a column matrix by 90 degrees.

Consider the following vector space:

Example 11.2.1 (Functions from $\mathbb{R}^3 \to \mathbb{R}$)
The set of real functions $f(x, y, z)$ is an infinite-dimensional real vector space.

Indeed, we can add two functions to get $f+g$, and we can think of functions like $2f$.

This is a terrifyingly large vector space, but you can do some reasonable reductions. For example, you can restrict your attention to just the *linear maps* from $\mathbb{R}^3$ to $\mathbb{R}$.

That's exactly what we're about to do. This definition might seem strange at first, but bear with me.

Definition 11.2.2. Let V be a k-vector space. Then $V^\vee$, the **dual space** of V, is defined as the vector space whose elements are *linear maps from V to k*.

The addition and multiplication are pointwise: it's the same notation we use when we write $cf+g$ to mean $c \cdot f(x) + g(x)$. The dual space itself is less easy to think about.

Let's try to find a basis for $V^\vee$. First, here is a very concrete interpretation of the vector space. Suppose for example $V = \mathbb{R}^3$. We can think of elements of V as column matrices, like

$$v = \begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix} \in V.$$

Then a linear map $f \colon V \to k$ can be interpreted as a *row matrix*:

$$f = \begin{bmatrix} 3 & 4 & 5 \end{bmatrix} \in V^\vee.$$

Then

$$f(v) = \begin{bmatrix} 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix} = 71.$$

More precisely: **to specify a linear map $V \to k$, I only have to tell you where each basis element of V goes**. In the above example, f sends e_1 to 3, e_2 to 4, and e_3 to 5. So f sends

$$2e_1 + 5e_2 + 9e_3 \mapsto 2 \cdot 3 + 5 \cdot 4 + 9 \cdot 5 = 71.$$

Let's make all this precise.

Proposition 11.2.3 (The dual basis for $V^\vee$)

Let V be a finite-dimensional vector space with basis $e_1, \dots, e_n$. For each i consider the function $e_i^\vee \colon V \to k$ defined by

$$e_i^\vee(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

In more humane terms, $e_i^\vee(v)$ gives the coefficient of e_i in v.
Then $e_1^\vee, e_2^\vee, \dots, e_n^\vee$ is a basis of $V^\vee$.

Example 11.2.4 (Explicit example of element in $V^\vee$)

In this notation, $f = 3e_1^\vee + 4e_2^\vee + 5e_3^\vee$. Do you see why the "sum" notation works

as expected here? Indeed

$$\begin{aligned} f(e_1) &= (3e_1^\vee + 4e_2^\vee + 5e_3^\vee)(e_1) \\ &= 3e_1^\vee(e_1) + 4e_2^\vee(e_1) + 5e_3^\vee(e_1) \\ &= 3 \cdot 1 + 4 \cdot 0 + 5 \cdot 0 = 3. \end{aligned}$$

That's exactly what we wanted.

You might be inclined to point out that $V \cong V^\vee$ at this point, with an isomorphism given by $e_i \mapsto e_i^\vee$. You might call it "rotating the column matrix by 90°".

This statement is technically true, but for a generic vector space V without any extra information, you can just think of this as an artifact of the $\dim V = \dim V^\vee$ (as *any* two vector spaces of equal dimension are isomorphic). Most importantly, the isomorphism given above depends on what basis you picked.

Remark 11.2.5 (Explicit example showing that the isomorphism $V \to V^\vee$ given above is unnatural) Alice or Bob are looking at the same two-dimensional real vector space

$$V = \{(x, y, z) \mid x + y + z = 0\}.$$

Also, let $v_{\text{example}} = (3, 5, -8)$ be an example of an arbitrary element of V for concreteness.
Suppose Alice chooses the following basis vectors for V.

$$\begin{aligned} e_1 &= (1, 0, -1) \\ e_2 &= (0, 1, -1). \end{aligned}$$

Alice uses this to construct an isomorphism $A \colon V \to V^\vee$ as described above, and considers $e_1^\vee = A(e_1)$. The element $e_1^\vee \in V^\vee$ is a function $e_1^\vee \colon V \to \mathbb{R}$, meaning Alice can plug any vector in V into it. As an example, for v_{example}

$$e_1^\vee(v_{\text{example}}) = e_1^\vee((3, 5, -8)) = e_1^\vee(3e_1 + 5e_2) = 3.$$

Meanwhile, Bob chooses the different basis vectors

$$\begin{aligned} f_1 &= (1, 0, -1) \\ f_2 &= (1, -1, 0). \end{aligned}$$

This gives Bob an isomorphism $B \colon V \to V^\vee$, and a corresponding $f_1^\vee = B(f_1)$. Bob can also evaluate it anywhere, e.g.

$$f_1^\vee(v_{\text{example}}) = f_1^\vee((3, 5, -8)) = f_1^\vee(8f_1 - 5f_2) = 8.$$

It follows that $e_1^\vee = A((1, 0, -1))$ and $f_1^\vee = B((1, 0, -1))$ are different elements of $V^\vee$. In other words Alice and Bob got different isomorphisms because they picked different bases.

§11.3 $V^\vee \otimes W$ gives matrices from V to W

Goal of this section:

If V and W are finite-dimensional k-vector spaces then $V^\vee \otimes W$ represents linear maps $V \to W$.

Here's the intuition. If V is three-dimensional and W is five-dimensional, then we can think of the maps $V \to W$ as a 5×3 array of numbers. We want to think of these maps as a vector space: (since one can add or scale matrices). So it had better be a vector space with dimension 15, but just saying "$k^{\oplus 15}$" is not really that satisfying (what is the basis?).

To do better, we consider the tensor product

$$V^\vee \otimes W$$

which somehow is a product of maps out of V and the target space W. We claim that this is in fact the space we want: i.e. **there is a natural bijection between elements of $V^\vee \otimes W$ and linear maps from V to W.**

First, how do we interpret an element of $V^\vee \otimes W$ as a map $V \to W$? For concreteness, suppose V has a basis e_1, e_2, e_3, and W has a basis f_1, f_2, f_3, f_4, f_5. Consider an element of $V^\vee \otimes W$, say

$$e_1^\vee \otimes (f_2 + 2f_4) + 4e_2^\vee \otimes f_5.$$

We want to interpret this element as a function $V \to W$: so given a $v \in V$, we want to output an element of W. There's really only one way to do this: feed in $v \in V$ into the $V^\vee$ guys on the left. That is, take the map

$$v \mapsto e_1^\vee(v) \cdot (f_2 + 2f_4) + 4e_2^\vee(v) \cdot f_5 \in W.$$

So, there's a natural way to interpret any element $\xi_1 \otimes w_1 + \cdots + \xi_m \otimes w_m \in V^\vee \otimes W$ as a linear map $V \to W$. The claim is that in fact, every linear map $V \to W$ has such an interpretation.

First, for notational convenience,

Definition 11.3.1. Let $\operatorname{Hom}(V, W)$ denote the set of linear maps from V to W (which one can interpret as matrices which send V to W), viewed as a vector space over k. (The "Hom" stands for homomorphism.)

Question 11.3.2. Identify $\operatorname{Hom}(V, k)$ by name.

We can now write down something that's more true generally.

Theorem 11.3.3 ($V^\vee \otimes W \iff$ linear maps $V \to W$)
Let V and W be finite-dimensional vector spaces. We described a map

$$\Psi \colon V^\vee \otimes W \to \operatorname{Hom}(V, W)$$

by sending $\xi_1 \otimes w_1 + \cdots + \xi_m \otimes w_m$ to the linear map

$$v \mapsto \xi_1(v)w_1 + \cdots + \xi_m(v)w_m.$$

Then Ψ is an isomorphism of vector spaces, i.e. every linear map $V \to W$ can be uniquely represented as an element of $V^\vee \otimes W$ in this way.

The above is perhaps a bit dense, so here is a concrete example.

Example 11.3.4 (Explicit example)
Let $V = \mathbb{R}^2$ and take a basis e_1, e_2 of V. Then define $T\colon V \to V$ by

$$T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Then we have

$$\Psi(e_1^\vee \otimes e_1 + 2e_2^\vee \otimes e_1 + 3e_1^\vee \otimes e_2 + 4e_2^\vee \otimes e_2) = T.$$

The beauty is that the Ψ definition is basis-free; thus even if we change the basis, although the above expression will look completely different, the *actual element* in $V^\vee \otimes V$ doesn't change.

Despite this, we'll indulge ourselves in using coordinates for the proof.

Proof of Theorem 11.3.3. This looks intimidating, but it's actually not difficult. We proceed in two steps:

1. First, we check that Ψ is *surjective*; every linear map has at least one representation in $V^\vee \otimes W$. To see this, take any $T\colon V \to W$. Suppose V has basis e_1, e_2, e_3 and that $T(e_1) = w_1$, $T(e_2) = w_2$ and $T(e_3) = w_3$. Then the element

$$e_1^\vee \otimes w_1 + e_2^\vee \otimes w_2 + e_3^\vee \otimes w_3$$

works, as it is contrived to agree with T on the basis elements e_i.

2. So it suffices to check now that $\dim V^\vee \otimes W = \dim \operatorname{Hom}(V, W)$. Certainly, $V^\vee \otimes W$ has dimension $\dim V \cdot \dim W$. But by viewing $\operatorname{Hom}(V, W)$ as $\dim V \cdot \dim W$ matrices, we see that it too has dimension $\dim V \cdot \dim W$. □

So there is a **natural isomorphism** $V^\vee \otimes W \cong \operatorname{Hom}(V, W)$. While we did use a basis liberally in the *proof that it works*, this doesn't change the fact that the isomorphism is "God-given", depending only on the spirit of V and W itself and not which basis we choose to express the vector spaces in.

§11.4 The trace

We are now ready to give the definition of a trace. Recall that a square matrix T can be thought of as a map $T\colon V \to V$. According to the above theorem,

$$\operatorname{Hom}(V, V) \cong V^\vee \otimes V$$

so every map $V \to V$ can be thought of as an element of $V^\vee \otimes V$. But we can also define an *evaluation map* $\operatorname{ev}\colon V^\vee \otimes V \to k$ by "collapsing" each pure tensor: $f \otimes v \mapsto f(v)$. So this gives us a composed map

$$\operatorname{Hom}(V, V) \xrightarrow{\cong} V^\vee \otimes V \xrightarrow{\operatorname{ev}} k.$$

This result is called the **trace** of a matrix T.

Example 11.4.1 (Example of a trace)
Continuing the previous example,

$$\operatorname{Tr} T = e_1^\vee(e_1) + 2e_2^\vee(e_1) + 3e_1^\vee(e_2) + 4e_2^\vee(e_2) = 1 + 0 + 0 + 4 = 5.$$

And that is why the trace is the sum of the diagonal entries.

§11.5 A few harder problems to think about

Problem 11A (Trace is sum of eigenvalues)**.** Let V be an n-dimensional vector space over an algebraically closed field k. Let $T \colon V \to V$ be a linear map with eigenvalues λ_1, λ_2, $\dots$, λ_n (counted with algebraic multiplicity). Show that $\operatorname{Tr} T = \lambda_1 + \cdots + \lambda_n$.

Problem 11B† (Product of traces)**.** Let $T \colon V \to V$ and $S \colon W \to W$ be linear maps of finite-dimensional vector spaces V and W. Define $T \otimes S \colon V \otimes W \to V \otimes W$ by $v \otimes w \mapsto T(v) \otimes S(w)$. Prove that

$$\operatorname{Tr}(T \otimes S) = \operatorname{Tr}(T) \operatorname{Tr}(S).$$

Problem 11C† (Traces kind of commute)**.** Let $T \colon V \to W$ and $S \colon W \to V$ be linear maps between finite-dimensional vector spaces V and W. Show that

$$\operatorname{Tr}(T \circ S) = \operatorname{Tr}(S \circ T).$$

Problem 11D (Putnam 1988)**.** Let V be an n-dimensional vector space. Let $T \colon V \to V$ be a linear map and suppose there exists $n+1$ eigenvectors, any n of which are linearly independent. Does it follow that T is a scalar multiple of the identity?

12 Determinant

The goal of this chapter is to give the basis-free definition of the determinant: that is, we're going to define $\det T$ for $T\colon V \to V$ without making reference to the encoding for T. This will make it obvious that the determinant of a matrix does not depend on the choice of basis, and that several properties are vacuously true (e.g. that the determinant is multiplicative).

The determinant is only defined for finite-dimensional vector spaces, so if you want you can restrict your attention to finite-dimensional vector spaces for this chapter. On the other hand we do not need the ground field to be algebraically closed.

§12.1 Wedge product

Prototypical example for this section: $\bigwedge^2(\mathbb{R}^2)$ *gives parallelograms.*

We're now going to define something called the wedge product. It will look at first like the tensor product $V \otimes V$, but we'll have one extra relation.

For simplicity, I'll first define the wedge product $\bigwedge^2(V)$. But we will later replace 2 with any n.

Definition 12.1.1. Let V be a k-vector space. The 2-wedge product $\bigwedge^2(V)$ is the abelian group generated by elements of the form $v \wedge w$ (where $v, w \in V$), subject to the same relations

$$\begin{aligned}(v_1 + v_2) \wedge w &= v_1 \wedge w + v_2 \wedge w \\ v \wedge (w_1 + w_2) &= v \wedge w_1 + v \wedge w_2 \\ (c \cdot v) \wedge w &= v \wedge (c \cdot w)\end{aligned}$$

plus two additional relations:

$$v \wedge v = 0 \quad \text{and} \quad v \wedge w = -w \wedge v.$$

As a vector space, its action is given by $c \cdot (v \wedge w) = (c \cdot v) \wedge w = v \wedge (c \cdot w)$.

> **Exercise 12.1.2.** Show that the condition $v \wedge w = -(w \wedge v)$ is actually extraneous: you can derive it from the fact that $v \wedge v = 0$. (Hint: expand $(v + w) \wedge (v + w) = 0$.)

This looks almost exactly the same as the definition for a tensor product, with two subtle differences. The first is that we only have V now, rather than V and W as with the tensor product.[1] Secondly, there is a new *mysterious* relation

$$v \wedge v = 0 \implies v \wedge w = -(w \wedge v).$$

What's that doing there? It seems kind of weird.

I'll give you a hint.

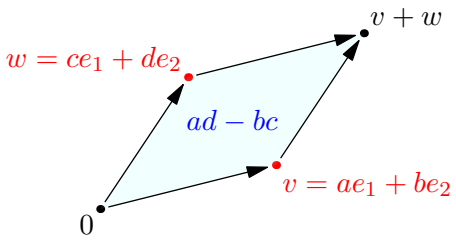
[1] So maybe the wedge product might be more accurately called the "wedge power"!

Example 12.1.3 (Wedge product explicit computation)
Let $V = \mathbb{R}^2$, and let $v = ae_1 + be_2$, $w = ce_1 + de_2$. Now let's compute $v \wedge w$ in $\bigwedge^2(V)$.

$$\begin{aligned} v \wedge w &= (ae_1 + be_2) \wedge (ce_1 + de_2) \\ &= ac(e_1 \wedge e_1) + bd(e_2 \wedge e_2) + ad(e_1 \wedge e_2) + bc(e_2 \wedge e_1) \\ &= ad(e_1 \wedge e_2) + bc(e_2 \wedge e_1) \\ &= (ad - bc)(e_1 \wedge e_2). \end{aligned}$$

What is $ad - bc$? You might already recognize it:

- You might know that the area of the parallelogram formed by v and w is $ad - bc$.
- You might recognize it as the determinant of $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$. In fact, you might even know that the determinant is meant to interpret hypervolumes.



This is absolutely no coincidence. The wedge product is designed to interpret signed areas. That is, $v \wedge w$ is meant to interpret the area of the parallelogram formed by v and w. You can see why the condition $(cv) \wedge w = v \wedge (cw)$ would make sense now. And now of course you know why $v \wedge v$ ought to be zero: it's an area zero parallelogram!

The **miracle of wedge products** is that the only additional condition we need to add to the tensor product axioms is that $v \wedge v = 0$. Then suddenly, the wedge will do all our work of interpreting volumes for us.

Remark 12.1.4 (Side digression on definitions in mathematics) This "property-based" philosophy is a common trope in modern mathematics. You have some intuition about an object you wish to define, and then you write down a wishlist of properties that "should" follow. But then it turns out the properties are sufficient to work with, and so for the definition, you just define an abstract object satisfying all the properties on your wishlist. Thereafter the intuition plays no "official" role; it serves only as cheerleading motivation for the wishlist.
For wedge products, the wishlist has only the single property $v \wedge v = 0$.

In analog to earlier:

Proposition 12.1.5 (Basis of $\bigwedge^2(V)$)
Let V be a vector space with basis $e_1, \dots, e_n$. Then a basis of $\bigwedge^2(V)$ is

$$e_i \wedge e_j$$

where $i < j$. Hence $\bigwedge^2(V)$ has dimension $\binom{n}{2}$.

Proof. Surprisingly slippery, and also omitted. (You can derive it from the corresponding theorem on tensor products.) □

Now I have the courage to define a multi-dimensional wedge product. It's just the same thing with more wedges.

Definition 12.1.6. Let V be a vector space and m a positive integer. The space $\bigwedge^m(V)$ is generated by wedges of the form

$$v_1 \wedge v_2 \wedge \cdots \wedge v_m$$

subject to relations

$$\begin{aligned}
\cdots \wedge (v_1 + v_2) \wedge \ldots &= (\cdots \wedge v_1 \wedge \ldots) + (\cdots \wedge v_2 \wedge \ldots) \\
\cdots \wedge (cv_1) \wedge v_2 \wedge \ldots &= \cdots \wedge v_1 \wedge (cv_2) \wedge \ldots \\
\cdots \wedge v \wedge v \wedge \ldots &= 0 \\
\cdots \wedge v \wedge w \wedge \ldots &= -(\cdots \wedge w \wedge v \wedge \ldots)
\end{aligned}$$

As a vector space

$$c \cdot (v_1 \wedge v_2 \wedge \cdots \wedge v_m) = (cv_1) \wedge v_2 \wedge \cdots \wedge v_m = v_1 \wedge (cv_2) \wedge \cdots \wedge v_m = \dots.$$

This definition is pretty wordy, but in English the three conditions say

- We should be able to add products like before,
- You can put constants onto any of the m components (as is directly pointed out in the "vector space" action), and
- Switching any two *adjacent* wedges negates the whole wedge.

So this is the natural generalization of $\bigwedge^2(V)$. You can convince yourself that any element of the form

$$\cdots \wedge v \wedge \cdots \wedge v \wedge \ldots$$

should still be zero.

Just like $e_1 \wedge e_2$ was a basis earlier, we can find the basis for general m and n.

Proposition 12.1.7 (Basis of the wedge product)
Let V be a vector space with basis $e_1, \dots, e_n$. A basis for $\bigwedge^m(V)$ consists of the elements

$$e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_m}$$

where

$$1 \le i_1 < i_2 < \cdots < i_m \le n.$$

Hence $\bigwedge^m(V)$ has dimension $\binom{n}{m}$.

Sketch of proof. We knew earlier that $e_{i_1} \otimes \cdots \otimes e_{i_m}$ was a basis for the tensor product. Here we have the additional property that (a) if two basis elements re-appear then the whole thing becomes zero, thus we should assume the i's are all distinct; and (b) we can shuffle around elements, and so we arbitrarily decide to put the basis elements in increasing order. □

§12.2 The determinant

Prototypical example for this section: $(ae_1 + be_2) \wedge (ce_1 + de_2) = (ad - bc)(e_1 \wedge e_2)$.

Now we're ready to define the determinant. Suppose $T \colon V \to V$ is a square matrix. We claim that the map $\bigwedge^m(V) \to \bigwedge^m(V)$ given on wedges by

$$v_1 \wedge v_2 \wedge \cdots \wedge v_m \mapsto T(v_1) \wedge T(v_2) \wedge \cdots \wedge T(v_m).$$

and extending linearly to all of $\bigwedge^m(V)$ is a well-defined linear map (Here "well-defined" means that equivalent elements of the domain get mapped to equivalent elements of the codomain. This, and linearity, both follow from T being a linear map.) We call that map $\bigwedge^m(T)$.

Example 12.2.1 (Example of $\bigwedge^m(T)$)
In $V = \mathbb{R}^4$ with standard basis e_1, e_2, e_3, e_4, let $T(e_1) = e_2$, $T(e_2) = 2e_3$, $T(e_3) = e_3$ and $T(e_4) = 2e_2 + e_3$. Then, for example, $\bigwedge^2(T)$ sends

$$\begin{aligned}(e_1 \wedge e_2) + (e_3 \wedge e_4) &\mapsto T(e_1) \wedge T(e_2) + T(e_3) \wedge T(e_4) \\ &= e_2 \wedge 2e_3 + e_3 \wedge (2e_2 + e_3) \\ &= 2(e_2 \wedge e_3 + e_3 \wedge e_2) \\ &= 0.\end{aligned}$$

Now here's something interesting. Suppose V has dimension n, and let $m = n$. Then $\bigwedge^n(V)$ has dimension $\binom{n}{n} = 1$ — it's a one dimensional space! Hence $\bigwedge^n(V) \cong k$.

So $\bigwedge^n(T)$ can be thought of as a linear map from k to k. But we know that *a linear map from k to k is just multiplication by a constant.* Hence $\bigwedge^n(T)$ is multiplication by some constant.

Definition 12.2.2. Let $T \colon V \to V$, where V is an n-dimensional vector space. Then $\bigwedge^n(T)$ is multiplication by a constant c; we define the **determinant** of T as $c = \det T$.

Example 12.2.3 (The determinant of a 2×2 matrix)
Let $V = \mathbb{R}^2$ again with basis e_1 and e_2. Let

$$T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

In other words, $T(e_1) = ae_1 + be_2$ and $T(e_2) = ce_1 + de_2$.
Now let's consider $\bigwedge^2(V)$. It has a basis $e_1 \wedge e_2$. Now $\bigwedge^2(T)$ sends it to

$$e_1 \wedge e_2 \xmapsto{\bigwedge^2(T)} T(e_1) \wedge T(e_2) = (ae_1 + be_2) \wedge (ce_1 + de_2) = (ad - bc)(e_1 \wedge e_2).$$

So $\bigwedge^2(T) \colon \bigwedge^2(V) \to \bigwedge^2(V)$ is multiplication by $\det T = ad - bc$, because it sent

$e_1 \wedge e_2$ to $(ad-bc)(e_1 \wedge e_2)$.

And that is the definition of a determinant. Once again, since we defined it in terms of $\bigwedge^n(T)$, this definition is totally independent of the choice of basis. In other words, the determinant can be defined based on $T\colon V \to V$ alone without any reference to matrices.

Question 12.2.4. Why does $\bigwedge^n(S \circ T) = \bigwedge^n(S) \circ \bigwedge^n(T)$?

In this way, we also get

$$\det(S \circ T) = \det(S)\det(T)$$

for free.

More generally if we replace 2 by n, an write out the result of expanding

$$(a_{11}e_1 + a_{21}e_2 + \cdots) \wedge \cdots \wedge (a_{1n}e_1 + a_{2n}e_2 + \cdots + a_{nn}e_n)$$

then you will get the formula

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \dots a_{n,\sigma(n)}$$

called the **Leibniz formula** for determinants. American high school students will recognize it; this is (unfortunately) taught as the definition of the determinant, rather than a corollary of the better definition using wedge products.

Exercise 12.2.5. Verify that expanding the wedge product yields the Leibniz formula for $n = 3$.

§12.3 Characteristic polynomials, and Cayley-Hamilton

Let's connect with the theory of eigenvalues. Take a map $T\colon V \to V$, where V is n-dimensional over an algebraically closed field, and suppose its eigenvalues are λ_1, λ_2, $\dots$, λ_n (with repetition). Then the **characteristic polynomial** is given by

$$p_T(X) = (X-\lambda_1)(X-\lambda_2)\dots(X-\lambda_n).$$

Note that if we've written T in Jordan form, that is,

$$T = \begin{bmatrix} \lambda_1 & * & 0 & \dots & 0 \\ 0 & \lambda_2 & * & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

(here each $*$ is either 0 or 1), then we can hack together the definition

$$p_T(X) \coloneqq \det\left(X \cdot \mathrm{id}_n - T\right) = \det \begin{bmatrix} X-\lambda_1 & * & 0 & \dots & 0 \\ 0 & X-\lambda_2 & * & \dots & 0 \\ 0 & 0 & X-\lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X-\lambda_n \end{bmatrix}.$$

The latter definition is what you'll see in most linear algebra books because it lets you define the characteristic polynomial without mentioning the word "eigenvalue" (i.e. entirely in terms of arrays of numbers). I'll admit it does have the merit that it means that given any matrix, it's easy to compute the characteristic polynomial and hence compute the eigenvalues; but I still think the definition should be done in terms of eigenvalues to begin with. For instance the determinant definition obscures the following theorem, which is actually a complete triviality.

Theorem 12.3.1 (Cayley-Hamilton)

Let $T\colon V \to V$ be a map of finite-dimensional vector spaces over an algebraically closed field. Then for any $T\colon V \to V$, the map $p_T(T)$ is the zero map.

Here, by $p_T(T)$ we mean that if

$$p_T(X) = X^n + c_{n-1}X^{n-1} + \cdots + c_0$$

then

$$p_T(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1 T + c_0 I$$

is the zero map, where T^k denotes T applied k times. We saw this concept already when we proved that T had at least one nonzero eigenvector.

Example 12.3.2 (Example of Cayley-Hamilton using determinant definition)

Suppose $T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Using the determinant definition of characteristic polynomial, we find that $p_T(X) = (X-1)(X-4) - (-2)(-3) = X^2 - 5X - 2$. Indeed, you can verify that

$$T^2 - 5T - 2 = \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - 5 \cdot \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

If you define p_T without the word eigenvalue, and adopt the evil view that matrices are arrays of numbers, then this looks like a complete miracle. (Indeed, just look at the terrible proofs on Wikipedia.)

But if you use the abstract viewpoint of T as a linear map, then the theorem is almost obvious:

Proof of Cayley-Hamilton. Suppose we write V in Jordan normal form as

$$V = J_1 \oplus \cdots \oplus J_m$$

where J_i has eigenvalue λ_i and dimension d_i. By definition,

$$p_T(T) = (T - \lambda_1)^{d_1}(T - \lambda_2)^{d_2} \dots (T - \lambda_m)^{d_m}.$$

By definition, $(T - \lambda_1)^{d_1}$ is the zero map on J_1. So $p_T(T)$ is zero on J_1. Similarly it's zero on each of the other J_i's — end of story. □

Remark 12.3.3 (Tensoring up) — The Cayley-Hamilton theorem holds without the hypothesis that k is algebraically closed: because for example any real matrix can be regarded as a matrix with complex coefficients (a trick we've mentioned before). I'll briefly hint at how you can use tensor products to formalize this idea.
Let's take the space $V = \mathbb{R}^3$, with basis e_1, e_2, e_3. Thus objects in V are of the form $r_1e_1 + r_2e_2 + r_3e_3$ where r_1, r_2, r_3 are real numbers. We want to consider essentially the same vector space, but with complex coefficients z_i rather than real coefficients r_i.
So here's what we do: view $\mathbb{C}$ as a $\mathbb{R}$-vector space (with basis $\{1, i\}$, say) and consider the **complexification**

$$V_{\mathbb{C}} \coloneqq \mathbb{C} \otimes_{\mathbb{R}} V.$$

Then you can check that our elements are actually of the form

$$z_1 \otimes e_1 + z_2 \otimes e_2 + z_3 \otimes e_3.$$

Here, the tensor product is over $\mathbb{R}$, so we have $z \otimes re_i = (zr) \otimes e_i$ for $r \in \mathbb{R}$. Then $V_{\mathbb{C}}$ can be thought as a three-dimensional vector space over $\mathbb{C}$, with basis $1 \otimes e_i$ for $i \in \{1, 2, 3\}$. In this way, the tensor product lets us formalize the idea that we "fuse on" complex coefficients.
If $T \colon V \to W$ is a map, then $T_{\mathbb{C}} \colon V_{\mathbb{C}} \to W_{\mathbb{C}}$ is just the map $z \otimes v \mapsto z \otimes T(v)$. You'll see this written sometimes as $T_{\mathbb{C}} = \mathrm{id} \otimes T$. One can then apply theorems to $T_{\mathbb{C}}$ and try to deduce the corresponding results on T.

§12.4 A few harder problems to think about

Problem 12A (Column operations)**.** Show that for any real numbers x_{ij} (here $1 \le i, j \le n$) we have

$$\det \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} = \det \begin{bmatrix} x_{11} + cx_{12} & x_{12} & \dots & x_{1n} \\ x_{21} + cx_{22} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} + cx_{n2} & x_{n2} & \dots & x_{nn} \end{bmatrix}.$$

Problem 12B (Determinant is product of eigenvalues)**.** Let V be an n-dimensional vector space over an algebraically closed field k. Let $T \colon V \to V$ be a linear map with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ (counted with algebraic multiplicity). Show that $\det T = \lambda_1 \dots \lambda_n$.

Problem 12C (Exponential matrix)**.** Let X be an $n \times n$ matrix with complex coefficients. We define the exponential map by

$$\exp(X) = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \cdots$$

(take it for granted that this converges to some $n \times n$ matrix). Prove that

$$\det(\exp(X)) = e^{\operatorname{Tr} X}.$$

Problem 12D (Extension to Problem 9B†)**.** Let $T \colon V \to V$ be a map of finite-dimensional vector spaces. Prove that T is an isomorphism if and only if $\det T \neq 0$.

Problem 12E (Based on Sweden 2010)**.** A herd of 1000 cows of nonzero weight is given. Prove that we can remove one cow such that the remaining 999 cows cannot be partitioned into two sets with equal sum of weights.

Problem 12F (Putnam 2015)**.** Define S to be the set of real matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ such that a, b, c, d form an arithmetic progression in that order. Find all $M \in S$ such that for some integer $k > 1$, $M^k \in S$.

Problem 12G. Let V be a finite-dimensional vector space over k and $T \colon V \to V$. Show that

$$\det(a \cdot \mathrm{id}_V - T) = \sum_{n=0}^{\dim V} a^{\dim V - n} \cdot (-1)^n \operatorname{Tr}\left(\bigwedge^n (T)\right)$$

where the trace is taken by viewing $\bigwedge^n(T) \colon \bigwedge^n(V) \to \bigwedge^n(V)$.

Problem 12H (Cauchy-Binet formula)**.** Let $n \geq s \geq 1$ be integers, and let A and B be an $s \times n$ matrix and $n \times s$ matrix, respectively. (Hence AB is an $s \times s$ matrix.) For any subset $S \subseteq \{1, 2, \dots, n\}$ with $|S| = s$, we let A_S be the $s \times s$ submatrix of A of the rows with indices in S and let B_S be the $s \times s$ submatrix of B of the columns with indices in S. Prove that

$$\det(AB) = \sum_{|S|=s} \det A_S \det B_S.$$

13 Inner product spaces

It will often turn out that our vector spaces which look more like $\mathbb{R}^n$ not only have the notion of addition, but also a notion of *orthogonality* and the notion of *distance*. All this is achieved by endowing the vector space with a so-called **inner form**, which you likely already know as the "dot product" for $\mathbb{R}^n$. Indeed, in $\mathbb{R}^n$ you already know that

- $v \cdot w = 0$ if and only if v and w are perpendicular, and
- $|v|^2 = v \cdot v$.

The purpose is to quickly set up this structure in full generality. Some highlights of the chapter:

- We'll see that the high school "dot product" formulation is actually very natural: it falls out from the two axioms we listed above. If you ever wondered why $\sum a_i b_i$ behaves as nicely as it does, now you'll know.
- We show how the inner form can be used to make V into a *metric space*, giving it more geometric structure.
- A few chapters later, we'll identify $V \cong V^\vee$ in a way that wasn't possible before, and as a corollary deduce the nice result that symmetric matrices with real entries always have real eigenvalues.

Throughout this chapter, *all vector spaces are over* $\mathbb{C}$ *or* $\mathbb{R}$, unless otherwise specified. We'll generally prefer working over $\mathbb{C}$ instead of $\mathbb{R}$ since $\mathbb{C}$ is algebraically closed (so, e.g. we have Jordan forms). Every real matrix can be thought of as a matrix with complex entries anyways.

§13.1 The inner product

Prototypical example for this section: Dot product in $\mathbb{R}^n$.

§13.1.i For real numbers: bilinear forms

First, let's define the inner form for real spaces. Rather than the notation $v \cdot w$ it is most customary to use $\langle v, w \rangle$ for general vector spaces.

Definition 13.1.1. Let V be a real vector space. A **real inner form**[1] is a function

$$\langle \bullet, \bullet \rangle : V \times V \to \mathbb{R}$$

which satisfies the following properties:

- The form is **symmetric**: for any $v, w \in V$ we have

$$\langle v, w \rangle = \langle w, v \rangle .$$

  Of course, one would expect this property from a product.

[1] Other names include "inner product", "dot product", "positive definite nondegenerate symmetric bilinear form", ...

- The form is **bilinear**, or **linear in both arguments**, meaning that $\langle -, v\rangle$ and $\langle v, -\rangle$ are linear functions for any fixed v. Spelled explicitly this means that
$$\begin{aligned} \langle cx, v\rangle &= c\langle x, v\rangle \\ \langle x+y, v\rangle &= \langle x, v\rangle + \langle y, v\rangle . \end{aligned}$$
and similarly if v was on the left. This is often summarized by the single equation $\langle cx+y, z\rangle = c\langle x, z\rangle + \langle y, z\rangle$.
- The form is **positive definite**, meaning $\langle v, v\rangle \geq 0$ is a nonnegative real number, and equality takes place only if $v = 0_V$.

Exercise 13.1.2. Show that linearity in the first argument plus symmetry already gives you linearity in the second argument, so we could edit the above definition by only requiring $\langle -, v\rangle$ to be linear.

Example 13.1.3 ($\mathbb{R}^n$)
As we already know, one can define the inner form on $\mathbb{R}^n$ as follows. Let $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $\ldots$, $e_n = (0, \ldots, 0, 1)$ be the usual basis. Then we let
$$\langle a_1e_1 + \cdots + a_ne_n, b_1e_1 + \cdots + b_ne_n\rangle \coloneqq a_1b_1 + \cdots + a_nb_n.$$
It's easy to see this is bilinear (symmetric and linear in both arguments). To see it is positive definite, note that if $a_i = b_i$ then the dot product is $a_1^2 + \cdots + a_n^2$, which is zero exactly when all a_i are zero.

§13.1.ii For complex numbers: sesquilinear forms

The definition for a complex product space is similar, but has one difference: rather than symmetry we instead have *conjugate symmetry* meaning $\langle v, w\rangle = \overline{\langle w, v\rangle}$. Thus, while we still have linearity in the first argument, we actually have a different linearity for the second argument. To be explicit:

Definition 13.1.4. Let V be a complex vector space. A **complex inner product** is a function
$$\langle \bullet, \bullet\rangle : V \times V \to \mathbb{C}$$
which satisfies the following properties:

- The form has **conjugate symmetry**, which means that for any $v, w \in V$ we have
$$\langle v, w\rangle = \overline{\langle w, v\rangle}.$$
- The form is **sesquilinear** (the name means "one-and-a-half linear"). This means that:
  - The form is **linear in the first argument**, so again we have
$$\begin{aligned} \langle x+y, v\rangle &= \langle x, v\rangle + \langle y, v\rangle \\ \langle cx, v\rangle &= c\langle x, v\rangle . \end{aligned}$$
Again this is often abbreviated to the single line $\langle cx+y, v\rangle = c\langle x, v\rangle + \langle y, v\rangle$ in the literature.

– However, it is now **anti-linear in the second argument**: for any complex number c and vectors x and y we have

$$\langle v, x+y\rangle = \langle v, x\rangle + \langle v, y\rangle$$
$$\langle v, cx\rangle = \overline{c}\,\langle v, x\rangle\,.$$

Note the appearance of the complex conjugate $\overline{c}$, which is new! Again, we can abbreviate this to just $\langle v, cx+y\rangle = \overline{c}\,\langle v, x\rangle + \langle v, y\rangle$ if we only want to write one equation.

- The form is **positive definite**, meaning $\langle v, v\rangle$ is a nonnegative real number, and equals zero exactly when $v = 0_V$.

Exercise 13.1.5. Show that anti-linearity follows from conjugate symmetry plus linearity in the first argument.

Example 13.1.6 ($\mathbb{C}^n$)
The dot product in $\mathbb{C}^n$ is defined as follows: let $\mathbf{e}_1$, $\mathbf{e}_2$, …, $\mathbf{e}_n$ be the standard basis. For complex numbers w_i, z_i we set

$$\langle w_1\mathbf{e}_1 + \cdots + w_n\mathbf{e}_n, z_1\mathbf{e}_1 + \cdots + z_n\mathbf{e}_n\rangle \coloneqq w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

Question 13.1.7. Check that the above is in fact a complex inner form.

§13.1.iii Inner product space

It'll be useful to treat both types of spaces simultaneously:

Definition 13.1.8. An **inner product space** is either a real vector space equipped with a real inner form, or a complex vector space equipped with a complex inner form.

A linear map between inner product spaces is a map between the underlying vector spaces (we do *not* require any compatibility with the inner form).

Remark 13.1.9 (Why sesquilinear?) — The above example explains one reason why we want to satisfy conjugate symmetry rather than just symmetry. If we had tried to define the dot product as $\sum w_i z_i$, then we would have lost the condition of being positive definite, because there is no guarantee that $\langle v, v\rangle = \sum z_i^2$ will even be a real number at all. On the other hand, with conjugate symmetry we actually enforce $\langle v, v\rangle = \overline{\langle v, v\rangle}$, i.e. $\langle v, v\rangle \in \mathbb{R}$ for every v.
Let's make this point a bit more forcefully. Suppose we tried to put a bilinear form $\langle -, -\rangle$, on a *complex* vector space V. Let e be any vector with $\langle e, e\rangle = 1$ (a unit vector). Then we would instead get $\langle ie, ie\rangle = -\langle e, e\rangle = -1$; this is a vector with length $\sqrt{-1}$, which is not okay! That's why it is important that, when we have a complex inner product space, our form is sesquilinear, not bilinear.

Now that we have a dot product, we can talk both about the norm and orthogonality.

§13.2 Norms

Prototypical example for this section: $\mathbb{R}^n$ *becomes its usual Euclidean space with the vector norm.*

The inner form equips our vector space with a notion of distance, which we call the norm.

Definition 13.2.1. Let V be an inner product space. The **norm** of $v \in V$ is defined by

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

This definition makes sense because we assumed our form to be positive definite, so $\langle v, v \rangle$ is a nonnegative real number.

Example 13.2.2 ($\mathbb{R}^n$ and $\mathbb{C}^n$ are normed vector spaces)
When $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$ with the standard dot product norm, then the norm of v corresponds to the absolute value that we are used to.

Our goal now is to prove that

With the metric $d(v, w) = \|v - w\|$, V becomes a metric space.

Question 13.2.3. Verify that $d(v, w) = 0$ if and only if $v = w$.

So we just have to establish the triangle inequality. Let's now prove something we all know and love, which will be a stepping stone later:

Lemma 13.2.4 (Cauchy-Schwarz)
Let V be an inner product space. For any $v, w \in V$ we have

$$|\langle v, w \rangle| \le \|v\| \, \|w\|$$

with equality if and only if v and w are linearly dependent.

Proof. The theorem is immediate if $\langle v, w \rangle = 0$. It is also immediate if $\|v\| \, \|w\| = 0$, since then one of v or w is the zero vector. So henceforth we assume all these quantities are nonzero (as we need to divide by them later).

The key to the proof is to think about the equality case: we'll use the inequality $\langle cv - w, cv - w \rangle \ge 0$. Deferring the choice of c until later, we compute

$$\begin{aligned}
0 &\le \langle cv - w, cv - w \rangle \\
&= \langle cv, cv \rangle - \langle cv, w \rangle - \langle w, cv \rangle + \langle w, w \rangle \\
&= |c|^2 \langle v, v \rangle - c \langle v, w \rangle - \overline{c} \langle w, v \rangle + \langle w, w \rangle \\
&= |c|^2 \|v\|^2 + \|w\|^2 - c \langle v, w \rangle - \overline{c \langle v, w \rangle} \\
2 \operatorname{Re} [c \langle v, w \rangle] &\le |c|^2 \|v\|^2 + \|w\|^2
\end{aligned}$$

At this point, a good choice of c is

$$c = \frac{\|w\|}{\|v\|} \cdot \frac{|\langle v, w\rangle|}{\langle v, w\rangle}$$

since then

$$c\langle v, w\rangle = \frac{\|w\|}{\|v\|} |\langle v, w\rangle| \in \mathbb{R}$$
$$|c| = \frac{\|w\|}{\|v\|}$$

whence the inequality becomes

$$2\frac{\|w\|}{\|v\|} |\langle v, w\rangle| \leq 2\|w\|^2$$
$$|\langle v, w\rangle| \leq \|v\| \|w\|.$$

□

Thus:

Theorem 13.2.5 (Triangle inequality)
We always have

$$\|v\| + \|w\| \geq \|v + w\|$$

with equality if and only if v and w are linearly dependent and point in the same direction.

Exercise 13.2.6. Prove this by squaring both sides, and applying Cauchy-Schwarz.

In this way, our vector space now has a topological structure of a metric space.

§13.3 Orthogonality

Prototypical example for this section: Still $\mathbb{R}^n$!

Our next goal is to give the geometric notion of "perpendicular". The definition is easy enough:

Definition 13.3.1. Two nonzero vectors v and w in an inner product space are **orthogonal** if $\langle v, w\rangle = 0$.

As we expect from our geometric intuition in $\mathbb{R}^n$, this implies independence:

Lemma 13.3.2 (Orthogonal vectors are independent)
Any set of pairwise orthogonal vectors $v_1, v_2, \ldots, v_n$, with $\|v_i\| \neq 0$ for each i, is linearly independent.

Proof. Consider a dependence

$$a_1 v_1 + \cdots + a_n v_n = 0$$

for a_i in $\mathbb{R}$ or $\mathbb{C}$. Then

$$0 = \left\langle v_1, \sum a_i v_i \right\rangle = \overline{a_1} \|v_1\|^2.$$

Hence $a_1 = 0$, since we assumed $\|v_1\| \neq 0$. Similarly $a_2 = \cdots = a_m = 0$. □

In light of this, we can now consider a stronger condition on our bases:

Definition 13.3.3. An **orthonormal** basis of a *finite-dimensional* inner product space V is a basis $e_1, \dots, e_n$ such that $\|e_i\| = 1$ for every i and $\langle e_i, e_j \rangle = 0$ for any $i \neq j$.

Example 13.3.4 ($\mathbb{R}^n$ and $\mathbb{C}^n$ have standard bases)
In $\mathbb{R}^n$ and $\mathbb{C}^n$ equipped with the standard dot product, the standard basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is also orthonormal.

This is no loss of generality:

Theorem 13.3.5 (Gram-Schmidt)
Let V be a finite-dimensional inner product space. Then it has an orthonormal basis.

Sketch of Proof. One constructs the orthonormal basis explicitly from any basis $e_1, \dots, e_n$ of V. Define $\text{proj}_u(v) = \frac{\langle v,u \rangle}{\langle u,u \rangle} u$. Then recursively define

$$\begin{aligned}
u_1 &= e_1 \\
u_2 &= e_2 - \text{proj}_{u_1}(e_2) \\
u_3 &= e_3 - \text{proj}_{u_1}(e_3) - \text{proj}_{u_2}(e_3) \\
&\vdots \\
u_n &= e_n - \text{proj}_{u_1}(e_n) - \cdots - \text{proj}_{u_{n-1}}(e_n).
\end{aligned}$$

One can show the u_i are pairwise orthogonal and not zero. □

Thus, we can generally assume our bases are orthonormal.
Worth remarking:

Example 13.3.6 (The dot product is the "only" inner form)
Let V be a finite-dimensional inner product space, and consider *any* orthonormal basis $e_1, \dots, e_n$. Then we have that

$$\langle a_1 e_1 + \cdots + a_n e_n, b_1 e_1 + \cdots + b_n e_n \rangle = \sum_{i,j=1}^{n} a_i \overline{b_j} \langle e_i, e_j \rangle = \sum_{i=1}^{n} a_i \overline{b_i}$$

owing to the fact that the $\{e_i\}$ are orthonormal.

And now you know why the dot product expression is so ubiquitous.

§13.4 Hilbert spaces

In algebra we are usually scared of infinity, and so when we defined a basis of a vanilla vector space many chapters ago, we only allowed finite linear combinations. However, if we have an inner product space, then it is a metric space and we *can* sometimes actually talk about convergence.

Here is how it goes:

Definition 13.4.1. A **Hilbert space** is an inner product space V, such that the corresponding metric space is complete.

In that case, it will now often make sense to take infinite linear combinations, because we can look at the sequence of partial sums and let it converge. Here is how we might do it. Let's suppose we have e_1, e_2, ... an infinite sequence of vectors with norm 1 and which are pairwise orthogonal. Suppose c_1, c_2, ..., is a sequence of real or complex numbers. Then consider the sequence

$$\begin{aligned} v_1 &= c_1 e_1 \\ v_2 &= c_1 e_1 + c_2 e_2 \\ v_3 &= c_1 e_1 + c_2 e_2 + c_3 e_3 \\ &\vdots \end{aligned}$$

Proposition 13.4.2 (Convergence criteria in a Hilbert space)
The sequence (v_i) defined above converges if and only if $\sum |c_i|^2 < \infty$.

Proof. This will make more sense if you read Chapter 26, so you could skip this proof if you haven't read the chapter. The sequence v_i converges if and only if it is Cauchy, meaning that when $i < j$,

$$\|v_j - v_i\|^2 = |c_{i+1}|^2 + \cdots + |c_j|^2$$

tends to zero as i and j get large. This is equivalent to the sequence $s_n = |c_1|^2 + \cdots + |c_n|^2$ being Cauchy.

Since $\mathbb{R}$ is complete, s_n is Cauchy if and only if it converges. Since s_n consists of nonnegative real numbers, converges holds if and only if s_n is bounded, or equivalently if $\sum |c_i|^2 < \infty$. □

Thus, when we have a Hilbert space, we change our definition slightly:

Definition 13.4.3. An **orthonormal basis** for a Hilbert space V is a (possibly infinite) sequence e_1, e_2, ..., of vectors such that

- $\langle e_i, e_i \rangle = 1$ for all i,
- $\langle e_i, e_j \rangle = 0$ for $i \neq j$, i.e. the vectors are pairwise orthogonal
- every element of V can be expressed uniquely as an infinite linear combination

$$\sum_i c_i e_i$$

where $\sum_i |c_i|^2 < \infty$, as described above.

That's the official definition, anyways. (Note that if $\dim V < \infty$, this agrees with our usual definition, since then there are only finitely many e_i.) But for our purposes you can mostly not worry about it and instead think:

A Hilbert space is an inner product space whose basis requires infinite linear combinations, not just finite ones.

The technical condition $\sum |c_i|^2 < \infty$ is exactly the one which ensures the infinite sum makes sense.

§13.5 A few harder problems to think about

Problem 13A (Pythagorean theorem)**.** Show that if $\langle v, w \rangle = 0$ in an inner product space, then $\|v\|^2 + \|w\|^2 = \|v+w\|^2$.

Problem 13B* (Finite-dimensional $\implies$ Hilbert)**.** Show that a finite-dimensional inner product space is a Hilbert space.

Problem 13C (Taiwan IMO camp)**.** In a town there are n people and k clubs. Each club has an odd number of members, and any two clubs have an even number of common members. Prove that $k \le n$.

Problem 13D* (Inner product structure of tensors)**.** Let V and W be finite-dimensional inner product spaces over k, where k is either $\mathbb{R}$ or $\mathbb{C}$.

(a) Find a canonical way to make $V \otimes_k W$ into an inner product space too.

(b) Let $e_1, \dots, e_n$ be an orthonormal basis of V and $f_1, \dots, f_m$ be an orthonormal basis of W. What's an orthonormal basis of $V \otimes W$?

Problem 13E (Putnam 2014)**.** Let n be a positive integer. What is the largest k for which there exist $n \times n$ matrices $M_1, \dots, M_k$ and $N_1, \dots, N_k$ with real entries such that for all i and j, the matrix product $M_i N_j$ has a zero entry somewhere on its diagonal if and only if $i \neq j$?

Problem 13F (Sequence space)**.** Consider the space ℓ^2 of infinite sequences of real numbers $a = (a_1, a_2, \dots)$ satisfying $\sum_i a_i^2 < \infty$. We equip it with the dot product

$$\langle a, b \rangle = \sum_i a_i b_i.$$

Is this a Hilbert space? If so, identify a Hilbert basis.

Problem 13G (Kuratowski embedding)**.** A **Banach space** is a normed vector space V, such that the corresponding metric space is complete. (So a Hilbert space is a special case of a Banach space.)

Let (M, d) be any metric space. Prove that there exists a Banach space X and an injective function $f \colon M \hookrightarrow X$ such that $d(x,y) = \|f(x) - f(y)\|$ for any x and y.

14 Bonus: Fourier analysis

Now that we've worked hard to define abstract inner product spaces, I want to give an (optional) application: how to set up Fourier analysis correctly, using this language.

For fun, I also prove a form of Arrow's Impossibility Theorem using binary Fourier analysis.

In what follows, we let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the "circle group", thought of as the additive group of "real numbers modulo 1". There is a canonical map $e \colon \mathbb{T} \to \mathbb{C}$ sending $\mathbb{T}$ to the complex unit circle, given by

$$e(\theta) = \exp(2\pi i\theta).$$

§14.1 Synopsis

Suppose we have a domain Z and are interested in functions $f \colon Z \to \mathbb{C}$. Naturally, the set of such functions form a complex vector space. We like to equip the set of such functions with a positive definite *inner product*.

The idea of Fourier analysis is to then select an *orthonormal basis* for this set of functions, say $(e_\xi)_\xi$, which we call the **characters**; the indexing ξ are called **frequencies**. In that case, since we have a basis, every function $f \colon Z \to \mathbb{C}$ becomes a sum

$$f(x) = \sum_\xi \widehat{f}(\xi) e_\xi$$

where $\widehat{f}(\xi)$ are complex coefficients of the basis; appropriately we call $\widehat{f}$ the **Fourier coefficients**. The variable $x \in Z$ is referred to as the **physical** variable. This is generally good because the characters are deliberately chosen to be nice "symmetric" functions, like sine or cosine waves or other periodic functions. Thus we decompose an arbitrarily complicated function into a sum of nice ones.

§14.2 A reminder on Hilbert spaces

For convenience, we record a few facts about orthonormal bases.

Proposition 14.2.1 (Facts about orthonormal bases)

Let V be a complex Hilbert space with inner form $\langle -, - \rangle$ and suppose $x = \sum_\xi a_\xi e_\xi$ and $y = \sum_\xi b_\xi e_\xi$ where e_ξ are an orthonormal basis. Then

$$\begin{aligned} \langle x, x \rangle &= \sum_\xi |a_\xi|^2 \\ a_\xi &= \langle x, e_\xi \rangle \\ \langle x, y \rangle &= \sum_\xi a_\xi \overline{b_\xi}. \end{aligned}$$

Exercise 14.2.2. Prove all of these. (You don't need any of the preceding section, it's only there to motivate the notation with lots of scary ξ's.)

In what follows, most of the examples will be of finite-dimensional inner product spaces (which are thus Hilbert spaces), but the example of "square-integrable functions" will actually be an infinite dimensional example. Fortunately, as I alluded to earlier, this is no cause for alarm and you can mostly close your eyes and not worry about infinity.

§14.3 Common examples

§14.3.i Binary Fourier analysis on $\{\pm 1\}^n$

Let $Z = \{\pm 1\}^n$ for some positive integer n, so we are considering functions $f(x_1, \dots, x_n)$ accepting binary values. Then the functions $Z \to \mathbb{C}$ form a 2^n-dimensional vector space $\mathbb{C}^Z$, and we endow it with the inner form

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in Z} f(x)\overline{g(x)}.$$

In particular,

$$\langle f, f \rangle = \frac{1}{2^n} \sum_{x \in Z} |f(x)|^2$$

is the average of the squares; this establishes also that $\langle -, - \rangle$ is positive definite.

In that case, the **multilinear polynomials** form a basis of $\mathbb{C}^Z$, that is the polynomials

$$\chi_S(x_1, \dots, x_n) = \prod_{s \in S} x_s.$$

Exercise 14.3.1. Show that they're actually orthonormal under $\langle -, - \rangle$. This proves they form a basis, since there are 2^n of them.

Thus our frequency set is actually the subsets $S \subseteq \{1, \dots, n\}$. Thus, we have a decomposition

$$f = \sum_{S \subseteq \{1, \dots, n\}} \widehat{f}(S)\chi_S.$$

Example 14.3.2 (An example of binary Fourier analysis)
Let $n = 2$. Then binary functions $\{\pm 1\}^2 \to \mathbb{C}$ have a basis given by the four polynomials

$$1, \quad x_1, \quad x_2, \quad x_1x_2.$$

For example, consider the function f which is 1 at $(1, 1)$ and 0 elsewhere. Then we can put

$$f(x_1, x_2) = \frac{x_1 + 1}{2} \cdot \frac{x_2 + 1}{2} = \frac{1}{4}\left(1 + x_1 + x_2 + x_1x_2\right).$$

So the Fourier coefficients are $\widehat{f}(S) = \frac{1}{4}$ for each of the four S's.

This notion is useful in particular for binary functions $f \colon \{\pm 1\}^n \to \{\pm 1\}$; for these functions (and products thereof), we always have $\langle f, f \rangle = 1$.

It is worth noting that the frequency $\varnothing$ plays a special role:

Exercise 14.3.3. Show that
$$\widehat{f}(\varnothing) = \frac{1}{|Z|} \sum_{x \in Z} f(x).$$

§14.3.ii Fourier analysis on finite groups Z

This time, suppose we have a finite abelian group Z, and consider functions $Z \to \mathbb{C}$; this is a $|Z|$-dimensional vector space. The inner product is the same as before:
$$\langle f, g \rangle = \frac{1}{|Z|} \sum_{x \in Z} f(x)\overline{g(x)}.$$

To proceed, we'll need to be able to multiply two elements of Z. This is a bit of a nuisance since it actually won't really matter what map I pick, so I'll move briskly; feel free to skip most or all of the remaining paragraph.

Definition 14.3.4. We select a *symmetric non-degenerate bilinear form*
$$\cdot \colon Z \times Z \to \mathbb{T}$$
satisfying the following properties:

- $\xi \cdot (x_1 + x_2) = \xi \cdot x_1 + \xi \cdot x_2$ and $(\xi_1 + \xi_2) \cdot x = \xi_1 \cdot x + \xi_2 \cdot x$ (this is the word "bilinear")
- $\cdot$ is symmetric,
- For any $\xi \neq 0$, there is an x with $\xi \cdot x \neq 0$ (this is the word "nondegenerate").

Example 14.3.5 (The form on $\mathbb{Z}/n\mathbb{Z}$)
If $Z = \mathbb{Z}/n\mathbb{Z}$ then $\xi \cdot x = (\xi x)/n$ satisfies the above.

In general, it turns out finite abelian groups decompose as the sum of cyclic groups (see Section 18.1), which makes it relatively easy to find such a $\cdot$; but as I said the choice won't matter, so let's move on.

Now for the fun part: defining the characters.

Proposition 14.3.6 (e_ξ are orthonormal)
For each $\xi \in Z$ we define the character
$$e_\xi(x) = e(\xi \cdot x).$$
The $|Z|$ characters form an orthonormal basis of the space of functions $Z \to \mathbb{C}$.

Proof. I recommend skipping this one, but it is:
$$\begin{aligned}
\langle e_\xi, e_{\xi'} \rangle &= \frac{1}{|Z|} \sum_{x \in Z} e(\xi \cdot x)\overline{e(\xi' \cdot x)} \\
&= \frac{1}{|Z|} \sum_{x \in Z} e(\xi \cdot x) e(-\xi' \cdot x) \\
&= \frac{1}{|Z|} \sum_{x \in Z} e\left((\xi - \xi') \cdot x\right).
\end{aligned}$$

□

In this way, the set of frequencies is also Z, but the $\xi \in Z$ play very different roles from the "physical" $x \in Z$. Here is an example which might be enlightening.

Example 14.3.7 (Cube roots of unity filter)
Suppose $Z = \mathbb{Z}/3\mathbb{Z}$, with the inner form given by $\xi \cdot x = (\xi x)/3$. Let $\omega = \exp(\frac{2}{3}\pi i)$ be a primitive cube root of unity. Note that

$$e_\xi(x) = \begin{cases} 1 & \xi = 0 \\ \omega^x & \xi = 1 \\ \omega^{2x} & \xi = 2. \end{cases}$$

Then given $f \colon Z \to \mathbb{C}$ with $f(0) = a$, $f(1) = b$, $f(2) = c$, we obtain

$$f(x) = \frac{a+b+c}{3} \cdot 1 + \frac{a+\omega^2 b+\omega c}{3} \cdot \omega^x + \frac{a+\omega b+\omega^2 c}{3} \cdot \omega^{2x}.$$

In this way we derive that the transforms are

$$\begin{aligned} \widehat{f}(0) &= \frac{a+b+c}{3} \\ \widehat{f}(1) &= \frac{a+\omega^2 b+\omega c}{3} \\ \widehat{f}(2) &= \frac{a+\omega b+\omega^2 c}{3}. \end{aligned}$$

Exercise 14.3.8. Show that in analogy to $\widehat{f}(\varnothing)$ for binary Fourier analysis, we now have

$$\widehat{f}(0) = \frac{1}{|Z|} \sum_{x \in Z} f(x).$$

Olympiad contestants may recognize the previous example as a "roots of unity filter", which is exactly the point. For concreteness, suppose one wants to compute

$$\binom{1000}{0} + \binom{1000}{3} + \cdots + \binom{1000}{999}.$$

In that case, we can consider the function

$$w \colon \mathbb{Z}/3 \to \mathbb{C}.$$

such that $w(0) = 1$ but $w(1) = w(2) = 0$. By abuse of notation we will also think of w as a function $w \colon \mathbb{Z} \twoheadrightarrow \mathbb{Z}/3 \to \mathbb{C}$. Then the sum in question is

$$\begin{aligned} \sum_n \binom{1000}{n} w(n) &= \sum_n \binom{1000}{n} \sum_{k=0,1,2} \widehat{w}(k) \omega^{kn} \\ &= \sum_{k=0,1,2} \widehat{w}(k) \sum_n \binom{1000}{n} \omega^{kn} \\ &= \sum_{k=0,1,2} \widehat{w}(k) (1+\omega^k)^{1000}. \end{aligned}$$

In our situation, we have $\widehat{w}(0) = \widehat{w}(1) = \widehat{w}(2) = \frac{1}{3}$, and we have evaluated the desired sum. More generally, we can take any periodic weight w and use Fourier analysis in order to interchange the order of summation.

Example 14.3.9 (Binary Fourier analysis)

Suppose $Z = \{\pm 1\}^n$, viewed as an abelian group under pointwise multiplication hence isomorphic to $(\mathbb{Z}/2\mathbb{Z})^{\oplus n}$. Assume we pick the dot product defined by

$$\xi \cdot x \coloneqq \frac{1}{2} \sum_i \frac{\xi_i - 1}{2} \cdot \frac{x_i - 1}{2}$$

where $\xi = (\xi_1, \dots, \xi_n)$ and $x = (x_1, \dots, x_n)$.
We claim this coincides with the first example we gave. Indeed, let $S \subseteq \{1, \dots, n\}$ and let $\xi \in \{\pm 1\}^n$ which is -1 at positions in S, and $+1$ at positions not in S. Then the character χ_S from the previous example coincides with the character e_ξ in the new notation. In particular, $\widehat{f}(S) = \widehat{f}(\xi)$.
Thus Fourier analysis on a finite group Z subsumes binary Fourier analysis.

§14.3.iii Fourier series for functions $L^2([-\pi, \pi])$

This is the most famous one, and hence the one you've heard of.

Definition 14.3.10. The space $L^2([-\pi, \pi])$ consists of all functions $f \colon [-\pi, \pi] \to \mathbb{C}$ such that the integral $\int_{[-\pi,\pi]} |f(x)|^2 \, dx$ exists and is finite, modulo the relation that a function which is zero "almost everywhere" is considered to equal zero.[1]

It is made into an inner product space according to

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{[-\pi,\pi]} f(x) \overline{g(x)} \, dx.$$

It turns out (we won't prove) that this is an (infinite-dimensional) Hilbert space!
Now, the beauty of Fourier analysis is that **this space has a great basis**:

Theorem 14.3.11 (The classical Fourier basis)

For each integer n, define

$$e_n(x) = \exp(inx).$$

Then e_n form an orthonormal basis of the Hilbert space $L^2([-\pi, \pi])$.

Thus this time the frequency set $\mathbb{Z}$ is infinite, and we have

$$f(x) = \sum_n \widehat{f}(n) \exp(inx) \quad \text{almost everywhere}$$

for coefficients $\widehat{f}(n)$ with $\sum_n \left|\widehat{f}(n)\right|^2 < \infty$. Since the frequency set is indexed by $\mathbb{Z}$, we call this a **Fourier series** to reflect the fact that the index is $n \in \mathbb{Z}$.

[1] We won't define this, yet, as it won't matter to us for now. But we will elaborate more on this in the parts on measure theory.

There is one point at which this is relevant. Often we require that the function f satisfies $f(-\pi) = f(\pi)$, so that f becomes a periodic function, and we can think of it as $f \colon \mathbb{T} \to \mathbb{C}$. This makes no essential difference since we merely change the value at one point.

Exercise 14.3.12. Show once again

$$\widehat{f}(0) = \frac{1}{2\pi} \int_{[-\pi,\pi]} f(x)\,dx.$$

§14.4 Summary, and another teaser

We summarize our various flavors of Fourier analysis in the following table.

Type	Physical var	Frequency var	Basis functions
Binary	$\{\pm 1\}^n$	Subsets $S \subseteq \{1,\dots,n\}$	$\prod_{s\in S} x_s$
Finite group	Z	$\xi \in Z$, choice of $\cdot$	$e(\xi \cdot x)$
Fourier series	$\mathbb{T}$ or $[-\pi,\pi]$	$n \in \mathbb{Z}$	$\exp(inx)$
Discrete	$\mathbb{Z}/n\mathbb{Z}$	$\xi \in \mathbb{Z}/n\mathbb{Z}$	$e(\xi x/n)$

I snuck in a fourth row with $Z = \mathbb{Z}/n\mathbb{Z}$, but it's a special case of the second row, so no cause for alarm.

Alluding to the future, I want to hint at how Chapter 39 starts. Each one of these is really a statement about how functions from $G \to \mathbb{C}$ can be expressed in terms of functions $\widehat{G} \to \mathbb{C}$, for some "dual" $\widehat{G}$. In that sense, we could rewrite the above table as:

Name	Domain G	Dual $\widehat{G}$	Characters
Binary	$\{\pm 1\}^n$	$S \subseteq \{1,\dots,n\}$	$\prod_{s\in S} x_s$
Finite group	Z	$\xi \in \widehat{Z} \cong Z$	$e(i\xi \cdot x)$
Fourier series	$\mathbb{T} \cong [-\pi,\pi]$	$n \in \mathbb{Z}$	$\exp(inx)$
Discrete	$\mathbb{Z}/n\mathbb{Z}$	$\xi \in \mathbb{Z}/n\mathbb{Z}$	$e(\xi x/n)$

It will turn out that in general we can say something about many different domains G, once we know what it means to integrate a measure. This is the so-called *Pontryagin duality*; and it is discussed as a follow-up bonus in Chapter 39.

§14.5 Parseval and friends

Here is a fun section in which you get to learn a lot of big names quickly. Basically, we can take each of the three results from Proposition 14.2.1, translate it into the context of our Fourier analysis (for which we have an orthonormal basis of the Hilbert space), and get a big-name result.

Corollary 14.5.1 (Parseval theorem)

Let $f\colon Z \to \mathbb{C}$, where Z is a finite abelian group. Then

$$\sum_{\xi} |\widehat{f}(\xi)|^2 = \frac{1}{|Z|} \sum_{x\in Z} |f(x)|^2.$$

Similarly, if $f\colon [-\pi,\pi] \to \mathbb{C}$ is square-integrable then its Fourier series satisfies

$$\sum_{n} |\widehat{f}(n)|^2 = \frac{1}{2\pi} \int_{[-\pi,\pi]} |f(x)|^2\,dx.$$

Proof. Recall that $\langle f, f\rangle$ is equal to the square sum of the coefficients. □

Corollary 14.5.2 (Fourier inversion formula)
Let $f\colon Z \to \mathbb{C}$, where Z is a finite abelian group. Then

$$\widehat{f}(\xi) = \frac{1}{|Z|} \sum_{x \in Z} f(x)\overline{e_\xi(x)}.$$

Similarly, if $f\colon [-\pi, \pi] \to \mathbb{C}$ is square-integrable then its Fourier series is given by

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{[-\pi,\pi]} f(x) \exp(-inx)\, dx.$$

Proof. Recall that in an orthonormal basis $(e_\xi)_\xi$, the coefficient of e_ξ in f is $\langle f, e_\xi \rangle$. □

Question 14.5.3. What happens when $\xi = 0$ above?

Corollary 14.5.4 (Plancherel theorem)
Let $f\colon Z \to \mathbb{C}$, where Z is a finite abelian group. Then

$$\langle f, g \rangle = \sum_{\xi \in Z} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}.$$

Similarly, if $f\colon [-\pi, \pi] \to \mathbb{C}$ is square-integrable then

$$\langle f, g \rangle = \sum_{n} \widehat{f}(n)\overline{\widehat{g}(n)}.$$

Question 14.5.5. Prove this one in one line (like before).

§14.6 Application: Basel problem

One cute application about Fourier analysis on $L^2([-\pi, \pi])$ is that you can get some otherwise hard-to-compute sums, as long as you are willing to use a little calculus.

Here is the classical one:

Theorem 14.6.1 (Basel problem)
We have

$$\sum_{n \geq 1} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

The proof is to consider the identity function $f(x) = x$, which is certainly square-integrable. Then by Parseval, we have

$$\sum_{n \in \mathbb{Z}} \left|\widehat{f}(n)\right|^2 = \langle f, f \rangle = \frac{1}{2\pi} \int_{[-\pi,\pi]} |f(x)|^2 \, dx.$$

A calculus computation gives

$$\frac{1}{2\pi} \int_{[-\pi,\pi]} x^2 \, dx = \frac{\pi^2}{3}.$$

On the other hand, we will now compute all Fourier coefficients. We have already that

$$\widehat{f}(0) = \frac{1}{2\pi}\int_{[-\pi,\pi]} f(x)\,dx = \frac{1}{2\pi}\int_{[-\pi,\pi]} x\,dx = 0.$$

For $n \neq 0$, we have by definition (or "Fourier inversion formula", if you want to use big words) the formula

$$\begin{aligned}\widehat{f}(n) &= \langle f, \exp(inx)\rangle \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} x \cdot \overline{\exp(inx)}\,dx \\ &= \frac{1}{2\pi}\int_{[-\pi,\pi]} x\exp(-inx)\,dx.\end{aligned}$$

The anti-derivative is equal to $\frac{1}{n^2}\exp(-inx)(1+inx)$, which thus with some more calculation gives that

$$\widehat{f}(n) = \frac{(-1)^n}{n}i.$$

So

$$\sum_n \left|\widehat{f}(n)\right|^2 = 2\sum_{n\geq 1}\frac{1}{n^2}$$

implying the result.

§14.7 Application: Arrow's Impossibility Theorem

As an application of binary Fourier analysis, we now prove a form of Arrow's theorem.

Consider n voters voting among 3 candidates A, B, C. Each voter specifies a tuple $v_i = (x_i, y_i, z_i) \in \{\pm 1\}^3$ as follows:

- $x_i = 1$ if person i ranks A ahead of B, and $x_i = -1$ otherwise.
- $y_i = 1$ if person i ranks B ahead of C, and $y_i = -1$ otherwise.
- $z_i = 1$ if person i ranks C ahead of A, and $z_i = -1$ otherwise.

Tacitly, we only consider $3! = 6$ possibilities for v_i: we forbid "paradoxical" votes of the form $x_i = y_i = z_i$ by assuming that people's votes are consistent (meaning the preferences are transitive).

For brevity, let $x_\bullet = (x_1, \dots, x_n)$ and define $y_\bullet$ and $z_\bullet$ similarly. Then, we can consider a voting mechanism

$$\begin{aligned} f&\colon \{\pm 1\}^n \to \{\pm 1\} \\ g&\colon \{\pm 1\}^n \to \{\pm 1\} \\ h&\colon \{\pm 1\}^n \to \{\pm 1\}\end{aligned}$$

such that

- $f(x_\bullet)$ is the global preference of A vs. B,
- $g(y_\bullet)$ is the global preference of B vs. C,
- and $h(z_\bullet)$ is the global preference of C vs. A.

We'd like to avoid situations where the global preference $(f(x_\bullet), g(y_\bullet), h(z_\bullet))$ is itself paradoxical.

Let $\mathbb{E}f$ denote the average value of f across all 2^n inputs. Define $\mathbb{E}g$ and $\mathbb{E}h$ similarly. We'll add an assumption that $\mathbb{E}f = \mathbb{E}g = \mathbb{E}h = 0$, which provides symmetry (and e.g. excludes the possibility that f, g, h are constant functions which ignore voter input). With that we will prove the following result:

Theorem 14.7.1 (Arrow Impossibility Theorem)

Assume that (f, g, h) always avoids paradoxical outcomes, and assume $\mathbb{E}f = \mathbb{E}g = \mathbb{E}h = 0$. Then (f, g, h) is either a dictatorship or anti-dictatorship: there exists a "dictator" k such that

$$f(x_\bullet) = \pm x_k, \qquad g(y_\bullet) = \pm y_k, \qquad h(z_\bullet) = \pm z_k$$

where all three signs coincide.

Unlike the usual Arrow theorem, we do *not* assume that $f(+1, \dots, +1) = +1$ (hence possibility of anti-dictatorship).

Proof. Suppose the voters each randomly select one of the $3! = 6$ possible consistent votes. In Problem 14B it is shown that the exact probability of a paradoxical outcome for any functions f, g, h is given exactly by

$$\frac{1}{4} + \frac{1}{4} \sum_{S \subseteq \{1,\dots,n\}} \left(-\frac{1}{3}\right)^{|S|} \left(\widehat{f}(S)\widehat{g}(S) + \widehat{g}(S)\widehat{h}(S) + \widehat{h}(S)\widehat{f}(S)\right).$$

Assume that this probability (of a paradoxical outcome) equals 0. Then, we derive

$$1 = \sum_{S \subseteq \{1,\dots,n\}} -\left(-\frac{1}{3}\right)^{|S|} \left(\widehat{f}(S)\widehat{g}(S) + \widehat{g}(S)\widehat{h}(S) + \widehat{h}(S)\widehat{f}(S)\right).$$

But now we can just use weak inequalities. We have $\widehat{f}(\varnothing) = \mathbb{E}f = 0$ and similarly for $\widehat{g}$ and $\widehat{h}$, so we restrict attention to $|S| \geq 1$. We then combine the famous inequality $|ab + bc + ca| \leq a^2 + b^2 + c^2$ (which is true across all real numbers) to deduce that

$$\begin{aligned}
1 &= \sum_{S \subseteq \{1,\dots,n\}} -\left(-\frac{1}{3}\right)^{|S|} \left(\widehat{f}(S)\widehat{g}(S) + \widehat{g}(S)\widehat{h}(S) + \widehat{h}(S)\widehat{f}(S)\right) \\
&\leq \sum_{S \subseteq \{1,\dots,n\}} \left(\frac{1}{3}\right)^{|S|} \left(\widehat{f}(S)^2 + \widehat{g}(S)^2 + \widehat{h}(S)^2\right) \\
&\leq \sum_{S \subseteq \{1,\dots,n\}} \left(\frac{1}{3}\right)^{1} \left(\widehat{f}(S)^2 + \widehat{g}(S)^2 + \widehat{h}(S)^2\right) \\
&= \frac{1}{3}(1 + 1 + 1) = 1.
\end{aligned}$$

with the last step by Parseval. So all inequalities must be sharp, and in particular $\widehat{f}$, $\widehat{g}$, $\widehat{h}$ are supported on one-element sets, i.e. they are linear in inputs. As f, g, h are ± 1 valued, each f, g, h is itself either a dictator or anti-dictator function. Since (f, g, h) is always consistent, this implies the final result. □

§14.8 A few harder problems to think about

Problem 14A (For calculus fans)**.** Prove that

$$\sum_{n \geq 1} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

Problem 14B. Let $f, g, h \colon \{\pm 1\}^n \to \{\pm 1\}$ be any three functions. For each i, we randomly select $(x_i, y_i, z_i) \in \{\pm 1\}^3$ subject to the constraint that not all are equal (hence, choosing among $2^3 - 2 = 6$ possibilities). Prove that the probability that

$$f(x_1, \dots, x_n) = g(y_1, \dots, y_n) = h(z_1, \dots, z_n)$$

is given by the formula

$$\frac{1}{4} + \frac{1}{4} \sum_{S \subseteq \{1,\dots,n\}} \left(-\frac{1}{3}\right)^{|S|} \left(\widehat{f}(S)\widehat{g}(S) + \widehat{g}(S)\widehat{h}(S) + \widehat{h}(S)\widehat{f}(S)\right)$$

15 Duals, adjoint, and transposes

This chapter is dedicated to the basis-free interpretation of the transpose and conjugate transpose of a matrix.

Poster corollary: we will see that symmetric matrices with real coefficients are diagonalizable and have real eigenvalues.

§15.1 Dual of a map

Prototypical example for this section: The example below.

We go ahead and now define a notion that will grow up to be the transpose of a matrix.

Definition 15.1.1. Let V and W be vector spaces. Suppose $T\colon V \to W$ is a linear map. Then we actually get a map

$$\begin{aligned} T^\vee\colon W^\vee &\to V^\vee \\ f &\mapsto f \circ T. \end{aligned}$$

This map is called the **dual map**.

Example 15.1.2 (Example of a dual map)

Work over $\mathbb{R}$. Let's consider V with basis e_1, e_2, e_3 and W with basis f_1, f_2. Suppose that

$$\begin{aligned} T(e_1) &= f_1 + 2f_2 \\ T(e_2) &= 3f_1 + 4f_2 \\ T(e_3) &= 5f_1 + 6f_2. \end{aligned}$$

Now consider $V^\vee$ with its dual basis $e_1^\vee$, $e_2^\vee$, $e_3^\vee$ and $W^\vee$ with its dual basis $f_1^\vee$, $f_2^\vee$. Let's compute $T^\vee(f_1^\vee) = f_1^\vee \circ T$: it is given by

$$\begin{aligned} f_1^\vee\left(T(ae_1 + be_2 + ce_3)\right) &= f_1^\vee\left((a+3b+5c)f_1 + (2a+4b+6c)f_2\right) \\ &= a + 3b + 5c. \end{aligned}$$

So accordingly we can write

$$T^\vee(f_1^\vee) = e_1^\vee + 3e_2^\vee + 5e_3^\vee$$

Similarly,

$$T^\vee(f_2^\vee) = 2e_1^\vee + 4e_2^\vee + 6e_3^\vee.$$

This determines $T^\vee$ completely.

If we write the matrices for T and $T^\vee$ in terms of our basis, we now see that

$$T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \quad \text{and} \quad T^\vee = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

So in our selected basis, we find that the matrices are **transposes**: mirror images of each other over the diagonal.

Of course, this should work in general.

Theorem 15.1.3 (Transpose interpretation of $T^\vee$)
Let V and W be finite-dimensional k-vector spaces. Then, for any $T\colon V \to W$, the following two matrices are transposes:

- The matrix for $T\colon V \to W$ expressed in the basis (e_i), (f_j).
- The matrix for $T^\vee\colon W^\vee \to V^\vee$ expressed in the basis $(f_j^\vee)$, $(e_i^\vee)$.

Proof. The (i,j)th entry of the matrix T corresponds to the coefficient of f_i in $T(e_j)$, which corresponds to the coefficient of $e_j^\vee$ in $f_i^\vee \circ T$. □

The nice part of this is that the definition of $T^\vee$ is basis-free. So it means that if we start with any linear map T, and then pick whichever basis we feel like, then T and $T^\vee$ will still be transposes.

§15.2 Identifying with the dual space

For the rest of this chapter, though, we'll now bring inner products into the picture.

Earlier I complained that there was no natural isomorphism $V \cong V^\vee$. But in fact, given an inner form we can actually make such an identification: that is we can naturally associate every linear map $\xi\colon V \to k$ with a vector $v \in V$.

To see how we might do this, suppose $V = \mathbb{R}^3$ for now with an orthonormal basis e_1, e_2, e_3. How might we use the inner product to represent a map from $V \to \mathbb{R}$? For example, take $\xi \in V^\vee$ by $\xi(e_1) = 3$, $\xi(e_2) = 4$ and $\xi(e_3) = 5$. Actually, I claim that

$$\xi(v) = \langle v, 3e_1 + 4e_2 + 5e_3 \rangle$$

for every v.

Question 15.2.1. Check this.

And this works beautifully in the real case.

Theorem 15.2.2 ($V \cong V^\vee$ for real inner forms)
Let V be a finite-dimensional *real* inner product space and $V^\vee$ its dual. Then the map $V \to V^\vee$ by

$$v \mapsto \langle -, v \rangle \in V^\vee$$

is an isomorphism of real vector spaces.

Proof. It suffices to show that the map is injective and surjective.

- Injective: suppose $\langle v, v_1 \rangle = \langle v, v_2 \rangle$ for every vector $v \in V$. This means $\langle v, v_1 - v_2 \rangle = 0$ for every vector $v \in V$. This can only happen if $v_1 - v_2 = 0$; for example, take $v = v_1 - v_2$ and use positive definiteness.

- Surjective: take an orthonormal basis $e_1, \dots e_n$ and let $e_1^\vee, \dots, e_n^\vee$ be the dual basis on $V^\vee$. Then e_1 maps to $e_1^\vee$, et cetera. □

Actually, since we already know $\dim V = \dim V^\vee$ we only had to prove one of the above. As a matter of personal taste, I find the proof of injectivity more elegant, and the proof of surjectivity more enlightening, so I included both. Thus

If a real inner product space V is given an inner form, then V and $V^\vee$ are canonically isomorphic.

Unfortunately, things go awry if V is complex. Here is the result:

Theorem 15.2.3 (V versus $V^\vee$ for complex inner forms)
Let V be a finite-dimensional *complex* inner product space and $V^\vee$ its dual. Then the map $V \to V^\vee$ by

$$v \mapsto \langle -, v \rangle \in V^\vee$$

is a bijection of sets.

Wait, what? Well, the proof above shows that it is both injective and surjective, but why is it not an isomorphism? The answer is that it is not a linear map: since the form is sesquilinear we have for example

$$iv \mapsto \langle -, iv \rangle = -i \langle -, v \rangle$$

which has introduced a minus sign! In fact, it is an *anti-linear* map, in the sense we defined before.

Eager readers might try to fix this by defining the isomorphism $v \mapsto \langle v, - \rangle$ instead. However, this also fails, because the right-hand side is not even an element of $V^\vee$: it is an "anti-linear", not linear.

And so we are stuck. Fortunately, we will only need the "bijection" result for what follows, so we can continue on anyways. (If you want to fix this, Problem 15D gives a way to do so.)

§15.3 The adjoint (conjugate transpose)

We will see that, as a result of the flipping above, the *conjugate transpose* is actually the better concept for inner product spaces: since it can be defined using only the inner product without making mention to dual spaces at all.

Definition 15.3.1. Let V and W be finite-dimensional inner product spaces, and let $T \colon V \to W$. The **adjoint** (or **conjugate transpose**) of T, denoted $T^\dagger \colon W \to V$, is defined as follows: for every vector $w \in W$, we let $T^\dagger(w) \in V$ be the unique vector with

$$\left\langle v, T^\dagger(w) \right\rangle_V = \langle T(v), w \rangle_W$$

for every $v \in V$.

Some immediate remarks about this definition:

- Our $T^\dagger$ is well-defined, because $v \mapsto \langle T(v), w \rangle_W$ is some function in $V^\vee$, and hence by the bijection earlier it should be uniquely of the form $\langle -, v \rangle$ for some $v \in V$.

- This map $T^\dagger$ is indeed a linear map (why?).
- The niceness of this definition is that it doesn't make reference to any basis or even $V^\vee$, so it is the "right" definition for a inner product space.

Example 15.3.2 (Example of an adjoint map)
We'll work over $\mathbb{C}$, so the conjugates are more visible. Let's consider V with orthonormal basis e_1, e_2, e_3 and W with orthonormal basis f_1, f_2. We put

$$\begin{aligned} T(e_1) &= if_1 + 2f_2 \\ T(e_2) &= 3f_1 + 4f_2 \\ T(e_3) &= 5f_1 + 6if_2. \end{aligned}$$

We compute $T^\dagger(f_1)$. It is the unique vector $x \in V$ such that

$$\langle v, x \rangle_V = \langle T(v), f_1 \rangle_W$$

for any $v \in V$. If we expand $v = ae_1 + be_2 + ce_3$ the above equality becomes

$$\begin{aligned} \langle ae_1 + be_2 + ce_3, x \rangle_V &= \langle T(ae_1 + be_2 + ce_3), f_1 \rangle_W \\ &= ia + 3b + 5c. \end{aligned}$$

However, since x is in the second argument, this means we actually want to take

$$T^\dagger(f_1) = -ie_1 + 3e_2 + 5e_3$$

so that the sesquilinearity will conjugate the i.

The pattern continues, though we remind the reader that we need the basis to be orthonormal to proceed.

Theorem 15.3.3 (Adjoints are conjugate transposes)
Fix an *orthonormal* basis of a finite-dimensional inner product space V. Let $T \colon V \to V$ be a linear map. If we write T as a matrix in this basis, then the matrix $T^\dagger$ (in the same basis) is the *conjugate transpose* of the matrix of T; that is, the (i,j)th entry of $T^\vee$ is the complex conjugate of the (j,i)th entry of T.

Proof. One-line version: take v and w to be basis elements, and this falls right out.

Full proof: let

$$T = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

in this basis $e_1, \dots, e_n$. Then, letting $w = e_i$ and $v = e_j$ we deduce that

$$\left\langle e_i, T^\dagger(e_j) \right\rangle = \langle T(e_i), e_j \rangle = a_{ji} \implies \left\langle T^\dagger(e_j), e_i \right\rangle = \overline{a_{ji}}$$

for any i, which is enough to deduce the result. □

§15.4 Eigenvalues of normal maps

We now come to the advertised theorem. Restrict to the situation where $T\colon V \to V$. You see, the world would be a very beautiful place if it turned out that we could pick a basis of eigenvectors that was also *orthonormal*. This is of course far too much to hope for; even without the orthonormal condition, we saw that Jordan form could still have 1's off the diagonal.

However, it turns out that there is a complete characterization of exactly when our overzealous dream is true.

Definition 15.4.1. We say a linear map T (from a finite-dimensional inner product space to itself) is **normal** if $TT^\dagger = T^\dagger T$.

We say a complex T is **self-adjoint** or **Hermitian** if $T = T^\dagger$; i.e. as a matrix in any orthonormal basis, T is its own conjugate transpose. For real T we say "self-adjoint", "Hermitian" or **symmetric**.

Theorem 15.4.2 (Normal $\iff$ diagonalizable with orthonormal basis)
Let V be a finite-dimensional complex inner product space. A linear map $T\colon V \to V$ is normal if and only if one can pick an orthonormal basis of eigenvectors.

Exercise 15.4.3. Show that if there exists such an orthonormal basis then $T\colon V \to V$ is normal, by writing T as a diagonal matrix in that basis.

Proof. This is long, and maybe should be omitted on a first reading. If T has an orthonormal basis of eigenvectors, this result is immediate.

Now assume T is normal. We first prove T is diagonalizable; this is the hard part.

Claim 15.4.4. If T is normal, then $\ker T = \ker T^r = \ker T^\dagger$ for $r \geq 1$. (Here T^r is T applied r times.)

Proof of Claim. Let $S = T^\dagger \circ T$, which is self-adjoint. We first note that S is Hermitian and $\ker S = \ker T$. To see it's Hermitian, note $\langle Sv, w\rangle = \langle Tv, Tw\rangle = \langle v, Sw\rangle$. Taking $v = w$ also implies $\ker S \subseteq \ker T$ (and hence equality since obviously $\ker T \subseteq \ker S$).

First, since we have $\langle S^r(v), S^{r-2}(v)\rangle = \langle S^{r-1}(v), S^{r-1}(v)\rangle$, an induction shows that $\ker S = \ker S^r$ for $r \geq 1$. Now, since T is normal, we have $S^r = (T^\dagger)^r \circ T^r$, and thus we have the inclusion

$$\ker T \subseteq \ker T^r \subseteq \ker S^r = \ker S = \ker T$$

where the last equality follows from the first claim. Thus in fact $\ker T = \ker T^r$.

Finally, to show equality with $\ker T^\dagger$ we

$$\begin{aligned} \langle Tv, Tv\rangle &= \left\langle v, T^\dagger T v\right\rangle \\ &= \left\langle v, TT^\dagger v\right\rangle \\ &= \left\langle T^\dagger v, T^\dagger v\right\rangle. \end{aligned}$$

■

Now consider the given T, and any λ.

Question 15.4.5. Show that $(T - \lambda\mathrm{id})^\dagger = T^\dagger - \overline{\lambda}\mathrm{id}$. Thus if T is normal, so is $T - \lambda\mathrm{id}$.

In particular, for any eigenvalue λ of T, we find that $\ker(T-\lambda\text{id}) = \ker(T-\lambda\text{id})^r$. This implies that all the Jordan blocks of T have size 1; i.e. that T is in fact diagonalizable. Finally, we conclude that the eigenvectors of T and $T^\dagger$ match, and the eigenvalues are complex conjugates.

So, diagonalize T. We just need to show that if v and w are eigenvectors of T with distinct eigenvalues, then they are orthogonal. (We can use Gram-Schmidt on any eigenvalue that appears multiple times.) To do this, suppose $T(v) = \lambda v$ and $T(w) = \mu w$ (thus $T^\dagger(w) = \overline{\mu} w$). Then

$$\lambda \langle v, w\rangle = \langle \lambda v, w\rangle = \langle Tv, w\rangle = \left\langle v, T^\dagger(w)\right\rangle = \langle v, \overline{\mu} w\rangle = \mu \langle v, w\rangle .$$

Since $\lambda \neq \mu$, we conclude $\langle v, w\rangle = 0$. □

This means that not only can we write

$$T = \begin{bmatrix} \lambda_1 & \dots & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

but moreover that the basis associated with this matrix happens to be orthonormal vectors.

As a corollary:

Theorem 15.4.6 (Hermitian matrices have real eigenvalues)
A Hermitian matrix T is diagonalizable, and all its eigenvalues are real.

Proof. Obviously Hermitian $\implies$ normal, so write it in the orthonormal basis of eigenvectors. To see that the eigenvalues are real, note that $T = T^\dagger$ means $\lambda_i = \overline{\lambda_i}$ for every i. □

§15.5 A few harder problems to think about

Problem 15A* (Double dual)**.** Let V be a finite-dimensional vector space. Prove that

$$\begin{aligned} V &\to (V^\vee)^\vee \\ v &\mapsto (\xi \mapsto \xi(v)) \end{aligned}$$

gives an isomorphism. (This is significant because the isomorphism is *canonical*, and in particular does not depend on the choice of basis. So this is more impressive.)

Problem 15B (Fundamental theorem of linear algebra)**.** Let $T\colon V \to W$ be a map of finite-dimensional k-vector spaces. Prove that

$$\dim \operatorname{im} T = \dim \operatorname{im} T^\vee = \dim V - \dim \ker T = \dim W - \dim \ker T^\vee.$$

Problem 15C† (Row rank is column rank)**.** A $m \times n$ matrix M of real numbers is given. The *column rank* of M is the dimension of the span in $\mathbb{R}^m$ of its n column vectors. The *row rank* of M is the dimension of the span in $\mathbb{R}^n$ of its m row vectors. Prove that the row rank and column rank are equal.

Problem 15D (The complex conjugate spaces)**.** Let $V = (V, +, \cdot)$ be a complex vector space. Define the **complex conjugate vector space**, denoted $\overline{V} = (V, +, *)$ by changing just the multiplication:

$$c * v = \overline{c} \cdot v.$$

Show that for any sesquilinear form on V, if V is finite-dimensional, then

$$\begin{aligned} \overline{V} &\to V^\vee \\ v &\mapsto \langle -, v \rangle \end{aligned}$$

is an isomorphism of complex vector spaces.

Problem 15E ($T^\dagger$ vs $T^\vee$)**.** Let V and W be real inner product spaces and let $T \colon V \to W$ be a linear map. Show that the following diagram commutes:

$$\begin{array}{ccc} W & \xrightarrow{T^\dagger} & V \\ \cong\downarrow & & \downarrow\cong \\ W^\vee & \xrightarrow[T^\vee]{} & V^\vee \end{array}$$

Here the isomorphisms are $v \mapsto \langle -, v \rangle$. Thus, for real inner product spaces, $T^\dagger$ is just $T^\vee$ with the duals eliminated (by Theorem 15.2.2).

Problem 15F (Polynomial criteria for normality)**.** Let V be a complex inner product space and let $T \colon V \to V$ be a linear map. Show that T is normal if and only if there is a polynomial[1] $p \in \mathbb{C}[t]$ such that

$$T^\dagger = p(T).$$

Problem 15G (Kronecker product of matrices)**.** Find an equivalence between the following two definitions of the Kronecker product, the former from a mathematician and the latter from a computer scientist:

- Suppose $A \colon V_1 \to W_1$ and $B \colon V_2 \to W_2$ are linear maps of finite-dimensional vector spaces over $\mathbb{R}$. Then we define $A \otimes B \colon V_1 \otimes V_2 \to W_1 \otimes W_2$ on simple tensors by $v_1 \otimes v_2 \mapsto A(v_1) \otimes B(v_2)$.
- Suppose A is an $m \times n$ matrix and B is a $p \times q$ matrix. Then $A \otimes B$ is an operator which takes a $q \times n$ matrix X and returns the $p \times m$ matrix $BXA^\top$.

[1] Here, we mean $p(T)$ in the same composition sense as in Cayley-Hamilton.