Linearity

Adam Layne

2025-01-14

Table of contents

What are these notes?

These are notes for a first course in Linear Algebra.

License

Preface

This section discusses why these notes exist. Students may skip this section.

Why publish a new set of Linear Algebra notes?

Linear Algebra, like Calculus, is one of the math subjects with the most textbooks, so it's reasonable to ask why a new set of notes is needed. Plainly, I looked at the six open-access books on the subject on the AIMath website and found that none of them were fit for my purpose (detailed below).

The perspective of these notes

These notes are constructed to vindicate the following objectives:

1. It is morally right for course materials to be free. Few existing books on this subject in English satisfy this criterion (eg Hefferon (2022) is GFDL or CC BY-SA 3.0 US, Treil (2024) is CC BY-NC-ND 3.0 with source unavailable). This book, and the source used to generate it are freely available with a permissive license.

2. In practice, scientists, engineers, etc. need to be able to recognize linearity so that they may choose the correct solution techniques. They also need to understand why linear problems are preferable to non-linear ones so that they might try to massage their current problem into a linear one.

3. When we say "solution techniques" as above, 99% of the time we mean software packages. Mathematicians and physicists teach linear algebra techniques in colleges and universities, and emphasize by-hand solution techniques for historical and cultural reasons. Most working people who encounter such problems do not use such techniques, they recognize that their problem is linear and offload the problem to a software package. Mathematicians and physicists generally get a second pass at learning linear algebra in a more theory-heavy context (at the very least when learning modules), and so do not need that approach in a first course.

4. The usefulness of linear algebra techniques stems wholly from the homomorphism property of linear maps:

$$L(aV + bW) = aL(V) + bL(W)$$

   No introductory, open-access, English language books on the topic that I am aware of motivate the study of the subject with this point. They traditionally begin with coordinate geometry or solving systems of linear equations. It is a very mathematicians' way of thinking to motivate study of a topic by identifying a class of equations and asking "How do we solve them? What properties do they have?" This is not a way of thinking that is useful for people encountering linearity in the wild.

   Many advanced undergraduate books like Axler (2024) begin with the definition of a vector space.

   It is also a very mathematician's way of thinking to begin with a definition of a set of objects which will be mapped into or out of (viz. most treatments of Abstract Algebra). To talk about maps (which are really the objects of interest), *surely* we must first talk about (co)domains!

   Coordinate geometry is, at least, a class of real problems where linear techniques naturally arise, but the relevance of this as an example from "the wild" has basically vanished in the last 70 years. Today's scientists and engineers are more likely to encounter linearity in optimization, data science, machine learning, or numerical PDEs.

So, for these reasons, I set out to write my own course notes.

Part I

Vector Spaces

Chapter 1

What is Linearity?

The function $C : \mathbb{R} \to \mathbb{R}$ given by

$$C(r) := 2\pi r$$

computes the circumference of a circle, given its radius.

> **Notation**
>
> First, let's talk about this notation.
> The notation $C : \mathbb{R} \to \mathbb{R}$ tells you about the inputs and outputs of the function C. When we write $f : A \to B$, we mean that f takes inputs from the set A and creates outputs in the set B. You can think of f as a machine transforming As into Bs.
> The notation $C(r) := 2\pi r$ tells you that the *definition* of the function C appears here. This is to avoid confusion like the following:
>
> $$f(x) = x^2$$
>
> If the function f hasn't appeared before, then this equation is probably a definition. But if we wrote "Given $f(x) = x - 2$, solve the equation $f(x) = x^2$", then the same equation is not a definition. To avoid this ambiguity, when we write an equals sign with a colon on one side like this $A := B$ or $B =: A$, we mean that the name on the side of the $=$ *with* the colon is defined to be the expression on the side without.

The circumference of a circle has a couple nice properties. First, the circumference of a circle of radius 14 is twice the circumference of a circle of radius 7:

$$C(7) = 2\pi(7) = 14\pi$$
$$C(14) = 2\pi(14) = 28\pi$$

This holds in general; multiplying the radius of a circle by k also changes the circumference by a factor of k:

$$C(kr) = kC(r).$$

Furthermore, adding any amount to the radius increases the circumference in a predictable way:

$$C(r_1 + r_2) = C(r_1) + C(r_2).$$

It's a bit remarkable that these two properties hold not just for circles; scaling any shape in the plane (with circumference c) by a factor of k multiplies its circumference by k (its new circumference is kc), and increasing the scale by a constant s increases its circumference by sc.

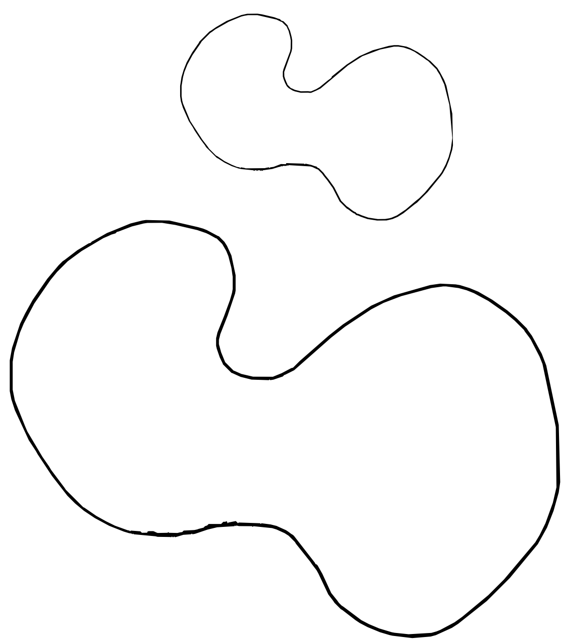

Figure 1.1: Scaling a curve by a factor of 2

1.1 Two properties

What other functions have the following two properties?

$$
\begin{aligned}
&\text{Property 1:} \quad f(ax) = af(x) \\
&\text{Property 2:} \quad f(x+y) = f(x) + f(y)
\end{aligned}
$$

Exercise 1.1. Demonstrate that the *area* of a circle as a function of its radius does not satisfy properties 1 and 2.

Exercise 1.2. Can you think of any shape in the plane whose area (as a function of scaling) satisfies properties 1 and 2? If yes, which? If no, why not?

Exercise 1.3. Can you think of any shape in the 3-dimensional space whose *volume* (as a function of scaling) satisfies properties 1 and 2? If yes, which? If no, why not?

Classification of functions satisfying Properties 1 and 2

So far, we have seen that some functions satisfy properties 1 and 2, and others do not.

Table 1.1: Functions that do and don't satisfy properties 1 and 2

do	don't
circumference of a circle as a function of radius	area of a circle as a function of radius
circumference of any shape in the plane as a function of scale	area of any shape in the plane as a function of scale
	volume of any shape in 3-dimensions as a function of scale

Why does scaling satisfy properties 1 & 2 for any shape, not just circles?

We asserted above that property 1 is not just satisfied by circles (when you scale the radius) but is satisfied by all curves. Why is this the case?

What do we mean by scaling a figure in the plane by a factor of 2? Well, a reasonable answer is to say that a figure is a set of points and each point has an x and y coordinate. For example, the circle of radius r is the set of points with coordinates given by

$$
\begin{aligned}
x =& r \cos\theta \\
y =& r \sin\theta
\end{aligned}
$$

where $\theta \in [0, 2\pi]$.

Notation

We introduced a symbol above: $\in$. This literally means "in" or "is in" depending on context.
For example, $\pi \in \mathbb{R}$ means "pi is in the real numbers".
Sometimes we will want to list the set first, like $\mathbb{R} \ni \pi$. In this case we read it as "the real numbers contain pi."

Notation

When we want to refer to a set, we will often use notation like the following:

$$\left\{ (x, y) \in \mathbb{R}^2 \middle| \begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \quad , \theta \in [0, 2\pi] \right\}.$$

In general, this notation has the form

$$B = \left\{ f(x_1, x_2, \ldots) \in A \middle| \begin{array}{l} \text{constraints involving} \\ x_1, x_2, \ldots \end{array} \right\}.$$

The process for constructing the set B is the following:

1. Find all the x_i that satisfy the constraints to the right of the $|$ symbol.
2. Plug all the x_i you found in the previous step into the function f.
3. The function f produces things in A, and so the set of all the things you produced in the last step is a collection of some (but not necessarily all) of the things in A.
   This is the set B.

Consider the point $(4, 2)$ in the plane. We can think of this point as the "sum" of its x and y coordinates:

$$(4, 2) = (4, 0) + (0, 2).$$

To scale this point by a factor of 2, it seems reasonable to multiply both coordinates by 2:

$$S_2((4, 2)) = (8, 4).$$

Notice that the function $S_2 : \mathbb{R}^2 \to \mathbb{R}^2$ "multiply coordinates by 2" has properties 1 and 2.

Exercise 1.4. Verify this.

Exercise 1.5. Check that applying S_2 to a circle of radius r produces a circle of radius $2r$.

Now, for any curve $\gamma : [a, b] \to \mathbb{R}^2$ given by

$$\gamma(t) = (x(t), y(t)),$$

its length can be computed by

$$\int_a^b \sqrt{x'(t)^2 + y'(t)^2} \quad dt.$$

(You may have seen this in multivariable calculus or physics.) Consider the composition

$$[a, b] \xrightarrow{\gamma} \mathbb{R}^2 \xrightarrow{S_2} \mathbb{R}^2$$

The formula is

$$(S_2 \circ \gamma)(x) = (2x(t), 2y(t))$$

and the length is given by

$$\begin{aligned}
&\int_a^b \sqrt{\left([2x(t)]'\right)^2 + \left([2y(t)]'\right)^2} \quad dt \\
=&\int_a^b \sqrt{\left(2x'(t)\right)^2 + \left(2y'(t)\right)^2} \quad dt \\
=&\int_a^b \sqrt{4\left(x'(t)\right)^2 + 4\left(y'(t)\right)^2} \quad dt \\
=&\int_a^b 2\sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \quad dt \\
=2&\int_a^b \sqrt{\left(x'(t)\right)^2 + \left(y'(t)\right)^2} \quad dt
\end{aligned}$$

Notice that what we obtain on the last line is exactly twice the length of the curve γ. Convince yourself that there is nothing special about the number 2 here; if we had replaced S_2 by S_{17}, then we would have obtained 17 times the length of γ in the last line.

More functions which break up over "sums"

Differentiation

Now consider how we compute the derivative of a function like the following:

$$\begin{aligned}\frac{d}{dx}\left[2x^2+x\right] &= \frac{d}{dx}\left[2x^2\right]+\frac{d}{dx}\left[x\right]\\ &= 2\frac{d}{dx}\left[x^2\right]+\frac{d}{dx}\left[x\right]\\ &= 4x+1\end{aligned}$$

If we let $D : \text{FUNCTIONS} \to \text{FUNCTIONS}$ be the operation of taking a derivative, then in the first line we used

$$D[f_1(x)+f_2(x)] = D[f_1(x)] + D[f_2(x)]$$

and in the second line we used the fact that, when k is constant,

$$D[kf(x)] = kD[f(x)].$$

Thus, D (that is, differentiation) satisfies properties 1 and 2. (Although you should probably be uncomfortable that we wrote $D : \text{FUNCTIONS} \to \text{FUNCTIONS}$ above. What is the set FUNCTIONS? Are all functions differentiable? We will address this later. For now, it suffices to replace FUNCTIONS above with P_n, the set of all polynomials of degree at most n. In fact, it is the case that $D : P_n \to P_{n-1}$ if $n \geq 1$.)

Definite integration

Fix an interval $[a, b]$ and consider

$$I(f) := \int_a^b f(x) \quad dx.$$

Exercise 1.6. Check that I has properties 1 and 2.

Why are properties 1 and 2 useful?

Why is it useful that D satisfies properties 1 and 2? It allows us to compute derivatives of complicated expressions like $2x^2 + x$ if we only know the computation on some simple parts of the expression. Knowing the derivative of x^2 and x is all that is needed.

Similarly, if we know the circumference of a shape in the plane at one scale, we can compute its circumference at all scales using property 1.

Not all functions are linear, but if a function is linear, it is much easier to compute with.

1.2 Linear functions

Definition 1.1 (linear function)**.** A function satisfying properties 1 and 2 is called *linear*.

> Warning
>
> We use the term "linear" for these functions, but we also use the word "line" for graphs in the plane with formula $f(x) = mx + b$. This term is overloaded and means different things in these two contexts.

Exercise 1.7. Show that $f(x) = mx + b$ is only a linear function when $b = 0$.

Chapter 2

The notion of a Vector Space

If $L : A \to B$ is linear, what must be true about A and B?

Let's go back to the definition of a linear function. A function is linear if and only if it satisfies the following two properties:

$$\begin{aligned} &\text{Property 1:} \quad L(ax) = aL(x) \\ &\text{Property 2:} \quad L(x+y) = L(x) + L(y). \end{aligned}$$

Let's list a few things that must be true to arrive at these expressions:

- there are terms (like a in property 1) that we can factor through L
- there are terms (like x, y in properties 1 and 2) that we cannot factor through L
- there is some kind of addition on the xs and ys

There are perhaps some more properties that would be nice, and that are true about all the domains and codomains of linear functions we have seen so far:

- the as (that we can pull through L) have nice algebraic properties ($+, -, \times$, division)
- the $+$ operation on the xs and ys has some nice properties, too (existence of an identity, commutativity, etc)

Eventually, mathematicians (who were working with linear functions intuitively) worked out the minimal set of facts that one needs about the domain and codomain of a linear function for everything to be coherent. Here it is

2.1 Definition

Definition 2.1. A *vector space* is a set V of *vectors* and a set F of *scalars* that satisfy the following properties

- the set of scalars F is a field (see the next section for more information on what we mean by field),
- there is a function $+\colon V \times V \to V$, called *vector addition* that
  - is associative: $(X+Y)+Z = X+(Y+Z)$,
  - is commutative: $X+Y = Y+X$,
  - has a 0 (an additive identity) which means that 0 satisfies $0+X = X$ for all $X \in V$,
  - has negatives (additive inverses) which means that for each $X \in V$ there is an element $Y \in V$ such that $X+Y = 0$ (one can prove that there is only one inverse of X, and this is usually written $-X$),
- there is a function $\cdot\colon F \times V \to V$, called *scalar multiplication* such that
  - scalar multiplication associates with field multiplication (that is, $\forall a, b \in F, X \in V$ it's true that $(ab)X = a(bX)$)
  - the multiplicative unit in F is a unit for scalar multiplication (that is, $\forall X \in V$ we have $1X = X$)
- the following distributive laws hold:
  - $(a+b)X = aX + bX, \quad \forall a, b \in F, X \in V$
  - $a(X+Y) = aX + aY, \quad \forall a \in F, X, Y \in V$

> Notation
>
> Above, we used the symbol $\forall$. This is a symbol that mathematicians use that literally just means "for all".
> So, as an example $x \in P_3, \forall x \in P_2$ is read as "for all x in the set of second degree polynomials, x is in the set of third degree polynomials", which is just a wordy way of saying that all second degree polynomials are third degree polynomials.

Note that the notation $+\colon V \times V \to V$ implies that V is, in particular, closed under addition (since V is the codomain of $+$; addition cannot by definition return an element not in V). Similarly we can conclude that V must be closed under scalar multiplication.

All of the properties above hold (with the field of scalars F set to $\mathbb{R}$) for all the spaces we have used so far as the domain and codomain of a linear function. (You may want to check this for yourself quickly. Most of these facts obviously hold for polynomials, functions, points in $\mathbb{R}^2$, and $\mathbb{R}$ itself. However, one must still check all of them before using a set as the domain or codomain of a linear function.)

Not all sets of mathematical objects satisfy these properties, though. An example is the set of points on the sphere. It turns out that there is no way (although the proof will have to wait for a future chapter) to turn that set into a vector space.

The way you should think about this definition is the following:

- we are **allowed to use** a set A satisfying all of the properties in Definition 2.1 as the domain or codomain of a linear function, but
- a set A that fails to satisfy any of the properties above **can never be used** as the domain or codomain of a linear function. If you tried to use it in this way, you would eventually run into statements that make no sense.[1]

An example of the kind of problem you can run into is the following: Let $L : A \to B$ be linear and suppose A is a vector space as defined above but that B fails to have an additive identity (that is, there is no zero vector in B). We can compute

$$L(0) = 0L(0) = 0(?) = 0$$

where the ? stands for whatever $L(0)$ maps to. Notice that in the end, it doesn't matter because we conclude that $L(0) = 0 \in B$. But this is nonsensical since we assumed there was no zero vector in B.

Exercise 2.1. Check that $\mathbb{R}$ forms a vector space (the set of scalars is $\mathbb{R}$).

Exercise 2.2. Check that $\mathbb{R}^2$ is a vector space.

2.2 Scalars

In the above definition, we said that the scalars for a vector space must come from a field. What does this mean?

For the purposes of this text, our field will always be $\mathbb{R}$ or the field of complex numbers:

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

where the symbol i has the property that $i^2 = -1$ (and so we sometimes write it as $\sqrt{-1}$). These two sets are the set of scalars for most vector spaces found in applications in the wild.

Exercise 2.3. Check that $\mathbb{C}$ can be thought of as a vector space with field of scalars $\mathbb{R}$.

[1] It is the case that some maps between spaces that are weaker than vector spaces have defining properties 1 and 2. The spaces are a generalization of vector spaces called modules and the maps are called module homomorphisms. If you look at the definition of module homomorphism on wikipedia, you'll see the same two equations we used to define linear functions.

Definition 2.2. Let P_n be the set (which we mentioned informally in the previous chapter)

$$\{a_n x^n + a_{n-1} x^n + \cdots + a_1 x + a_0 \mid a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{R}\}.$$

This is called the *set of polynomials of degree n*.

Exercise 2.4. Is P_n a vector space for each $n \in \mathbb{N}$?

From the previous two examples, you might notice that the definitions

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

and

$$P_n := \{a_n x^n + a_{n-1} x^n + \cdots + a_1 x + a_0 \mid a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{R}\}$$

look similar, and they give you a hint about how to think of these as vector spaces; in both cases, they are constructed as the sum of things with coefficients in $\mathbb{R}$. We will investigate this further in the next chapter.

Fields

In case you are interested, the full definition of a field can be found on Wikipedia. A field that you already know about, but that is not $\mathbb{R}$ or $\mathbb{C}$ is the field of rational numbers:

$$\mathbb{Q} := \left\{ \frac{a}{b} \middle| a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

There are esoteric examples of fields, and also algebraic structures stranger than fields. If this is interesting to you, you might try to take a course in Abstract Algebra. I learned this topic from Liz Stanhope and Dummit and Foote (2004).

In computing, real numbers (which do form a field) are often represented by 32 or 64 bit floating point numbers. It is perhaps interesting to know that floating point numbers (in any number of bits) (which are encountered **often** in the wild) do not form a field because addition of floating point numbers is not associative.

2.3 Subspaces

We have seen in this chapter that if you want to do something with a linear function, you better check first that its domain and codomain are vector spaces. It's quite a lengthy process to verify all the properties in Definition 2.1, even if in many cases it is not difficult to verify each one individually. We are in need of machines that allow us to know that something is a vector space more easily. That is what this and the next section are about.

How do we know that a set $W \subset V$ is a vector space?

> **Notation**
>
> Above, we used the symbol $\subset$. This just means "is a sub set of", for example $\mathbb{Z} \subset \mathbb{Q}$ means "the set of integers is a subset of the set of rational numbers."
> When we say "A is a subset of B" or, equivalently write $A \subset B$, we mean that everything in A is also in B.
> We will sometimes write $A \supset B$ to mean that B is a subset of A or that "A has B as a subset."

We've already seen some examples of this. For example, we know that differentiable functions form a vector space, and also the set P_n forms a vector space. Since every polynomial is differentiable, it follows that $P_n \subset \{\text{differentiable functions}\}$. So P_n is an example of a subset of a vector space that turns out to be a vector space itself.

Let's introduce some terminology for this situation.

Definition 2.3. A subset W of a vector space V is called a (vector) *subspace* if W is itself a vector space (with field of scalars F, addition, multiplication, zero vector inherited from V).

We have already seen an example of a subspace above. A nonexample is the set of points in the plane at distance 1 from the origin. This set doesn't satisfy the closure assumptions of Definition 2.1.

It is a bit remarkable (although not difficult to prove) that we actually only need to check the closure assumptions. The following theorem encodes this fact.

Theorem 2.1. *A subset W of a vector space V is itself a vector space (with field of scalars F, addition, multiplication, zero vector inherited from V) iff the following two properties hold:*

1. *W is closed under scalar multiplication: $aX \in W \quad \forall a \in F, X \in W$*
2. *W is closed under vector addition: $X + Y \in W \quad \forall X, Y \in W$*

We won't do the proof here (you can go to e.g. Hefferon (2022) to see it), but we will sketch it.

Since the scalars, and the operations $\cdot, +$ are inherited from the ambient space V, they are known to satisfy associativity, commutativity, existence of inverses and zeros, etc (because we had to prove those things when we showed that V was a vector space). However, in Definition 2.1, we wrote $+ \colon V \times V \to V$. That is, $+$ takes two vectors in V and returns a vector in V. Since $W \subset V$, we know that we can use elements of W as inputs to $+$. That is, we know $+ \colon W \times W \to V$, but we cannot know addition of vectors always lands in the subset W. The first assumption above tells us this. Similarly for scalar multiplication.

For the proof in the opposite direction, notice that for W to be a vector space we must have $+\colon W \times W \to W$. If addition is not closed, then $+$ has the wrong codomain, and W isn't a vector space. Similarly if W is not closed under scalar multiplication.

Easier subspaces

Notice that we can use Theorem 2.1 to quite easily prove (assuming we know differentiable functions form a vector space) that P_n is a vector space. We can also conclude easily that, for example,

$$\{a \sin x + b \cos x \mid a, b \in \mathbb{R}\}$$

is a vector space.

Exercise 2.5. Is the set of arrows that are either fully horizontal or fully vertical a vector subspace of the vector space of arrows in 2D? Why?

Exercise 2.6. Consider the set of monomials

$$\{1, x, x^2, x^3, ...\}.$$

If I pick any two of these (say x^k and x^l with $k \neq l$, eg x^{17} and x^{2025}) and form the set

$$\{ax^k + bx^l \mid a, b \in \mathbb{R}\},$$

is this a vector subspace of the space of differentiable functions? Why?

2.4 The space of lists of length n

We have seen that the following are vector spaces: $\mathbb{R}$ over $\mathbb{R}$, $\mathbb{R}^2$ over $\mathbb{R}$, $\mathbb{C}$ over $\mathbb{C}$.

There is a very powerful theorem that we can prove to generate many vector spaces that are of a similar form.

Theorem 2.2. *Let F be any field. Define the following set*

$$F^n := \{(a_1, a_2, \cdots, a_n) \mid a_i \in F\}.$$

This is the space of lists of length n, with members from F.

Let $+ : F^n \times F^n \to F^n$ *be defined by*

$$(a_1, a_2, \cdots, a_n) + (b_1, b_2, \cdots, b_n) = (a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n)$$

and let $\cdot : F \times F^n \to F^n$ *be defined by*

$$x(a_1, a_2, \cdots, a_n) = (xa_1, xa_2, \cdots, xa_n).$$

Then F^n *with these two operations (and field of scalars* F*) is a vector space.*

The proof is a straightforward check of all the items in Definition 2.1.

Note that $\mathbb{R}, \mathbb{R}^2, \mathbb{C}$ can be proved to be vector spaces using this theorem.

To develop your intuition for what this looks like, let's do a multiplication in $\mathbb{C}^4$:

$$\begin{aligned} & i(1, i, 1 + i, 2 - 3i) \\ =& (i, i^2, i + i^2, 2i - 3i^2) \\ =& (i, -1, i - 1, 2i + 3). \end{aligned}$$

Note that this theorem only works if F^n is a vector space over F. So, above, $\mathbb{C}^4$ is a vector space with scalars $\mathbb{C}$.

Exercise 2.7. Is $\mathbb{R}^n$ a vector subspace of $\mathbb{C}^n$? Think very carefully about Definition 2.3.

Chapter 3

Bases

Remember that the derivative function $D(f) := \frac{df}{dx}$ allows us to compute the derivative of, for example

$$f(x) = x^2 + 4\sin x$$

if we know how to compute D on only x^2 and $\sin x$. In fact, knowing just how to compute D on these two functions and knowing that D is linear allows us to compute the derivative of any function of the form

$$x \mapsto ax^2 + b\sin x, \quad a, b \in \mathbb{R}.$$

Notation

We just introduced the notation $\mapsto$. This is a mathematician's way of referring to a function without giving it a name.
As an example both of

$$x \mapsto x^2 \quad \text{and} \quad f(x) = x^2$$

are the same function. On the right, we choose to give the function a name "f". On the left, we describe the function by saying how inputs are transformed into outputs, but we don't give it a name. The function on the left is just "the function that squares its input". It doesn't have a name, so this is called an *anonymous function.*
Mathematicians use an anonymous function when they won't want to refer to a function again later, and so have no need of giving it a name. If you'll want to refer to a function later, it's useful to give it a name when you define it, so you'll write $f(x) = x^2$ (or use some other letter if f is already

used for something else).
This pattern also shows up in computer science, although they often call anonymous functions *lambdas* and instead of writing $x \mapsto x^2$ they will write `x.x^2` or `fun x => x^2` or some variation thereof.

From a very small amount of information, we actually know a lot about D. Thus, we begin this chapter straightforwardly with a question, which is the title of the next section:

3.1 What is the minimal amount of information needed to unambiguously describe a linear function?

Consider the following example:

Assume that $L : P_2 \to P_0$ is linear and that

$$L(x^2 + 3x) = 4.$$

Can we determine L on every element of P_2? That is, can we compute

$$L(ax^2 + bx + c)?$$

Let's try our best. We would like to isolate a term that looks like x^2+3x because that is something we have information about.

$$\begin{aligned} L(ax^2 + bx + c) =& L(ax^2 + bx) + L(c) && \text{by property 1} \\ =& L\left(a\left[x^2 + \frac{b}{a}x\right]\right) + L(c) && \text{assuming } a \neq 0 \\ =& aL\left(x^2 + \frac{b}{a}x\right) + L(c) && \text{by property 2} \end{aligned}$$

then we are basically stuck. Although we can force x^2 to show up, we can't *at the same time* force $3x$ to show up; forcing a coefficient of 1 on x^2 will always effect the coefficient of x. Furthermore, we have no information about how to deal with $L(c)$.

So, in this case, knowing that L is linear and knowing its value on *one* input is insufficient.

Exercise 3.1. In the example above, is it enough to know that

$$\begin{aligned} L(x^2) =& 1, \\ L(x) =& 1, \end{aligned}$$

to be able to determine L on all of P_2?

What if we know

$$\begin{aligned} L(x^2) =& 1, \\ L(x) =& 1, \quad ? \\ L(1) =& -2 \end{aligned}$$

3.2 Writing vectors relative to a set of vectors

Let $X, Y \in \mathbb{R}^2$ be given by $X := (1,1)$ and $Y := (1,-1)$. Suppose $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and also that we know

$$\begin{aligned} L(X) =& 1, \\ L(Y) =& -2 \end{aligned}.$$

Can we compute L on the vector $(2,3)$? If the answer to this question were yes then there would have to be a way to express $(2,3)$ in terms of X and Y (since those are the only two values we know anything about). That is, there would be $a, b \in \mathbb{R}$ such that

$$(2,3) = aX + bY = a(1,1) + b(1,-1) = (a,a) + (b,-b) = (a+b, a-b).$$

So, L can be computed on $(2,3)$ if and only if the following system has a solution:

$$\begin{aligned} 2 =& a + b, \\ 3 =& a - b \end{aligned}.$$

We can solve this system (solving for, say, a in the first and substituting the value in the second). We find that there is only one solution: $a = \frac{5}{2}, b = -\frac{1}{2}$.

Using this information, how do we compute $L(2,3)$? Well, now we know that

$$(2,3) = \frac{5}{2}(1,1) + \left(-\frac{1}{2}\right)(1,-1)$$

so let's just apply the function L to both sides of this equation:

$$
\begin{aligned}
L(2,3) =& L\left[\frac{5}{2}(1,1)+\left(-\frac{1}{2}\right)(1,-1)\right] \\
=& L\left[\frac{5}{2}(1,1)\right]+L\left[\left(-\frac{1}{2}\right)(1,-1)\right] \\
=& \frac{5}{2}L(1,1)+\left(-\frac{1}{2}\right)L(1,-1) \\
=& \frac{5}{2}1+\left(-\frac{1}{2}\right)(-2) \\
=& \frac{5}{2}+1 \\
=& \frac{7}{2}
\end{aligned}
$$

Exercise 3.2. For each =-symbol in the computation above, write the assumption, property, or rule that tells us we are allowed to conclude the left hand side is equal to the right hand side.

The above computation shows that the answer to our question is yes, this amount of information about L is sufficient to compute $L(2,3)$. Notice that this is the case even though we were not given the formula for L.

You may want to convince yourself that the vector $(2,3)$ above was not special; given the information we have about L, we can compute $L(k,l)$ for any $(k,l) \in \mathbb{R}^2$.

Exercise 3.3. If we know that $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and that

$$L(0,1) = 1, \quad L(0,2) = 2,$$

can we compute $L(1,2)$? If yes, compute it. If no, why not?

Exercise 3.4. If we know that $L : \mathbb{R}^2 \to \mathbb{R}$ is linear and that

$$L(0,1) = 2, \quad L(1,1) = 2,$$

can we compute $L(1,2)$? If yes, compute it. If no, why not?

Linear combinations

The trick to computing $L(2,3)$ in the previous section was to rewrite $(2,3)$ in the form

$$(2,3) = a(1,1) + b(1,-1).$$

This is a good ansatz[1] because we are able to split over the + and pull the a and b out (using the fact that L is linear). The following definition captures this pattern in general.

Definition 3.1. If $X_1, X_2, \dots, X_n$ is a collection of vectors in the vector space V, then any expression of the form

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n, \quad a_i \in F$$

is called a *linear combination* of $X_1, X_2, \dots, X_n$.

For some of the following exercises, we need to know about the following definition:

Definition 3.2. We use $C^k(\mathbb{R}, \mathbb{R})$ to denote the set of functions $\mathbb{R} \to \mathbb{R}$ such that, $\forall f \in C^k(\mathbb{R}, \mathbb{R})$

- f has at least k derivatives and
- all of f's k derivatives are continuous.

Sometimes we write just C^k rather than $C^k(\mathbb{R}, \mathbb{R})$ if it is implicit from the context that we are talking about functions $\mathbb{R} \to \mathbb{R}$.

Notice, in particular, that C^0 is just the set of continuous functions $\mathbb{R} \to \mathbb{R}$. C^1 is the set of functions with at least one derivative which is continuous.

C^∞ is used to denote the set of functions $\mathbb{R} \to \mathbb{R}$ that have infinitely many derivatives, all of which are continuous (many "nice" familiar functions are in C^∞, for example $x \mapsto \sin x$, $x \mapsto e^x$ and any polynomial).

Lemma 3.1. *For any k, the set $C^k(\mathbb{R}, \mathbb{R})$ is a vector space over $\mathbb{R}$.*

The proof of this lemma follows from the fact that $\mathbb{R}$ is a vector space over $\mathbb{R}$, and when we do operations on functions, we really do operations inside the codomain of the function.

Exercise 3.5. Is

$$\pi \sin x - 100x^3$$

a linear combination in C^0? If no, why not? If yes, what is a set of vectors in C^0 that this is a linear combination of?

[1]This is a German word that, in math, means a guess with some free parameters (in this case, a and b) in it. We compute with this guess, and then later find values of the parameters that make the computation valid and also solve our problem. If we cannot find such values, the we have essentially proved that there is no solution to our problem with the form given in the ansatz (although there may be a solution of another form if our ansatz is not general enough).

Exercise 3.6. Is

$$7\cos x\sin x$$

a linear combination in C^0? If no, why not? If yes, what is a set of vectors in C^0 that this is a linear combination of?

We then have the following generalization of the computation we did in the example above:

Theorem 3.1. *Let $L : A \to B$ be linear. These two statements are equivalent:*

- *We know the value of L on a set of vectors $X_1, X_2, \dots, X_n \in A$. That is,*

$$\begin{aligned} L(X_1) =& Y_1 \\ L(X_2) =& Y_2 \\ & \vdots \\ L(X_n) =& Y_n \end{aligned}$$

  for some $Y_i \in B$.

- *We know the value of L on any linear combination of $X_1, X_2, \dots, X_n$.*

The proof $\implies$ is a straightforward application of properties 1 and 2 of linear functions. You might try to generate the justification yourself. The proof of $\impliedby$ is an application of proof by contradiction.

> **Note**
>
> Theorem 3.1 is not a tautology!
> The set $\{X_1, X_2, \dots, X_n\}$ is a finite set (think of $\{(1,1),(1,-1)\}$ which contains only 2 elements).
> The set of linear combinations of $\{X_1, X_2, \dots, X_n\}$ is *infinite* (if the field of scalars is infinite, which in the case of $\mathbb{R}$ or $\mathbb{C}$ it is).
> The set of linear combinations of $\{(1,1),(1,-1)\}$ contains, for example,
>
> $$\begin{aligned} &(1,1),(2,2),(a,a)\forall a \in \mathbb{R}, \\ &(\pi, e), \left(\sqrt{2}, 2025^{1/3}\right), \\ &a(1,1)+b(1,-1) = (a+b, a-b)\forall a, b \in \mathbb{R}. \end{aligned}$$
>
> In particular, this set is infinite.

The span of a set of vectors

We will use the words "is a linear combination of" or "can be expressed as a linear combination of" quite often, so we introduce some terminology to shorthand this.

Definition 3.3. If V is a vector space over a field F, $X_1, \dots, X_n \in V$, and W is expressible as a linear combination of $\{X_1, \dots, X_n\}$ we say that W is in the *span* of $\{X_1, \dots, X_n\}$. We can write this

$$W \in \operatorname{Span}\{X_1, \dots, X_n\}.$$

We also refer to the following set as *the span* of $\{X_1, \dots, X_n\}$:

$$\operatorname{Span}\{X_1, \dots, X_n\} := \{a_1 X_1 + \cdots + a_n X_n \mid a_i \in F\}.$$

So, for example, $(0, 2025) \in \operatorname{Span}\{(0,1), (0,6)\}$, but the vector $(17, 9) \notin \operatorname{Span}\{(0,1), (0,6)\}$.

Exercise 3.7. Let $(a, b) \in \mathbb{R}^2$. Is $(a, b) \in \operatorname{Span}\{(1,1), (1,-1)\}$? Why or why not?

Exercise 3.8. Let $(a, b) \in \mathbb{R}^2$. Is $(a, b) \in \operatorname{Span}\{(1,2), (1,5), (1,-3)\}$? Why or why not?

Given the above definition, we can rephrase Theorem 3.1:

Theorem 3.2. *Let $L : A \to B$ be linear. These two statements are equivalent:*

- *We know the value of L on a set of vectors $X_1, X_2, \dots, X_n \in A$.*
- *We know the value of L on $\operatorname{Span}\{X_1, X_2, \cdots, X_n\}$.*

There isn't really anything to prove here; this is exactly Theorem 3.1 except with "any linear combination of" replaced by Span, which is Definition 3.3.

Exercise 3.9. Is $\operatorname{Span}\{1 + x, x + x^2\} = P_2$? Why or why not?

Exercise 3.10. Show that $\operatorname{Span}\{1 + x, x + x^2, x^2, x^3\} = P_3$ by finding a representation of

$$ax^3 + bx^2 + cx + d$$

(which is the form of a generic element of P_3) as a linear combination of $\{1 + x, x + x^2, x^2, x^3\}$.

Exercise 3.11. Suppose we know that $F'(x) = f(x)$ and $G'(x) = g(x)$ (that is, F, G have derivatives, and we give the names f, g to those derivatives). Recall the definition of the operator D for differentiation:

$$D(h) := h'(x),$$

and that we already know $D : \{\text{differentiable functions}\} \to \{\text{functions}\}$ is linear. (In view of the notation introduced in Definition 3.2, we can now write this as $D : C^1 \to C^0$.)

Is it true that

$$D\left[F(x) + 14G(x)\right] = f(x) + 14g(x)?$$

Did you use any properties about D *other than* linearity to conclude that?

Suppose $G(x) > 0, \forall x$. Is it true that

$$D\left[\frac{F(x)}{G(x)}\right] = \frac{f(x)G(x) - F(x)g(x)}{[G(x)]^2}?$$

Did you use any properties about D *other than* linearity to conclude that?

Linear independence

Suppose we have a collection of vectors $\{X_1, \ldots, X_n\} \in \mathbb{R}^2$. We have seen examples where

1. $\{X_1, \ldots, X_n\}$ consists of two vectors and $\text{Span}\{X_1, \ldots, X_n\}$ **is not all of** $\mathbb{R}^2$

   (this happened when $\{X_1, \ldots, X_n\} = \{(0,1), (0,2)\}$),

2. $\{X_1, \ldots, X_n\}$ consists of two vectors and $\text{Span}\{X_1, \ldots, X_n\}$ **is equal to** $\mathbb{R}^2$

   (this happened when $\{X_1, \ldots, X_n\} = \{(1,1), (1,-1)\}$),

3. $\{X_1, \ldots, X_n\}$ consists of three vectors and $\text{Span}\{X_1, \ldots, X_n\}$ **is equal to** $\mathbb{R}^2$

   (this happened when $\{X_1, \ldots, X_n\} = \{(1,2), (1,6), (1,-3)\}$).

In the last case, we got something extra; there were infinitely many ways of representing an arbitrary element $(a, b) \in \mathbb{R}^2$ as a linear combination of $\{(1,2), (1,6), (1,-3)\}$. In the second case, we get a unique way of representing each vector as a linear combination, and in the first case we don't even get any way to represent some vectors as a linear combination.

From Definition 3.3 we now can say that the distinction between the first case and the other two is that

$$\text{Span}\{(1,1), (1,-1)\} = \text{Span}\{(1,2), (1,6), (1,-3)\} = \mathbb{R}^2$$

but that

$$\text{Span}\{(0,1), (0,2)\} \neq \mathbb{R}^2.$$

From Theorem 3.2, we know that if we know a linear function on the two vectors

$$(1,1), (1,-1)$$

or on the three vectors

$$(1,2),(1,6),(1,-3)$$

then we know its value on all inputs $(a,b)\in\mathbb{R}^2$.

What is the criterion that distinguishes the second case from the last one? That is, is there a property we can measure about the set of vectors

$$\{(1,2),(1,6),(1,-3)\}$$

that is different when we measure it about

$$\{(1,1),(1,-1)\}$$

that tells is that we get a unique linear combination in the latter case, but a nonunique linear combination in the former case? It turns out to be the following definition, and we will see why in a later part of the book.

Definition 3.4. Let V be a vector space and $\{X_1,\dots,X_n\}\subset V$. We say that the set $\{X_1,\dots,X_n\}$ is *linearly independent* if, for each $i\in\{1,\dots,n\}$,

$$X_i\notin \operatorname{Span}\{X_1,X_2,\cdots,X_{i-1},X_{i+1},\cdots,X_n\}.$$

This is another way of saying that X_i is **not** in the span of all the $\{X_j\}$ with X_i removed.

Exercise 3.12. Is the set $\{(1,0),(0,1)\}\subset\mathbb{R}^2$ linearly independent? Explain why.

Exercise 3.13. Is the set $\{(2,1),(2,-1)\}\subset\mathbb{R}^2$ linearly independent? Explain why.

Exercise 3.14. Is the set $\{(1,2),(3,4),(-1,4)\}\subset\mathbb{R}^2$ linearly independent? Explain why.

Although we chose to define linear independence using Definition 3.4, there are equivalent statements that we could have used. We summarize them in the following theorem.

Theorem 3.3. *Let V be a vector space and $\{X_1,\dots,X_n\}\subset V$. The following statements are equivalent:*

1. *$\{X_1,\dots,X_n\}$ is linearly independent in the sense of Definition 3.4.*
2. *$0=a_1X_1+\cdots+a_nX_n$ has only one solution (when $a_i=0\,\forall i$).*
3. *For every $v\in\operatorname{Span}\{X_1,\dots,X_n\}$, the equation $v=a_1X_1+\cdots+a_nX_n$ has **exactly** one solution.*

When we say "the following statements are equivalent", we mean that among (1.), (2.), and (3.) above, there are proofs

$$(1.) \Leftrightarrow (2.) \Leftrightarrow (3.).$$

So, we chose to use Definition 3.4 as the definition, and (in the proof of Theorem 3.3) would prove that (2.) is equivalent to Definition 3.4. But we could have just as well used (2.) as the definition, and proved the statement in Definition 3.4 as a theorem. If you look at different books on this subject, you will see different approaches; all of them are equally correct.

We omit the proof of Theorem 3.3; the equivalences are not difficult to prove.

From this, we get the following lemma quite easily:

Lemma 3.2. *Let V be a vector space and $S = \{X_1, \dots, X_n\} \subset V$.*

- *If S contains the zero vector, it is linearly dependent.*
- *If S contains the same vector twice, it is linearly dependent.*

Note that this lemma **is not an iff** statement. For example, not every linearly dependent set contains the zero vector. But if a set does contain the zero vector, this lemma tells you that that set is linearly dependent.

Necessity and sufficiency

We are finally in a position to answer the question that began this chapter:

> What is the minimal amount of information needed to unambiguously describe a linear function?

Note that there are two parts of this statement: how much information suffices to define a linear function, and how much information minimally suffices. These criteria are different.

For example, in Chapter 1 we saw that the formula

$$C(r) := 2\pi r$$

is enough to know C on all its inputs (after all, this formula tells you how to compute C for any input you could pick). So this formula **suffices** to define the linear function C.

However, we saw later in that chapter that, once we know C is linear, this is more information than we need. It is enough to know far less information. For example, suppose I tell you that, in some other universe there are creatures called pumans who have a shape they call a pircle whose pircumference (as they call it) is a function of some measurement they refer to as the pradius of the pircle. Both pradius and pircumference are nonnegative real numbers, thankfully.

I don't know the formula for pircumference as a function of pradius (communication with the pumans is very slow and difficult), but I have learned from them two things:

- pircumference as a function of pradius is linear
- the pircumference of a pircle of pradius 2025 is 45.

It turns out that this is enough information to discover the formula for pircumference! Let $k \geq 0$ be any arbitrary pradius, and call the pircumference function P. We can compute

$$P(k) = P\left(\frac{k}{2025}2025\right) = \frac{k}{2025}P\,(2025) = \frac{k}{2025}45 = \frac{k}{45}.$$

Verify for yourself that you understand why each equality in this computation is true. For the second equality, we used our knowledge that pircumference is linear and for the third equality we used our knowledge that $P(2025) = 45$. If we had been missing either of those facts, we could not conclusively determine the formula for P. That means that, once we know P is linear, the knowledge of P on at least one nonzero input is **necessary**.

The following theorem gives two criteria addressing our motivating question at the start of this section: the first criteron describes when we have sufficient information to determine a linear function, the second describes when we have only the necessary (minimal) amount of information.

Theorem 3.4. *Let $L : A \to B$ be linear. Suppose we know the value of L on a set of vectors $\{V_1, \dots, V_n\}$.*

- *If* $\mathrm{Span}\{V_1, \dots V_n\} = A$, *we can compute the value of L on all inputs.*
- *If $\{V_1, \dots, V_n\}$ is linearly independent, then $\{V_1, \dots, V_n\}$ is a minimal set of information defining L on* $\mathrm{Span}\{V_1, \dots V_n\}$.

Note that, if $\mathrm{Span}\{V_1, \dots V_n\} = A$ and $\{V_1, \dots, V_n\}$ is linearly independent, this theorem implies that $\{V_1, \dots, V_n\}$ is a minimal choice of information defining L everywhere on A.

This theorem is very powerful. For example, consider a linear function $F : \mathbb{R}^3 \to \mathbb{R}^{2025}$. Theorem 3.4 tells us that if we know only three facts, e.g. the values of $L(1, 0, 0)$, $L(0, 1, 0)$, and $L(0, 0, 1)$, then we are able to use this information to find the value of L on **any** of the infinitely many vectors in $\mathbb{R}^3$. Of course, we would first have to show that $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ spans $\mathbb{R}^3$.

The situation of having a set of vectors that both spans an entire vector space and is linearly independent is so common, we give it the following name:

A basis for a vector space

Definition 3.5. Let V be a vector space and let B be a finite, ordered list of vectors in V. If

- Span $B = V$ and
- B is linearly independent, we say that B is *a basis for* V.

We write list here because order and repetition matter, and these are not qualities of sets.

Notation

We write lists like $[1, 2, 2, 3]$ with square brackets rather than curly ones.[2] Note that for lists, $[1, 1, 1] \neq [1]$ while for sets $\{1, 1, 1\} = \{1\}$, and also, for lists $[1, 2] \neq [2, 1]$ while for sets $\{1, 2\} = \{2, 1\}$.

Exercise 3.15. Is the list $[(1, 0), (0, 1)] \subset \mathbb{R}^2$ a basis? Explain why.

Exercise 3.16. Is the set $[(2, 1), (2, -1)] \subset \mathbb{R}^2$ a basis? Explain why.

Note

The preceding exercises demonstrate a very important point: if V has a basis, it is not unique. There are actually many bases for a vector space, and which one you should choose to use **depends on the problem at hand**.
For example, both $[(1, 0), (0, 1)]$ and $[(1, 1), (1, -1)]$ are bases for $\mathbb{R}^2$. If we are dealing with a linear function $L : \mathbb{R}^2 \to \mathbb{R}$ where we know $L(1, 1) = 5$ and $L(1, -1) = 3$, then it is not useful to work with the basis $[(1, 0), (0, 1)]$.

We now have a restatement of Theorem 3.4 that uses this definition:

Theorem 3.5. *Let $L : A \to B$ be linear. Suppose we know the value of L on a set of vectors $\{V_1, \dots, V_n\}$. If $[V_1, \dots, V_n]$ is a basis for A, then we know the value of L on all its inputs and $\{V_1, \dots, V_n\}$ is a minimal choice of data defining L.*

We finally answered the two mysteries that have been motivating us for a few chapters, so we're done, right?

[2] This is the notation convention used by many modern programming languages.

Chapter 4

Vector spaces isomorphic to Euclidean space

No, there are still more mysteries!

One example is the difference between P_1 and $\mathbb{R}^2$. We have the following ordered bases for these

$$[x, 1] \text{ for } P_1 \text{ and } [(1,0),(0,1)] \text{ for } \mathbb{R}^2.$$

When we have fixed a basis for a vector space, Definition 3.5 together with Theorem 3.3 and Definition 3.3 tell us that, for each vector in our vector space, there is a exactly one way to write that vector as a linear combination of the basis vectors. So, for example, the vector $x - 3 \in P_1$ can be thought of as $(1) \underbrace{(x)}_{\text{1st basis vector}} +(-3) \underbrace{(1)}_{\text{2nd basis vector}}$.

If you and I are thinking of the same basis $[x, 1]$ for P_1, then there is no confusion if we just list the coefficients of this vector relative to this basis. That is, we use the notation $(1, -3)$ to *refer* to the vector $x - 3 \in P_1$.

But $(1, -3)$ is a list of two real numbers, so isn't this an element in $\mathbb{R}^2$? Is P_1 the same as $\mathbb{R}^2$?

4.1 Dimension

One remarkable fact about P_1 is that, although $[x, 1]$ isn't the only basis for it (the list $[x + 1, x - 1]$, we have seen, is another), **all** the bases of P_1 have something in common: they have exactly two elements. This can be generalized to the following lemma:

Lemma 4.1. *Let V be a vector space with at least one finite basis and let B be that basis. Then all bases of V have the same number of elements as B.*[1]

This immediately tells us that all bases for P_1 have two elements, as well as all bases for $\mathbb{R}^2$.

Since all bases for a vector space are the same length we can refer to this length as a property of the vector space, rather than a property of a particular basis.

Definition 4.1. Let V be a vector space with a finite basis of size n. Then we say that the *dimension* of V is n and we write this like

$$\dim V = n.$$

So, using this notation, $\dim P_1 = 2 = \dim \mathbb{R}^2$.

Exercise 4.1. What is the dimension of P_2? What about P_3? What about P_n?

Exercise 4.2. What is the dimension of the vector space of arrows drawn on a piece of paper? What is the dimension of the vector space of arrows created in a space like a room? (You can think of bits of string of different lengths, with arrows on one end of them, held in space by two people. Addition and multiplication is as for arrows on paper.)

Exercise 4.3. What is the dimension of $\mathbb{C}$ as a vector space over itself?

Exercise 4.4. What is the dimension of $\mathbb{C}$ as a vector space over $\mathbb{R}$?

The following lemma says that if one vector space is inside another, then the dimension of the one on the inside can't be larger than the one on the outside:

Lemma 4.2. *Let $W \subset V$ be a vector subspace of the vector space V. Suppose W had finite dimension. Then*

$$\dim W \leq \dim V.$$

Exercise 4.5. The dimension of the space of continuous functions $\mathbb{R} \to \mathbb{R}$ is not finite. Use the result of Exercise 4.1 to explain why.

Exercise 4.6.

> Recall that we know that the function $D : C^1 \to C^0$ "take the derivative" is linear. Let K be the set of all vectors in C^1 which are taken to the 0 vector in C^0 by D. That is, K is the set of all differentiable functions whose first derivative is 0 everywhere. As an equation,
>
> $$K := \{ f : \mathbb{R} \to \mathbb{R} \,|\, D(f) = f'(x) = 0 \}.$$

[1] For readers familiar with the notion of cardinality, this lemma can be generalized to infinite bases where it states that any two bases have the same cardinality.

K is a vector subspace of C^1 (you might want to convince yourself that this is true using Theorem 2.1, but you don't have to for this exercise). What is $\dim K$?

4.2 The notion of isomorphism

We have identified one property that P_1 and $\mathbb{R}^2$ have in common: they have the same dimension. There is a more general notion of "sameness" that is true for these two vector spaces. We will introduce the word for it here, before explaining precisely what it means in the following section.

For the purposes of this book, V and W are *isomorphic* if they are "the same" once we fix some piece of data.

In mathematics more generally, if two objects are isomorphic, it means that they "behave identically" but might need an extra piece of information to actually identify one with the other. The full notion is extremely broad.[2]

4.3 Isomorphism to $\mathbb{R}^n$ (given a basis)

Recall Theorem 2.2, that $\mathbb{R}^n$ is the set of lists of length n, where the elements of the list are real numbers. Recall that $\mathbb{R}^n$ is a vector space, and notice that we can demonstrate at least one basis for $\mathbb{R}^n$:

$$\begin{gathered}(1,0,0,\cdots,0,0),\\(0,1,0,\cdots,0,0),\\\vdots\\(0,0,\cdots,0,1,0),\\(0,0,0,\cdots,0,1)\end{gathered}$$

Notice that this basis has exactly n elements.

Now, suppose we have another vector space V over $\mathbb{R}$ with a basis of n elements $[b_1,\dots,b_n]$. Given a vector $v \in V$, there is a unique list of n real numbers $v_1,\dots,v_n$ such that

$$v = v_1b_1 + v_2b_2 + \cdots + v_nb_n.$$

As an example, consider P_1 with basis $[x+1, x-1]$. The vector $12\pi x - 2025$ can only be written in one way relative to this basis:

$$12\pi x - 2025 = \left(\frac{12\pi - 2025}{2}\right)(x+1) + \left(\frac{12\pi + 2025}{2}\right)(x-1).$$

[2] https://en.wikipedia.org/wiki/Isomorphism

Once we choose a basis, every vector can be "identified" with its list of cooefi-cients relative to that basis. In the example above, we have

$$12\pi x - 2025 \quad '' ='' \quad \left(\frac{12\pi - 2025}{2}, \frac{12\pi + 2025}{2}\right).$$

We put the equals sign in quotes here because these two vectors are not identical (they're not even in the same vector space), but once we choose a basis, we can think of the left as the right.

Exercise 4.7.

> Choose a different basis for P_1 and write the coordinates of the vector $12\pi x - 2025$ relative to that basis.

Exercise 4.8.

> Choose any basis you want for P_2.
>
> The set $W := \operatorname{Span}\{(1,0,0,0),(0,0,1,0),(0,0,0,1),(1,0,1,0)\} \subset \mathbb{R}^4$ is a vector subspace (But you shouldn't believe me! You can check for yourself!). Choose a basis for this subspace.
>
> Using your choses bases, what vector is $(1,2,3) \in \mathbb{R}^3$ when interpreted as coordinates for a vector in P_3? How about when interpreted as coordinates for a vector in W?

To make this notion precise, we notice that there is a function (which depends on the basis we choose for P_1) which we'll name $C : P_1 \to \mathbb{R}^2$. It turns out that this function is a linear function between vector spaces.

In particular, adding in P_1 and converting to coordinates, adding, and converting back to a polynomial produce the same value:

$$(ax+b)+(cx+d) = (a+c)x+(b+d)$$

and

$$\underbrace{\left(\frac{a+b}{2}, \frac{a-b}{2}\right)}_{C(ax+b)} + \underbrace{\left(\frac{c+d}{2}, \frac{c-d}{2}\right)}_{C(cx+d)} = \left(\frac{a+b+c+d}{2}, \frac{a+c-(b+d)}{2}\right)$$

$$\begin{aligned} C^{-1}\left(\frac{a+b+c+d}{2}, \frac{a+c-(b+d)}{2}\right) &= \frac{a+b+c+d}{2}(x+1) + \frac{a+c-(b+d)}{2}(x-1) \\ &= (a+c)x+(b+d) \end{aligned}$$

where C^{-1} is the inverse function.

A similar property holds for scalar multiplication.

It is perhaps clear that these properties are not unique to P_1 and $\mathbb{R}^2$, nor to the basis we chose for P_1 in the above example. Generalizing this example, we arrive at the following theorem.

Theorem 4.1. *Let V be a finite dimensional vector space over the field F. Let $n = \dim V$, and let B be a choice of basis for V. Then the function $V \to F^n$ taking a vector to its list of coordinates relative to B is a linear function which is injective and surjective. (That is, the function is an isomorphism of vector spaces.)*

Exercise 4.9.

> Consider two bases $B_1 := [x^2, x, 1]$ and $B_2 := [x^2, x, -1]$ for P_2.
>
> Consider the vector $(17, 5, 2) \in \mathbb{R}^3$. What vector in P_2 does this represent relative to B_1? What about relative to B_2?

What this theorem says is that, once we choose a basis for P_1, we can treat it *the same as* $\mathbb{R}^2$. We can work with polynomials or coordinates in $\mathbb{R}^2$ and we will get the same results. The same thing is true for all vector spaces over $\mathbb{R}$ with dimension 2. This fact is generalized in the following corollary:

Corollary 4.1. *Let V and W be two vector spaces over the same field of scalars F. If $\dim V$ and $\dim W$ are finite and equal, then V and W are "the same" (isomorphic) once we choose bases for each of them.*

The proof of this fact is that, since they have the same dimension (call it n), then both V and W can be thought of as lists of coordinates in F^n.

4.4 Basis dependence

[TODO]

Chapter 5

Ways of writing vectors (relative to a basis)

[TODO]

5.1 Columns

[TODO]

5.2 Linear combinations

[TODO]

Summation notation

[TODO]

5.3 Other ways you might encounter

[TODO]

Rows

[TODO]

Einstein summation notation

[TODO]

bra-ket

[TODO]

Chapter 6

Ways of writing linear maps

[TODO]

6.1 Maps from 1-dimensional spaces

[TODO]

6.2 Maps to 1-dimensional spaces

[TODO]

6.3 Maps from 2-dimensions to 2-dimensions

[TODO]

6.4 Maps in finite dimensions

[TODO]

6.5 Maps involving infinite dimensional spaces

[TODO]

6.6 Composing maps

[TODO]

Noncommutativity

[TODO]

6.7 Adding maps

[TODO]

6.8 Matrices are sort of like numbers

[TODO]

Numbers are 1×1 matrices

[TODO]

References

Axler, Sheldon. 2024. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, Cham. https://doi.org/10.1007/978-3-031-41026-0.

Dummit, David S., and Richard M. Foote. 2004. *Abstract Algebra*. Third Ed. John Wiley & Sons, Inc., Hoboken, NJ.

Hefferon, Jim. 2022. *Linear Algebra*. Fourth Ed. https://hefferon.net/linearalgebra/.

Treil, Sergei. 2024. *Linear Algebra Done Wrong*. https://sites.google.com/a/brown.edu/sergei-treil-homepage/linear-algebra-done-wrong.