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 $e_1, \dots, e_n$ 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: it's a way of writing a linear map in terms of bases.

Suppose we have a finite-dimensional vector space V with basis $e_1, \dots, e_m$ and a vector space W with basis $w_1, \dots, w_n$. I also 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. In fact, 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 9.6.1 (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

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

Example 9.6.2 (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 9.6.3 (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.

Exercise 9.6.4. Check this for yourself! For a concrete example let $\mathbb{R}^2 \xrightarrow{T} \mathbb{R}^2 \xrightarrow{S} \mathbb{R}^2$ by $T(e_1) = 2e_1 + 3e_2$ and $T(e_2) = 4e_1 + 5e_2$, $S(e_1) = 6e_1 + 7e_2$ and $S(e_2) = 8e_1 + 9e_2$. Compute $S(T(e_1))$ and $S(T(e_2))$ and see how it compares to multiplying the matrices associated to S and T.

In particular, since function composition is associative, it follows that matrix multiplication is as well. To drive this point home,

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

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

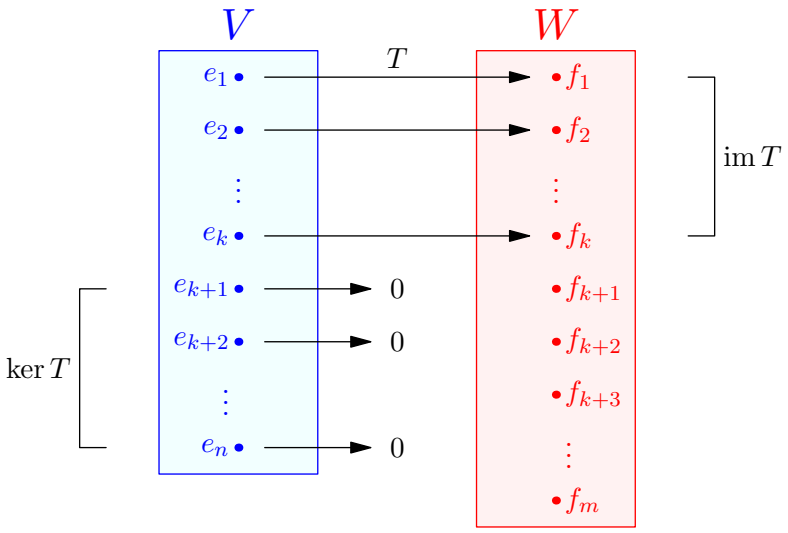
$$T(v_i) = \begin{cases} w_i & \text{if } i \le k \\ 0_W & \text{if } i > k. \end{cases}$$

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.

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

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 (Digression) Arrays of numbers are evil

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

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

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 (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 9D (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 9E (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 9F (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 9G* (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 9H*. 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

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