Linear Algebra W214

Bruce Bartlett
Afrikaans translation by Murray Heymann

February 26, 2019

Contents

Note for the student

You are about to meet linear algebra for the second time. In the first year, we focused on systems of linear equations, matrices, and their determinants. That was good and well, but the time has come for you to return to these topics from a more abstract, mathematical viewpoint.

You should not be scared by *abstraction*. It simply means getting rid of extraneous detail and limiting oneself exclusively to the most important features of a problem. This allows you to understand the problem better. There are less things to worry about! Moreover, if you encounter another problem, which is superficially different but shares the same important features as the original problem, then you could understand it in the same way. This is the power of abstraction.

When we study abstract mathematics, we use the language of *definitions*, *theorems* and *proofs*. Learning to think along these lines (developing abstract mathematical thinking) can be a daunting task at first. But keep trying. One day it will 'click' into place and you will realize it is all much more simple than you had first imagined.

You cannot read mathematics the way you read a novel. You need to have a *pencil and notepad* with you, and you need to actively *engage* with the material. For instance, if you encounter a definition, start by writing down the definition on your notepad. Just the act of writing it out can be therapeutic!

If you encounter a worked example, try to write out the example yourself. Perhaps the example is trying to show that A equals B. Start by asking yourself: Do I understand what 'A' actually means? Do I understand what 'B' actually means? Only then are you ready to consider the question of whether A is equal to B !

Good luck in this new phase of your mathematical training. Enjoy the ride!

Chapter 1

Abstract vector spaces

1.1 Introduction

1.1.1 Three different sets

We start by playing a game. Recall that in mathematics, a *set X* is just a collection of distinct objects. We call these objects the *elements* of X.

I am going to show you three different sets, and you need to tell me the properties that they all have in common.

The first set, A, is defined to be the set of all ordered pairs (x, y) where x and y are real numbers.

Let us pause for a second and translate this definition from English into mathematical symbols. The translation is:

$$A := \{(a_1,\, a_2) : a_1, a_2 \in \mathbb{R}\}. \tag{1.1}$$

The := stands for 'is defined to be'. The { and } symbols stand for 'the set of all'. The lone colon : stands for 'where' or 'such that'. The comma in between a and b stands for 'and'. The $\in$ stands for 'an element of'. And $\mathbb{R}$ stands for the set of all real numbers.

Well done — you are learning the language of mathematics!

An element of A is an arbitrary pair of real numbers $\mathbf{a} = (a_1,\, a_2)$. For instance, $(1,\, 2) \in A$ and $(3.891,\, e^{\pi})$ are elements of A. Note also that I am using a boldface $\mathbf{a}$ to refer to an element of A. This is so that we can distinguish $\mathbf{a}$ from its *components* a_1 and a_2, which are just ordinary numbers (not elements of A).

We can visualize an element $\mathbf{a}$ of A as a point in the Cartesian plane whose x-coordinate is a_1 and whose y-coordinate is a_2: tikzalign
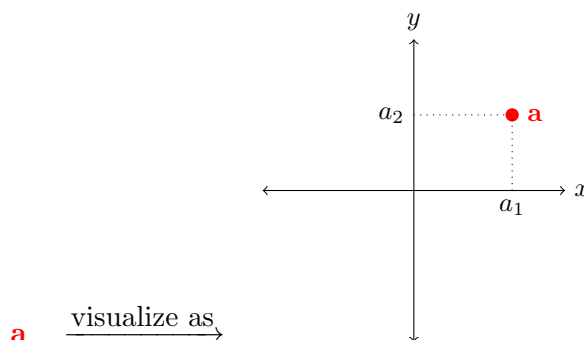
$$\mathbf{a} \xrightarrow{\text{ visualize as }}$$

The second set, B, is defined to be the set of all ordered real triples $(b_1,\, b_2,\, b_3)$ satisfying $b_1 - b_2 + b_3 = 0$. Translated into mathematical symbols,

$$B := \{(b_1,\, b_2,\, b_3) : b_1, b_2, b_3 \in \mathbb{R} \text{ and } b_1 - b_2 + b_3 = 0\}. \qquad (1.2)$$

For instance, $(2, 3, 1) \in B$ but $(1, 1, 1) \notin B$. We can visualize an element $\mathbf{b}$ of B as a point in the plane in 3-dimensional space carved out by the equation $x - y + z = 0$: tikzalign

$$\mathbf{b} \xrightarrow{\text{ visualize as }}$$
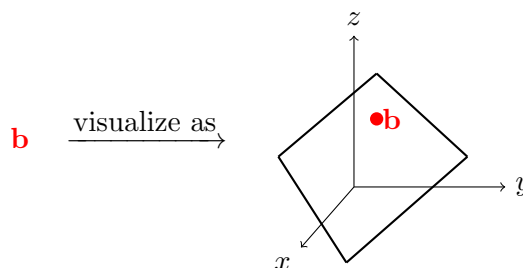
The third set, C, is the set of all polynomials of degree 4. Translated into mathematical symbols,

$$C := \{\text{polynomials of degree} \leq 4\}. \qquad (1.3)$$

Recall that the *degree* of a polynomial is the highest power of x which occurs. For instance, $\mathbf{c} = x^4 - 3x^3 + 2x^2$ is a polynomial of degree 4, and so is $\mathbf{p} = 2x^3 + \pi x$. So $\mathbf{c}$ and $\mathbf{p}$ are elements of C. But $\mathbf{r} = 8x^5 - 7$ and $\mathbf{s} = \sin(x)$ are not elements of C. We can visualize an element $\mathbf{c} \in C$ (i.e. a polynomial of degree 4) via its *graph*. For instance, the polynomial $\mathbf{c} = x^4 - 3x^3 + 2x^2 \in C$ can be visualized as: tikzalign
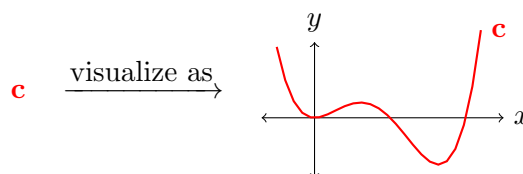
$$\mathbf{c} \xrightarrow{\text{ visualize as }}$$

There you have it. I have defined three different sets: A, B and C, and I have explained how to visualize the elements of each of these sets. On the face of it, the sets are quite different. Elements of A are arbitrary points in $\mathbb{R}^2$. Elements of B are points in $\mathbb{R}^3$ satisfying a certain equation. Elements of C are polynomials.

What features do these sets have in common?

1.1.2 Features the sets have in common

I want to focus on two features that the sets A, B and C have in common.
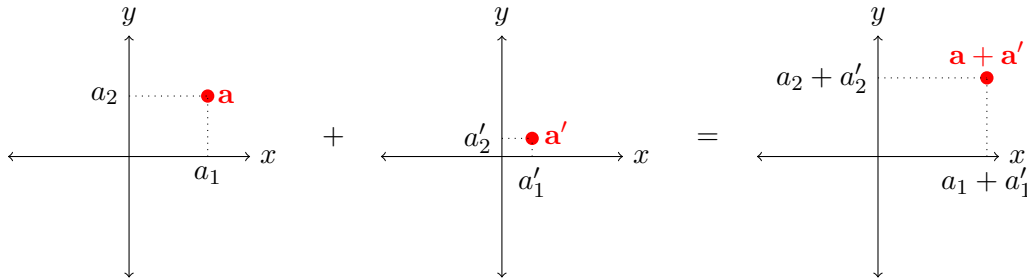
Addition

Firstly, in each of these sets, there is a natural *addition operation*. We can add two elements of the set to get a third element.

In set A, we can add two elements $\mathbf{a} = (a_1,\, a_2)$ and $\mathbf{a}' = (a_1',\, a_2')$ together by adding their components together, to form a new element $\mathbf{a} + \mathbf{a}' \in A$:

$$\underbrace{(a_1,\, a_2)}_{\mathbf{a}} + \underbrace{(a_1',\, a_2')}_{\mathbf{a}'} := \underbrace{(a_1 + a_1',\, a_2 + a_2')}_{\mathbf{a}+\mathbf{a}'} \tag{1.4}$$

For instance, $(1,\, 3) + (2,\, -1.6) = (3,\, 1.4)$. We can visualize this addition operation as follows: tikzalign

We can do a similar thing in set B. Suppose we have two elements of B, $\mathbf{b} = (b_1,\, b_2,\, b_3)$ and $\mathbf{b}' = (b_1',\, b_2',\, b_3')$. Note that, since $\mathbf{b} \in B$, its components satisfy $b_1 - b_2 + b_3 = 0$. Similarly the components of b' satisfy and $b_1 - b_2 + b_3 = 0$. We can add $\mathbf{b}$ and $\mathbf{b}'$ together to get a new element $\mathbf{b} + \mathbf{b}'$ of B, by adding their components together as before:

$$\underbrace{(b_1,\, b_2,\, b_3)}_{\mathbf{b}} + \underbrace{(b_1',\, b_2',\, b_3')}_{\mathbf{b}'} := \underbrace{(b_1 + b_1',\, b_2 + b_2',\, b_3 + b_3')}_{\mathbf{b}+\mathbf{b}'} \tag{1.5}$$

We should be careful here. How do we know that the expression on the right hand side is really an element of B? We need to check that it satisfies the

equation 'the first component minus the second component plus the third component equals zero'. Let us check that formally:

$$
\begin{aligned}
(\mathbf{b} + \mathbf{b}')_1 - (\mathbf{b} + \mathbf{b}')_2 + (\mathbf{b} + \mathbf{b}')_3 &= (b_1 + b_1') - (b_2 + b_2') + (b_3 + b_3') \\
&= (b_1 - b_2 + b_3) + (b_1' - b_2' + b_3') \\
&= 0 + 0 \\
&= 0.
\end{aligned}
$$

We can visualize this addition operation in B in the same way as we did for A.

There is also an addition operation in set C. We can add two polynomials together algebraically by adding their corresponding coefficients:

$$
\begin{aligned}
&[c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x^1 + c_0] + [d_4 x^4 + d_3 x^3 + d_2 x^2 + d_1 x^1 + d_0] \\
&:= (c_4 + d_4)x^4 + (c_3 + d_3)x^3 + (c_2 + d_2)x^2 + (c_1 + d_1)x^1 + (c_0 + d_0) \quad (1.6)
\end{aligned}
$$

For instance,

$$
[2x^4 + x^2 - 3x + 2] + [2x^3 - 7x^2 + x] = 2x^4 + 2x^3 - 6x^2 - 2x + 2.
$$

There is another way to think about the addition of polynomials. Each polynomial $\mathbf{c}$ can be thought of as a *function*, in the sense that we can substitute an arbitrary value of x into the polynomial $\mathbf{c}$, and it will output a number $\mathbf{c}(x)$. For instance, if $\mathbf{c}(x) = 3x^2 - 1$, then $\mathbf{c}(2) = 11$. If we think of polynomials as functions in this way, then the addition $\mathbf{c} + \mathbf{d}$ of two polynomials can be thought of as the new function which, when you substitute some number x into it, outputs $\mathbf{c}(x) + \mathbf{d}(x)$. Written mathematically,

$$
(\mathbf{c} + \mathbf{d})(x) := \mathbf{c}(x) + \mathbf{d}(x) \tag{1.7}
$$

Thinking in this way, we can visualize the graph of $\mathbf{c} + \mathbf{d}$ as the graph of $\mathbf{c}$ added to the graph of $\mathbf{d}$: tikzalign
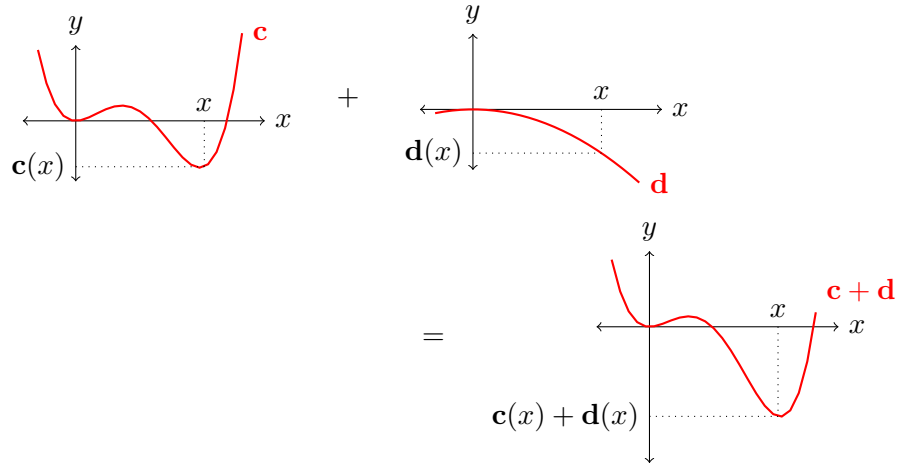
Zero element

In all three sets A, B and C, there is a specific element (the *zero element*) $\mathbf{0}$ which, when you add it to another element, leaves that element unchanged.

In A, the zero element $\mathbf{0}$ is defined by

$$\mathbf{0} := (0,0) \in A. \tag{1.8}$$

When you add this point to another point $(a_1, a_2) \in A$, nothing happens!

$$(0,\, 0) + (a_1,\, a_2) = (a_1,\, a_2).$$

Do not confuse the zero element $\mathbf{0} \in A$ with the real number zero, $0 \in \mathbb{R}$. This is another reason why I am using boldface! (You should use underline to distinguish them.)

In B, the zero element $\mathbf{0}$ is the point $(0,0,0) \in B$. When you add this point to another point $(u_1, u_2, u_3) \in B$, nothing happens!

$$(0,\, 0,\, 0) + (u_1,\, u_2,\, u_3) = (u_1,\, u_2,\, u_3).$$

In C, the zero element $\mathbf{0}$ is the *zero polynomial*. If we think algebraically, this is the degree polynomial whose coefficients are all zero:

$$\mathbf{0} = 0x^4 + 0x^3 + 0x^2 + 0x + 0 \tag{1.9}$$

If we think of the polynomial as a function, then the zero polynomial $\mathbf{0}$ is the function which returns zero for all values of x, that is $\mathbf{0}(x) = 0$ for all x. Whichever way we think of it, when we add the zero polynomial to another polynomial, nothing happens!

$$[0x^4 + 0x^3 + 0x^2 + 0x + 0] + [c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0]$$
$$= [c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0]$$

Multiplication by scalars

The last feature all the sets A, B and C have in common is that in each set, you can *multiply* elements of the set by real numbers.

For instance, if $\mathbf{a} = (a_1, a_2)$ is an element of A, then we can multiply it by some arbitrary real number, say 9, to get a new element $9 \cdot a$ of A. We do this multiplication component-wise:

$$9 \cdot (a_1,\, a_2) := (9a_1,\, 9a_2). \tag{1.10}$$

In general, if $k \in \mathbb{R}$ is an arbitrary real number, then we can multiply elements $\mathbf{a} \in A$ by k to get a new element $k \cdot \mathbf{a} \in A$ by multiplying each component of $\mathbf{a}$ by k:

$$\underbrace{k \cdot (a_1,\, a_2)}_{\text{Multiplying a vector by a scalar}} := (\; \underbrace{ka_1}_{\text{Multiplying two numbers together}} ,\, \underbrace{ka_2}{})$$

Just be careful to distinguish scalar multiplication $k \cdot \mathbf{a}$ (written with a $\cdot$) from ordinary multiplication of real numbers ka_1 (written with no symbol, just using juxtaposition). Later on, because we are lazy, we will stop writing the $\cdot$ explicitly — you have been warned!

Visually, this multiplcation operation *scales* $\mathbf{a}$ by a factor of k. That is why we call it *scalar multiplication*.

There is a similar scalar multiplication in B:

$$k(u_1, \, u_2, \, u_3) := (ku_1, \, ku_2, \, ku_3) \tag{1.11}$$

There is also a scalar multiplication operation in C. We simply multiply each coefficient of a polynomial $\mathbf{c} \in C$ by k:

$$k \cdot [c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x + c_0] = kc_0 x^4 + kc_3 x^3 + kc_2 x^2 + kc_1 x + kc_0 \tag{1.12}$$

If we think of the polynomial $\mathbf{c}$ as a function, then this corresponds to *scaling* the graph of the function vertically by a factor of k.

1.1.3   Features that the sets do *not* have

Let us mention a few features that the sets do *not* have, or at least do not have in common.

- Set $A = \mathbb{R}^2$ has a *multiplication operation*. Because we can think of $\mathbb{R}^2$ as the complex plane $\mathbb{C}$, and we know how to multiply complex numbers. There is no clear choice of a multiplication operation on B. The same for C: if you try to multiply two degree 4 polynomials in C together, you will get out a polynomial of degree 8, which does not live in C!

- There is a *'take the derivative'* operation on C,

$$\mathbf{c} \mapsto \frac{d}{dx} \mathbf{c}$$

  which we will meet again later. Note that taking the derivative decreases the degree of a polynomial by 1, so the result remains in C, and so this is a well defined map from C to C. There is no analogue of this operation in A and B.

  Note that there is no *integration* map from C to C, because integrating a polynomial *increases* the degree by 1, so the result might be a polynomial of degree 5, which does not live in C!

1.1.4 Rules

We have found that each of our three sets A, B and C have an *addition operation* $+$, a *zero vector* $\mathbf{0}$, and a *scalar multiplication* operation $\cdot$. Do these operations satisfy any rules, common to all three sets?

For instance, we can think of the addition operation in A as a function which assigns to each pair of elements $\mathbf{a}$ and $\mathbf{a}'$ in A a new element $\mathbf{a} + \mathbf{a}'$ in A. Does this operation satisfy any rules?

Let us see. Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{a}' = (a_1', a_2')$ be elements of A. We can add them in two different orders,

$$\mathbf{a} + \mathbf{a}' = (a_1 + a_1', a_2 + a_2')$$

and

$$\mathbf{a}' + \mathbf{a} = (a_1' + a_1, a_2' + a_2).$$

Are these the same? In other words, does the rule

$$\mathbf{a} + \mathbf{a}' = \mathbf{a}' + \mathbf{a} \tag{1.13}$$

hold? The answer is *yes*, but why? To check whether two elements of A are equal, we have to check whether each of their components are equal. The first component of $\mathbf{a} + \mathbf{a}'$ is $a_1 + a_1'$. The first component of $\mathbf{a}' + \mathbf{a}$ is $a_1' + a_1$. Is $a_1 + a_1' = a_1' + a_1$? Yes — because these are just ordinary real numbers (not elements of A anymore), and we know that for ordinary real numbers, you can add them together in either order and get the same result. So the first component of $\mathbf{a} + \mathbf{a}'$ is equal to the first component of $a' + a$. Similarly, we can check that the second component of $\mathbf{a} + \mathbf{a}'$ is equal to the second component of $\mathbf{a}' + \mathbf{a}$. So all the components of $\mathbf{a} + \mathbf{a}'$ are equal to all the components of $\mathbf{a}' + \mathbf{a}$. So, finally, we conclude that $\mathbf{a} + \mathbf{a}' = \mathbf{a}' + \mathbf{a}$.

Does this rule (1.13) also hold for the addition operations in B and C? Yes. For instance, let us check that it holds in C. Suppose that $\mathbf{c}$ and $\mathbf{d}$ are polynomials in C. Does the rule

$$\mathbf{c} + \mathbf{d} = \mathbf{d} + \mathbf{c} \tag{1.14}$$

hold?

The left and right hand sides of (1.14) are elements of C. And elements of C are polynomials. To check if two polynomials are equal, we need to check if they are equal *as functions*, in other words, if you get identical results output from both functions no matter what input value of x you substitute in.

At an arbitrary input value x, the left hand side computes as $(\mathbf{c}+\mathbf{d})(x) = \mathbf{c}(x)+\mathbf{d}(x)$. On the other hand, the right hand side computes as $(\mathbf{d}+\mathbf{c})(x) = \mathbf{d}(x) + \mathbf{c}(x)$. Now, remember that $\mathbf{c}(x)$ and $\mathbf{d}(x)$ are just ordinary numbers (not polynomials). So $\mathbf{c}(x) + \mathbf{d}(x) = \mathbf{d}(x) + \mathbf{c}(x)$, because this is true for

ordinary numbers. So for each input value x, $(\mathbf{c} + \mathbf{d})(x) = (\mathbf{d} + \mathbf{c})(x)$. Therefore the polynomials $\mathbf{c} + \mathbf{d}$ and $\mathbf{d} + \mathbf{c}$ are equal, because they output the same values for all numbers x.

There are other rules that also hold in all three sets. For instance, in all three sets, the rule

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) \tag{1.15}$$

holds for any three elements $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Can you find the other common rules?

1.2   Definition of an abstract vector space

Mathematics is about identifying patterns. We have found three different sets, A, B and C, which look very different on the surface but have much in common. In each set, there is an addition operation, a zero vector, and a scalar multiplication operation. Moreover, in each set, these operations satisfy the same rules. Let us now record this pattern by giving it a name and writing down the rules explicitly.

Definition 1. A *vector space* is a set V equipped with the following data:

D1. An *addition operation*. (That is, for every pair of elements $\mathbf{u}, \mathbf{v} \in V$, a new element $\mathbf{u} + \mathbf{v} \in V$ is defined.)

D2. A *zero vector*. (That is, a special vector $\mathbf{0} \in V$ is marked out.)

D3. A *scalar multiplication operation*. (That is, for each real number k and each element $\mathbf{v} \in V$, a new element $k \cdot \mathbf{v} \in V$ is defined).

This data should satisfy the following rules for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and for all real numbers k and l:

R1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$

R2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

R3. a) $\mathbf{0} + \mathbf{v} = \mathbf{v}$
     b) $\mathbf{v} + \mathbf{0} = \mathbf{v}$

R4. $k \cdot (\mathbf{v} + \mathbf{w}) = k \cdot \mathbf{v} + k \cdot \mathbf{w}$

R5. $(k + l) \cdot \mathbf{v} = k \cdot \mathbf{v} + l \cdot \mathbf{v}$

R6. $k \cdot (l \cdot \mathbf{v}) = (kl) \cdot \mathbf{v}$

R7. $1 \cdot \mathbf{v} = \mathbf{v}$

R8. $0 \cdot \mathbf{v} = \mathbf{0}$

We will call the elements of a vector space *vectors*, and we will write them in boldface eg. $\mathbf{v} \in V$. We do this to distinguish vectors from real numbers, which we will call *scalars*, and which don't have a boldface. It is difficult to use boldface in handwriting, so you should write them with an arrow on top, like so: $\vec{v}$.

Also, in this chapter we will write scalar multiplication with a $\cdot$, for instance $k \cdot \mathbf{v}$, but in later chapters we will simply write it as $k\mathbf{v}$ for brevity, so be careful!

To prove that a certain set can be given the structure of a vector space, one therefore needs to do the following:

1. Define a set V.

2. Define the data of an addition operation (D1), a zero vector (D2), and a scalar multiplication operation (D3) on V.

3. Check that this data satisfies the rules (R1) - (R8).

1.3 First example of a vector space

We were led to the definition (Definition 1) of an abstract vector space by considering the properties of sets A, B and C in Section **??**. Let us check, for instance, that B indeed satisfies Definition 1. The others will be left as exercises. tikc

---

Example 1.1. The set B is a vector space.

1. Define a set B

We define

$$B := \{(u_1,\, u_2,\, u_3) : u_1, u_2, u_3 \in \mathbb{R} \text{ and } u_1 - u_2 + u_3 = 0\}. \qquad (1.16)$$

2. Define addition, the zero vector, and scalar multiplication

D1. We define addition as follows. Suppose $\mathbf{u} = (u_1,\, u_2,\, u_3)$ and $\mathbf{v} = (v_1,\, , v_2,\, v_3)$ are elements of B. Note that in particular this means $u_1 - u_2 + u_3 = 0$ and $v_1 - v_2 + v_3 = 0$. We define $\mathbf{u} + \mathbf{v}$ by:

$$\mathbf{u} + \mathbf{v} := (u_1 + v_1,\, u_2 + v_2,\, u_3 + v_3). \qquad (1.17)$$

We need to check that this makes sense. We are supposed to have that $\mathbf{u} + \mathbf{v}$ is also an element of B. We can't just write down any definition! To check if $\mathbf{u} + \mathbf{v}$ is an element of B, we need to check if it satisfies equation (1.16). Let us check:

$(u_1 + v_1) - (u_2 + v_2) + (u_3 + v_3)$
$= (u_1 - u_2 + u_3) + (v_1 - v_2 + v_3)$ (this algebra step is true for ordinary numbers)
$= 0 + 0$ (since $\mathbf{u}$ and $\mathbf{v}$ are in B)
$= 0.$

Therefore, $\mathbf{u} + \mathbf{v}$ is indeed an element of B, so we have written down a well-defined addition operation on B, which takes two arbitrary elements of B and outputs another element of B.

D2. We define the zero vector $\mathbf{0} \in B$ as

$$\mathbf{0} := (0, 0, 0). \tag{1.18}$$

We need to check that this makes sense. Does $(0, 0, 0)$ really belong to B, in other words, does it satisfy equation (1.16)? Yes, since $0 - 0 + 0 = 0$. So we have a well-defined zero vector.

D3. We define scalar multiplication on B as follows. Let k be a real number and $\mathbf{u} = (u_1,\, u_2,\, u_3)$ be an element of B. We define

$$k \cdot \mathbf{u} := (ku_1,\, ku_2,\, ku_3). \tag{1.19}$$

We need to check that this makes sense. When I multiply a multiply a vector $\mathbf{u}$ in B by a scalar k, the result $k \cdot \mathbf{u}$ is supposed to be an element of B. Does $(ku_1,\, ku_2,\, ku_3)$ really belong to B? Let us check that it satisfies the defining equation (1.16):

$ku_1 - ku_2 + ku_3$
$= k(u_1 - u_2 + u_3)$ (this algebra step is true for ordinary numbers)
$= k0$ (since $\mathbf{u}$ is in B)
$= 0.$

Therefore, $k \cdot \mathbf{u}$ is indeed an element of B, so we have written down a well-defined scalar multiplication operation on B.

3. Check that the data satisfies the rules

We must check that our data D1, D2 and D3 satisfies the rules R1 – R8. So, suppose $\mathbf{u} = (u_1,\, u_2,\, u_3)$, $\mathbf{v} = (v_1,\, v_2,\, v_3)$ and $\mathbf{w} = (w_1,\, w_2,\, w_3)$ are

in B, and suppose that k and l are real numbers.

R1. We check:

$\mathbf{v} + \mathbf{w}$ (R1.)

$= (v_1 + w_1,\, v_2 + w_2,\, v_3 + w_3)$ (by defn of addition in B)

$= (w_1 + v_1,\, w_2 + v_2,\, w_3 + v_3)$ (because $x + y = y + x$ is true for real numbers)

$= \mathbf{w} + \mathbf{v}.$ (by defn of addition in B)

R2. We check:

$(\mathbf{u} + \mathbf{v}) + \mathbf{w}$

$= (u_1 + v_1,\, u_2 + v_2,\, u_3 + v_3) + \mathbf{w}$ (by defn of addition in B)

$= ((u_1 + v_1) + w_1,\, (u_2 + v_2) + w_2,\, (u_3 + v_3) + w_3)$ (by defn of addition in B)

$= (u_1 + (v_1 + w_1),\, u_2 + (v_2 + w_2),\, u_3 + (v_3 + w_3))$ (since $(x + y) + z = x + (y + z)$ is true for real numbers)

$= \mathbf{u} + (v_1 + w_1,\, v_2 + w_2,\, v_3 + w_3)$ (by defn of addition in B)

$= \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (by defn of addition in B).

R3. We check:

$\mathbf{0} + \mathbf{v}$

$= (0,\, 0,\, 0) + (v_1,\, v_2,\, v_3)$ (by defn of the zero vector in B

$= (0 + v_1,\, 0 + v_2,\, 0 + v_3)$ (by defn of addition in B)

$= (v_1,\, v_2,\, v_3)$ (because $x + 0 = x$ is true for real numbers)

$= \mathbf{v}.$

By the same reasoning, we can check that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

R4. We check:

$k \cdot (\mathbf{v} + \mathbf{w})$

$= k \cdot (v_1 + w_1,\, v_2 + w_2,\, v_3 + w_3)$ (by defn of addition in B

$= (k(v_1 + w_1),\, k(v_2 + w_2),\, k(v_3 + w_3))$ (by defn of scalar multipliation in B)

$= (kv_1 + kw_1,\, kv_2 + kw_2,\, kv_3 + kw_3)$ (since $k(x + y) = kx + ky$ for real numbers x, y)

$= (kv_1,\, kv_2,\, kv_3) + (kw_1,\, kw_2,\, kw_3)$ (by defn of addition in B)

$= k \cdot \mathbf{v} + k \cdot \mathbf{w}$ (by defn of scalar multiplication in B)

R5. We check:

$(k + l) \cdot \mathbf{v}$
$= ((k + l)v_1,\ (k + l)v_2,\ (k + l)v_3)$ (by defn of scalar multiplication in B)
$= (kv_1 + lv_1,\ kv_2 + lv_2,\ kv_3 + lv_3)$ (since $(k + l)x = kx + lx$ for real numbers)
$= (kv_1,\ kv_2,\ kv_3) + (lv_1,\ lv_2,\ lv_3)$ (by defn of addition in B)
$= k \cdot \mathbf{v} + l \cdot \mathbf{v}$ (by defn of scalar multiplication in B)

R6. We check:

$k \cdot (l \cdot \mathbf{v})$
$= k \cdot (lv_1,\ lv_2,\ lv_3)$ (by defn of scalar multiplication in B)
$= (k(lv_1),\ k(lv_2),\ k(lv_3))$ (by defn of scalar multiplication in B)
$= ((kl)v_1,\ (kl)v_2,\ (kl)v_3)$ (since $k(lx) = (kl)x$ for real numbers)
$= (kl) \cdot \mathbf{v}$ (by defn of scalar multiplication in B).

R7. We check:

$1 \cdot \mathbf{v} = (1v_1,\ 1v_2,\ 1v_3)$ (by defn of scalar multiplication in B)
$= (v_1,\ v_2,\ v_3)$ (since $1x = x$ for real numbers x)
$= \mathbf{v}$.

R8. We check:

$0 \cdot \mathbf{v} = (0v_1,\ 0v_2,\ 0v_3)$ (by defn of scalar multiplication in B)
$= (0,\ 0,\ 0)$ (since $0x = 0$ for real numbers
$= \mathbf{0}$ (by defn of the zero vector in B).

Further exercises for 1.3

Exercise 1. Prove that set A from Section **??** together with the addition operation (1.4), the zero vector (1.8) and the scalar multiplication operation (1.10) forms a vector space.

Exercise 2. Prove that set C from Section **??** together with the addition operation (1.6), the zero vector (1.9) and the scalar multiplication operation (1.12) forms a vector space.

1.4 More examples and non-examples

Example 1.2 (A non-example)**.** Define the set V by

$$V := \{\mathbf{a}, \mathbf{b}\}. \tag{1.20}$$

Define the addition operation by

$$\mathbf{a} + \mathbf{a} := \mathbf{a} \qquad\qquad \mathbf{a} + \mathbf{b} := \mathbf{a} \tag{1.21}$$
$$\mathbf{b} + \mathbf{a} := \mathbf{b} \qquad\qquad \mathbf{b} + \mathbf{b} := \mathbf{c} \tag{1.22}$$

For this to be a well-defined addition operation, we need to check that whenever you add two elements of V together, you get out a well-defined element of V. But $\mathbf{b} + \mathbf{b} = \mathbf{c}$, so adding $\mathbf{b} \in V$ to itself outputs something, namely $\mathbf{c}$, which is not an element of V. So V does not form a vector space since it does not even have a well-defined addition operation.

Example 1.3 (Another non-example)**.** Define the set V by

$$V := \{\mathbf{a}, \mathbf{b}\}. \tag{1.23}$$

Define the addition operation by

$$\mathbf{a} + \mathbf{a} := \mathbf{a} \qquad\qquad \mathbf{a} + \mathbf{b} := \mathbf{b} \tag{1.24}$$
$$\mathbf{b} + \mathbf{a} := \mathbf{b} \qquad\qquad \mathbf{b} + \mathbf{b} := \mathbf{a} \tag{1.25}$$

This is a well-defined addition operation, since whenever you add two elements of V together, you get out a well-defined element of V.

Define the zero vector by

$$\mathbf{0} := \mathbf{a}. \tag{1.26}$$

This is well-defined, since $\mathbf{a}$ is indeed an element of V.

Define scalar multiplication by a real number $k \in \mathbb{R}$ by

$$k \cdot \mathbf{a} := \mathbf{a} \quad \text{and} \quad k \cdot \mathbf{b} := \mathbf{b}. \tag{1.27}$$

This is a well-defined scalar multiplication, since it allows us to multiply any element $\mathbf{v} \in V$ by a scalar k and it outputs a well-defined element $k \cdot \mathbf{v} \in V$.

movebackintoexample

Exercise 3. Show that these operations satisfy rules R1, R2, R3, R4, R6 and R7, but not R5 and R8.

Example 1.4 (The zero vector space)**.** Define the set Z by

$$Z := \{\mathbf{z}\}. \tag{1.28}$$

Note that it contains just a single element, $\mathbf{z}$. Define the addition operation as

$$\mathbf{z} + \mathbf{z} := \mathbf{z} \tag{1.29}$$

Define the zero element as

$$\mathbf{0} := \mathbf{z}. \tag{1.30}$$

Finally define scalar multiplication by a scalar $k \in \mathbb{R}$ as:

$$k \cdot \mathbf{z} := \mathbf{z}. \tag{1.31}$$

---

movebackintoexample

Exercise 4. Show that this data satisfies the rules R1 to R8.

Example 1.5 ($\mathbb{R}^n$)**.** Define the set $\mathbb{R}^n$ by

$$\mathbb{R}^n := \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R} \text{ for all } i = 1 \ldots n\}. \tag{1.32}$$

Define the addition operation as

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{1.33}$$

Define the zero element as

$$\mathbf{0} := (0, 0, \ldots, 0). \tag{1.34}$$

Define scalar multiplication by

$$k \cdot (x_1, x_2, \ldots, x_n) := (kx_1, kx_2, \ldots, kx_n). \tag{1.35}$$

---

move back into example

Exercise 5. Show that this data satisfies the rules R1 to R8.

Example 1.6 ($\mathbb{R}^\infty$)**.** Define the set $\mathbb{R}^\infty$ by

$$\mathbb{R}^\infty := \{(x_1, x_2, x_3, \ldots,) : x_i \in \mathbb{R} \text{ for all } i = 1, 2, 3, \ldots\} \qquad (1.36)$$

So an element $\mathbf{x} \in \mathbb{R}^\infty$ is an infinite sequence of real numbers. Define the addition operation componentwise:

$$(x_1, x_2, x_3, \ldots) + (y_1, y_2, y_3, \ldots) := (x_1 + y_1, x_2 + y_2, x_3 + y_3, \ldots).$$
$$(1.37)$$

Define the zero element as

$$\mathbf{0} := (0, 0, 0, \ldots), \qquad (1.38)$$

the infinite sequence whose components are all zero. Finally, define scalar multiplication componentwise:

$$k \cdot (x_1, x_2, x_3, \ldots) := (kx_1, kx_2, kx_3, \ldots) \qquad (1.39)$$

> Thinking about infinity is an important part of mathematics. Have you watched the movie about me called *The man who knew infinity*?

movebackintoexample

Exercise 6. Show that this data satisfies the rules R1 to R8.

Example 1.7 (Functions on a set)**.** Let X be any set. Define the set $\text{Fun}(X)$ of *real-valued functions on X* by

$$\text{Fun}(X) := \{\mathbf{f} : X \to \mathbb{R}\}. \qquad (1.40)$$

Note that the functions can be arbitrary; there is no requirement for them to be continuous, or differentiable. Such a requirement would not make sense, since X could be an arbitrary set. For instance X could be the set $\{a, b, c\}$ —without any further informaiton, it does not make sense to say that a function $\mathbf{f} : X \to \mathbb{R}$ is continuous.

Define the addition operation by

$$(\mathbf{f} + \mathbf{g})(x) := \mathbf{f}(x) + \mathbf{g}(x), \quad x \in X \qquad (1.41)$$

Make sure you understand what this formula is saying! We start with two functions $\mathbf{f}$ and $\mathbf{g}$, and we are defining their sum $\mathbf{f} + \mathbf{g}$. This is supposed to be another function on X. To define a function on X, I am supposed to write down what value it outputs for each $x \in X$. And that is what the formula says: the value that the function $\mathbf{f} + \mathbf{g}$ assigns to the element $x \in X$ is defined to be the number $\mathbf{f}(x)$ plus the number $\mathbf{g}(x)$. Remember: $\mathbf{f}$ is a function, while $\mathbf{f}(x)$ is a number!

Define the zero vector $\mathbf{0}$ to be the function which outputs the number 0 for every input value of $x \in X$:

$$\mathbf{0}(x) := 0 \quad \text{for all } x \in X. \tag{1.42}$$

Define scalar multiplication by

$$(k \cdot \mathbf{f})(x) := k\mathbf{f}(x). \tag{1.43}$$

movebackintoexample (all 3)

Exercise 7. Notation quiz! Say whether the following combination of symbols represents a real number or a function.

1. $\mathbf{f}$

2. $\mathbf{f}(x)$

3. $k \cdot \mathbf{f}$

4. $(k \cdot \mathbf{f})(x)$

Exercise 8. Let $X = \{a, b, c\}$.

1. Write down three different functions $\mathbf{f}, \mathbf{g}, \mathbf{h}$ in $\text{Fun}(X)$.

2. For the functions you constructed in (i), calculate $\mathbf{f} + \mathbf{g}$ as well as $3 \cdot \mathbf{h}$.

Exercise 9. Show that the data above satisfies the rules R1 to R8, so that $\text{Fun}(X)$ is a vector space.

Example 1.8 (Matrices)**.** the set $\text{Mat}_{n,m}$ of all $n \times m$ matrices is a vector space. See Appendix **??** for a reminder about matrices.

movebackintoexample

Exercise 10. Show that when equipped with the addition operation, zero vector, and scalar multiplication operation as defined in Appendix **??**, the set $\mathrm{Mat}_{n,m}$ of all $n \times m$ matrices is a vector space.

Example 1.9. We will write Col_n for the vector space $\mathrm{Mat}_{n,1}$ of n-dimensional *column vectors*,

$$\mathrm{Col}_n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}$$

So, Col_n 'is' just $\mathbb{R}^n$, but we make explicit the fact that the components of the vectors are arranged in a column.

Further exercises for 1.4

Exercise 11. Define an addition operation on the set $X := \{\mathbf{0}, \mathbf{a}, \mathbf{b}\}$ by the following table:

+	**0**	**a**	**b**
0	**0**	**a**	**b**
a	**a**	**0**	**a**
b	**b**	**a**	**0**

This table works as follows. To calculate, for example, $\mathbf{b} + \mathbf{a}$, find the intersection of the row labelled by $\mathbf{b}$ with the column labelled by $\mathbf{a}$. We see that $\mathbf{b} + \mathbf{a} := \mathbf{a}$.

Prove that this addition operation satisfies R1.

Exercise 12. Prove that the addition operation from Exercise 11 does not satisfy R2.

Exercise 13. Define a strange new addition operation $\hat{+}$ on $\mathbb{R}$ by

$$\mathbf{x} \hat{+} \mathbf{y} := \mathbf{x} - \mathbf{y}, \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}.$$

Does $\hat{+}$ satisfy R2? If it does, prove it. If it does not, give a counterexample.

Exercise 14. Let $\mathbb{R}^+$ be the set of positive real numbers. Define an addition operation, a zero vector and a scalar multiplication on $\mathbb{R}^+$ by

$$\mathbf{x} + \mathbf{y} := \mathbf{xy}$$
$$\mathbf{0} := 1$$
$$k \cdot \mathbf{x} := \mathbf{x}^k$$

where $\mathbf{x}$ and $\mathbf{y}$ are positive real numbers, and k is a scalar (i.e. an arbitrary real number).

1. Check that these operations are well-defined. For instance, is $\mathbf{x} + \mathbf{y} \in \mathbb{R}^+$, as it should be?

2. Check that they satisfy R1 to R8, so that $\mathbb{R}^+$ is a vector space with these operations.

1.5   Some results about abstract vector spaces

It is time to use the rules of a vector space to prove some general results.

We are about to do our first formal proof in the course!

Lemma 2. *Suppose V is a vector spaces with zero vector $\mathbf{0}$. If $\mathbf{0}'$ is another vector satisfying R3(a),*

$$\mathbf{0}' + \mathbf{v} = \mathbf{v} \ \textit{for all } \mathbf{v} \in V \tag{1.44}$$

then $\mathbf{0} = \mathbf{0}'$.

Proof.

$$\begin{aligned}
\mathbf{0} &= \mathbf{0}' + \mathbf{0} && \text{(using (1.44)) for } \mathbf{v} = \mathbf{0}) \\
&= \mathbf{0}' && \text{(using R3(b) for } \mathbf{v} = \mathbf{0}')
\end{aligned}$$

$\square$

Definition 3. If V is a vector space, we define the *additive inverse* of a vector $\mathbf{v} \in V$ as

$$-\mathbf{v} := (-1) \cdot \mathbf{v}$$

Lemma 4. *If V is a vector space, then for all $\mathbf{v} \in V$,*

$$-\mathbf{v} + \mathbf{v} = \mathbf{0} \ \textit{and } \mathbf{v} + (-\mathbf{v}) = \mathbf{0}. \tag{1.45}$$

Proof.

$$\begin{aligned}
-\mathbf{v} + \mathbf{v} &= (-1) \cdot \mathbf{v} + \mathbf{v} && \text{(using defn of } -\mathbf{v}) \\
&= (-1) \cdot \mathbf{v} + 1 \cdot \mathbf{v} && \text{(R7)} \\
&= (-1 + 1) \cdot \mathbf{v} && \text{(R5)} \\
&= 0 \cdot \mathbf{v} && \\
&= \mathbf{0} && \text{(R8)}
\end{aligned}$$

In addition,

$$\mathbf{v} + (-\mathbf{v}) = -\mathbf{v} + \mathbf{v} \qquad\qquad \text{(R1)}$$
$$= \mathbf{0} \qquad\qquad\qquad\quad \text{(by previous proof)}$$

$\square$

Lemma 5. *Suppose that two vectors* $\mathbf{w}$ *and* $\mathbf{v}$ *in a vector space satisfy* $\mathbf{w} + \mathbf{v} = \mathbf{0}$. *Then* $\mathbf{w} = -\mathbf{v}$.

Proof.

$$\mathbf{w} = \mathbf{w} + \mathbf{0} \qquad\qquad\qquad \text{(R3b)}$$
$$= \mathbf{w} + (\mathbf{v} + -\mathbf{v}) \qquad\quad \text{(by Lemma 4)}$$
$$= (\mathbf{w} + \mathbf{v}) + -\mathbf{v}) \qquad \text{(R2)}$$
$$= \mathbf{0} + -\mathbf{v} \qquad\qquad\quad \text{(by assumption)}$$
$$= -\mathbf{v} \qquad\qquad\qquad\; \text{(R3a)}.$$

$\square$

Example 1.10. Let us practice using the rules of a vector space to perform everyday calculations. For instance, suppose that we are trying to solve for the vector $\mathbf{x}$ appearing in the following equation:

$$\mathbf{v} + 7 \cdot \mathbf{x} = \mathbf{w} \qquad\qquad (1.46)$$

We do this using the rules as follows:

$$\mathbf{v} + 7 \cdot \mathbf{x} = \mathbf{w}$$
$$\therefore -\mathbf{v} + (\mathbf{v} + 7 \cdot \mathbf{x}) = -\mathbf{v} + \mathbf{w} \qquad \text{(Added } -\mathbf{v} \text{ on left to both sides)}$$
$$\therefore (-\mathbf{v} + \mathbf{v}) + 7 \cdot \mathbf{x} = -\mathbf{v} + \mathbf{w} \qquad \text{(used R2 on LHS)}$$
$$\therefore \mathbf{0} + 7 \cdot \mathbf{x} = -\mathbf{v} + \mathbf{w} \qquad \text{(used Lemma 4 on LHS)}$$
$$\therefore 7 \cdot \mathbf{x} = -\mathbf{v} + \mathbf{w} \qquad \text{(used R3a on LHS)}$$
$$\therefore \frac{1}{7} \cdot (7 \cdot \mathbf{x}) = \frac{1}{7} \cdot (-\mathbf{v} + \mathbf{w}) \qquad \text{(scalar multiplied both sides by } \tfrac{1}{7})$$
$$\therefore (\frac{1}{7} 7) \cdot \mathbf{x} = \frac{1}{7} \cdot (-\mathbf{v} + \mathbf{w}) \qquad \text{(used R6 on LHS)}$$
$$\therefore 1 \cdot \mathbf{x} = \frac{1}{7} \cdot (-\mathbf{v} + \mathbf{w}) \qquad \text{(multiplied } \tfrac{1}{7} \text{ with 7)}$$
$$\therefore \mathbf{x} = \frac{1}{7} \cdot (-\mathbf{v} + \mathbf{w}) \qquad \text{(R7)}$$

As the course goes on we will leave out all these steps. But it is important for you to be able to reproduce them all, if asked to do so!

We will need one final lemma.

Lemma 6. *Suppose that* $\mathbf{v}$ *is a vector in a vector space* V *and that* k *is a scalar. Then*

$$k \cdot \mathbf{v} = \mathbf{0} \quad \Leftrightarrow \quad k = 0 \ or \ \mathbf{v} = \mathbf{0}.$$

Proof. (Proof of $\Leftarrow$). Suppose $k = 0$. Then $k \cdot \mathbf{v} = 0 \cdot \mathbf{v} = \mathbf{0}$ by R8 of a vector space. On the other hand, suppose $\mathbf{v} = \mathbf{0}$. Then $k \cdot \mathbf{v} = k \cdot \mathbf{0} = \mathbf{0}$ by Exercise 16.

(Proof of $\Rightarrow$). Suppose $k \cdot \mathbf{v} = \mathbf{0}$. There are two possibilities: either $k = 0$, or $k \neq 0$. If $k = 0$, then we are done. If $k \neq 0$, then $\frac{1}{k}$ exists and we can multiply both sides by it:

$$k \cdot \mathbf{v} = \mathbf{0}$$

$$\therefore \frac{1}{k} \cdot (k \cdot \mathbf{v}) = \frac{1}{k} \cdot \mathbf{0} \qquad \text{(Multiplied both sides by } \tfrac{1}{k}\text{)}$$

$$\therefore (\frac{1}{k}k) \cdot \mathbf{v} = \mathbf{0} \qquad \begin{array}{l}\text{(On the LHS, we used R6.}\\ \text{On the RHS, we used Exercise}\\ \text{16.)}\end{array}$$

$$\therefore 1 \cdot \mathbf{v} = \mathbf{0} \qquad \text{(using } \tfrac{1}{k}k = 1\text{)}$$

$$\therefore \mathbf{v} = \mathbf{0} \qquad \text{(R7)}$$

Hence in the case $k \neq 0$ we must have $\mathbf{v} = \mathbf{0}$, which is what we wanted to show. $\qquad\qquad\square$

Further exercises for 1.5

Exercise 15. Prove that for all vectors v in a vector space, $-(-\mathbf{v}) = \mathbf{v}$.

Exercise 16. Let V be a vector space with zero vector $\mathbf{0}$. Prove that for all scalars k, $k \cdot \mathbf{0} = \mathbf{0}$.

Exercise 17. Suppose that two vectors $\mathbf{x}$ and $\mathbf{w}$ in a vector space satisfy $2\mathbf{x} + 6\mathbf{w} = \mathbf{0}$. Solve for $\mathbf{x}$, showing explicitly how you use the rules of a vector space, as in Example 1.10.

1.6   Subspaces

The notion of a subspace will allow us to quickly establish many more examples of vector spaces.

Definition 7. A subset $U \subseteq V$ of a vector space V is called a *subspace of* V if:

- For all $\mathbf{u}, \mathbf{u}' \in U$, $\mathbf{u} + \mathbf{u}' \in U$

- $\mathbf{0} \in U$

- For all scalars k and all vectors $\mathbf{u} \in U$, $k \cdot \mathbf{u} \in U$

Lemma 8. *If U is a subspace of a vector space V, then U is also a vector space, when we equip it with the same addition operation, zero vector and scalar multiplication as in V.*

Proof. Since U is a subspace, we know that it actually makes sense to "equip it with the same addition operation, zero vector and scalar multiplication as in V". (If U was *not* a subspace, then we might have for instance $\mathbf{u}, \mathbf{u}' \in U$ but $\mathbf{u} + \mathbf{u}' \notin U$, so the addition operation would not make sense.)

So we simply need to check the rules R1 to R8. Since these rules hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V, they certainly hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in U. So R1 to R8 are satisfied. $\qquad\square$

Example 1.11 (Line in $\mathbb{R}^2$)**.** A line L through the origin in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$:

$$(1.47)$$
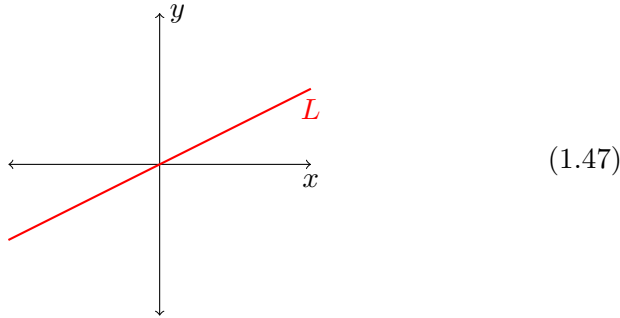
Indeed, recall that L can be specified by a homogenous linear equation of the form:

$$L = \{(x, y) \in \mathbb{R}^2 : ax + by = 0\} \qquad (1.48)$$

for some constants a and b. So, if $\mathbf{v} = (x, y)$ and $\mathbf{v}' = (x', y')$ lie on L, then their sum $\mathbf{v} + \mathbf{v}' = (x + x', y + y')$ also lies on L, because its components satisfy the defining equation (1.48):

$$
\begin{aligned}
& a(x + x') + b(y + y') \\
&= (ax + by) + (ax' + by') \\
&= 0 + 0 \qquad\qquad \text{(since } ax + by = 0 \text{ and } ax' + by' = 0) \\
&= 0.
\end{aligned}
$$

This also makes sense geometrically: if you look at the picture (1.47), then you will see that adding two vectors $\mathbf{v}, \mathbf{v}'$ on L by the head-to-tail method will produce another vector on L.

movebackintoexample

Exercise 18. Complete the proof that L is a subspace of $\mathbb{R}^2$ by checking that the zero vector is in L, and that multiplying a vector in L by a scalar outputs a vector still in L.

Example 1.12 (Lines and planes in $\mathbb{R}^3$). Similarly a line L or a plane P through the origin in $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$:

$$\tag{1.49}$$
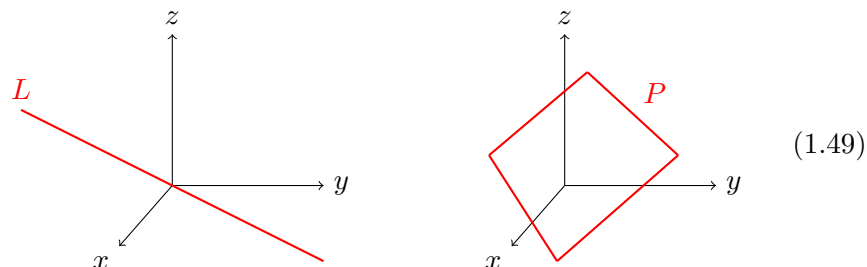
Example 1.13 (Zero vector space). If V is a vector space, the the set $\{\mathbf{0}\} \subseteq V$ containing just the zero vector $\mathbf{0}$ is a subspace of V.

---

movebackintoexample

Exercise 19. Check this.

Example 1.14 (Non-example: Line not through origin). Be careful though — not *every* line $L \subset \mathbb{R}^2$ is a subspace of $\mathbb{R}^2$. If L does not go through the origin, then $\mathbf{0} \notin L$, so L is not a subspace.

Another reason that L is not a subspace is that it is not closed under addition: when we add two nonzero vectors $\mathbf{v}$ and $\mathbf{v}'$ on L, we end up

with a vector $\mathbf{v} + \mathbf{v}'$ which does not lie on L:

$$(1.50)$$
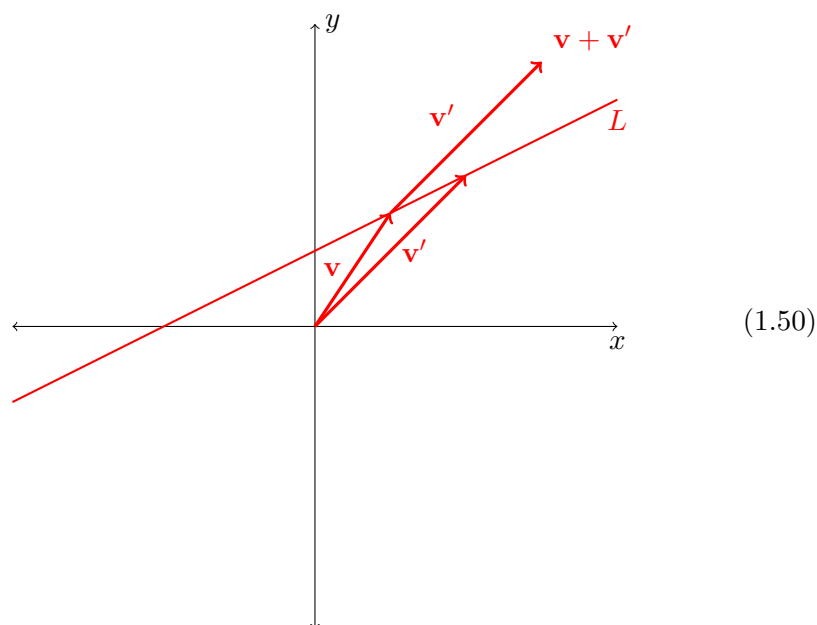
Example 1.15 (Continuous functions as a subspace)**.** The set

$$\mathrm{Cont}(I) := \{\mathbf{f} : I \to \mathbb{R}, \mathbf{f} \text{ continuous}\}$$

of all continuous functions on an interval I is a subspace of the set $\mathrm{Fun}(I)$ of *all* functions on I. Let us check that it satisfies the definition. You know from earlier courses that:

- If $\mathbf{f}$ and $\mathbf{g}$ are continuous functions on I, then $\mathbf{f} + \mathbf{g}$ is also a continuous function.

- The zero function $\mathbf{0}$ defined by $\mathbf{0}(x) = 0$ for all $x \in I$ is a continuous function.

- If $\mathbf{f}$ is a continuous function, and k is a scalar, then $k \cdot \mathbf{f}$ is also continuous.

Hence, by Lemma 8, $\mathrm{Cont}(I)$ is a vector space in its own right.

Example 1.16 (Differentiable functions as a subspace)**.** Similarly, the set

$$\text{Diff}(I) := \{\mathbf{f} : I \to \mathbb{R}, \mathbf{f} \text{ differentiable}\}$$

of differentiable functions on an open interval I is a subspace of $\text{Fun}(I)$.

movebackintoexample

Exercise 20. Check this. Also, is $\text{Diff}(I)$ a subspace of $\text{Cont}(I)$?

Example 1.17 (Vector spaces of polynomials)**.** A *polynomial* is a function $\mathbf{p} : \mathbb{R} \to \mathbb{R}$ of the form

$$\mathbf{p}(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0. \qquad (1.51)$$

for some fixed real coefficients $a_0, \ldots, a_n$. Two polynomials $\mathbf{p}$ and $\mathbf{q}$ are *equal* if they are equal *as functions*, that is if $\mathbf{p}(x) = \mathbf{q}(x)$ for all $x \in \mathbb{R}$. The *degree* of a polynomial is the highest power of x which occurs in its formula.

For example, $2x^3 - x + 7$ is a polynomial of degree 3, while $x^5 - 2$ is a polynomial of degree 5. We write the set of *all* polynomials as Poly and the set of all polynomials having degree less than or equal to n as Poly_n.

movebackintoexample

Exercise 21. Check that Poly and Poly_n are subspaces of $\text{Cont}(\mathbb{R})$.

Example 1.18 (Trigonometric polynomials)**.** A *trigonometric polynomial* is a function $\mathbf{T} : \mathbb{R} \to \mathbb{R}$ of the form

$$\mathbf{T}(x) = a_0 + a_1 \cos(x) + b_1 \sin(x) + a_2 \cos(2x) + b_2 \sin(2x) + \cdots$$
$$+ a_n \cos(nx) + b_n \sin(nx). \quad (1.52)$$

The *degree* of a trigonometric polynomial is the highest multiple of x which occurs inside one of the sines or cosines in its formula. For instance,

$$3 - \cos(x) + 6 \sin(3x)$$

is a trigonometric polynomial of degree 3. We write the set of *all* trigonometric polynomials as Trig and the set of all trigonometric polynomials having degree less than or equal to n as Trig_n.

movebackintoexample (both)

Exercise 22. Show that Trig and Trig_n are subspaces of $\text{Cont}(\mathbb{R})$.

Exercise 23. Consider the function $\mathbf{f}(x) = \sin^3(x)$. Show that $\mathbf{f} \in \text{Trig}_3$ by writing it in the form (1.52). Hint: use the identities

$$\sin(A)\sin(B) = \frac{1}{2}(\cos(A-B) - \cos(A+B))$$
$$\sin(A)\cos(B) = \frac{1}{2}(\sin(A-B) + \sin(A+B))$$
$$\cos(A)\cos(B) = \frac{1}{2}(\cos(A-B) + \cos(A+B))$$

which follow easily from the addition formulae

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$$
$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B.$$

Example 1.19 (Solutions to linear differential equations)**.** Fix constants a, b and c. Let $V \subseteq \text{Diff}(\mathbb{R})$ be the set of all solutions to the following differential equation:

$$a\frac{d^2 f}{dx^2} + b\frac{df}{dx} + cf = 0$$

movebackintoexample

Exercise 24. Show that V is a subspace of $\text{Diff}(\mathbb{R})$.

Further exercises for 1.6

Exercise 25. Show that the set

$$V := \{(a, -a, b, -b) : a, b \in \mathbb{R}\}$$

is a subspace of $\mathbb{R}^4$.

Exercise 26. Show that the set

$$V := \{\text{polynomials of the form } \mathbf{p}(x) = ax^3 + bx^2 - cx + a, \ a, b, c \in \mathbb{R}\}$$

is a subspace of Poly_3.

Exercise 27. Let $b \in \mathbb{R}$. Prove that

$$V := \{(x_1, \ x_2, \ x_3) \in \mathbb{R}^3 : 2x_1 - 3x_2 + 5x_3 = b\}$$

is a subspace of $\mathbb{R}^3$ if and only if $b = 0$.

Exercise 28. Consider the set

$$V := \{\mathbf{f} \in \text{Diff}((-1, 1)) : f'(0) = 2\}$$

Is V a subspace of $\text{Diff}((-1, 1))$? If you think it is, *prove* that it is. If you think it is not, *prove* that it is not!

Exercise 29. Consider the set

$$V := \{(x_1, \ x_2, \ x_3, \ \ldots) \in \mathbb{R}^\infty : \lim_{n \to \infty} x_n = 0\}$$

Is V a subspace of $\mathbb{R}^\infty$? If you think it is, *prove* that it is. If you think it is not, *prove* that it is not!

Exercise 30. Is $\mathbb{R}^+ := \{\mathbf{x} \in \mathbb{R} : \mathbf{x} \geq 0\}$ a subspace of $\mathbb{R}$? If you think it is, *prove* that it is. If you think it is not, *prove* that it is not!

Exercise 31. Give an example of a nonempty subset U of $\mathbb{R}^2$ which is closed under addition and under taking additive inverses (i.e. if $\mathbf{u}$ is in U then $-\mathbf{u}$ is in V), but U is not a subspace of $\mathbb{R}^2$.

Exercise 32. Give an example of a nonempty subset V of $\mathbb{R}^2$ which is closed under scalar multiplication, but V is not a subspace of $\mathbb{R}^2$.

Chapter 2

Finite-dimensional vector spaces

In this course we concentrate on *finite-dimensional* vector spaces, which we will define in this chapter.

Warning: From now on, I will use shorthand and write scalar multiplication $k \cdot \mathbf{v}$ simply as $k\mathbf{v}$!

2.1 Linear combinations and span

We start with some basic definitions.

Definition 9. A *linear combination* of a finite collection $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of vectors in a vector space V is a vector of the form

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_b\mathbf{v}_n \qquad (2.1)$$

where $a_1, a_2, \ldots, a_n$ are scalars. If all the scalars a_i are 0, we say that it is the *trivial* linear combination.

Example 2.1. In $\mathbb{R}^3$, $(6, 2, -14)$ is a linear combination of $(-3, 1, 2)$ and $(-2, 0, 3)$ because

$$(6, 2, -14) = 2(-3, 1, 2) - 6(-2, 0, 3).$$

Example 2.2. In $\mathbb{R}^4$, $(2, -1, 3, 0)$ is not a linear combination of $(1, 3, 2, 0)$, $(5, 1, 2, 4)$ and $(-1, 0, 2, 1)$. If it *was*, then we would have

$$(-2, -1, 3, 0) = a_1(1, 3, 2, 0) + a_2(5, 1, 2, 4) + a_3(-1, 0, 2, 1)$$

for some scalars a_1, a_2, a_3. In other words, we would have

$$(-2, -1, 3, 0) = (a_1 + a_2 + a_3,\ 3a_1 + a_2,\ 2a_1 + 2a_2 + 2a_3,\ 4a_2 + a_3)$$

which is equivalent to the system of equations

$$\begin{cases} a_1 + 5a_2 - \ a_3 = -2 \\ 3a_1 + \ a_2 \qquad\quad = -1 \\ 2a_1 + 2a_2 + 2a_3 = 3 \\ \qquad\quad 4a_2 + \ a_3 = 0 \end{cases} \tag{2.2}$$

movebackintoexample

Exercise 33. Check that the system of equations (2.2) has no solution.

Example 2.3. Define the functions $\mathbf{f}, \mathbf{f}_1, \mathbf{f}_2 \in \mathrm{Diff}$ by

$$\mathbf{f}(x) = \cos^3 x, \quad \mathbf{f}_1(x) = \cos(x), \quad \mathbf{f}_2(x) = \cos(3x).$$

Then $\mathbf{f}$ is a linear combination of $\mathbf{f}_1$ and $\mathbf{f}_2$ because of the identity $\cos(3x) = \frac{1}{4}(3\cos x + \cos(3x))$. See Example 1.18. In other words,

$$\mathbf{f} = \frac{3}{4}\mathbf{f}_1 + \frac{1}{4}\mathbf{f}_2.$$

This example shows that $\mathbf{f}$ is also a trigonometric polynomial, even though its original formula $\mathbf{f}(x) = cos(3x)$ was not in the form (1.52).

Definition 10. We say that a list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space V *spans* V if every vector $\mathbf{v} \in V$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

Example 2.4. $\mathbb{R}^2$ is spanned by

$$\mathbf{e}_1 := (1, 0),\ \mathbf{e}_2 := (0, 1)$$
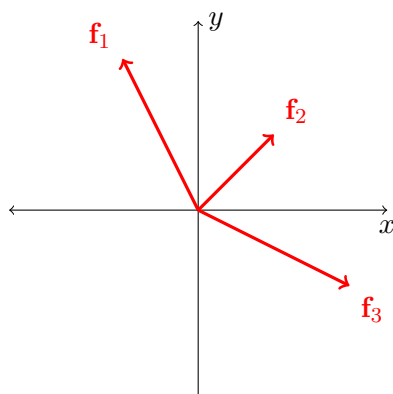
Figure 2.1: A list of vectors which spans $\mathbb{R}^2$.

because every vector $\mathbf{v} = (a_1, a_2)$ can be written as the linear combination

$$\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2.$$

Example 2.5. $\mathbb{R}^2$ is also spanned by

$$\mathbf{f}_1 := (-1, 2),\ \mathbf{f}_2 := (1, 1),\ \mathbf{f}_3 := (2, -1)$$

See Figure 2.1. Note that every vector $\mathbf{v}$ can be written as a linear combination of $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ in more than one way (in fact infinitely many!). For instance,

$$(2, 3) = 3\mathbf{f}_1 - \frac{1}{3}\mathbf{f}_2 + \frac{8}{3}\mathbf{f}_3 \tag{2.3}$$

$$(2, 3) = 2\mathbf{f}_1 + \frac{2}{3}\mathbf{f}_2 + \frac{5}{3}\mathbf{f}_3 \tag{2.4}$$

movebackintoexample

Exercise 34. Check equations (2.3) and (2.4).

Example 2.6. $\mathbb{R}^n$ is spanned by

$$\mathbf{e}_1 := (1, 0, \ldots, 0),\ \ \mathbf{e}_2 := (0, 1, \ldots, 0),\ \ \ldots,\ \ \mathbf{e}_n := (0, 0, \ldots, 0, 1) \tag{2.5}$$

because every vector $\mathbf{v} = (a_1, a_2, \ldots, a_n)$ can be written as the linear combination

$$\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n. \tag{2.6}$$

movebackintoexampel

Exercise 35. Check equation (2.6).

Further exercises for 2.1

Exercise 36. Consider the following polynomials in Poly_2:

$$\mathbf{r}_1(x) := 3x^2 - 2, \ \mathbf{r}_2(x) := x^2 + x, \ \mathbf{r}_3(x) := x + 1, \ \mathbf{r}_4(x) := x - 1$$

1. Can the polynomial $\mathbf{p}$ with $\mathbf{p}(x) = x^2 + 1$ be written as a linear combination of $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$?

2. If so, in *how many ways* can this be done?

Exercise 37. Suppose that the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and $\mathbf{e}_4$ span a vector space V. Show that the vectors $\mathbf{f}_1 := \mathbf{e}_2 - \mathbf{e}_1$, $\mathbf{f}_2 := \mathbf{e}_3 - \mathbf{e}_2$, $\mathbf{f}_3 := \mathbf{e}_4 - \mathbf{e}_3$, $\mathbf{f}_4 := \mathbf{e}_4$ also span V.

Exercise 38. Show that the polynomials

$$\mathbf{q}_0(x) := 1, \ \mathbf{q}_1(x) := x, \ \mathbf{q}_2(x) := 2x^2 - 1, \ \mathbf{q}_3(x) := 4x^3 - 3x$$

span Poly_3.

2.2   Linear independence

Definition 11. A list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space V is called *linearly independent* if the equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots k_n\mathbf{v}_n = \mathbf{0} \tag{2.7}$$

has only the trivial solution $k_1 = k_2 = \cdots = k_n = 0$. Otherwise, the list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is called *linearly dependent*.

Remark 12. Suppose one of the vectors in the collection $\mathbf{v}_1$, $\mathbf{v}_2$, $\ldots$, $\mathbf{v}_n$ is the zero vector $\mathbf{0}$; for instance suppose $\mathbf{v}_1 = \mathbf{0}$. Then the vectors must be linearly dependent, since the equation (2.7) has the nontrivial solution $k_1 = 1$, $k_2 = k_2 = \cdots = k_n = 0$. So: *a list of linearly independent vectors never contains the zero vector!*

Remark 13. The criterion for a list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to be linearly independent,

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots k_n\mathbf{v}_n = \mathbf{0} \quad \Rightarrow \quad k_1 = k_2 = \cdots = k_n = 0$$

is clearly equivalent to the criterion that

> All of the $\mathbf{v}_i$ are nonzero and none of them can be expressed as a nontrivial linear combination of the others.

Exercise 39. Prove this last remark.

Example 2.7. The list of vectors $\mathbf{f}_1 = (-1, 2)$, $\mathbf{f}_2 = (1, 1)$ from Example 2.5 is linearly independent, because the equation

$$k_1(-1, 2) + k_2(1, 1) = (0, 0)$$

is equivalent to the system of equations

$$\begin{cases} -k_1 + k_2 = 0 \\ 2k_1 + k_2 = 0 \end{cases} \tag{2.8}$$

which has only the trivial solution $k_1 = 0$ and $k_2 = 0$.

---

movebackintoexample

Exercise 40. Check that (2.8) has only the trivial solution.

Example 2.8. The list of vectors $\mathbf{f}_1 = (-1, 2)$, $\mathbf{f}_2 = (1, 1)$, $\mathbf{f}_3 = (2, -1)$ from Example 2.5 is linearly dependent, because the equation

$$k_1(-1, 2) + k_2(1, 1) + k_3(2, -1) = (0, 0) \tag{2.9}$$

is equivalent to the system of equations

$$\begin{cases} -k_1 + k_2 + 2k_3 = 0 \\ 2k_1 + k_2 - k_3 = 0 \end{cases} \tag{2.10}$$

which has a one-dimensional vector space of solutions parameterized by t,

$$k_1 = t, \, k_2 = -t, \, k_3 = t, \quad t \in \mathbb{R}. \tag{2.11}$$

For instance, for $t = 2$, we have

$$2(-1, 2) - 2(1, 1) + 2(2, -1) = (0, 0)$$

so that (2.9) has nontrivial solutions.

movebackintoexample

Exercise 41. Check that (2.10) has the solution set (2.11).

Example 2.9. The list of polynomials

$$\mathbf{q}_0(x) := 1, \ \mathbf{q}_1(x) := x, \ \mathbf{q}_2(x) := 2x^2 - 1, \ \mathbf{q}_3(x) := 4x^3 - 3x$$

from Example 38 is linearly independent in Poly_3. This is because the equation

$$k_0\mathbf{q}_0 + k_1\mathbf{q}_1 + k_2\mathbf{q}_2 + k_3\mathbf{q}_3 = \mathbf{0}$$

becomes the following equation between polynomials:

$$4k_3x^3 + 2k_2x^2 + (-3k_3 + k_1)x + (-k_2 + k_0) = 0$$

This is equivalent to the following system of equations,

$$\begin{cases} \qquad\qquad\quad 4k_3 = 0 \\ \qquad\quad 2k_2 \qquad\;\; = 0 \\ \quad k_1 \qquad - 3k_3 = 0 \\ k_0 \quad - \quad k_2 \qquad\;\; = 0 \end{cases}$$

which has only the trivial solution $k_0 = k_1 = k_2 = k_3 = 0$.

At the moment, the definition of a list of vectors being 'linearly dependent' is that the list is 'not linearly independent'! That might get your head in a knot. Here is a more straightforward way to think about linearly depdendent lists of vectors.

Proposition 14 (Linear Combination of Preceding Vectors). *Let $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ be a list of vectors in a vector space V. The following statements are equivalent:*

1. *The list of vectors $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ is linearly dependent.*

2. *Either $\mathbf{v}_1 = \mathbf{0}$, or for some $r \in \{2, 3, \ldots, n\}$, $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_{r-1}$.*

Proof. $(1) \Rightarrow (2)$. Suppose that the list $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ is linearly dependent. This means that there are scalars k_1, k_2, ..., k_n, not all zero, such that

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = \mathbf{0}. \tag{2.12}$$

Let $r \in \{1, 2, \ldots, n\}$ be the largest index such that $k_r \neq 0$. (We are told that not all the k_i are zero, so this makes sense.) If $r = 1$, then (2.12) is simply the equation

$$k_1 \mathbf{v}_1 = \mathbf{0}, \quad \text{where } k_1 \neq 0.$$

Therefore $\mathbf{v}_1 = \mathbf{0}$ by Lemma 6, and we are done. On the other hand, suppose $r \neq 1$. Then (2.12) becomes the equation

$$k_1 \mathbf{v}_1 + k_2 \mathbf{v}_2 + \cdots + k_r \mathbf{v}_r = \mathbf{0}, \quad \text{where } k_r \neq 0.$$

By dividing by k_r, we can now solve for $\mathbf{v}_r$ in terms of the preceding vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r-1}$:

$$\therefore \mathbf{v}_r = -\frac{k_1}{k_r}\mathbf{v}_1 - \frac{k_2}{k_r}\mathbf{v}_2 - \cdots - \frac{k_{r-1}}{k_r}\mathbf{v}_{r-1}.$$

$(2) \Rightarrow (1)$. If $\mathbf{v}_1 = \mathbf{0}$, then the list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is certainly linearly dependent by Remark 12. On the other hand, suppose that for some $r \in \{2, 3, \ldots, n\}$, $\mathbf{v}_r$ is a linear combination of the preceding vectors, that is

$$\mathbf{v}_r = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_{r-1} \mathbf{v}_{r-1}.$$

Then clearly

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_{r-1} \mathbf{v}_{r-1} - \mathbf{v}_r + 0\mathbf{v}_{r+1} + \cdots + 0\mathbf{v}_n = \mathbf{0}$$

which shows that the list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is not linearly independent, since not all the coefficients are zero above — in particular the coefficient of $\mathbf{v}_r$ is -1. $\qquad \square$

I call this the 'Linear Combination of Preceding Vectors Proposition'. You should give all your theorems names, it helps you to remember them and become familiar with them.

Example 2.10. We saw in Example 2.8 that the list of vectors $\mathbf{f}_1 = (-1, 2)$, $\mathbf{f}_2 = (1, 1)$, $\mathbf{f}_3 = (2, -1)$ in $\mathbb{R}^3$ is linearly dependent. Therefore, by Proposition 14, we must have either:

- $\mathbf{f}_1 = \mathbf{0}$ (not the case), or

- $\mathbf{f}_2$ is a scalar multiple of $\mathbf{f}_1$ (not the case), or

- $\mathbf{f}_3$ is a linear combination of $\mathbf{f}_1$ and $\mathbf{f}_2$, which is true:

$$\mathbf{f}_3 = -\mathbf{f}_1 + \mathbf{f}_2$$

Proposition 15 (Bumping Off Proposition). *Suppose* $\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m\}$ *is a linearly independent list of vectors in a vector space V, and that $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ spans V. Then $m \leq n$.*

Proof. Start with the original spanning list of vectors

$$\mathcal{S} = \{\, \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\} \tag{2.13}$$

and consider the 'bloated' list

$$\mathcal{S}' = \{\mathbf{l}_1, \mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n\} \tag{2.14}$$

Since $\mathbf{l}_1$ is a linear combination of the $\mathbf{s}_i$, this list is linearly dependent by Remark 13. Thus, by Proposition 14 (Linear Combination of Preceding Vectors), one of the $\mathbf{s}$-vectors, say $\mathbf{s}_r$, can be expressed as a linear combination of the preceding vectors. We can then remove $\mathbf{s}_r$ from the list ('bump it off'), and the resulting list

$$\mathcal{S}_1 := \{\mathbf{l}_1, \mathbf{s}_1, \mathbf{s}_2, \cdots, \hat{\mathbf{s}_r}, \cdots, \mathbf{s}_n\} \qquad (\mathbf{s}_r \text{ omitted}) \tag{2.15}$$

still spans V. We can go on in this way, each time transferring another one of the $\mathbf{l}$-vectors into the list, and removing another one of the $\mathbf{s}$-vectors, to arrive at a new spanning list $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$, and so on, where list $\mathcal{S}_k$ has $(n-k)$ $\mathbf{s}$-vectors.

Suppose $m > n$. Then the list $\mathcal{S}_n$ will consist only of $\mathbf{l}$-vectors,

$$\mathcal{S}_n : \; \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_n$$

and it spans V by construction. But there are still some $\mathbf{l}$-vectors $\mathbf{l}_{n+1}, \ldots \mathbf{l}_m$ left over! This means that $\mathbf{l}_{n+1}$ is a linear combination of the vectors in $\mathcal{S}_n$, which contradicts the fact that the list $\{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m\}$ was linearly independent. So we cannot have $m > n$ and hence $m \leq n$. $\qquad \square$

Further exercises for 2.2

Exercise 42. Show that the list of vectors $(2, 3, 1)$, $(1, -1, 2)$, $(7, 3, c)$ is linearly dependent in $\mathbb{R}^3$ if and only if $c = 8$.

Exercise 43. The list of vectors in $\text{Mat}_{2,2}$ given by

$$\mathbf{v}_1 = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \, \mathbf{v}_2 = \begin{bmatrix} 0 & 1 \\ -2 & 1 \end{bmatrix}, \, \mathbf{v}_3 = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, \, \mathbf{v}_4 = \begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix}, \, \mathbf{v}_5 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is linearly independent (you will prove this in Exercise 52, but for the sake of this question you may assume it to be true). Go through the same steps as in Example 2.10 to find the first vector in the list which is either the zero vector or a linear combination of the preceding vectors.

2.3 Basis and dimension

Definition 16. A list of vectors $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ in a vector space V is called a *basis* for V if it is linearly independent and spans V.

There is a more direct way to think about a basis.

Proposition 17 (Bases give coordinates). *A list of vectors* $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ *in a vector space* V *is a basis for* V *if and only if every vector* $\mathbf{v} \in V$ *can be written as a linear combination*

$$\mathbf{v} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n \tag{2.16}$$

in precisely one *way. (That is, for each* $\mathbf{v} \in V$ *there* **exist** *scalars* a_1, a_2, $\ldots$, a_n *satisfying* (2.16), *and that moreover these scalars are* **unique**).

It is important to understand the mathematical phrase 'There exists a unique X satisfying Y'. It means two things. Firstly, that *there does exist an X* which satisfies Y. And secondly, that there is *no more than one X* which satisfies Y.

We call the scalars $a_1, a_2, \ldots, a_n$ appearing in (2.16) the *coordinates* of $\mathbf{v}$ in the basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$.

Proof. $\Rightarrow$. Suppose that the list of vectors $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis for V. Suppose $\mathbf{v} \in V$. Since the list of vectors $\mathcal{B}$ spans V, we know that we *can* write $\mathbf{v}$ as a linear combination of the vectors in the list in at least one way,

$$\mathbf{v} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n. \tag{2.17}$$

We need to show that this is the *only* way to express $\mathbf{v}$ as a linear combination of the vectors $\mathbf{e}_i$. Indeed, suppose that we also have

$$\mathbf{v} = b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2 + \cdots + b_n \mathbf{e}_n. \tag{2.18}$$

Subtracting these two equations gives

$$\mathbf{0} = (a_1 - b_1) \mathbf{e}_1 + (a_2 - b_2) \mathbf{e}_2 + \cdots + (a_n - b_n) \mathbf{e}_n.$$

Since the list of vectors $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_n$ is linearly independent, we conclude that

$$a_1 - b_1 = 0, \ a_2 - b_2 = 0, \ \cdots, \ a_n - b_n = 0.$$

That is, $a_1 = b_1$, $a_2 = b_2$, and so on up to $a_n = b_n$, and hence the expansion (2.17) is unique.

$\Leftarrow$. Conversely, suppose that every vector $\mathbf{v}$ can be written as a unique linear combination

$$\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n.$$

The fact that each $\mathbf{v}$ *can* be written as a linear combination of the vectors $\mathbf{e}_1$, $\mathbf{e}_2$, ..., $\mathbf{e}_n$ means that $\mathcal{B}$ spans V. We still need to show that this list $\mathcal{B}$ is linearly independent. So, suppose that there exist scalars b_1, b_2, ..., b_n such that

$$b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \cdots + b_n\mathbf{e}_n = \mathbf{0}. \tag{2.19}$$

We need to show that all the b_i must equal zero. We already know *one* possible solution of (2.19) : simply set each $b_i = 0$. But we are told that each vector (in particular, the vector $\mathbf{0}$) can be expressed as a linear combination of the $\mathbf{e}_i$ in exactly one way. Hence this must be the only solution, i.e. we must have $b_1 = b_2 = \cdots = b_n = 0$, and so the list $\mathcal{B}$ is linearly independent.

$\square$

Theorem 18 (Invariance of dimension)**.** *If $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ are bases of a vector space V, then $m = n$.*

Proof. This is a consequence of Proposition 15 (the Bumping Off Proposition). Since the $\mathbf{b}$-vectors are linearly independent and the $\mathbf{c}$-vectors span V, we have $m \leq n$. On the other hand, since the $\mathbf{c}$-vectors are linearly independent and the $\mathbf{b}$-vectors span V, we have $n \leq m$. Hence $m = n$. $\square$

Definition 19. A vector space V is *finite-dimensional* if it has a basis. In that case, the *dimension* of V is the number of elements in a basis for V. A vector space is *infinte-dimensional* if it is not finite-dimensional.

Note that the concept of 'dimension of a vector space' is only well-defined because of Theorem 18.

Example 2.11 (Standard basis for $\mathbb{R}^n$)**.** The list of vectors

$$\mathbf{e}_1 := (1, 0, \ldots, 0), \ \ \mathbf{e}_2 := (0, 1, \ldots, 0), \ \ \ldots, \ \ \mathbf{e}_n := (0, 0, \ldots, 0, 1)$$

is a basis for $\mathbb{R}^n$. We already saw in Example 2.6 that this list spans $\mathbb{R}^n$. We need to check that it is linearly independent. So, suppose that

$$a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n = \mathbf{0}.$$

Expanding out the left hand side in components using the definition of the standard basis vectors $\mathbf{e}_i$, this becomes the equation

$$(a_1, 0, 0, \ldots, 0) + (0, a_2, 0, \ldots, 0) + \cdots + (0, 0, 0, \ldots, a_n) = (0, 0, 0, \ldots, 0).$$

In other words, we have

$$(a_1, a_2, a_3, \ldots, a_n) = (0, 0, 0, \ldots, 0)$$

which says precisely that $a_1 = a_2 = a_3 = \cdots = a_n = 0$, which is what we needed to prove. Thus the list of vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ is linearly independent, and is hence a basis for $\mathbb{R}^n$. So $\mathbb{R}^n$ has dimension n.

Example 2.12. The list of polynomials

$$\mathbf{p}_0(x) := 1, \ \mathbf{p}_1(x) := x, \ \mathbf{p}_2(x) := x^2, \ \ldots, \ \mathbf{p}_n(x) := x^n$$

is a basis for Poly_n, so $\mathrm{Dim}\,\mathrm{Poly}_n = n + 1$. Indeed, this list spans Poly_n by definition, so we just need to check that it is linearly independent. Suppose that

$$a_0 \mathbf{p}_0 + a_1 \mathbf{p}_1 + a_2 \mathbf{p_2} + \cdots + a_n \mathbf{p}_n = \mathbf{0}.$$

This is an equation between functions, so it holds for all $x \in \mathbb{R}$! In other words, for all $x \in \mathbb{R}$, the following equation holds:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0 \tag{2.20}$$

But, we know from algebra that a polynomial equation of the form (2.20) with nonzero coefficients has *at most n* roots $x_1, x_2, \ldots, x_n$. So, in order for (2.20) to hold for *all* real numbers x, the coefficients must be zero, i.e. $a_0 = a_1 = a_2 = \cdots = a_n = 0$, which is what we needed to show.

Example 2.13. Suppose X is a finite set. Then $\mathrm{Fun}(X)$ is finite-dimensional, with dimension $|X|$, with basis given by the functions $\mathbf{f}_a$, $a \in X$, defined by:

$$\mathbf{f}_a(x) := \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

We will prove this in a series of exercises.

The formula on the right hand side of (2.21) occurs so often in mathematics we give it a symbol of its own, δ_{ab} (the 'Kronecker delta'). This symbol stands for the formula: "If $a = b$, return a 1. If $a \neq b$, return a 0". In this language, we can rewrite the definition of the functions $\mathbf{f}_a$ as

$$\mathbf{f}_a(x) := \delta_{ax}. \tag{2.22}$$

movebackintoexample (both)

Exercise 44. Suppose $X = \left\{ \text{🐦}, \text{🐭}, \text{✝} \right\}$.

1. Evaluate the function $\mathbf{f}_{🐦}$ at each $x \in X$.

2. Show that every function $\mathbf{f} \in \mathrm{Fun}(X)$ can be written as a linear combination of $\mathbf{f}_{🐦}$, $\mathbf{f}_{🐭}$ and $\mathbf{f}_{✝}$.

Exercise 45. Now let X be an arbitrary finite set. Consider the collection of functions

$$\mathbf{f}_a, \ a \in X$$

Show that this collection (i) spans $\mathrm{Fun}(X)$ and (ii) is linearly independent.

Example 2.14. Trig_n is $(2n + 1)$-dimensional, with basis

$$\mathbf{T}_0(x) := 1, \ \ \mathbf{T}_1(x) := \cos x, \ \mathbf{T}_2(x) := \sin x, \ \mathbf{T}_3(x) := \cos 2x,$$
$$\mathbf{T}_4(x) := \sin 2x, \ \ldots, \ \mathbf{T}_{2n-1}(x) := \cos nx, \ \mathbf{T}_{2n}(x) := \sin nx. \tag{2.23}$$

You know that these functions span Trig_n, by definition. They are also linearly independent, though we will not prove this.

Example 2.15. The dimension of $\mathrm{Mat}_{n,m}$ is nm, with basis given by the matrices
$$\mathsf{E}_{ij}, \quad i = 1 \ldots n, \ j = 1 \ldots m$$
which have a 1 in the ith row and jth column and zeroes everywhere else.

Usually A is a matrix, and A_{ij} is the element of the matrix at position (i,j). But now E_{ij} is a matrix in its own right! Its element at position (k,l) will be written as $(\mathsf{E}_{ij})_{kl}$. I hope you don't find this too confusing. In fact, we can write down an elegant formula for the elements of E_{ij} using the Kronecker delta symbol:

$$(\mathsf{E}_{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{2.24}$$

movebackintoramanujansays

movebackintoexample

Exercise 46. Check that (2.24) is indeed the correct formula for the matrix elements of E_{ij}.

Example 2.16. The standard basis of $\mathrm{Mat}_{2,2}$ is
$$\mathsf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Example 2.17. The standard basis of Col_n is
$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots, \ \mathbf{e}_n := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

We now consider dimensions of subspaces of vector spaces.

Proposition 20. *Let W be a subspace of a finite-dimensional vector space V. Then W is finite-dimensional, and $\mathrm{Dim}(W) \leq \mathrm{Dim}(V)$.*

Proof. Let $n = \mathrm{Dim}(V)$. If $W = \{\mathbf{0}\}$, then the statement is clearly true. Suppose $W \neq \{\mathbf{0}\}$. Choose a nonzero vector $\mathbf{e}_1 \in W$ and form the list $\mathcal{B}_1 = \{\mathbf{e}_1\}$. If it spans W, then we are done. If not, there exists $\mathbf{e}_2 \in W$ which is not a scalar multiple of $\mathbf{e}_1$. Now consider the list $\mathcal{B}_2 = \{\mathbf{e}_1, \mathbf{e}_2\}$. If it spans W, we are done. If not, there exists a vector $\mathbf{e}_3 \in W$ which is not a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. Consider the new list $\mathcal{B}_3 = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. And so on.

This process must eventually terminate. That is, there must exist some $k \geq 1$ such that $\mathcal{B}_k = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ spans W, and hence is a basis for W. This is because each list $\mathcal{B}_i$ is linearly independent (by construction, no vector is a linear combination of the preceding vectors). And it is impossible to construct a linearly independent list of vectors in W (and hence in V) which has more than n elements, by the Bumping Off Proposition. This also shows that $k \leq n$.

$\square$

It is good to have an example of an infinite-dimensional vector space.

Proposition 21. Poly *is infinite-dimensional.*

Proof. Suppose Poly is finite-dimensional. This means there exists a finite collection of polynomials $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ which spans Poly. But, let d be the highest degree of all the polynomials in the list $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$. Then $\mathbf{p} := x^{d+1}$ is a polynomial which is not in the span of $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, since adding polynomials together and multiplying them by scalars can never increase the degree. We have arrived at a contradiction. So our initial assumption cannot be correct, i.e. Poly cannot be finite-dimensional. $\square$

Example 2.18. We will not prove this here, but the following vector spaces are also infinite-dimensional:

- $\mathbb{R}^\infty$,

- $\mathrm{Fun}(X)$ where X is an infinite set,

- $\mathrm{Cont}(I)$ for any nonempty interval I, and

- $\mathrm{Diff}(I)$ for any open interval I.

2.3.1 Sifting

If we consider the proof of Proposition 15 (the 'Bumping off' Proposition) carefully, we find that it makes use of a *sifting algorithm*. This algorithm can actually be applied to *any* list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space. Consider each vector $\mathbf{v}_i$ in the list consecutively. If $\mathbf{v}_i$ is the zero vector, or if it is a linear combination of the preceding vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$, remove it from the list.

Example 2.19. Sift the following list of vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = (1, 2, -1), \qquad \mathbf{v}_2 = (0, 0, 0), \qquad \mathbf{v}_3 = (3, 6, -3)$$
$$\mathbf{v}_4 = (1, 0, 5), \qquad \mathbf{v}_5 = (5, 4, 13), \qquad \mathbf{v}_6 = (1, 1, 0).$$

We start with $\mathbf{v}_1$. Since it is not the zero vector, and is not a linear combination of any preceding vectors, it remains. We move on to $\mathbf{v}_2$, which is zero, so we remove it. We move on to $\mathbf{v}_3$, which by inspection is equal to $3\mathbf{v}_1$, so we remove it. We move on to $\mathbf{v}_4$. It is not zero, and cannot be expressed as a multiple of $\mathbf{v}_1$ (check this), so it remains. We move on to $\mathbf{v}_5$. We check if it can be written as a linear combination

$$\mathbf{v}_5 = a\mathbf{v}_1 + b\mathbf{v}_4$$

and find the solution $a = 2, b = 3$ (check this), so we remove it. Finally we move on to $\mathbf{v}_6$. We check if it can be written as a linear combination

$$\mathbf{v}_6 = a\mathbf{v}_1 + b\mathbf{v}_4$$

and find no solutions (check this), so it remains. Our final sifted list is

$$\mathbf{v}_1, \mathbf{v}_4, \mathbf{v}_6.$$

---

movebackintoexample

Exercise 47. Do the three 'check this' calculations above.

Sifting is a very useful way to construct a basis of a vector space!

Lemma 22. *If a list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ spans a vector space V, then sifting the list will result in a basis for V.*

Proof. At each step, the vector that is removed from the list is either the zero vector, or a linear combination of the vectors before it. So if we remove this vector, the resulting list will still span V. Thus by the end of the process, the final sifted list of vectors still spans V.

To see that the final sifted list is linearly independent, we can apply Proposition 14. By construction, no vector in the final sifted list is a linear combination of the preceding vectors (if it was, it would have been removed!). Hence the final sifted list is not linearly dependent, so it must be linearly independent! □

Corollary 23. *Any linearly independent list of vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ *in a finite-dimensional vector space* V *can be extended to a basis of* V.

Proof. Since V is finite-dimensional, it has a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Now consider the list

$$L : \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$$

which clearly spans V. By sifting this list, we will arrive at a basis for V, by Lemma 22. Some of the $\mathbf{e}$-vectors may have been removed. But none of the $\mathbf{v}$-vectors will have been removed, since that would mean some $\mathbf{v}_i$ is a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$, which is impossible, as $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is linearly independent list. Hence after sifting the list L we indeed extend our original list $\mathbf{v}_1, \ldots, \mathbf{v}_k$ to a basis of V. □

Corollary 24. *If* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *is a linearly independent list of* n *vectors in an* n-*dimensional vector space* V, *then it is a basis.*

Proof. By Corollary 23, we can extend $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ to a basis for V. But V has dimension n, so the basis must contain only n vectors by Theorem 18 (Invariance of Dimension). So we have not added any vectors at all! Hence our original list was already a basis. □

Example 2.20. In Example 2.9 we showed that the list of polynomials

$$\mathbf{q}_0(x) := 1, \ \mathbf{q}_1(x) := x, \ \mathbf{q}_2(x) := 2x^2 - 1, \ \mathbf{q}_3(x) := 4x^3 - 3x$$

is linearly independent in Poly_3. Since $\mathrm{Dim} \, \mathrm{Poly}_3 = 4$, we see that it is a basis of Poly_3.

In Exercise 38, you showed that $\mathbf{q}_0, \ldots, \mathbf{q}_3$ is a basis for Poly_3 by 'brute force'. This new method is *different*!

Further exercises for 2.3

Exercise 48. Sift the list of vectors

$$\mathbf{v}_1 = (0,0,0), \qquad \mathbf{v}_2 = (1,0,-1), \qquad \mathbf{v}_3 = (1,2,3)$$
$$\mathbf{v}_4 = (3,4,5), \qquad \mathbf{v}_5 = (4,8,12), \qquad \mathbf{v}_6 = (1,1,0).$$

Exercise 49. Let V be a vector space of dimension n. State whether each of the following statements is true or false. If it is true, prove it. If it is false, give a counterexample.

(a) Any linearly independent list of vectors in V contains at most n vectors.

(b) Any list of vectors which spans V contains at least n vectors.

Exercise 50. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a linearly independent list of vectors in a vector space V. Suppose that $\mathbf{v}$ is a vector in V which cannot be written as a linear combination of the vectors from $\mathcal{B}$. Show that the list $\mathcal{B}' = \{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}\}$ is still linearly independent. (Hint: Use the Linear Combination of Preceding Vectors Proposition.)

Exercise 51. Complete the proof of the following lemma.

Lemma. Suppose V is a vector space of dimension n. Then any linearly independent set of n vectors in V is a basis for V.

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a linearly independent set of vectors in V.

Suppose that $\mathcal{B}$ is *not* a basis for V.
Therefore, $\mathcal{B}$ does not span V, since ... (a)
Therefore, there exists $\mathbf{v} \in V$ such that ... (b)
Now, add $\mathbf{v}$ to the list $\mathcal{B}$ to obtain a new list $\mathcal{B}' := \ldots$ (c)
The new list $\mathcal{B}'$ is linearly independent because ... (d)
This is a contradiction because ... (e)
Hence, $\mathcal{B}$ must be a basis for V.

Exercise 52. Use Exercise 49(a) to show that the list of matrices in $\mathrm{Mat}_{2,2}$ in Exercise 43 is linearly dependent.

Exercise 53. In each case, use the results in Exercises 49 and 51 to determine if $\mathcal{B}$ is a basis for V.

(a) $V = \mathrm{Poly}_2$, $\mathcal{B} = \{2 + x^2,\ 1 - x,\ 1 + x - 3x^2,\ x - x^2\}$

(b) $V = \mathrm{Mat}_{2,2}$, $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 3 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \right\}$

(c) $V = \mathrm{Trig}_2$, $\mathcal{B} = \left\{ \sin^2 x,\ \cos^2 x,\ 1 - \sin 2x,\ \cos 2x + 3\sin 2x \right\}$

Exercise 54. Let $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ be a linearly independent list of vectors in a vector space V. State whether each of the following statements is true or false. If it is true, prove it. If it is false, give a counterexample. (Hint: Use the definition of linear independence.)

(a) The list $\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{u} + \mathbf{w}\}$ is linearly independent.

(b) The list $\{\mathbf{u} - \mathbf{v}, \mathbf{v} - \mathbf{w}, \mathbf{u} - \mathbf{w}\}$ is linearly independent.

Exercise 55. For each of the following, show that V is a subspace of Poly_2, find a basis for V, and compute $\mathrm{Dim}\, V$.

(a) $V = \{p \in \mathrm{Poly}_2 : p(2) = 0\}$

(b) $V = \{p \in \mathrm{Poly}_2 : xp'(x) = p(x)\}$

2.4 Coordinate vectors

Definition 25. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V, and let $\mathbf{v} \in V$. Write

$$\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + \mathbf{a}_n\mathbf{b}_n\,.$$

Then the column vector

$$[\mathbf{v}]_{\mathcal{B}} := \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathrm{Col}_n$$

is called the *coordinate vector of* $\mathbf{v}$ *with respect to the basis* $\mathcal{B}$.

I indicate that a collection of things is a *list* (where the order matters) and not merely a *set* (where the order does not matter) using my own home-made symbols {}. A basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ is a list of vectors. The order of the vectors matters because it affects the coordinate vector $[\mathbf{v}]_{\mathcal{B}}$.

Example 2.21. Find the coordinate vector of $\mathbf{p} = 2x^2 - 2x + 3$ with respect to the basis $\mathcal{B} = \{1 + x, x^2 + x - 1, x^2 + x + 1\}$ of Poly_3.

Solution. We write $\mathbf{p}$ as a linear combination of polynomials from the basis $\mathcal{B}$:

$$2x^2 - 2x + 3 = -4(1 + x) - \frac{5}{2}(x^2 + x - 1) + \frac{9}{2}(x^2 + x + 1)$$

$$[\mathbf{p}]_{\mathcal{B}} = \begin{bmatrix} -4 \\ -\frac{5}{2} \\ \frac{9}{2} \end{bmatrix}$$

movebackintoexample

Exercise 56. Check this!

Example 2.22. Find the coordinate vectors of $\mathbf{v}$ and $\mathbf{w}$ in Figure 2.2 with respect to the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$.

Solution. By inspection, we see that $\mathbf{v} = 2\mathbf{b}_1 - \mathbf{b}_2$, so that

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Also by inspection, we see that $\mathbf{w} = -3\mathbf{b}_1 + 2\mathbf{b}_2$, so that

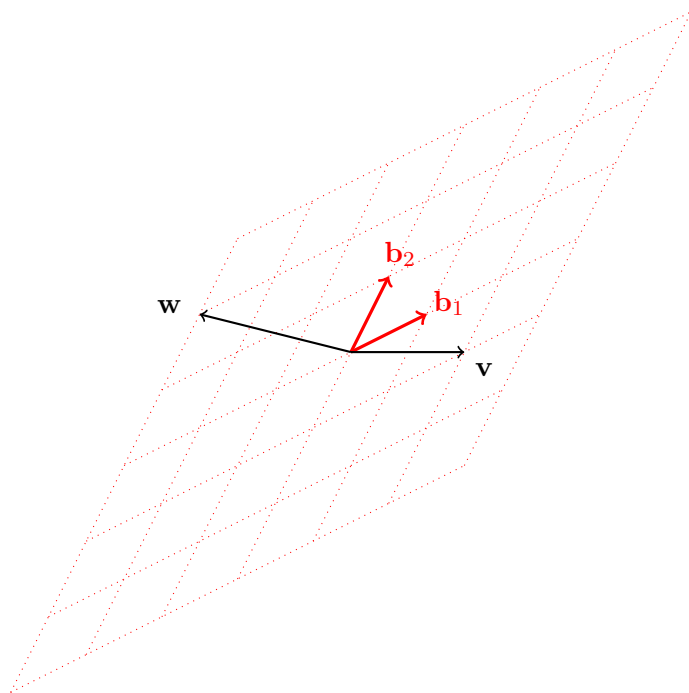$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$$

Example 2.23. Find the coordinate vector of the function $\mathbf{f}$ given by

$$\mathbf{f}(x) = \sin^2 x - \cos^3 x$$

with respect to the standard basis

$$\mathcal{S} = \{1, \cos x, \sin x, \cos 2x, \sin 2x, \cos 3x, \sin 3x\}$$

of Trig_3.

Figure 2.2: The basis B for $\mathbb{R}^2$.

Solution. Using the addition formulae for sin and cos as in Exercise 23, we compute

$$\sin^2 x - \cos^3 x = \frac{1}{2} - \frac{3}{4}\cos x - \frac{1}{2}\cos 2x - \frac{1}{4}\cos 3x.$$

Hence

$$[\mathbf{f}]_{\mathcal{S}} = \begin{bmatrix} \frac{1}{2} \\ -\frac{3}{4} \\ 0 \\ -\frac{1}{2} \\ 0 \\ -\frac{1}{4} \\ 0 \end{bmatrix}$$

.

movebackintoexample

Exercise 57. Check this!

Lemma 26. *Let $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be a basis for a vector space V. Then for all vectors $\mathbf{v}, \mathbf{w} \in V$ and all scalars k we have*

(a) $[\mathbf{v} + \mathbf{w}]_\mathcal{B} = [\mathbf{v}]_\mathcal{B} + [\mathbf{w}]_\mathcal{B}$

(b) $[k\mathbf{v}]_\mathcal{B} = k[\mathbf{v}]_\mathcal{B}$

Proof. (a) Suppose that

$$\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n$$

and

$$\mathbf{w} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \cdots + b_n\mathbf{e}_n.$$

Then, using the rules of a vector space, we compute

$$\mathbf{v} + \mathbf{w} = (a_1 + b_1)\mathbf{e}_1 + (a_2 + b_2)\mathbf{e}_2 + \cdots + (a_n + b_n)\mathbf{e}_n.$$

From this we read off that

$$[\mathbf{v} + \mathbf{w}]_\mathcal{B} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}.$$

The proof of (b) is similar. $\qquad\qquad\square$

Further exercises for 2.4

Exercise 58. Prove Lemma 26(b) in the case where V is two-dimensional, so that $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$. Justify each step using the rules of a vector space.

Exercise 59. Let $\mathcal{B}$ be the basis of $\mathrm{Mat}_{2,2}$ given by

$$\mathsf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathsf{B}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathsf{B}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathsf{B}_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Determine $[\mathsf{A}]_\mathcal{B}$, where

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

Exercise 60. In Exercise 55, you were asked to find a basis $\mathcal{B}$ for the vector space

$$V := \{p \in \mathrm{Poly}_2 : p(2) = 0\}.$$

Consider $p(x) = x^2 + x - 6$.

(a) Show that $p \in V$.

(b) Determine the coordinate vector of p with respect to your basis $\mathcal{B}$, i.e. determine $[p]_\mathcal{B}$.

2.5   Change of basis

2.5.1   Coordinate vectors are different in different bases

Suppose that $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ are two different bases for $\mathbb{R}^2$, shown below:
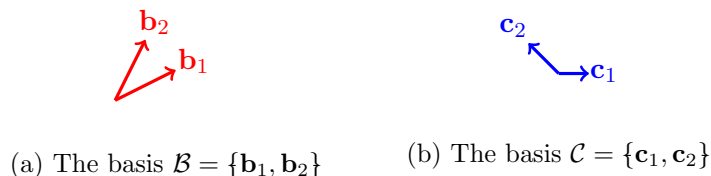

(a) The basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ (b) The basis $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$

Figure 2.3: Two different bases for $\mathbb{R}^2$

Suppose we are given a vector $\mathbf{w} \in \mathbb{R}^2$:

We would like to compute the coordinate vector of the *same* vector $\mathbf{w}$ with respect to the two different bases $\mathcal{B}$ and $\mathcal{C}$.

For this particular $\mathbf{w}$, from Figure 2.3(a), we see that in the basis $\mathcal{B}$, we have

$$\mathbf{w} = -3\mathbf{b}_1 + 2\mathbf{b}_2 \qquad \therefore \ [\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}. \tag{2.25}$$

On the other hand, in the basis $\mathcal{C}$, we have

$$\mathbf{w} = \mathbf{c}_1 - 3\mathbf{c}_2 \qquad \therefore \ [\mathbf{w}]_{\mathcal{C}} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}. \tag{2.26}$$

tikzalign

So, the *same* vector $\mathbf{w}$ has different coordinate vectors $[\mathbf{w}]_{\mathcal{B}}$ and $[\mathbf{w}]_{\mathcal{C}}$ with respect to the bases $\mathcal{B}$ and $\mathcal{C}$!

2.5.2   Changing from one basis to another

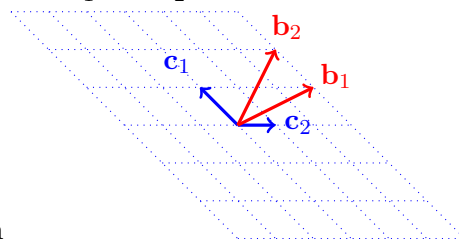Now, suppose we only knew $[\mathbf{w}]_{\mathcal{B}}$, the coordinate vector of $\mathbf{w}$ in the basis $\mathcal{B}$. In other words, suppose we only knew that

$$[\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix},$$

that is, $\mathbf{w} = -3\mathbf{b}_1 + 2\mathbf{b}_2$. How could we compute $[\mathbf{w}]_{\mathcal{C}}$, the coordinate vector of $\mathbf{w}$ in the basis $\mathcal{C}$?

(b) $\mathbf{w} = \mathbf{c}_1 - 3\mathbf{c}_2$

(a) $\mathbf{w} = -3\mathbf{b}_1 + 2\mathbf{b}_2$
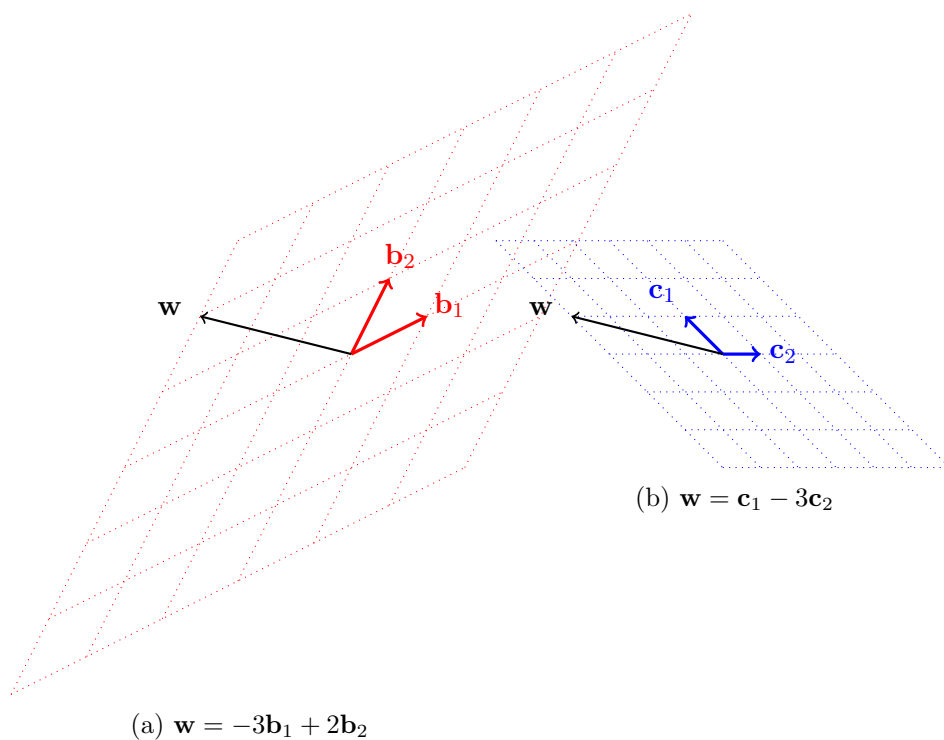
The best way is to express each vector in the basis $\mathcal{B}$ as a linear combination of the basis vectors in $\mathcal{C}$. In the next figure, the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are dis-

played against the background of the basis $\mathcal{C}$: tikzalign
We read off that:

$$\mathbf{b}_1 = \mathbf{c}_1 + 3\mathbf{c}_2 \qquad (2.27)$$
$$\mathbf{b}_2 = 2\mathbf{c}_1 + 3\mathbf{c}_2 \qquad (2.28)$$

Therefore, we compute:

$$\begin{aligned} \mathbf{w} &= -3\mathbf{b}_1 + 2\mathbf{b}_2 \\ &= -3(\mathbf{c}_1 + 3\mathbf{c}_2) + 2(2\mathbf{c}_1 + 3\mathbf{c}_2) \\ &= \mathbf{c}_1 - 3\mathbf{c}_2 \end{aligned}$$

From this we read off that

$$[\mathbf{w}]_{\mathcal{C}} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \qquad (2.29)$$

which is the right answer, as we know from (2.26).

In fact, this calculation can be phrased in terms of matrices.

Definition 27. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases for a vector space V. The *change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$* is the $n \times n$ matrix whose columns are the coordinate vectors $[\mathbf{b}_1]_\mathcal{C}, \ldots, [\mathbf{b}_n]_\mathcal{C}$:

$$\mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}} := \left[\left[\mathbf{b}_1 \right]_\mathcal{C} \left[\mathbf{b}_2 \right]_\mathcal{C} \cdots \left[\mathbf{b}_n \right]_\mathcal{C} \right]$$

Example 2.24. In our running example, we see from (2.27) and (2.28) that

$$[\mathbf{b}_1]_\mathcal{C} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad [\mathbf{b}_2]_\mathcal{C} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Hence the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$ is

$$\mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix}.$$

Before we move on, we need to recall something about matrix multiplication. Suppose you collect together m column vectors to form a matrix:

$$\left[\mathsf{C}_1 \, \mathsf{C}_2 \, \cdots \, \mathsf{C}_m \right]$$

(For instance, our change-of-basis matrix $\mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}}$ was formed in this way.) Then the product of this matrix with a column vector can be computed as follows:

$$\left[\mathsf{C}_1 \, \mathsf{C}_2 \, \cdots \, \mathsf{C}_m \right] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = a_1 \mathsf{C}_1 + a_2 \mathsf{C}_2 + \cdots + a_n \mathsf{C}_m \qquad (2.30)$$

Exercise 61. Prove the above formula!

We can now prove the following theorem.

Theorem 28 (Change of basis). *Suppose that $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ are bases for a vector space V, and let $\mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}}$ be the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$. Then for all vectors $\mathbf{v}$ in V,*

$$[\mathbf{v}]_\mathcal{C} = \mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}} [\mathbf{v}]_\mathcal{B}. \qquad (2.31)$$

Proof. Let $\mathbf{v} \in V$. Expand it in the basis $\mathcal{B}$:

$$\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n, \quad \text{i.e. } [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

Then,

$$\begin{aligned} [\mathbf{v}]_{\mathcal{C}} &= [a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n]_{\mathcal{C}} \\ &= a_1[\mathbf{b}_1]_{\mathcal{C}} + \cdots + a_n[\mathbf{b}_n]_{\mathcal{C}} \qquad \text{(by Lemma 26)} \\ &= \begin{bmatrix} \begin{bmatrix} \\ \mathbf{b}_1 \\ \\ \end{bmatrix}_{\mathcal{C}} \begin{bmatrix} \\ \mathbf{b}_2 \\ \\ \end{bmatrix}_{\mathcal{C}} \cdots \begin{bmatrix} \\ \mathbf{b}_n \\ \\ \end{bmatrix}_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \qquad \text{(by (2.30))} \\ &= \mathsf{P}_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}. \end{aligned}$$

$\square$

Example 2.25. In our running example, the theorem says that for *any* vector $\mathbf{v} \in \mathbb{R}^2$,

$$[\mathbf{v}]_{\mathcal{C}} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix} [\mathbf{v}]_{\mathcal{B}}.$$

In particular, this must hold for our vector $\mathbf{w}$, whose coordinate vector in the basis $\mathbf{B}$ was:

$$[\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}.$$

So in this case, the theorem is saying that

$$[\mathbf{w}]_{\mathcal{C}} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

which agrees with our previous calculation (2.29)!

Further exercises for 2.5

Exercise 62. This is a continuation of Exercise 59. Consider the following two bases for $\text{Mat}_{2,2}$:

$$\mathcal{B} = \left\{ \mathsf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathsf{B}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \mathsf{B}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathsf{B}_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

$$\mathcal{C} = \left\{ \mathsf{C}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{C}_2 = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, \mathsf{C}_3 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \mathsf{C}_4 = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \right\}$$

(a) Determine the change-of-basis matrices $\mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}}$ and $\mathsf{P}_{\mathcal{B} \leftarrow \mathcal{C}}$.

(b) Determine $[\mathsf{A}]_{\mathcal{B}}$ and $[\mathsf{A}]_{\mathcal{C}}$ where

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

(c) Check that $[\mathsf{A}]_{\mathcal{C}} = \mathsf{P}_{\mathcal{C} \leftarrow \mathcal{B}}[\mathsf{A}]_{\mathcal{B}}$ and that $[\mathsf{A}]_{\mathcal{B}} = \mathsf{P}_{\mathcal{B} \leftarrow \mathcal{C}}[\mathsf{A}]_{\mathcal{C}}$.

Exercise 63. Compute the change-of-basis matrix $\mathsf{P}_{\mathcal{B} \leftarrow \mathcal{S}}$ from the standard basis

$$\mathcal{S} = \{1, \cos x, \sin x, \cos 2x, \sin 2x\}$$

of Trig_2 to the basis

$$\mathcal{B} = \{1, \cos x, \sin x, \cos^2 x, \sin^2 x\}.$$

Chapter 3

Linear maps

3.1   Definition and examples

Recall that a *function* (or a *map*) $f : X \to Y$ from a set X to a set Y is simply a rule which assigns to each element of X an element $f(x)$ of Y. We write $x \mapsto f(x)$ to indicate that an element $x \in X$ maps to $f(x) \in Y$ under the function f. See Figure 3.1. Two functions $f, g : X \to Y$ are *equal* if $f(x) = g(x)$ for all x in X.
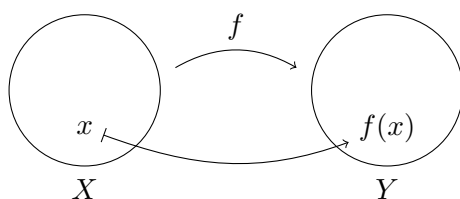
Figure 3.1:   A function $f : X \to Y$.

Definition 29. Let V and W be vector spaces. A *linear map* from V to W is a function $T : V \to W$ satisfying:

- $T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}')$ for all vectors $\mathbf{v}, \mathbf{v}' \in V$

- $T(k\mathbf{v}) = kT(\mathbf{v})$ for all vectors $\mathbf{v} \in V$ and scalars $k \in \mathbb{R}$

Another name for a linear map is a *linear transformation.*

Example 3.1 (Identity map)**.** Let V be a vector space. The function

$$\mathrm{id}_V : V \to V$$
$$\mathbf{v} \mapsto \mathbf{v}$$

is called the *identity map* on V. It is clearly a linear map, since

$$\mathrm{id}_V(\mathbf{v} + \mathbf{w}) = \mathbf{v} + \mathbf{w}$$
$$= \mathrm{id}_V(\mathbf{v}) + \mathrm{id}_V(\mathbf{w})$$

and

$$\mathrm{id}_V(k\mathbf{v}) = k\mathbf{v}$$
$$= k\,\mathrm{id}_V(\mathbf{v}).$$

Example 3.2 (Projection)**.** The function

$$T : \mathbb{R}^2 \to \mathbb{R}$$
$$(x, y) \mapsto x$$

which projects vectors onto the x-axis is a linear map.

Let us check additivity algebraically:

$$T\left((x_1, y_1)\right) + \left((x_2, y_2)\right) \overset{?}{=} T\left((x_1, y_1)\right) + T\left((x_2, y_2)\right)$$

$$\mathrm{LHS} = T\left((x_1 + x_2, y_1 + y_2)\right) \qquad \mathrm{RHS} = x_1 + x_2$$
$$= x_1 + x_2$$

Here is a a graphical version of this proof: tikzalign
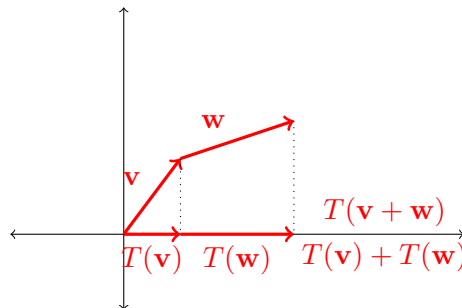
movebackintoexample

Exercise 64. Prove algebraically that $T(k\mathbf{v}) = kT(\mathbf{v})$, so that T is a linear map.

Example 3.3 (Rotation). Fix an angle θ. The function

$$R : \mathbb{R}^2 \to \mathbb{R}^2$$

$$\mathbf{v} \mapsto \text{rotation of } \mathbf{v} \text{ counterclockwise through angle } \theta$$

is a linear map, by a similar graphical argument as in Example 3.2. tikzalign

---
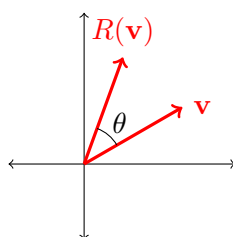
Example 3.4. (Cross product with a fixed vector) Fix a vector $\mathbf{w} \in \mathbb{R}^3$. The function

$$C : \mathbb{R}^3 \to \mathbb{R}^3$$

$$\mathbf{v} \mapsto \mathbf{w} \times \mathbf{v}$$

is a linear map because of the properties of the cross-product,

$$\mathbf{w} \times (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{w} \times \mathbf{v}_1 + \mathbf{w} \times \mathbf{v}_2$$

$$\mathbf{w} \times (k\mathbf{v}) = k\mathbf{w} \times \mathbf{v}.$$

Example 3.5. (Dot product with a fixed vector) Fix a vector $\mathbf{u} \in \mathbb{R}^3$. The function

$$D : \mathbb{R}^3 \to \mathbb{R}$$

$$\mathbf{v} \mapsto \mathbf{u} \cdot \mathbf{v}$$

(here · is the dot product of vectors, not scalar multiplication!) is a linear map, because of the properties

$$\mathbf{u} \cdot (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{u} \cdot \mathbf{v}_1 + \mathbf{u} \cdot \mathbf{v}_2$$
$$\mathbf{u} \cdot (k\mathbf{v}) = k\mathbf{u} \cdot \mathbf{v}.$$

We will soon see that *all* linear maps $\mathbb{R}^3 \to \mathbb{R}$ (indeed all linear maps $\mathbb{R}^n \to \mathbb{R}$) are of this form.

Example 3.6. (Differentiation) The operation 'take the derivative' can be interpreted as a linear map

$$D : \text{Poly}_n \to \text{Poly}_{n-1}$$
$$\mathbf{p} \mapsto \mathbf{p}'.$$

For example, $D(2x^3 - 6x + 2) = 6x - 6$.

movebackintoexample

Exercise 65. (a) Why is D a map from Poly_n to Poly_{n-1}? (b) Check that D is linear.

Example 3.7. (Antiderivative) The operation 'find the unique antiderivative with zero constant term' can be interpreted as a linear map

$$A : \text{Poly}_n \to \text{Poly}_{n+1}$$
$$\mathbf{p} \mapsto \int_0^x \mathbf{p}(t)\, dt$$

For example, $A(2x^3 - 6x + 2) = 4x^4 - 3x^2 + 2x$.

movebackintoexample

Exercise 66. (a) Check this last sentence. (b) Why is A a map from Poly_n to Poly_{n+1}? (c) Check that A is linear.

Example 3.8 (Shift map)**.** Define the 'shift forward by 1' map

$$S : \text{Poly}_n \to \text{Poly}_n$$
$$\mathbf{p} \mapsto S(\mathbf{p})$$

by $S(\mathbf{p})(x) = \mathbf{p}(x - 1)$. *** Consider the case $n = 3$. In terms of the standard basis

$$\mathbf{p}_0(x) = 1, \ \mathbf{p}_1(x) = x, \ \mathbf{p}_2(x) = x^2, \ \mathbf{p}_3(x) = x^3$$

of Poly_3, we have:

$$S(\mathbf{p}_0) = \mathbf{p}_0$$
$$S(\mathbf{p}_1) = \mathbf{p}_1 - \mathbf{p}_0$$
$$S(\mathbf{p}_2) = \mathbf{p}_2 - 2\mathbf{p}_1 + \mathbf{p}_0$$
$$S(\mathbf{p}_3) = \mathbf{p}_3 - 3\mathbf{p}_2 + 3\mathbf{p}_1 - \mathbf{p}_0$$

---

movebackintoexample (both)

Exercise 67. Check this.

***

Exercise 68. Check that S is a linear map.

Example 3.9 (Matrices give rise to linear maps)**.** Every $n \times m$ matrix A induces a *linear map*

$$T_{\mathsf{A}} : \text{Col}_m \to \text{Col}_n$$
$$\mathbf{v} \mapsto \mathsf{A}\mathbf{v} .$$

That is, $T_{\mathsf{A}}(\mathbf{v}) := \mathsf{A}\mathbf{v}$ is the matrix product of A with the column vector $\mathbf{v}$. The fact that T_{A} is indeed a linear map follows from the linearity of matrix multiplication (Proposition 69 parts 2 and 3).

Note that an $n \times m$ matrix gives a linear map from Col_m to Col_n!

Lemma 30. *Suppose* $T : V \to W$ *is a linear map. Then*

1. $T(\mathbf{0}_V) = \mathbf{0}_W$

2. $T(-\mathbf{v}) = -T(\mathbf{v})$ *for all vectors* $\mathbf{v} \in V$.

Proof.

$$
\begin{aligned}
\text{(i)} \quad T(\mathbf{0}_V) &= T(0\mathbf{0}_V) && \text{(R8 applied to } \mathbf{v} = \mathbf{0}_V \in V) \\
&= 0T(\mathbf{0}_V) && \text{(T is linear)} \\
&= \mathbf{0}_W && \text{(R8 applied to } \mathbf{v} = T(\mathbf{0}_V) \in W)
\end{aligned}
$$

$$
\begin{aligned}
\text{(ii)} \quad T(-\mathbf{v}) &= T((-1)\mathbf{v}) && \text{(defn of } -\mathbf{v} \text{ in } V) \\
&= (-1)T(\mathbf{v}) && \text{(T is linear)} \\
&= -T(\mathbf{v}) && \text{(defn of } -T(\mathbf{v}) \text{ in } W)
\end{aligned}
$$

$\square$

The next result is very important. It tells us that if we know how a linear map T acts on a basis, then we know how it acts on the whole vector space (this is the 'uniqueness' part). Moreover, we are free to write down any willy-nilly formula for what T does on the basis vectors, and we are guaranteed that this will always extend to a linear map defined on the whole vector space (this is the 'existence' part).

Proposition 31 (Sufficient to Define a Linear Map on a Basis). *Suppose* $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ *is a basis for* V *and* $\mathbf{w}_1, \ldots, \mathbf{w}_m$ *are vectors in* W. *Then there exists a unique linear map* $T : V \to W$ *such that*

$$
T(\mathbf{e}_i) = \mathbf{w}_i, \quad i = 1 \ldots m.
$$

Proof. Proof of existence. To define a linear map T, we must define $T(\mathbf{v})$ for each vector $\mathbf{v}$. We can write $\mathbf{v}$ in terms of its coordinate vector $[\mathbf{v}]_{\mathcal{B}}$ with respect to the basis $\mathcal{B}$ as

$$
\mathbf{v} = [\mathbf{v}]_{\mathcal{B},1}\mathbf{e}_1 + \cdots + [\mathbf{v}]_{\mathcal{B},m}\mathbf{e}_m \tag{3.1}
$$

where $[\mathbf{v}]_{\mathcal{B},i}$ is the entry at row i of the coordinate vector $[\mathbf{v}]_{\mathcal{B}}$. We define

$$
T(\mathbf{v}) := [\mathbf{v}]_{\mathcal{B},1}\mathbf{w}_1 + [\mathbf{v}]_{\mathcal{B},2}\mathbf{w}_2 + \cdots [\mathbf{v}]_{\mathcal{B},m}\mathbf{w}_m. \tag{3.2}
$$

We clearly have $T(\mathbf{e}_i) = \mathbf{w}_i$. To complete the proof of existence, we must show that T is linear:

$$
\begin{aligned}
T(\mathbf{v} + \mathbf{v}') &= [\mathbf{v} + \mathbf{v}']_{\mathcal{B},1}\mathbf{w}_1 + \cdots + [\mathbf{v} + \mathbf{v}']_{\mathcal{B},m}\mathbf{w}_m \\
&= \big([\mathbf{v}]_{\mathcal{B},1} + [\mathbf{v}']_{\mathcal{B},1}\big)\mathbf{w}_1 + \cdots + \big([\mathbf{v}]_{\mathcal{B},m} + [\mathbf{v}']_{\mathcal{B},m}\big)\mathbf{w}_m \quad ([\mathbf{v} + \mathbf{v}']_{\mathcal{B}} = [\mathbf{v}]_{\mathcal{B}} + [\mathbf{v}']_{\mathcal{B}}) \\
&= [\mathbf{v}]_{\mathcal{B},1}\mathbf{w}_1 + \cdots + [\mathbf{v}]_{\mathcal{B},m}\mathbf{w}_m + \\
&\qquad [\mathbf{v}']_{\mathcal{B},1}\mathbf{w}_1 + \cdots + [\mathbf{v}']_{\mathcal{B},m}\mathbf{w}_m \\
&= T(\mathbf{v}) + T(\mathbf{v}').
\end{aligned}
$$

Similarly, one can check that $T(k\mathbf{v}) = kT(\mathbf{v})$, which completes the proof of existence.

Proof of uniqueness. Suppose that $S, T : V \to W$ are linear maps with

$$S(\mathbf{e}_i) = \mathbf{w}_i, \quad \text{and} \quad T(\mathbf{e}_i) = \mathbf{w}_i, \quad i = 1 \dots m. \tag{3.3}$$

Then,

$$
\begin{aligned}
S(\mathbf{v}) &= S\big([\mathbf{v}]_{\mathcal{B},1}\mathbf{e}_1 + \cdots + [\mathbf{v}]_{\mathcal{B},m}\mathbf{e}_m\big) \\
&= [\mathbf{v}]_{\mathcal{B},1}S(\mathbf{e}_1) + \cdots + [\mathbf{v}]_{\mathcal{B},m}S(\mathbf{e}_m) && (S \text{ is linear}) \\
&= [\mathbf{v}]_{\mathcal{B},1}\mathbf{w}_1 + \cdots + [\mathbf{v}]_{\mathcal{B},m}\mathbf{w}_m && (\text{since } S(\mathbf{e}_i) = \mathbf{w}_i) \\
&= [\mathbf{v}]_{\mathcal{B},1}T(\mathbf{e}_1) + \cdots + [\mathbf{v}]_{\mathcal{B},m}T(\mathbf{e}_m) && (\text{since } T(\mathbf{e}_i) = \mathbf{w}_i) \\
&= T([\mathbf{v}]_{\mathcal{B},1}\mathbf{e}_1 + \cdots + [\mathbf{v}]_{\mathcal{B},m}\mathbf{e}_m) && (T \text{ is linear}) \\
&= T(\mathbf{v}).
\end{aligned}
$$

Hence $S = T$, in other words the linear map satisfying (3.3) is unique. $\quad\square$

Example 3.10. As an application of Proposition 31, we can define a linear map

$$T : \mathrm{Col}_2 \to \mathrm{Fun}(\mathbb{R})$$

simply by defining its action on the standard basis of Col_2. For instance, we may set

$$
\begin{aligned}
\mathbf{e}_1 &\to f_1 \\
\mathbf{e}_2 &\to f_2
\end{aligned}
$$

The point is that we are free to send $\mathbf{e}_1$ and $\mathbf{e}_2$ to *any functions* f_1 and f_2 we like, and we are assured that this will give a well-defined linear map $T : \mathrm{Col}_2 \to \mathrm{Fun}(\mathbb{R})$. For instance, we might set $f_1(x) = \sin x$ and $f_2(x) = |x|$. Then the general formula for T is

$$T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right)(x) = a\sin x + b|x|$$

Example 3.11 (Rotation map on standard basis). Let us compute the action of the 'counterclockwise rotation by θ' map R from Example 3.3

with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2$ of $\mathbb{R}^2$. tikzalign
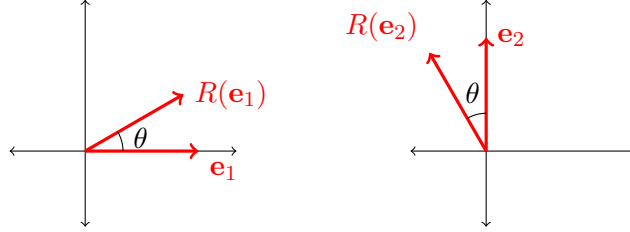
From the figure, we have:

$$R(\mathbf{e}_1) = (\cos\theta, \sin\theta) \qquad R(\mathbf{e}_2) = (-\sin\theta, \cos\theta)$$

so that

$$R(\mathbf{e}_1) = \cos\theta\,\mathbf{e}_1 + \sin\theta\,\mathbf{e}_2, \quad R(\mathbf{e}_2) = -\sin\theta\,\mathbf{e}_1 + \cos\theta\,\mathbf{e}_2$$

Now that we know the action of R on the standard basis vectors, we can compute its action on an arbitrary vector $(x, y) \in \mathbb{R}^2$:

$$\begin{aligned}
R\left((x,y)\right) &= R(x\mathbf{e}_1 + y\mathbf{e}_2) \\
&= xR(\mathbf{e}_1) + yR(\mathbf{e}_2) \\
&= x(\cos\theta, \sin\theta) + y(-\sin\theta, \cos\theta) \\
&= (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta).
\end{aligned}$$

Further exercises for 3.1

Exercise 69. Let V be a vector space, and let $\mathbf{a} \neq \mathbf{0}$ be a fixed vector. Let Define the map T as follows:

$$T : V \to V$$
$$\mathbf{v} \mapsto \mathbf{a} + \mathbf{v}$$

(a) Is T a linear map? (Yes or no)

(b) Prove your assertion from (a).

Exercise 70. Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T\left((x,y,z)\right) = (z, x, y)$.

(a) Is T a linear map? (Yes or no)

(b) Prove your assertion from (a).

Exercise 71. Define the 'multiplication by x^2' map

$$M : \text{Poly}_n \to \text{Poly}_{n+2}$$
$$\mathbf{p} \mapsto M(\mathbf{p})$$

where $M(\mathbf{p})(x) = x^2 \mathbf{p}(x)$.

(a) Why does M map from Poly_n to Poly_{n+2}?

(b) Prove that M is linear.

(c) Compute the action of M in the standard basis for Poly_3, as in Example 3.8.

Exercise 72. Define the 'integrate over the interval $[-1, 1]$' map

$$I : \text{Poly}_n \to \mathbb{R}$$
$$\mathbf{p} \mapsto \int_{-1}^{1} \mathbf{p}(x) \, dx$$

(a) Prove that I is linear.

(b) Compute the action of I with respect to the standard basis $\mathbf{p}_0, \ldots, \mathbf{p}_3$ for Poly_3.

(c) Compute the action of I with respect to the basis $\mathbf{q}_0, \ldots, \mathbf{q}_3$ for Poly_3 from Example 2.9.

Exercise 73. Compute the action of the differentiation map $D : \text{Poly}_4 \to \text{Poly}_3$ from Example 3.6 with respect to the standard bases of these two vector spaces.

Exercise 74. Consider the cross-product linear map $C : \mathbb{R}^3 \to \mathbb{R}^3$ from Example 3.4 in the case $\mathbf{w} = (1, 2, 3)$. Compute the action of C with respect to the standard basis of $\mathbb{R}^3$.

3.2 Composition of linear maps

Definition 32. If $S : U \to V$ and $T : V \to W$ are linear maps, then the *composition of T with S* is the map $T \circ S : U \to W$ defined by

$$(T \circ S)(\mathbf{u}) := T(S(\mathbf{u}))$$

where $\mathbf{u}$ is in U.

See Figure 3.2. tikzalign
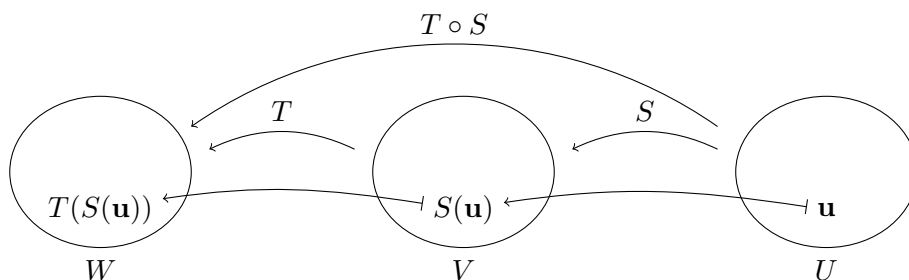
Figure 3.2: Composition of linear maps.

The standard convention in mathematics is to write the evaluation of a function from right-to-left, eg. $f(x)$. That is, you start with the right hand symbol x, then you apply f. So the most natural way to draw these pictures is from right-to-left!

Example 3.12. Let $S : \mathbb{R}^3 \to \text{Poly}_2$ and $T : \text{Poly}_2 \to \text{Poly}_4$ be the linear maps defined by

$$S((a,b,c)) := ax^2 + (a-b)x + c, \quad T(p(x)) = x^2 p(x)$$

Then $T \circ S$ can be computed as follows:

$$\begin{aligned}
(T \circ S)((a,b,c)) &= T(S((a,b,c))) \\
&= T(ax^2 + (a-b)x + c) \\
&= x^2(ax^2 + (a-b)x + c) \\
&= ax^4 + (a-b)x^3 + cx^2.
\end{aligned}$$

Proposition 33. *If $S : U \to V$ and $T : V \to W$ are linear maps, then $T \circ S : U \to W$ is also a linear map.*

Proof. Let $\mathbf{u_1}, \mathbf{u}_2 \in U$. Then:

$$\begin{aligned}
(T \circ S)(\mathbf{u}_1 + \mathbf{u}_2) &= T(S(\mathbf{u}_1 + \mathbf{u}_2)) && \text{(defn of } T \circ S) \\
&= T(S(\mathbf{u}_1) + S(\mathbf{u}_2)) && (S \text{ is linear}) \\
&= T(S(\mathbf{u}_1)) + T(S(\mathbf{u}_2)) && (T \text{ is linear}) \\
&= (T \circ S)(\mathbf{u}_1) + (T \circ S)(\mathbf{u}_2) && \text{(defn of } T \circ S)
\end{aligned}$$

Similarly,

$$
\begin{aligned}
(T \circ S)(k\mathbf{u}) &= T(S(k\mathbf{u})) &&\text{(defn of } T \circ S\text{)} \\
&= T(kS(\mathbf{u})) &&\text{(}S\text{ is linear)} \\
&= kT(S(\mathbf{u})) &&\text{(}T\text{ is linear)} \\
&= k(T \circ S)(\mathbf{u}) &&\text{(defn of } T \circ S\text{)}
\end{aligned}
$$

$\square$

Example 3.13. Consider the antiderivative (A) and derivative (D) linear maps

$$
A : \mathrm{Poly}_n \to \mathrm{Poly}_{n+1}
$$
$$
D : \mathrm{Poly}_{n+1} \to \mathrm{Poly}_n .
$$

Is $D \circ A = \mathrm{id}_{\mathrm{Poly}_n}$?

Solution. We compute the action of $D \circ A$ on the basis x^k, $k = 0 \ldots n$ of Poly_n:

$$
x^k \overset{A}{\mapsto} \frac{x^{k+1}}{k+1} \overset{D}{\mapsto} \frac{k+1}{k+1}x^k = x^k
$$

Hence for $k = 0 \ldots n$,

$$
\begin{aligned}
(D \circ A)(x^k) &= x^k \\
&= \mathrm{id}_{\mathrm{Poly}_n}(x^k).
\end{aligned}
$$

Since $D \circ A$ and $\mathrm{id}_{\mathrm{Poly}_n}$ agree on a basis for Poly_n, they agree on all vectors $\mathbf{p} \in \mathrm{Poly}_n$ by Proposition 31. Hence $D \circ A = \mathrm{id}_{\mathrm{Poly}_n}$.

In fact, the statement that $D \circ A = \mathrm{id}_{\mathrm{Poly}_n}$ is precisely Part I of the Fundamental Theorem of Calculus, applied to the special case of polynomials!

---

movebackintoexample

Exercise 75. Is $A \circ D = \mathrm{id}_{\mathrm{Poly}_{n+1}}$? If it is, prove it. If it is not, give an explicit counterexample.

Further exercises for 3.2

Exercise 76. Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the 'rotation by θ' map from Example 3.11,

$$R((x,y)) = (x\cos\theta - y\sin\theta, \, x\sin\theta + y\cos\theta).$$

Check algebraically that $R_\phi \circ R_\theta = R_{\phi+\theta}$ by computing the action of the linear maps on both sides of this equation on an arbitrary vector $(x,y) \in \mathbb{R}^2$.

Exercise 77. Let $M : \mathrm{Poly}_3 \to \mathrm{Poly}_4$ be the 'multiplication by x' map, $M(p)(x) = xp(x)$. Let $S : \mathrm{Poly}_4 \to \mathrm{Poly}_4$ be the map $S(p(x)) = p(x-1)$. Similarly let $T : \mathrm{Poly}_3 \to \mathrm{Poly}_3$ be the map $T(p(x)) = p(x-1)$. Compute $S \circ M$ and $M \circ T$. Are they equal?

3.3   Isomorphisms of vector spaces

Suppose you have two sets,

$$A = \{\text{bird}, \text{eye}, \text{person}\} \quad \text{and} \quad B = \left\{ 鸟, 目, 人 \right\}.$$

The elements of A and B are not *the same*, so A is not *equal* to B. But this is unsatisfactory — clearly the elements of A are just English versions of the Chinese symbols in B. How can we make this mathematically precise?

   We could define two maps, say

$$S : A \to B$$
$$\text{bird} \mapsto 鸟$$
$$\text{eye} \mapsto 目$$
$$\text{cross} \mapsto 人$$

and

$$T : B \to A$$
$$鸟 \mapsto \text{bird}$$
$$目 \mapsto \text{eye}$$
$$人 \mapsto \text{cross}.$$

Then we observe that

$$T \circ S = \mathrm{id}_A \quad \text{and} \quad S \circ T = \mathrm{id}_B. \tag{3.4}$$

A pair of maps $S : A \to B$ and $T : B \to A$ satisfying (3.4) is called an *isomorphism of sets* between A and B. If you like, you can rename T as S^{-1}

since $S^{-1} \circ S = \text{id}_A$ and $S \circ S^{-1} = \text{id}_B$. (Calling T by the name S^{-1} from the beginning would have been presumptive of me. I needed to first define it, and then check that it satisfied (3.4). Only then did I have the right to call it S^{-1}!)

Perhaps you are somewhat of a penny-pincher. You see the need for the English-to-Chinese map S, but not the need for a Chinese-to-English map T. After all, you say, since no two different English symbols in A get mapped to the same Chinese symbol in B ('S is one-to-one') and every Chinese symbol $y \in B$ is equal to $S(x)$ for some $x \in A$ ('S is onto'), we have no need for T. It is an extravagance!

To this I respond: you are right, but is it not useful to have the explicit Chinese-to-English map T? In bookshops, cross-language dictionaries like this most often come bundled as a pair, in a single volume. After all, if one needs to look up the English word for 的, it is a nuisance to have to traverse through the entire English-to-Chinese dictionary, trying to find the English word which translates to 的!

This motivates the following definition.

Definition 34. We say that a linear map $S : V \to W$ is an *isomorphism* if there exists a linear map $T : W \to V$ such that

$$T \circ S = \text{id}_V \quad \text{and} \quad S \circ T = \text{id}_W . \tag{3.5}$$

Lemma 35 (Uniqueness of Inverses)**.** *If $S : V \to W$ is an isomorphism and $T, T' : W \to V$ both satisfy*

$$T \circ S = \text{id}_V, \qquad\qquad S \circ T = \text{id}_W \tag{3.6}$$
$$T' \circ S = \text{id}_V, \qquad\qquad S \circ T' = \text{id}_W \tag{3.7}$$

then $T = T'$.

This lemma justifies us calling T *the* inverse of S (as opposed to *an* inverse of S), and also justifies us writing $T = S^{-1}$.

Proof. To show that $T = T'$, we must show that for all $\mathbf{w} \in W$, $T(\mathbf{w}) = T'(\mathbf{w})$. Indeed:

$$
\begin{aligned}
T(\mathbf{w}) &= T(\text{id}_W(\mathbf{w})) & &(\text{Defn of id}_W) & &(3.8)\\
&= T((S \circ T')(\mathbf{w})) & &(S \circ T' = \text{id}_W) & &(3.9)\\
&= T(S(T'(\mathbf{w}))) & &(\text{Defn of } S \circ T') & &(3.10)\\
&= (T \circ S)(T'(\mathbf{w})) & &(\text{Defn of } T \circ S) & &(3.11)\\
&= \text{id}_V(T'(\mathbf{w})) & &(T \circ S = \text{id}_V) & &(3.12)\\
&= T'(\mathbf{w}) & &(\text{Defn of id}_V). & &(3.13)
\end{aligned}
$$

$\square$

Definition 36. We say that two vector spaces V and W are *isomorphic* if there exists an isomorphism between them.

Example 3.14. Show that $\mathbb{R}^n$ is isomorphic to Poly_{n-1}.

Solution. We define a pair of linear maps

$$S : \mathbb{R}^n \rightleftarrows \mathrm{Poly}_{n-1} : T$$

as follows:

$$(a_1, a_2, \ldots, a_n) \overset{S}{\mapsto} a_1 + a_2 x + \cdots + a_n x^{n-1}$$
$$(a_1, a_2, \ldots, a_n) \overset{T}{\leftarrow\!\shortmid} a_1 + a_2 x + \cdots + a_n x^{n-1}$$

We clearly have $T \circ S = \mathrm{id}_{\mathbb{R}^n}$ and $S \circ T = \mathrm{id}_{\mathrm{Poly}_{n-1}}$.

---

movebackintoexample

Exercise 78. Check that these maps are linear.

We will now show that up to isomorphism, there is only one vector space of each dimension!

Theorem 37. *Two finite-dimensional vector spaces V and W are isomorphic if and only if they have the same dimension.*

Proof. $\Rightarrow$. Suppose V and W are isomorphic, via a pair of linear maps $S : V \rightleftarrows W : T$. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ be a basis for V. Then I claim that $\mathcal{C} = \{S(\mathbf{e}_1), \ldots, S(\mathbf{e}_m)\}$ is a basis for W. Indeed, the list of vectors $\mathcal{C}$ is linearly independent, since if

$$a_1 S(\mathbf{e}_1) + a_2 S(\mathbf{e}_2) + \cdots + a_m S(\mathbf{e}_m) = \mathbf{0}_W,$$

then applying T to both sides we obtain

$$T(a_1 S(\mathbf{e}_1) + a_2 S(\mathbf{e}_2) + \cdots + a_m S(\mathbf{e}_m)) = T(\mathbf{0}_W)$$
$$\therefore a_1 T(S(\mathbf{e}_1)) + a_2 T(S(\mathbf{e}_2)) + \cdots + a_m T(S(\mathbf{e}_m)) = \mathbf{0}_V \qquad (T \text{ is linear})$$
$$\therefore a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_m \mathbf{e}_m = \mathbf{0}_V \qquad (T \circ S = \mathrm{id}_V)$$

which implies that $a_1 = a_2 = \cdots = a_m = 0$, since $\mathcal{B}$ is linearly independent. Moreover, the list of vectors $\mathcal{C}$ spans W, for if $\mathbf{w} \in W$, then applying T, we can write

$$T(\mathbf{w}) = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_m \mathbf{e}_m$$

for some scalars a_i since $\mathcal{B}$ spans V. But then

$$
\begin{aligned}
\mathbf{w} &= S(T(\mathbf{w})) && (\text{since } S \circ T = \mathrm{id}_W) \\
&= S(a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_m\mathbf{e}_m) \\
&= a_1 S(\mathbf{e}_1) + a_2 S(\mathbf{e}_2) + \cdots + a_m S(\mathbf{e}_m) && (S \text{ is linear})
\end{aligned}
$$

so that $\mathcal{C}$ spans W. Hence $\mathcal{C}$ is a basis for V, so $\mathrm{Dim}\, V$ = number of vectors in $\mathcal{B}$ $= m$, while $\mathrm{Dim}\, W$ = number of vectors in $\mathcal{C} = m$.

$\Leftarrow$. Suppose $\mathrm{Dim}\, V = \mathrm{Dim}\, W$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be a basis for V, and let $\mathbf{f}_1, \ldots, \mathbf{f}_m$ be a basis for W. (We know that the number of basis vectors is the same since $\mathrm{Dim}\, V = \mathrm{Dim}\, W$.)

To define linear maps

$$
S : V \rightleftarrows W : T
$$

it is sufficient, by Proposition 31 (Sufficient to Define a Linear Map on a Basis), to define the action of S and T on the basis vectors. We set:

$$
\mathbf{e}_i \overset{S}{\mapsto} \mathbf{f}_i
$$

$$
\mathbf{e}_i \overset{T}{\mapsfrom} \mathbf{f}_i
$$

Clearly we have $T \circ S = \mathrm{id}_V$ and $S \circ T = \mathrm{id}_W$. $\square$

Example 3.15. Show that $\mathrm{Mat}_{n,m}$ is isomorphic to $\mathbb{R}^{mn}$.

Solution. We simply observe that by Example 2.15, $\mathrm{Dim}\,\mathrm{Mat}_{n,m} = mn$ while from Example 2.11, $\mathrm{Dim}\,\mathbb{R}^{mn}$ is also equal to mn.

There is one very important isomorphism which we will use over and over. Let V be a vector space with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$. Consider the map

$$
[\cdot]_{\mathcal{B}} : V \to \mathrm{Col}_m
$$

$$
\mathbf{v} \mapsto [\mathbf{v}]_{\mathcal{B}}
$$

which sends a vector $\mathbf{v} \in V$ to its corresponding coordinate vector $[\mathbf{v}]_{\mathcal{B}} \in \mathrm{Col}_m$. Lemma 26 says precisely that $[\cdot]_{\mathcal{B}}$ is a linear map. We will now describe its inverse.

Definition 38. Let V be an m-dimensional vector space with basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$. Let $\mathsf{c} \in \mathrm{Col}_m$ be an m-dimensional column vector. Then the *vector in V corresponding to c with respect to the basis $\mathcal{B}$* is

$$
\mathbf{vec}_{V,\mathcal{B}}(\mathsf{c}) := \mathsf{c}_1\mathbf{e}_1 + \mathsf{c}_2\mathbf{e}_2 + \cdots + \mathsf{c}_m\mathbf{e}_m.
$$

Example 3.16. The polynomials $\mathcal{B} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ where

$$\mathbf{p}_1 := 1 + x, \quad \mathbf{p}_2 := 1 + x + x^2, \mathbf{p}_3 := 1 - x^2$$

are a basis of Poly_2 (check this). Then, for instance,

$$\mathbf{vec}_{\mathrm{Poly}_3, \mathcal{B}}\left(\begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}\right) = 2(1 + x) - 3(1 + x + x^2) + 3(1 - x^2)$$

$$= 2 - x - 6x^2 \in \mathrm{Poly}_3 \ .$$

Exercise 79. Show that:

(a) $\mathbf{vec}_{V,\mathcal{B}}(\mathsf{c} + \mathsf{c}') = \mathbf{vec}_{V,\mathcal{B}}(\mathsf{c}) + \mathbf{vec}_{V,\mathcal{B}}(\mathsf{c}')$

(b) $\mathbf{vec}_{V,\mathcal{B}}(k\mathsf{c}) = k\,\mathbf{vec}_{V,\mathcal{B}}(\mathsf{c})$.

This means that $\mathbf{vec}_{V,\mathcal{B}} : \mathrm{Col}_m \to V$ is a linear map.

Theorem 39. *Let V be a vector space with basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$. The maps*

$$[\cdot]_{\mathcal{B}} : V \rightleftarrows \mathrm{Col}_m : \mathbf{vec}_{V,\mathcal{B}}$$

$$\mathbf{v} \mapsto [\mathbf{v}]_{\mathcal{B}}$$

$$\mathbf{vec}_{V,\mathcal{B}}(\mathsf{c}) \hookleftarrow \mathsf{c}$$

are an isomorphism of vector spaces.

Proof. Given $\mathbf{v} \in V$, expand it in the basis $\mathcal{B}$:

$$\mathbf{v} = a_1\mathbf{e}_1 + \ldots + a_m\mathbf{e}_m.$$

Then

$$\left(\mathbf{vec}_{V,\mathcal{B}} \circ [\cdot]_{\mathcal{B}}\right)(\mathbf{v}) = \mathbf{vec}_{V,\mathcal{B}}\left([\mathbf{v}]_{\mathcal{B}}\right)$$

$$= \mathbf{vec}_{V,\mathcal{B}}\left(\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}\right)$$

$$= a_1\mathbf{e}_1 + \cdots + a_m\mathbf{e}_m$$

$$= \mathbf{v}$$

so that $\mathbf{vec}_{V,\mathcal{B}} \circ [\cdot]_{\mathcal{B}} = \mathrm{id}_V$. Conversely, given

$$\mathsf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \in \mathrm{Col}_m,$$

we have

$$\begin{aligned}
\left([\cdot]_{\mathcal{B}} \circ \mathbf{vec}_{V,\mathcal{B}}\right)(\mathsf{c}) &= \left[\mathbf{vec}_{V,\mathcal{B}}(\mathsf{c})\right]_{\mathcal{B}} \\
&= \left[c_1 \mathbf{e}_1 + \cdots + c_m \mathbf{e}_m\right] \\
&= \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \\
&= \mathsf{c}
\end{aligned}$$

where the second last step uses the *definition* of the coordinate vector of $\mathbf{v} = c_1 \mathbf{e}_1 + \cdots + c_m \mathbf{e}_m$. Hence $[\cdot]_{\mathcal{B}} \circ \mathbf{vec}_{V,\mathcal{B}} = \mathrm{id}_{\mathrm{Col}_m}$. $\square$

> The result above is very important in linear algebra. It says that, once we have chosen a basis for an abstract finite-dimensional vector space V, we can treat the elements of V as if they were column vectors!

Further exercises for 3.3

Exercise 80. Are the following vector spaces isomorphic?

$$V = \left\{ \mathsf{v} \in \mathrm{Col}_4 : \begin{bmatrix} 1 & 2 & 0 & -1 \\ -1 & 1 & 1 & 0 \end{bmatrix} \mathsf{v} = 0 \right\}$$

$$W = \left\{ p \in \mathrm{Poly}_2 : \int_0^2 p(x)\,dx = 0 \right\}.$$

If they are, construct an explicit isomorphism between them. If not, prove that they are not isomorphic.

Exercise 81. Are the following vector spaces isomorphic?

$$V = \{\mathbf{v} \in \mathbb{R}^3 : \mathbf{v} \times (1,2,3) = \mathbf{0}\}$$

$$W = \left\{ M \in \mathrm{Mat}_{2,2} : M^T = -M \right\}.$$

If they are, construct an explicit isomorphism between them. If not, prove that they are not isomorphic.

3.4   Linear maps and matrices

Definition 40. Let $T : V \to W$ be a linear map from a vector space V to a vector space W. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis for V and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be a basis for W. The *matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$* is defined as the $n \times m$ matrix whose columns are the coordinate vectors of $T(\mathbf{b}_i)$ with respect to the basis $\mathcal{C}$:

$$[T]_{\mathcal{C} \leftarrow \mathcal{B}} := \left[\left[T(\mathbf{b}_1) \right]_{\mathcal{C}} \quad \left[T(\mathbf{b}_2) \right]_{\mathcal{C}} \quad \cdots \quad \left[T(\mathbf{b}_m) \right]_{\mathcal{C}} \right]$$

Do you understand why $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$ is an $n \times m$ matrix?

Example 3.17. Example from class!

Theorem 41. *Let $T : V \to W$ be a linear map from a vector space V to a vector space W. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a basis for V and $\mathcal{C}$ be a basis for W. Then for all vectors $\mathbf{v}$ in V,*

$$[T(\mathbf{v})]_{\mathcal{C}} = [T]_{\mathcal{C} \leftarrow \mathcal{B}} [\mathbf{v}]_{\mathcal{B}} \tag{3.14}$$

where the right hand side is the product of the matrix $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$ with the column vector $\mathbf{v}]_{\mathcal{B}}$.

Proof. Similar to the proof of the Change-Of-Basis Theorem (Theorem 28). Let $\mathbf{v} \in V$. Expand it in the basis $\mathcal{B}$:

$$\mathbf{v} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + \cdots + a_m \mathbf{b}_m, \quad \text{i.e. } [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}.$$

Then,

$$
\begin{aligned}
[T(\mathbf{v})]_{\mathcal{C}} &= [T(a_1\mathbf{b}_1 + \cdots + a_m\mathbf{b}_m)]_{\mathcal{C}} \\
&= [a_1T(\mathbf{b}_1) + \cdots + a_nT(\mathbf{b}_m)]_{\mathcal{C}} && (T \text{ is linear}) \\
&= a_1[T(\mathbf{b}_1)]_{\mathcal{C}} + \cdots + a_n[T(\mathbf{b}_m)]_{\mathcal{C}} && (\text{by Lemma 26}) \\
&= \left[\begin{bmatrix} T(\mathbf{b}_1) \end{bmatrix}_{\mathcal{C}} \begin{bmatrix} T(\mathbf{b}_2) \end{bmatrix}_{\mathcal{C}} \cdots \begin{bmatrix} T(\mathbf{b}_m) \end{bmatrix}_{\mathcal{C}} \right] \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} && (\text{by (2.30)}) \\
&= [T]_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.
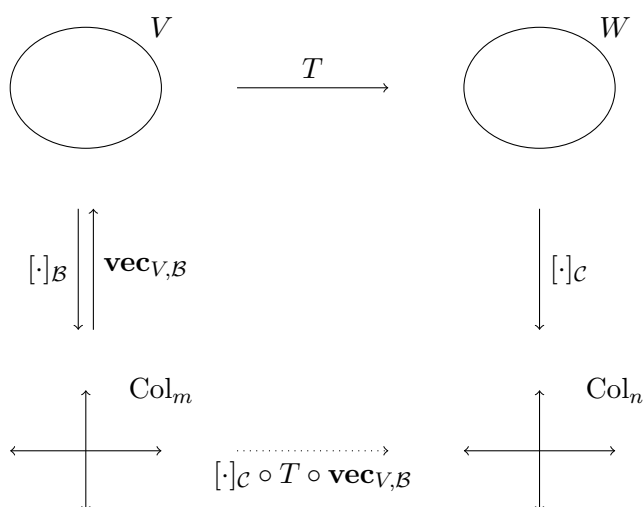\end{aligned}
$$

$\square$

If we could give this theorem a name, it would be "The Relationship Between Linear Maps, Coordinate Vectors, and Matrix Multiplication of Column Vectors Theorem"!

Example 3.18. Continuation of example from class!

We can interpret Theorem 41 in a more abstract way as follows. We have the following diagram of linear maps of vector spaces:

The map at the top is the linear map $T : V \to W$. The map on the left from V to Col_m is the coordinate vector map $[\cdot]_{\mathcal{B}}$ associated to the basis $\mathcal{B}$. Its inverse map $\mathbf{vec}_{V,\mathcal{B}} : \mathrm{Col}_m \to V$ is also drawn. The map on the right is the coordinate vector map $[\cdot]_{\mathcal{C}}$ from W to Col_n associated to the basis C. The dotted arrow on the bottom is the composite map, and can be computed explicitly as follows.

Lemma 42. *The composite map*

$$[\cdot]_{\mathcal{C}} \circ T \circ \mathbf{vec}_{V,\mathcal{B}} : \mathrm{Col}_m \to \mathrm{Col}_n$$

is multiplication by the matrix $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$. That is, for all column vectors u in Col_m,

$$\left([\cdot]_{\mathcal{C}} \circ T \circ \mathbf{vec}_{V,\mathcal{B}}\right)(\mathsf{u}) = [T]_{\mathcal{C} \leftarrow \mathcal{B}}\,\mathsf{u}.$$

Proof. Let u be a column vector in Col_m. Define $\mathbf{v} := \mathbf{vec}_{V,\mathcal{B}}(\mathsf{u})$. Then $\mathbf{v}$ is the vector in V whose coordinate vector with respect to the basis $\mathcal{B}$ is u. That is, $\mathsf{u} = [\mathbf{v}]_{\mathcal{B}}$. So,

$$
\begin{aligned}
\left([\cdot]_{\mathcal{C}} \circ T \circ \mathbf{vec}_{V,\mathcal{B}}\right)(\mathsf{u}) &= [\cdot]_{\mathcal{C}}\left(T\left(\mathbf{vec}_{V,\mathcal{B}}(\mathsf{u})\right)\right) && \text{(Defn of composite map)}\\
&= [\cdot]_{\mathcal{C}}(T(\mathbf{v})) && \text{(Defn of } \mathbf{v})\\
&= [T(\mathbf{v})]_{\mathcal{C}} && \text{(Defn of } [\cdot]_{\mathcal{C}})\\
&= [T]_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{v}]_{\mathcal{B}} && \text{(Theorem 41)}.
\end{aligned}
$$

$\square$

Before we move on, we need to recall another thing about matrices. Suppose A is a matrix with n rows. Let

$$
\mathsf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \mathsf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \; \dots, \; \mathsf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}
$$

be the standard basis for Col_n. Then the ith column of A can be obtained by multiplying A with e_i:

$$i\text{th column of } \mathsf{A} = A\mathsf{e}_i. \tag{3.15}$$

Exercise 82. Check this!

Now we can prove the following important Theorem.

Theorem 43 (Functoriality of the Matrix of a Linear Map)**.** *Let $S : U \to V$ and $T : V \to W$ be linear maps between finite-dimensional vector spaces. Let $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ be bases for U, V and W respectively. Then*

$$[T \circ S]_{\mathcal{D} \leftarrow \mathcal{B}} = [T]_{\mathcal{D} \leftarrow \mathcal{C}}[S]_{\mathcal{C} \leftarrow \mathcal{B}}$$

where the right hand side is the product of the matrices $[T]_{\mathcal{D} \leftarrow \mathcal{C}}$ and $[S]_{\mathcal{C} \leftarrow \mathcal{B}}$.

Proof. We have:

ith column of $[T \circ S]_{\mathcal{D} \leftarrow \mathcal{B}}$

$$\begin{aligned}
&= [(T \circ S)(\mathbf{b}_i)]_{\mathcal{D}} && \text{(Defn of } [T \circ S]_{\mathcal{D} \leftarrow \mathcal{B}}) \\
&= [T(S(\mathbf{b}_i))]_{\mathcal{D}} && \text{(Defn of } T \circ S) \\
&= [T]_{\mathcal{D} \leftarrow \mathcal{C}}\, [S(\mathbf{b}_i)]_{\mathcal{C}} && \text{(Theorem 41)} \\
&= [T]_{\mathcal{D} \leftarrow \mathcal{C}}\, [S]_{\mathcal{C} \leftarrow \mathcal{B}}\, [\mathbf{b}_i]_{\mathcal{B}} && \text{(Theorem 41)} \\
&= [T]_{\mathcal{D} \leftarrow \mathcal{C}}\, [S]_{\mathcal{C} \leftarrow \mathcal{B}}\, \mathsf{e}_i && \left(\begin{array}{l}\text{since } [\mathbf{b}_i]_{\mathcal{B}} = \mathsf{e}_i \text{, the } i\text{th} \\ \text{standard basis column} \\ \text{vector.}\end{array}\right) \\
&= i\text{th column of } [T]_{\mathcal{D} \leftarrow \mathcal{C}}[S]_{\mathcal{C} \leftarrow \mathcal{B}} && \text{(by (3.15)).}
\end{aligned}$$

$\square$

Corollary 44. *Let $T : V \to W$ be a linear map, and suppose $\mathcal{B}$ is a basis for V, and $\mathcal{C}$ is a basis for W. Then*

$$T \text{ is an isomorphism} \iff [T]_{\mathcal{C} \leftarrow \mathcal{B}} \text{ is invertible.}$$

Proof. $\Rightarrow$. Suppose the linear map T is an isomorphism. This means there exists a linear map $S : W \to V$ such that

$$S \circ T = \mathrm{id}_V \quad \text{and} \quad T \circ S = \mathrm{id}_W /$$

Therefore,

$$[S \circ T]_{\mathcal{B} \leftarrow \mathcal{B}} = [\mathrm{id}_V]_{\mathcal{B} \leftarrow \mathcal{B}} \quad \text{and} \quad [T \circ S]_{\mathcal{C} \leftarrow \mathcal{C}} = [\mathrm{id}_W]_{\mathcal{C} \leftarrow \mathcal{C}}.$$

Therefore, by the Functoriality of the Matrix of a Linear Map (Theorem 43),

$$[S]_{\mathcal{B} \leftarrow \mathcal{C}}[T]_{\mathcal{C} \leftarrow \mathcal{B}} = I \quad \text{and} \quad [T]_{\mathcal{C} \leftarrow \mathcal{B}}[S]_{\mathcal{B} \leftarrow \mathcal{C}} = I$$

Therefore the matrix $[T]_{\mathcal{C} \leftarrow \mathcal{B}}$ is invertible, with inverse given by

$$[T]_{\mathcal{C} \leftarrow \mathcal{B}}^{-1} = [S]_{\mathcal{B} \leftarrow \mathcal{C}}.$$

$\Leftarrow$. Suppose the matrix $[T] \equiv [T]_{\mathcal{C} \leftarrow \mathcal{B}}$ is invertible. Define the linear map

$$S : W \to V$$

by firstly defining it on the basis vectors in $\mathcal{C}$ by

$$S(\mathbf{c}_i) := \sum_{p=1}^{\dim V} [T]_{pi}^{-1} \mathbf{b}_p$$

and then extending to all of W by linearity. Then we have

$$(T \circ S)(\mathbf{c}_i) = T(S(\mathbf{c}_i))$$

$$= T\left(\sum_{p=1}^{\dim V} [T]_{pi}^{-1}\mathbf{b}_p\right)$$

$$= \sum_{p=1}^{\dim V}\sum_{q=1}^{\dim W} [T]_{pi}^{-1}[T]_{qp}\mathbf{c}_q$$

$$= \sum_{q=1}^{\dim W}\left(\sum_{p=1}^{\dim V} [T]_{qp}[T]_{pi}^{-1}\right)\mathbf{c}_q$$

$$= \sum_{q=1}^{\dim W}\left([T][T]^{-1}\right)_{qi}\mathbf{c}_q$$

$$= \sum_{q=1}^{\dim W} I_{qi}\mathbf{c}_q$$

$$= \sum_{q=1}^{\dim W} \delta_{qi}\mathbf{c}_q$$

$$= \mathbf{c_i}.$$

Therefore, $T \circ S = \mathrm{id}_W$. In a similar way, we can prove that $S \circ T = \mathrm{id}_V$. Therefore the linear map T is an isomorphism, with inverse map $T^{-1} = S$. $\qquad\square$

We can refine this a bit further. Explicitly, 'the inverse of the matrix of a linear map equals the matrix of the inverse of the linear map'.

Corollary 45. *Suppose $\mathcal{B}$ and $\mathcal{C}$ are bases for vector spaces V and W respectively. Suppose a linear map $T : V \to W$ has inverse $T^{-1} : W \to V$. Then*

$$[T]_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = [T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}.$$

Proof. We have

$$\begin{aligned}
[T]_{\mathcal{C}\leftarrow\mathcal{B}}[T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}} &= [T \circ T^{-1}]_{\mathcal{C}\leftarrow\mathcal{C}} && \text{(by Theorem 43)}\\
&= [\mathrm{id}_W]_{\mathcal{C}\leftarrow\mathcal{C}} && (T \circ T^{-1} = \mathrm{id}_W)\\
&= I
\end{aligned}$$

and

$$\begin{aligned}
[T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}} &= [T^{-1} \circ T]_{\mathcal{B}\leftarrow\mathcal{B}} && \text{(by Theorem 43)}\\
&= [\mathrm{id}_V]_{\mathcal{B}\leftarrow\mathcal{B}} && (T^{-1} \circ T = \mathrm{id}_V)\\
&= I.
\end{aligned}$$

$\square$

The next Lemma says that the 'change-of-basis matrix' from Section **??**
is just the matrix of the identity linear map with respect to the relevant
bases.

Lemma 46. *Let $\mathcal{B}$ and $\mathcal{C}$ be bases for an m-dimensional vector space V.
Then*

$$\mathsf{P}_{\mathcal{C}\leftarrow\mathcal{B}} = [\mathrm{id}]_{\mathcal{C}\leftarrow\mathcal{B}}.$$

Proof.

$$
\begin{aligned}
\mathsf{P}_{\mathcal{C}\leftarrow\mathcal{B}} &= [[\mathbf{b}_1]_{\mathcal{C}}\cdots[\mathbf{b}_m]_{\mathcal{C}}] && (\text{Defn of } \mathsf{P}_{\mathcal{C}\leftarrow\mathcal{B}}) \\
&= [[\mathrm{id}(\mathbf{b}_1)]_{\mathcal{C}}\cdots[\mathrm{id}(\mathbf{b}_m)]_{\mathcal{C}}] \\
&= [\mathrm{id}]_{\mathcal{C}\leftarrow\mathcal{B}}. && (\text{Defn of } [\mathrm{id}]_{\mathcal{C}\leftarrow\mathcal{B}})
\end{aligned}
$$

$\square$

The next Theorem tells us how the matrix of a linear operator changes
when we change the bases used in computing it.

Theorem 47. *Let $\mathcal{B}$ and $\mathcal{C}$ be bases for a vector space V, and let $T : V \to V$
be a linear operator on V. Then*

$$[T]_{\mathcal{C}\leftarrow\mathcal{C}} = \mathsf{P}^{-1}[T]_{\mathcal{B}\leftarrow\mathcal{B}}P$$

where $P \equiv P_{\mathcal{B}\leftarrow\mathcal{C}}$.

Proof.

$$
\begin{aligned}
\mathrm{RHS} &= \mathsf{P}^{-1}[T]_{\mathcal{B}\leftarrow\mathcal{B}}P \\
&= [\mathrm{id}]^{-1}_{\mathcal{B}\leftarrow\mathcal{C}}[T]_{\mathcal{B}\leftarrow\mathcal{B}}[\mathrm{id}]_{\mathcal{B}\leftarrow\mathcal{C}} && (\text{Lemma 46}) \\
&= [\mathrm{id}]_{\mathcal{C}\leftarrow\mathcal{B}}[T]_{\mathcal{B}\leftarrow\mathcal{B}}[\mathrm{id}]_{\mathcal{B}\leftarrow\mathcal{C}} && (\text{Corollary 45}) \\
&= [\mathrm{id}\circ T \circ \mathrm{id}]_{\mathcal{C}\leftarrow\mathcal{C}} && (\text{Theorem 43}) \\
&= [T]_{\mathcal{C}\leftarrow\mathcal{C}} \\
&= \mathrm{LHS}.
\end{aligned}
$$

$\square$

Further exercises for 3.4

Exercise 83. Let

$$T : \mathrm{Trig}_1 \to \mathrm{Trig}_2$$

be the 'multiply with $\sin x$' linear map, $T(f)(x) = \sin x f(x)$. Compute
$[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ with respect to the standard basis $\mathcal{B}$ of Trig_1 and $\mathcal{C}$ of Trig_2.

Exercise 84. Let

$$S : \mathrm{Trig}_2 \to \mathrm{Trig}_2$$

be the 'shift by $\frac{\pi}{6}$' map, $S(f)(x) = f(x - \frac{\pi}{6})$. Compute $[S]_{\mathcal{C} \leftarrow \mathcal{C}}$ with respect to the standard basis $\mathcal{C}$ of Trig_2.

Exercise 85. Verify Theorem 41 for the linear map $S : \mathrm{Mat}_{2,2} \to \mathrm{Mat}_{2,2}$ given by $S(\mathsf{M}) = \mathsf{M}^T$, for the vector $\mathsf{A} \in \mathrm{Mat}_{2,2}$ given by

$$\mathsf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

and using the following bases of $\mathrm{Mat}_{2,2}$:

$$\mathcal{B} = \mathcal{C} = \{\mathsf{M}_1 = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, \mathsf{M}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathsf{M}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathsf{M}_4 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}\}.$$

Exercise 86. Check that the linear maps T and S from Exercises 83 and 84 satisfy $[S \circ T]_{\mathcal{C} \leftarrow \mathcal{B}} = [S]_{\mathcal{B} \leftarrow \mathcal{B}}[T]_{\mathcal{C} \leftarrow \mathcal{B}}$.

3.5   Kernel and range of a linear map

Definition 48. Let $T : V \to W$ be a linear map between vector spaces V and W. The *kernel* of T, written $\mathrm{Ker}(T)$, is the set of all vectors $\mathbf{v} \in V$ such that are mapped to $\mathbf{0}_W$ by T. That is,

$$\mathrm{Ker}(T) := \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}_W\}.$$

The *image* of T, written $\mathrm{Im}(T)$, is the set of all vectors $\mathbf{w} \in W$ such that $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v} \in V$. That is,

$$\mathrm{Im}(T) := \{\mathbf{w} \in W : \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$$

See Figure 3.3 for a schematic representation.

Sometimes, to be absolutely clear, I will put a subscript on the zero vector to indicate which vector space it belongs to, eg. $\mathbf{0}_W$ refers to the zero vector in W, while $\mathbf{0}_V$ refers to the zero vector in V.

(a) Kernel of T (b) Image of T
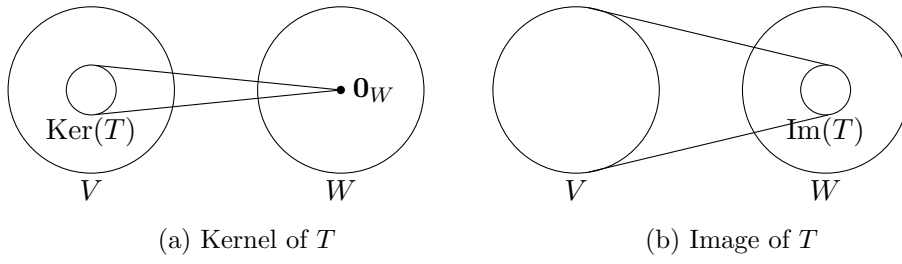
Figure 3.3

Another name for the kernel of T is the *nullspace* of T, and another name for the image of T is the *range* of T.

Lemma 49. *Let $T : V \to W$ be a linear map. Then:*

(i) $\mathrm{Ker}(T)$ is a subspace of V

(ii) $\mathrm{Im}(T)$ is a subspace of W

Proof. (i) We must check the three requirements for being a subspace.

    a. $\mathrm{Ker}(T)$ is closed under addition.

        Suppose $\mathbf{v}$ and $\mathbf{v}'$ are in $\mathrm{Ker}(T)$. In other words, $T(\mathbf{v}) = \mathbf{0}$ and $T(\mathbf{v}') = \mathbf{0}$. We need to show that $\mathbf{v} + \mathbf{v}'$ is in $\mathrm{Ker}(T)$, in other words, that $T(\mathbf{v} + \mathbf{v}') = \mathbf{0}$. Indeed,

$$T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}') = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

    b. $\mathbf{0}_V \in \mathrm{Ker}(T)$.

        To show that $\mathbf{0}_V$ is in $\mathrm{Ker}(T)$, we need to show that $T(\mathbf{0}_V) = \mathbf{0}_W$. Indeed, this is true since T is a linear map, by Lemma 30.

    c. $\mathrm{Ker}(T)$ is closed under scalar multiplication.

        Suppose $\mathbf{v} \in \mathrm{Ker}(T)$ and $k \in \mathbb{R}$ is a scalar. We need to show that $k\mathbf{v} \in \mathrm{Ker}(T)$, that is, we need to show that $T(k\mathbf{v}) = \mathbf{0}$. Indeed,
$$T(k\mathbf{v}) = kT(\mathbf{v}) = k\mathbf{0} = \mathbf{0}.$$

(ii) Again, we must check the three requirements for being a subspace.

a. $\text{Im}(T)$ is closed under addition.

Suppose $\mathbf{w}$ and $\mathbf{w}'$ are in $\text{Im}(T)$. In other words, there exist vectors $\mathbf{v}$ and $\mathbf{v}'$ in V such that $T(\mathbf{v}) = \mathbf{w}$ and $T(\mathbf{v}') = \mathbf{w}'$. We need to show that $\mathbf{w}+\mathbf{w}'$ is also in $\text{Im}(T)$, in other words, that there exists a vector $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{w} + \mathbf{w}'$. Indeed, set $\mathbf{u} := \mathbf{v} + \mathbf{v}'$. Then,

$$T(\mathbf{u}) = T(\mathbf{v} + \mathbf{v}') = T(\mathbf{v}) + T(\mathbf{v}') = \mathbf{w} + \mathbf{w}'.$$

b. $\mathbf{0}_W \in \text{Im}(T)$.

To show that $\mathbf{0}_W \in \text{Im}(T)$, we need to show that there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{0}_W$. Indeed, choose $\mathbf{v} = \mathbf{0}_V$. Then $T(\mathbf{v}) = T(\mathbf{0}_V) = \mathbf{0}_W$ by Lemma 30.

c. $\text{Im}(T)$ is closed under scalar multiplication.

Suppose $\mathbf{w} \in \text{Im}(T)$ and k is a scalar. We need to show that $k\mathbf{w} \in \text{Im}(T)$. The fact that $\mathbf{w}$ is in $\text{Im}(T)$ means that there exists a $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{w}$. We need to show that there exists a $\mathbf{u} \in V$ such that $T(\mathbf{u}) = k\mathbf{w}$. Indeed, set $\mathbf{u} := k\mathbf{v}$. Then

$$T(\mathbf{u}) = T(k\mathbf{v}) = kT(\mathbf{v}) = k\mathbf{w}.$$

$\square$

Now that we know that the kernel and image of a linear map are subspaces, and hence vector spaces in their own right, we can make the following definition.

Definition 50. Let $T : V \to W$ be a linear map from a finite-dimensional vector space V to a vector space W. The *nullity* of T is the dimension of $\text{Ker}(T)$, and the *rank* of T is the dimension of $\text{Im}(T)$:

$$\text{Nullity}(T) := \text{Dim}(\text{Ker}(T))$$
$$\text{Rank}(T) = \text{Dim}(\text{Im}(T))$$

The 'dimension of $\text{Ker}(T)$' makes sense because $\text{Ker}(T)$ is a subspace of a finite-dimensional vector space V, and hence is finite-dimensional by Proposition 20. We do not yet know that $\text{Im}(T)$ is finite-dimensional, but this will follow from the Rank-Nullity Theorem (Theorem 51).

Example 3.19. Let $\mathbf{a} \in \mathbb{R}^3$ be a fixed nonzero vector. Consider the 'cross product with $\mathbf{a}$' linear map from Example 3.4,

$$C : \mathbb{R}^3 \to \mathbb{R}^3$$

$$\mathbf{v} \mapsto \mathbf{a} \times \mathbf{v}$$

Determine the kernel, image, nullity and rank of C.

Solution. The kernel of C is the subspace of $\mathbb{R}^3$ consisting of all vectors $\mathbf{v} \in V$ such that $\mathbf{a} \times \mathbf{v} = \mathbf{0}$. From the geometric formula for the cross-product,

$$|\mathbf{a} \times \mathbf{v}| = |\mathbf{a}||\mathbf{v}| \sin \theta$$

where θ is the angle from $\mathbf{a}$ to θ, we see that

$$\mathbf{a} \times \mathbf{v} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{v} = \mathbf{0} \text{ or } \theta = 0 \text{ or } \theta = \pi.$$

In other words, $\mathbf{v}$ must be a scalar multiple of $\mathbf{a}$. So,

$$\mathrm{Ker}(C) = \{k\mathbf{a}, k \in \mathbb{R}\}.$$

I claim that the *image* of C is the subspace of *all* vectors perpendicular to $\mathbf{a}$, i.e.

$$\mathrm{Im}(C) := \{\mathbf{u} \in \mathbb{R}^3 : \mathbf{u} \cdot \mathbf{a} = 0\}. \tag{3.16}$$

If you believe me, then the picture is as follows:
tikzalign
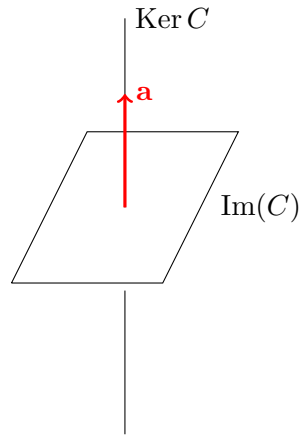
Let me prove equation (3.16). By definition, the image of C is the subspace of $\mathbb{R}^3$ consisting of all vectors $\mathbf{w}$ of the form $\mathbf{w} = \mathbf{a} \times \mathbf{v}$ for some $\mathbf{v} \in \mathbb{R}^3$. This implies that $\mathbf{w}$ is perpendicular to $\mathbf{a}$. This was the 'easy' part. The 'harder' part is to show the converse. That is, we need

to show that if $\mathbf{u}$ is perpendicular to $\mathbf{a}$, then $\mathbf{u}$ is in the image of C, i.e. there exists a vector $\mathbf{v}$ such that $C(\mathbf{v}) = \mathbf{u}$.

Indeed, we can choose $\mathbf{v}$ to be the vector obtained by rotating $\mathbf{u}$ by 90 degrees clockwise in the plane I, and scaling it appropriately:
tikzalign

In terms of a formula, we have

$$\mathbf{v} = \frac{|\mathbf{u}|}{|\mathbf{a}|} \mathbf{u} \times \mathbf{a}.$$

Note that this is not the *only* vector $\mathbf{v}$ such that $C(\mathbf{v}) = \mathbf{u}$. Indeed, if we add to $\mathbf{v}$ any vector that lies on the line through $\mathbf{a}$, the resulting vector

$$\tilde{\mathbf{v}} = \mathbf{v} + k\mathbf{a}$$

also satisfies $C(\tilde{\mathbf{v}}) = \mathbf{u}$, since

$$C(\tilde{\mathbf{v}}) = C(\mathbf{v} + k\mathbf{a}) = C(\mathbf{v}) + C(k\mathbf{a}) = \mathbf{u} + \mathbf{0} = \mathbf{u}.$$
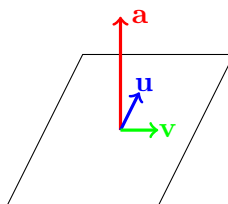

Example 3.20. Determine the kernel, image, nullity and rank of the linear map

$$I : \mathrm{Trig}_2 \to \mathbb{R} \tag{3.17}$$

$$f \mapsto \int_0^\pi f(x)dx. \tag{3.18}$$

Solution. The kernel of I consists of all degree 2 trigonometric polynomials

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x$$

such that

$$\int_0^\pi (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x)\, dx = 0.$$

Performing the integrals, this becomes the equation

$$\pi a_0 + 2b_1 = 0$$

with no constraints on the other constants a_1, a_2, b_2. In other words,

$$\text{Ker}(I) = \left\{ \begin{array}{l} \text{all trigonometric polynomials of the form} \\ a_0(1 - \frac{\pi}{2}\sin x) + a_1 \cos x + a_2 \cos 2x + b_2 \sin 2x, \\ \text{where } a_0, a_1, a_2, b_2 \in \mathbb{R}. \end{array} \right\}$$

Hence $\text{Nullity}(I) = \text{Dim}(\text{Ker}(I)) = 4$.

The image of I consists of all real numbers $p \in \mathbb{R}$ such that there exists a $f \in \text{Trig}_2$ such that $I(f) = p$. I claim that

$$\text{Im}(I) = \mathbb{R}.$$

Indeed, given $p \in \mathbb{R}$, we may choose $f(x) = \frac{p}{2}\sin x$, since

$$I(f) = \frac{p}{2} \int_0^\pi \sin x \, dx = p.$$

Hence $\text{Im}(I) = \mathbb{R}$, and $\text{Rank}(I) = 1$.

Note that the choice of $f(x) = \frac{p}{2}\sin(x)$ satisfying $I(f) = p$ is not unique. We could set $\tilde{f} = f + g$ where $g \in \text{Ker}(I)$ and we would still have $I(\tilde{f}) = p$:

$$I(\tilde{f}) = I(f + g) = I(f) + I(g) = p + 0 = p.$$

Example 3.21. Consider the function

$$T : \text{Poly}_2 \to \mathbb{R}^2$$
$$p \mapsto (p(1), p'(1)).$$

Show that T is a linear map, and determine its kernel, image, rank and nullity.

Solution. We first show that T is a linear map. Let $p, q \in \text{Poly}_2$. Then

$$\begin{aligned} T(p + q) &= ((p+q)(1), \, (p+q)'(1)) & &\text{(Defn of } T\text{)} \\ &= (p(1) + q(1), \, (p+q)'(1)) & &\text{(Defn of the function } p+q\text{)} \\ &= (p(1) + q(1), \, (p'+q')(1)) & &((p+q)' = p' + q') \\ &= (p(1) + q(1), \, p'(1) + q'(1)) & &\text{(Defn of the function } p' + q'\text{)} \\ &= (p(1), \, p'(1)) + (q(1), \, q'(1)) & &\text{(Defn of addition in } \mathbb{R}^2\text{)} \\ &= T(p) + T(q). \end{aligned}$$

The proof of $T(kp) = kT(p)$ is similar.

The kernel of T is the set of all polynomials

$$p(x) = a_0 + a_1x + a_2x^2$$

such that $T(p) = (0,0)$. This translates into the equation

$$(a_0 + a_1 + a_2, \, a_1 + 2a_2) = (0,0).$$

This in turn translates into two equations:

$$\begin{cases} a_0 + a_1 + a_2 = 0 \\ \phantom{a_0 + {}} a_1 + 2a_2 = 0 \end{cases}$$

whose solution is $a_2 = t$, $a_1 = -2t$, $a_0 = -t$, where $t \in \mathbb{R}$. Hence

$$\text{Ker}(T) = \left\{ \begin{array}{l} \text{all polynomials of the form} \\ -t - 2tx + tx^2 \text{ where } t \in \mathbb{R} \end{array} \right\}.$$

Hence $\text{Nullity}(T) = 1$.

The image of T is the set of all $(v, w) \in \mathbb{R}^2$ such that there exists a polynomial $p = a_0 + a_1x + a_2x^2$ in Poly_2 such that $T(p) = (v, w)$. So, (v, w) is in the image of T if and only if we can find a polynomial $p = a_0 + a_1x + a_2x^2$ such that

$$(a_0 + a_1 + a_2, a_1 + 2a_2) = (v, w).$$

In other words, (v, w) is in the image of T if and only if the equations

$$\begin{cases} a_0 + a_1 + a_2 = v \\ \phantom{a_0 + {}} a_1 + 2a_2 = w \end{cases}$$

have a solution for some a_0, a_1, a_2. But these equations *always* have a solution, for *all* $(v, w) \in \mathbb{R}^2$ is. For instance, one solution is

$$a_2 = 0, \, a_1 = w, \, a_0 = v - w$$

which corresponds to the polynomial

$$p(x) = v - w + wx. \tag{3.19}$$

Note that indeed $T(p) = (v, w)$. Hence,

$$\text{Im}(T) = \{\text{all } (v, w) \in \mathbb{R}^2\} = \mathbb{R}^2.$$

Hence $\text{Rank}(T) = \text{Dim}(\text{Im}(T)) = 2$.

Note that the choice of the polynomial $p(x) = v - w + wx$ from (3.19) which satisfies $T(p) = (v, w)$ is not the *only* possible choice. Indeed, any polynomial of the form $\tilde{p} = p + q$ where $q \in \text{Ker}(T)$ will also satisfy $T(\tilde{p}) = (v, w)$, since

$$T(\tilde{p}) = T(p + q) = T(p) + T(q) = (v, w) + (0, 0) = (v, w).$$

Theorem 51 (Rank-nullity theorem). *Let $T : V \to W$ be a linear map from a finite-dimensional vector space V to a vector space W. Then*

$$\text{Nullity}(T) + \text{Rank}(T) = \text{Dim}(V).$$

Proof. Let $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ be a basis for $\text{Ker}(T)$. Since $\mathcal{B}$ is a list of linearly independent vectors in V, we can extend it to a basis $\mathcal{C} = \{\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{f}_1, \ldots, \mathbf{f}_p\}$ for V, by Corollary 23. I claim that

$$\mathcal{D} := \{T(\mathbf{f}_1), \ldots, T(\mathbf{f}_p)\}$$

is a basis for $\text{Im}(T)$. If I can prove that, then we are done, since then we have

$$\text{Nullity}(T) + \text{Rank}(T) = k + p$$
$$= \text{Dim}(V).$$

Let us prove that $\mathcal{D}$ is a basis for $\text{Im}(T)$.

$\mathcal{D}$ is linearly independent. Suppose

$$b_1 T(\mathbf{f}_1) + \cdots + b_p T(\mathbf{f}_p) = \mathbf{0}_W.$$

We recognize the left hand side as $T(b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p)$. Hence

$$b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p \in \text{Ker}(T)$$

which means we can write it as a linear combination of the vectors in $\mathcal{B}$,

$$b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p = a_1 \mathbf{e}_1 + \cdots + a_k \mathbf{e}_k.$$

Bringing all terms onto one side, this becomes the equation

$$-a_1 \mathbf{e}_1 - \cdots - a_k \mathbf{e}_k + b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p = \mathbf{0}_V.$$

We recognize the left hand side as a linear combination of the $\mathcal{C}$ basis vectors. Since they are are linearly independent, all the scalars must be zero. In particular, $b_1 = \cdots = b_p = 0$, which is what we wanted to prove.

$\mathcal{D}$ spans W. Suppose $\mathbf{w} \in \text{Im}(T)$. We need to show that $\mathbf{w}$ is a linear combination of the vectors from $\mathcal{D}$. Since $\mathbf{w}$ is in the image of T, there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. Since $\mathcal{C}$ is a basis for V, we can write

$$\mathbf{v} = a_1 \mathbf{e}_1 + \cdots + a_k \mathbf{e}_k + b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p$$

for some scalars $a_1, \ldots, a_k, b_1, \ldots, b_p$. Then

$$\begin{aligned} \mathbf{w} &= T(\mathbf{v}) \\ &= T(a_1 \mathbf{e}_1 + \cdots + a_k \mathbf{e}_k + b_1 \mathbf{f}_1 + \cdots + b_p \mathbf{f}_p) \\ &= a_1 T(\mathbf{e}_1) + \cdots + a_k T(\mathbf{e}_k) + b_1 T(\mathbf{f}_1) + \cdots + b_p T(\mathbf{f}_p) \\ &= b_1 T(\mathbf{f}_1) + \cdots + b_p T(\mathbf{f}_p) \qquad\qquad (\mathbf{e}_i \in \text{Ker}(T)) \end{aligned}$$

so that $\mathbf{w}$ is indeed a linear combination of the vectors from $\mathcal{D}$. $\qquad \square$

Further exercises for 3.5

Exercise 87. Verify the Rank-Nullity theorem for the following linear maps. That is, for each map T, (a) determine $\mathrm{Ker}(T)$ and $\mathrm{Im}(T)$ explicitly, (b) determine the dimension of $\mathrm{Ker}(T)$ and $\mathrm{Im}(T)$, (c) check that these numbers satisfy the Rank-Nullity theorem.

1. The identity map $\mathrm{id}_V : V \to V$ on a finite-dimensional vector space V.

2. The zero map

$$Z : V \to V$$
$$\mathbf{v} \mapsto \mathbf{0}$$

   on a finite-dimensional vector space V.

3. The map

$$T : \mathrm{Poly}_3 \to \mathrm{Col}_3$$

$$p \mapsto \begin{bmatrix} p(1) \\ p(2) \\ p(3) \end{bmatrix}$$

4. The map

$$S : \mathrm{Trig}_2 \to \mathrm{Col}_2$$

$$f \mapsto \begin{bmatrix} \int_0^\pi f(x) \cos x \, dx \\ \int_0^\pi f(x) \sin x \, dx \end{bmatrix}$$

Exercise 88. Given an example of a linear map $T : \mathrm{Col}_4 \to \mathrm{Col}_4$ such that $\mathrm{Rank}(T) = \mathrm{Nullity}(T)$.

Exercise 89. For each of the following assertions, state whether it is **true** or **false**. If it is true, prove it. If it is false, prove that it is false.

1. There exists a linear map $T : \mathbb{R}^5 \to \mathbb{R}^2$ such that

$$\mathrm{Ker}(T) = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 : x_1 = 3x_2 \text{ and } x_3 = x_4 = x_5\}.$$

2. There exists a linear map $F : \mathrm{Trig}_3 \to \mathrm{Trig}_3$ such that $\mathrm{Rank}(T) = \mathrm{Nullity}(T)$.

Exercise 90. Let $f(x, y, z)$ be a function on $\mathbb{R}^3$ and fix a point $\mathbf{p} = (x_0, y_0, z_0) \in \mathbb{R}^3$. For each vector $\mathbf{u} \in \mathbb{R}^3$, we can regard the derivative of f in the direction of $\mathbf{u}$ at $\mathbf{p}$ as a map

$$D_{\mathbf{p}} : \mathbb{R}^3 \to \mathbb{R}$$
$$\mathbf{u} \mapsto (\nabla f)(\mathbf{p}) \cdot \mathbf{u}.$$

1. Show that $D_{\mathbf{p}}$ as defined above is a linear map.

2. Consider the example of $f(x, y, z) = x^2 + y^2 + z^2$. Determine $\mathrm{Ker}(D_{\mathbf{p}})$ for all points $\mathbf{p} \in \mathbb{R}^3$.

Exercise 91. Using the Rank-Nullity theorem, give a different proof of the fact that the image of the map C in Example 3.19 is $\{\mathbf{u} \in \mathbb{R}^3 : \mathbf{u} \cdot \mathbf{a} = 0\}$.

3.6 Injective and surjective linear maps

Definition 52. A function $f : X \to Y$ from a set X to a set Y is called *one-to-one* (or *injective*) if whenever $f(x) = f(x')$ for some $x, x' \in X$ it necessarily holds that $x = x'$. The function f is called *onto* (or *surjective*) if for all $y \in Y$ there exists an $x \in X$ such that $f(x) = y$.

If f is a *linear map* between vector spaces (and not just an arbitrary function between sets), there is a simple way to check if f is injective.

Lemma 53. *Let $T : V \to W$ be a linear map between vector spaces. Then:*

$$T \text{ is injective} \iff \mathrm{Ker}(T) = \{\mathbf{0}_V\}.$$

Proof. $\Rightarrow$. Suppose $T : V \to W$ is one-to-one. We already know one element in $\mathrm{Ker}(T)$, namely $\mathbf{0}_V$, since $T(\mathbf{0}_V) = \mathbf{0}_W$ as T is linear. Since T is one-to-one, this must be the only element in $\mathrm{Ker}(T)$.
$\Leftarrow$. Suppose $\mathrm{Ker}(T) = \{\mathbf{0}_V\}$. Now, suppose that

$$T(\mathbf{v}) = T(\mathbf{v}')$$

for some vectors $\mathbf{v}, \mathbf{v}' \in V$. Then we have $T(\mathbf{v}) - T(\mathbf{v}') = \mathbf{0}_W$, and since T is linear, this means $T(\mathbf{v} - \mathbf{v}') = \mathbf{0}_W$. Hence $\mathbf{v} - \mathbf{v}' \in \mathrm{Ker}(T)$, and so $\mathbf{v} - \mathbf{v}' = \mathbf{0}_V$, in other words, $\mathbf{v} = \mathbf{v}'$, which is what we wanted to show. $\square$

Another simplification occurs if T is a linear map from a vector space V to itself (i.e. T is a linear operator on V), and V is finite-dimensional.

Lemma 54. *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V. Then:*

$$T \text{ is injective} \iff T \text{ is surjective}.$$

Proof. $\Rightarrow$. Suppose T is injective.

$$\therefore \text{Ker}(T) = \{\mathbf{0}_V\} \qquad\qquad \text{(by Lemma 53)}$$
$$\therefore \text{Nullity}(T) = 0$$
$$\therefore \text{Rank}(T) = \text{Dim}(V) \qquad \text{(by Rank-Nullity Theorem)}$$
$$\therefore \text{Im}(T) = V \qquad\qquad \text{(by Corollary 24)}$$

Hence T is surjective.

$\Leftarrow$. Suppose T is surjective.

$$\therefore \text{Im}(T) = V$$
$$\therefore \text{Rank}(T) = \text{Dim}(V)$$
$$\therefore \text{Nullity}(T) = 0 \qquad \text{(by Rank-Nullity Theorem)}$$
$$\therefore \text{Ker}(T) = \{\mathbf{0}_V\}$$
$$\therefore T \text{ is injective.} \qquad \text{(by Lemma 53)}$$

$\square$

Proposition 55. *A linear map $T : V \to W$ is an isomorphism if and only if T is injective and surjective.*

Proof. $\Rightarrow$. Suppose V and W are isomorphic. That is, there exists a pair of linear maps $T : V \rightleftarrows W : S$ such that $T \circ S = \text{id}_W$ and $S \circ T = \text{id}_C$. We will show that T is injective and surjective.

$$\text{Suppose that } T(\mathbf{v}_1) = T(\mathbf{v}_2).$$
$$\therefore S(T(\mathbf{v}_1)) = S(T(\mathbf{v}_2))$$
$$\therefore \text{id}_V(\mathbf{v}_1) = \text{id}_V(\mathbf{v}_2) \quad \text{(using } S \circ T = \text{id}_V)$$
$$\therefore \mathbf{v}_1 = \mathbf{v}_2$$

which shows that T is injective. To show that T is surjective, let $\mathbf{w} \in W$. We must show that there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. Indeed, set $\mathbf{v} := S(\mathbf{w})$. Then

$$T(\mathbf{v}) = T(S(\mathbf{w}))$$
$$= \text{id}_W(\mathbf{w}) \quad \text{(using } T \circ S = \text{id}_W)$$
$$= \mathbf{w}.$$

$\Leftarrow$. Suppose that there exists a linear map $T : V \to W$ which is injective and surjective. We will show that there exists a linear map $S : W \to V$ such that $S \circ T = \text{id}_V$ and $T \circ S = \text{id}_W$, which will prove that V and W are isomorphic.

We define the inverse map S as follows:

$$S : W \to V$$
$$\mathbf{w} \mapsto \text{the unique } \mathbf{v} \in V \text{ such that } T(\mathbf{v}) = \mathbf{w}.$$

This map is well-defined. Indeed, given $\mathbf{w} \in W$, the fact that T is surjective means there does exist *some* $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. The fact that T is injective implies that this $\mathbf{v}$ is unique. For if there exists another $\mathbf{v}' \in V$ with $S(\mathbf{v}') = \mathbf{w}$, then we have $\mathbf{v}' = \mathbf{v}$ since T is injective.

At this point we have a well-defined *function* $S : W \to V$ which satisfies $T \circ S = \mathrm{id}_W$ and $S \circ T = \mathrm{id}_V$. We only need to check that S is linear.

Let $\mathbf{w}_1, \mathbf{w}_2 \in W$. Then

$$
\begin{aligned}
S(a\mathbf{w}_1 + b\mathbf{w}_2) &= S\big(aT(S(\mathbf{w}_1)) + bT(S(\mathbf{w}_2))\big) && (\text{using } T \circ S = \mathrm{id}_W) \\
&= S\big(aT(\mathbf{v}_1) + bT(\mathbf{v}_2)\big) && \left(\begin{matrix} \text{setting } \mathbf{v}_1 := S(\mathbf{w}_1), \\ \mathbf{v}_2 := S(\mathbf{w}_2) \end{matrix}\right) \\
&= S\big(T(a\mathbf{v}_1 + b\mathbf{v}_2)\big) && (T \text{ is linear}) \\
&= a\mathbf{v}_1 + b\mathbf{v}_2 && (S \circ T = \mathrm{id}_V)
\end{aligned}
$$

Hence S is linear, which completes the proof. $\qquad\square$

Proposition 56. *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V. The following statements are equivalent:*

1. *T is injective.*

2. *T is surjective.*

3. *T is an isomorphism.*

Proof. (1) is equivalent to (2) by Lemma 54. On the other hand, (1) and (2) is equivalent to (3) by Proposition 55. $\qquad\square$

Chapter 4

Eigenvalues and eigenvectors

4.1 Eigenvalues

Definition 57. Let $T : V \to V$ be a linear operator on a vector space V. We say that $\lambda \in \mathbb{R}$ is an *eigenvalue of T* if there exists a nonzero vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$.

Proposition 58. *Suppose that $T : V \to V$ is a linear map, and V is finite-dimensional. Then the following are equivalent:*

1. *λ is an eigenvalue of T.*

2. *$\lambda\,\mathrm{id}_V - T$ is not injective.*

3. *$\lambda\,\mathrm{id}_V - T$ is not surjective.*

4. *$\lambda\,\mathrm{id}_V - T$ is not invertible.*

I hope you understand the notation here. If $S, T : V \to W$ are linear maps, we can *add them* together to get a new linear map

$$S + T : V \to W$$

defined by $(S + T)(\mathbf{v}) := S(\mathbf{v}) + T(\mathbf{v})$. So when I write '$T - \lambda\,\mathrm{id}_V$', I am referring to the linear map $V \to V$ defined by

$$(\lambda\,\mathrm{id}_V - T)(\mathbf{v}) := \lambda\,\mathrm{id}_V(\mathbf{v}) - T(\mathbf{v}) = \lambda\mathbf{v} - T(\mathbf{v}).$$

Proof. $(1) \Rightarrow (2)$. Suppose λ is an eigenvalue of T.

$\therefore$ There exists a nonzero $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$.

$\therefore \lambda\mathbf{v} - T(\mathbf{v}) = \mathbf{0}$

$\therefore (\lambda\operatorname{id}_V - T)(\mathbf{v}) = \mathbf{0}$

$\therefore \mathbf{v} \in \operatorname{Ker}(\lambda\operatorname{id}_V - T)$

But $\mathbf{v}$ is nonzero. So we have a nonzero vector in the kernel of $\lambda\operatorname{id}_V - T$, so $\lambda\operatorname{id}_V - T$ is not injective, by Lemma 53.

$(2) \Rightarrow (1)$. Suppose $\lambda\operatorname{id}_V - T$ is not injective.

$\therefore \operatorname{Ker}(\lambda\operatorname{id}_V - T) \neq \{\mathbf{0}\}$ (by Lemma 53)

$\therefore$ there exists a nonzero $\mathbf{v} \in V$ such that

$(\lambda\operatorname{id}_V - T)(\mathbf{v}) = \mathbf{0}$.

That is, $T(\mathbf{v}) = \lambda\mathbf{v}$, so λ is an eigenvalue of T.

Clearly $(2) \Leftrightarrow (3) \Leftrightarrow (4)$ by Proposition 56. $\square$

We have a simple test for when a *matrix A* is invertible: simply compute the determinant $\det(A)$. If $\det(A) = 0$, then the matrix is *not* invertible. If $\det(A) \neq 0$, the matrix is invertible.

Indeed, this is where the word 'determinant' comes from! It *determines* whether a matrix is invertible or not.

But how do we know when a *linear operator* is invertible?

Definition 59. The *determinant* of a linear operator $T : V \to V$ on a finite-dimensional vector space V is the determinant of the matrix of T relative to any basis $\mathcal{B}$ of V. That is,

$$\det(T) := \det\left([T]_{\mathcal{B}\leftarrow\mathcal{B}}\right), \quad \mathcal{B} \text{ any basis for } V.$$

Lemma 60. *The determinant of a linear operator, as defined above, is well-defined. That is, if $\mathcal{B}$ and $\mathcal{C}$ are bases for V, then*

$$\det\left([T]_{\mathcal{C}\leftarrow\mathcal{C}}\right) = \det\left([T]_{\mathcal{B}\leftarrow\mathcal{B}}\right).$$

Proof. We know from Theorem 47 that

$$[T]_{\mathcal{C}\leftarrow\mathcal{C}} = \mathsf{P}^{-1}[T]_{\mathcal{B}\leftarrow\mathcal{B}}\mathsf{P}$$

where $\mathsf{P} = \mathsf{P}_{\mathcal{B}\leftarrow\mathcal{C}}$. Therefore,

$$\begin{aligned} \det\left([T]_{\mathcal{C}\leftarrow\mathcal{C}}\right) &= \det\left(\mathsf{P}^{-1}[T]_{\mathcal{B}\leftarrow\mathcal{B}}\mathsf{P}\right) \\ &= \det(\mathsf{P}^{-1})\det\left([T]_{\mathcal{B}\leftarrow\mathcal{B}}\right)\det(\mathsf{P}) \quad (\det(\mathsf{AB}) = \det(\mathsf{A})\det(\mathsf{B})) \\ &= \frac{1}{\det(\mathsf{P})}\det\left([T]_{\mathcal{B}\leftarrow\mathcal{B}}\right)\det(\mathsf{P}) \\ &= \det\left([T]_{\mathcal{B}\leftarrow\mathcal{B}}\right). \end{aligned}$$

$\square$

Now that we know how to define the determinant of a linear operator, we can make the following definition.

Definition 61. The *characteristic polynomial* χ_T of a linear operator $T : V \to V$ on a finite-dimensional vector space V is defined by

$$\chi_T(\lambda) := \det(\lambda\,\mathrm{id} - T).$$

The reason that χ_T is a *polynomial* (and not just an arbitrary function) comes from the formula for the determinant. More on this later.

In summary, we see that λ is an eigenvalue of a linear operator T if and only if λ is a root of the characteristic polynomial of T. Let us record this formally as an extension of Proposition 58.

Proposition 62. *Let T be a linear operator on a finite-dimensional vector space V, and $\lambda \in \mathbb{R}$. Then the following are equivalent:*

1. *λ is an eigenvalue of T.*

2. *$\lambda\,\mathrm{id} - T$ is not invertible.*

3. *$\chi_T(\lambda) = 0$.*

Proof. The only thing left to show is that $(2) \Leftrightarrow (3)$. Indeed, if $\mathcal{B}$ is a basis for V, then

$\lambda\,\mathrm{id} - T$ is not invertible
$\Leftrightarrow$ the matrix $[\lambda\,\mathrm{id} - T]_{\mathcal{B}\leftarrow\mathcal{B}}$ is not invertible $\qquad$ (Corollary 44)
$\Leftrightarrow \det\left([\lambda\,\mathrm{id} - T]_{\mathcal{B}\leftarrow\mathcal{B}}\right) = 0$ $\qquad$ (Property of det)
$\Leftrightarrow \chi_T(\lambda) = 0$ $\qquad$ (Defn of $\chi_T(\lambda)$)

$\square$

Example 4.1. Find the eigenvalues of the linear operator

$$T : \text{Poly}_2 \to \text{Poly}_2$$

defined by

$$T(p)(x) := p(2x + 3)$$

Solution. We first need to determine the matrix of T relative to some basis for Poly_2. Let

$$\mathcal{B} := \{p_0, p_1, p_2\}$$

be the standard basis for Poly_2, that is

$$p_0(x) = 1, \quad p_1(x) = x, \quad p_2(x) = x^2.$$

Then

$$\begin{aligned} T(p_0)(x) &= p_0(2x + 3) \\ &= 1 \\ \therefore T(p_0) &= p_0 \end{aligned} \tag{4.1}$$

$$\begin{aligned} T(p_1)(x) &= p_1(2x + 3) \\ &= 2x + 3 \\ \therefore T(p_1) &= 3p_0 + 2p_1 \end{aligned} \tag{4.2}$$

$$\begin{aligned} T(p_2)(x) &= p_2(2x + 3) \\ &= (2x + 3)^2 \\ &= 4x^2 + 12x + 9 \\ \therefore T(p_2) &= 9p_0 + 12p_1 + 4p_0. \end{aligned} \tag{4.3}$$

So the matrix of T relative to $\mathcal{B}$ is

$$[T]_{\mathcal{B} \leftarrow \mathcal{B}} = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 2 & 12 \\ 0 & 0 & 4 \end{bmatrix}$$

So the characteristic polynomial of T is

$$
\begin{aligned}
\chi_T(\lambda) &= \det(\lambda\,\mathrm{id} - T) && \text{(defn of } \chi_T) && (4.4)\\
&= \det\left([\lambda\,\mathrm{id} - T]_{\mathcal{B}\leftarrow\mathcal{B}}\right) && \text{(defn of det)} && (4.5)\\
&= \det\left([\lambda\,\mathrm{id}]_{\mathcal{B}\leftarrow\mathcal{B}} - [T]_{\mathcal{B}\leftarrow\mathcal{B}}\right) &&&& (4.6)\\
&= \det\left(\begin{bmatrix} \lambda - 1 & -3 & -9 \\ 0 & \lambda - 2 & -12 \\ 0 & 0 & 4 \end{bmatrix}\right) &&&& (4.7)\\
&= (\lambda - 1)(\lambda - 2)(\lambda - 4). &&&& (4.8)
\end{aligned}
$$

The eigenvalues of T are the roots of the characteristic polynomial of T. So the eigenvalues of T are 1, 2 and 4.

When we speak about the eigenvalues and eigenvectors of an *$n \times n$ matrix* A, we are referring to the eigenvalues and eigenvectors of the associated linear operator

$$
\begin{aligned}
T_\mathsf{A} : \mathrm{Col}_n &\to \mathrm{Col}_n \\
\mathbf{v} &\mapsto \mathsf{A}\mathbf{v}.
\end{aligned}
$$

Now Col_n has the standard basis

$$
\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},\ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix},\ \cdots,\ \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.
$$

The matrix of T_A with respect to this standard basis $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is just A itself.

Exercise 92. Check this. In other words, check that for an $n \times n$ matrix A, $[T_\mathsf{A}]_{\mathcal{B}\leftarrow\mathcal{B}} = \mathsf{A}$.

So the eigenvalues of a matrix A are just the solutions of the equation $\det(\lambda I - \mathsf{A}) = 0$.

The following examples come from the online Desmos worksheet on eigenvalues at `student.desmos.com`, class code `RGJ93S`. In the Desmos worksheet, you find the eigenvalues graphically, by inspection. Work through that worksheet first, and only then follow the algebraic method presented below.

Example 4.2. Find the eigenvalues of $A = \begin{bmatrix} 2 & 1 \\ \frac{1}{2} & \frac{5}{2} \end{bmatrix}$.

Solution. The characteristic polynomial of A is

$$
\begin{aligned}
\chi_A(\lambda) &= \det\left(\begin{bmatrix} \lambda - 2 & -1 \\ -\frac{1}{2} & \lambda - \frac{5}{2} \end{bmatrix}\right) \\
&= (\lambda - 2)(\lambda - \frac{5}{2}) - \frac{1}{2} \\
&= \Leftrightarrow \lambda^2 - \frac{9}{2}\lambda + \frac{9}{2}.
\end{aligned}
$$

The eigenvalues are the roots of the characteristic polynomial. So:

$$
\chi_A(\lambda) = 0 \Leftrightarrow (\lambda - \frac{3}{2})(\lambda - 3) = 0
$$
$$
\Leftrightarrow \lambda = \frac{3}{2} \text{ or } \lambda = 3.
$$

So the eigenvalues of A are $\lambda = \frac{3}{2}$ and $\lambda = 3$. This can be checked visually in the Desmos worksheet: for certain vectors $\mathbf{v}$, $A\mathbf{v} = \frac{3}{2}\mathbf{v}$ and for certain other vectors $\mathbf{v}$, $A\mathbf{v} = 3\mathbf{v}$.

Example 4.3. Find the eigenvalues of $A = \begin{bmatrix} 1 & 3 \\ -2 & 2 \end{bmatrix}$.

Solution. The characteristic polynomial of A is

$$
\det(\lambda I - A) = \det\begin{bmatrix} \lambda - 1 & 3 \\ \lambda + 2 & 2 \end{bmatrix} = \lambda^2 - 3\lambda + 8.
$$

The equation for eigenvalues is thus

$$
\lambda^2 - 3\lambda + 8 = 0
$$

which has no real roots. So A has no eigenvalues. This can be confirmed visually on the Desmos worksheet: roughly speaking, $A\mathbf{v}$ rotates the vector $\mathbf{v}$ clockwise by a certain amount. So $A\mathbf{v}$ can never be a multiple of $\mathbf{v}$.

4.2 Eigenvectors

In the previous section we focused on eigenvalues. Now we turn our focus to eigenvectors.

Definition 63. An *eigenvector* of a linear operator $T : V \to V$ is a nonzero vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda \mathbf{v}$ for some scalar $\lambda \in \mathbb{R}$ (the associated eigenvalue).

Note that we demand that $\mathbf{v}$ be nonzero in order to call it an eigenvector of T. The reason for this convention is that, otherwise, the zero vector will be an eigenvector of *every* operator T (with eigenvalue 0), since we always have $T(\mathbf{0}) = 0\mathbf{0}$. In that case 'being an eigenvector' wouldn't be a very special property of T.

> It is perfectly fine for a linear operator T to have a zero *eigenvalue*. But our convention is that an *eigenvector* cannot be the zero vector.

Instead of considering just a *single* vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda \mathbf{v}$, it is often more natural to consider the set of *all* such vectors.

Definition 64. Let $T : V \to V$ be a linear operator on a vector space V, and let λ be an eigenvalue of T. The *eigenspace of T corresponding to λ* is the set

$$E_\lambda := \{\mathbf{v} \in V : T(\mathbf{v}) = \lambda \mathbf{v}.\}$$

> Note that if λ is an eigenvalue of T, then the zero vector $\mathbf{0}$ *is* in the eigenspace E_λ of T associated to λ. But it is not an *eigenvector* of T. That is just terminology.

Note that we can also express E_λ as

$$E_\lambda = \mathrm{Ker}(\lambda \, \mathrm{id} - T) \tag{4.9}$$

since $T(\mathbf{v}) = \lambda \mathbf{v}$ is equivalent to saying that $(\lambda \, \mathrm{id} - T)(\mathbf{v}) = \mathbf{0}$.

Proposition 65. *E_λ is a subspace of V.*

Proof. This follows immediately from the fact (equation (4.9)) that we can write E_λ as a kernel of a certain linear map. The kernel of any linear operator on V is a subspace of V, by Lemma 49. $\qquad\square$

Proposition 66 (Eigenvectors from Distinct Eigenspaces are Linearly Independent). *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V, and suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are eigenvectors of T with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent.*

Proof. Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is *not* linearly independent, in other words that it linearly dependent. Then, by the Linear Combination of Preceding Vectors Proposition (Proposition 14), either $\mathbf{v}_1 = \mathbf{0}$ (impossible, since eigenvectors must be nonzero), or one of the $\mathbf{v}_i$ is a linear combination of the preceding vectors. Let i be the *smallest* index for which this is true. So, for some i with $2 \le i \le k$, we have

$$\mathbf{v}_i = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_{i-1} \mathbf{v}_{i-1} \tag{4.10}$$

for some scalars $a_1, \ldots, a_{i-1}$, not all zero, and moreover the list of vectors

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}\}$$

is linearly independent (otherwise, i would not be the *smallest* index for which $\mathbf{v}_i$ is a linear combination of the preceding vectors). Applying T to both sides of (4.10) gives us a new equation:

$$\begin{aligned} T(\mathbf{v}_i) &= T(a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_{i-1} \mathbf{v}_{i-1}) \\ &= a_1 T(\mathbf{v}_1) + a_2 T(\mathbf{v}_2) + \cdots + a_{i-1} T(\mathbf{v}_{i-1}) \quad (T \text{ is linear}) \\ \therefore \lambda_i \mathbf{v}_i &= a_1 \lambda_1 \mathbf{v}_1 + a_2 \lambda_2 \mathbf{v}_2 + \cdots + a_{i-1} \lambda_{i-1} \mathbf{v}_{i-1} \quad (T(\mathbf{v}_r) = \lambda_r \mathbf{v}_r) \end{aligned} \tag{4.11}$$

Multiplying equation (4.10) by λ_i and subtracting (4.11) gives:

$$\mathbf{0} = (\lambda_i - \lambda_1) a_1 \mathbf{v}_1 + (\lambda_i - \lambda_2) a_2 \mathbf{v}_2 + \cdots + (\lambda_i - \lambda_{i-1}) a_{i-1} \mathbf{v}_{i-1}.$$

Since the list of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}\}$ is linearly independent, we have

$$(\lambda_i - \lambda_1) a_1 = 0, \ (\lambda_i - \lambda_2) a_2 = 0 \ \ldots, \ (\lambda_i - \lambda_{i-1}) a_{i-1} = 0.$$

Now all the eigenvalues are distinct, so we cannot have $\lambda_i - \lambda_r = 0$ for $r \ne i$. Hence $a_1 = 0$, $a_2 = 0$, …, $a_{i-1} = 0$. But this contradicts (4.10) which implied these scalars are *not* all zero. So our initial assumption must be false. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly independent. $\square$

Example 4.4. Determine the eigenvalues and corresponding eigenspaces of the linear operator

$$T : \mathrm{Poly}_4 \to \mathrm{Poly}_4$$

$$T(p)(x) = x^2 \frac{d^2}{dx^2}(p).$$

Solution. Computing the action of T with respect to the standard basis of Poly_2,

$$p_0(x) = 1,\ p_1(x) = x,\ p_2(x) = x^2,\ p_3(x) = x^3,\ p_4(x) = x^4$$

gives:

$$p_0 \mapsto 0$$
$$p_1 \mapsto 0$$
$$p_2 \mapsto 2p_2$$
$$p_3 \mapsto 6p_3$$
$$p_4 \mapsto 12p_4$$

By inspection, we see that p_0 and p_1 are eigenvectors with eigenvalue 0, p_1 is an eigenvector with eigenvalue 2, p_3 is an eigenvector with eigenvalue 6, and p_4 is an eigenvector with eigenvalue 12. So the eigenvalues and associated eigenspaces of T are:

$$\lambda = 0, \qquad E_0 = \{sp_0 + tp_1,\ s,t \in \mathbb{R}\}$$
$$\lambda = 2, \qquad E_1 = \{tp_2,\ t \in \mathbb{R}\}$$
$$\lambda = 6, \qquad E_2 = \{tp_3,\ t \in \mathbb{R}\}$$
$$\lambda = 12, \qquad E_1 2 = \{tp_4,\ t \in \mathbb{R}\}.$$

Example 4.5. Determine the eigenspaces of the linear operator

$$T : \mathrm{Poly}_2 \to \mathrm{Poly}_2$$
$$T(p)(x) = p(2x + 3)$$

from Example 4.1.

Solution. The eigenvalues of T are $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = 4$. Let us think for a second. Since the eigenvalues are all distinct, if we find eigenvectors q_1, q_2 and q_3 with eigenvalues λ_1, λ_2 and λ_3 respectively, then $\mathcal{Q} = \{q_1, q_2, q_3\}$ will be linearly independent by Proposition 66. This means that the eigenspaces E_1, E_2 and E_4 associated with these eigenvalues will be 1-dimensional. This is useful information. Ok, now we are ready to start computing these eigenspaces.

- $\lambda = 1$:

We seek polynomials q such that

$$T(q) = 1q$$

By inspection of the formulas (4.1)-(4.2) for how T works, we already see one such polynomial, namely $q = p_0$, since $T(p_0) = p_0$. Since E_1 is 1-dimensional, we conclude that p_0 is a basis for E_1, so

$$E_1 = \{tp_0,\, t \in \mathbb{R}\}.$$

- $\lambda = 2$:

  We seek polynomials q such that

  $$T(q) = 2q.$$

  Expand q in the standard basis as

  $$q = a_0 p_0 + a_1 p_1 + a_2 p_2.$$

  So we seek scalars a_0, a_1, a_2 such that

  $$T(a_0 p_0 + a_1 p_1 + a_2 p_2) = 2(a_0 p_0 + a_1 p_1 + a_2 p_2). \qquad (4.12)$$

  Using the formulas (4.9) for how T acts on the standard basis, (4.12) becomes

  $$(-a_0 + 3a_1 + 9a_2)p_0 + 12a_2 p_1 + 2a_2 p_2 = 0.$$

  Since $\{p_0, p_1, p_2\}$ is linearly independent, this becomes the following set of equations:

  $$\begin{cases} -a_0 + 3a_1 + \ 9a_2 = 0 \\ \qquad\qquad\quad 12a_2 = 0 \\ \qquad\qquad\quad 2a_2 = 0 \end{cases} \qquad (4.13)$$

  The general solution of these equations is

  $$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix},\ t \in \mathbb{R}.$$

  In other words, an eigenvector of T with eigenvalue 2 is

  $$q_2(x) = 3 + x,$$

and the whole eigenspace corresponding to the eigenvalue 2 is

$$E_2 = \{tq_2,\ t \in \mathbb{R}\}.$$

Let us check this. Consider computing $T(q_2)$ using the *definition* of T, namely $T(p)(x) = p(2x + 3)$ for any polynomial p. We get:

$$
\begin{aligned}
T(q_2)(x) &= q_1(2x + 3) \\
&= 3 + 3(2x + 3) \\
&= 6 + 6x \\
&= 2(3 + 3x) \\
&= (2q_2)(x) \\
\therefore T(q_2) &= 2q_1 \ \text{(as expected!)}
\end{aligned}
$$

- $\lambda = 4$:

  We seek polynomials q such that

  $$T(q) = 4q.$$

  Expand q in the standard basis as

  $$q = a_0 p_0 + a_1 p_1 + a_2 p_2$$

  So we seek scalars a_0, a_1, a_2 such that

  $$T(a_0 p_0 + a_1 p_1 + a_2 p_2) = 4(a_0 p_0 + a_1 p_1 + a_2 p_2). \qquad (4.14)$$

  Using the formulas (4.9) for how T acts on the standard basis, (4.14) becomes

  $$(-3a_0 + 3a_1 + 9a_2)p_0 + (-2a_1 + 12a_2)p_1 + 0p_2 = 0.$$

  Since $\{p_0, p_1, p_2\}$ is linearly independent, this becomes the following set of equations:

  $$
  \begin{cases}
  -3a_0 + 3a_1 + \ 9a_2 = 0 \\
  \qquad\quad - 2a_1 + 12a_2 = 0
  \end{cases}
  \qquad (4.15)
  $$

  The general solution of these equations is

  $$
  \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 9 \\ 6 \\ 1 \end{bmatrix},\ t \in \mathbb{R}.
  $$

In other words, an eigenvector of T with eigenvalue 4 is

$$q_4(x) = 9 + 6x + x^2,$$

and the whole eigenspace corresponding to the eigenvalue 4 is

$$E_2 = \{tq_4, \, t \in \mathbb{R}\}.$$

Let us check this. Consider computing $T(q_4)$ using the *definition* of T, namely $T(p)(x) = p(2x + 3)$ for any polynomial p. We get:

$$
\begin{aligned}
T(q_4)(x) &= q_4(2x + 3) \\
&= 9 + 6(2x + 3) + (2x + 3)^2 \\
&= 36 + 24x + 4x^2 \\
&= 4(9 + 6x + x^2) \\
&= (4q_4)(x) \\
\therefore T(q_4) &= 4q_4 \text{ (as expected!)}
\end{aligned}
$$

Example 4.6. Determine the eigenspaces of the matrix

$$\mathsf{A} = \begin{bmatrix} 2 & 1 \\ \frac{1}{2} & \frac{5}{2} \end{bmatrix}$$

from Example 4.2.

Solution. From (4.9), the eigenspace E_λ can be computed as

$$E_\lambda = \mathrm{Ker}(\mathsf{A} - \lambda I) = \mathrm{Ker}\left(\begin{bmatrix} 2 - \lambda & 1 \\ \frac{1}{2} & \frac{5}{2} - \lambda \end{bmatrix} \right)$$

The eigenvalues of A are $\lambda = \frac{3}{2}$ and $\lambda = 3$.

- $\lambda = \frac{3}{2}$: We must compute

$$E_{\frac{3}{2}} = \mathrm{Ker}\left(\begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{bmatrix} \right).$$

In other words, we must find those column vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ such that

$$\begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This boils down to the single equation

$$\frac{1}{2}v_1 + v_2 = 0$$

which has general solution $v_2 = t$, $v_1 = -2t$, $t \in \mathbb{R}$. Therefore,

$$E_{\frac{3}{2}} = \left\{ t \begin{bmatrix} -2 \\ 1 \end{bmatrix}, t \in \mathbb{R} \right\}.$$

You can verify visually that this line is indeed an eigenspace of A with corresponding eigenvalue $\frac{3}{2}$ on the Desmos worksheet. When you multiply vectors on this line with A, they get scaled by a factor of $\frac{3}{2}$.

- $\lambda = 3$: We must compute

$$E_3 = \mathrm{Ker}\left(\begin{bmatrix} -1 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \right).$$

In other words, we must find those column vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ such that

$$\begin{bmatrix} -1 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This boils down to the single equation

$$-v_1 + v_2 = 0$$

which has general solution $v_2 = t$, $v_1 = t$, $t \in \mathbb{R}$. Therefore,

$$E_{\frac{3}{2}} = \left\{ t \begin{bmatrix} -2 \\ 1 \end{bmatrix}, t \in \mathbb{R} \right\}.$$

You can verify visually that this line is indeed an eigenspace of A with corresponding eigenvalue 3 on the Desmos worksheet. When you multiply vectors on this line with A, they get scaled by a factor of 3.

Note that the 'kernel of a matrix' method we used to compute the eigenspaces in Example 4.6 can be used to find the eigenvectors of *any* linear operator.

Example 4.7. Determine the eigenspaces of the linear operator T from Example 4.5 using the 'kernel-of-a-matrix' method.

Solution. The eigenspace E_λ corresponding to the eigenvalue λ is the kernel of the linear operator $T - \lambda\,\mathrm{id}$. That is, we seek polynomials $q \in \mathrm{Poly}_2$ such that

$$(T - \lambda\,\mathrm{id})(q) = 0. \tag{4.16}$$

We can expand q in the standard basis $\mathcal{B}$:

$$q = a_0 p_0 + a_1 p_1 + a_2 p_2 \tag{4.17}$$

We can also compute the matrix of T with respect to the basis $\mathcal{B}$:

$$[T]_{\mathcal{B}\leftarrow\mathcal{B}} = \begin{bmatrix} 1 & 3 & 9 \\ 0 & 2 & 12 \\ 0 & 0 & 4 \end{bmatrix}.$$

Then equation (4.16) can be written in matrix form:

$$[T - \lambda\,\mathrm{id}]_{\mathcal{B}\leftarrow\mathcal{B}} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

In other words, for each eigenvalue λ, to compute E_λ we must compute the kernel of the *matrix*

$$\begin{bmatrix} 1-\lambda & 3 & 9 \\ 0 & 2-\lambda & 12 \\ 0 & 0 & 4-\lambda \end{bmatrix}$$

and then rewrite the resulting column vectors as polynomials, in the form (4.17).

- $\lambda = 1$:

  We seek scalars a_0, a_1, a_2 such that

  $$\begin{bmatrix} 0 & 3 & 9 \\ 0 & 1 & 12 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

  This equation has general solution

  $$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; t \in \mathbb{R}.$$

So,
$$E_1 = \{tp_0, \, t \in \mathbb{R}\}.$$

- $\lambda = 2$:

  We seek scalars a_0, a_1, a_2 such that
  $$\begin{bmatrix} -1 & 3 & 9 \\ 0 & 0 & 12 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

  This equation has general solution
  $$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \, t \in \mathbb{R}.$$

  So,
  $$E_2 = \{tq_2, \, t \in \mathbb{R}\}$$
  where $q_2(x) = 3 + x$.

- $\lambda = 4$:

  We seek scalars a_0, a_1, a_2 such that
  $$\begin{bmatrix} -3 & 3 & 9 \\ 0 & -2 & 12 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

  This equation has general solution
  $$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = t \begin{bmatrix} 9 \\ 6 \\ 1 \end{bmatrix}, \, t \in \mathbb{R}.$$

  So,
  $$E_2 = \{tq_4, \, t \in \mathbb{R}\}$$
  where $q_4(x) = 9 + 6x + 1$.

Further exercises for 4.2

Exercise 93. Determine the eigenvalues and eigenspaces of the linear operator

$$T : \mathrm{Poly}_2 \to \mathrm{Poly}_2$$

$$T(p)(x) := \frac{d^2}{dx^2} p(x^2 + x - 2).$$

Exercise 94. Determine the eigenvalues and eigenspaces of the linear operator

$$T : \mathrm{Poly}_2 \to \mathrm{Poly}_2$$

$$T(p)(x) := x^2 p\left(\frac{1}{x}\right)$$

4.3 Diagonalizing matrices

Definition 67. We say that an $n \times n$ matrix A is *diagonalizable* if there exists an invertible $n \times n$ matrix P such that $\mathsf{P}^{-1}\mathsf{AP}$ is a diagonal matrix.

Example 4.8. $\mathsf{A} = \begin{bmatrix} 2 & 1 \\ \frac{1}{2} & \frac{5}{2} \end{bmatrix}$ is diagonalizable, since if we set

$$\mathsf{P} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

then

$$\mathsf{P}^{-1} = -\frac{1}{3} \begin{bmatrix} 1 & -1 \\ -1 & -2 \end{bmatrix}$$

and we check,

$$\mathsf{P}^{-1}\mathsf{AP} = -\frac{1}{3} \begin{bmatrix} 1 & -1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ \frac{1}{2} & \frac{5}{3} \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= -\frac{1}{3} \begin{bmatrix} \frac{3}{2} & -\frac{3}{2} \\ -3 & -6 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= -\frac{1}{3} \begin{bmatrix} -\frac{9}{2} & 0 \\ 0 & -9 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 3 \end{bmatrix}$$

---

which is indeed a diagonal matrix. More on this in the second semester! For now this is all you need to know.

Appendices

Appendix A

Reminder about matrices

Let us recall a few things about matrices, and set up our notation.

An $n \times m$ *matrix* A is just a rectangular array of numbers, with n rows and m columns:

$$\mathsf{A} = \begin{bmatrix} \mathsf{A}_{11} & \mathsf{A}_{12} & \cdots & \mathsf{A}_{1m} \\ \mathsf{A}_{21} & \mathsf{A}_{22} & \cdots & \mathsf{A}_{2m} \\ \vdots & & \ddots & \vdots \\ \mathsf{A}_{n1} & \mathsf{A}_{n2} & \cdots & \mathsf{A}_{nm} \end{bmatrix}$$

I will always write matrices in 'sans serif' font, eg. A. It is difficult to 'change fonts' in handwritten text, but I encourage you to at least reserve the letters A, B, C, etc. for matrices, and S, T, etc. for linear maps!

Two $n \times m$ matrices A and B can be added, to get a new $n \times m$ matrix $\mathsf{A} + \mathsf{B}$:

$$(\mathsf{A} + \mathsf{B})_{ij} := \mathsf{A}_{ij} + \mathsf{B}_{ij}$$

There is the *zero $n \times m$ matrix*:

$$\mathsf{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

You can also multiply an $n \times m$ matrix A by a scalar k, to get a new $n \times m$ matrix $k\mathsf{A}$:

$$(k\mathsf{A})_{ij} := k\mathsf{A}_{ij}$$

109

Lemma 68.   *1. Equipped with these operations, the set* $\mathrm{Mat}_{n,m}$ *of all* $n \times m$ *matrices is a vector space.*

   2. The dimension of Mat_{nm} *is* nm, *with basis given by the matrices*

$$\mathsf{E}_{ij}, \quad i = 1 \ldots n, \, j = 1 \ldots m$$

   which have a 1 in the ith row and jth column and zeroes everywhere else.

Proof. Left as an exercise.   □

Example A.1. $\mathrm{Mat}_{2,2}$ has basis

$$\mathsf{E}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathsf{E}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \mathsf{E}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Usually A is a matrix, and A_{ij} is the element of the matrix at position (i, j). But now E_{ij} is a matrix in its own right! Its element at position (k, l) will be written as $(\mathsf{E}_{ij})_{kl}$. I hope you don't find this too confusing. In fact, we can write down an elegant formula for the elements of E_{ij} using the Kronecker delta symbol:

$$(\mathsf{E}_{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{A.1}$$

movebackintoramanujansays

Exercise 95. Check that (A.1) is indeed the correct formula for the matrix elements of E_{ij}.

Example A.2. We will write Col_n for the vector space $\mathrm{Mat}_{n,1}$ of n-dimensional *column vectors*, and we will write the standard basis vectors E_{i1} of Col_n more simply as $\mathbf{e}_i$:

$$\mathbf{e}_1 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 := \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Vectors in Col_n will be written in bold sans-serif font eg. $\mathbf{v} \in \mathrm{Col}_n$.

A.0.1 Matrix multiplication

Equipped with these operations, the set $\text{Mat}_{n,m}$ of all $n \times m$ matrices is a vector space (see Example 1.8), with dimension nm (see Example 2.15). We write Col_n for the vector space $\text{Mat}_{n,1}$ of n-dimensional *column vectors*.

The most important operation is *matrix multiplication*. An $n \times k$ matrix A can be multiplied from the right with a $k \times m$ matrix B to get a $n \times m$ matrix AB,

$$
\begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1k} \\
A_{21} & A_{22} & \cdots & A_{2k} \\
\vdots & & \ddots & \vdots \\
A_{n1} & A_{n2} & \cdots & A_{nk}
\end{bmatrix}
\begin{bmatrix}
B_{11} & B_{12} & \cdots & B_{1m} \\
B_{21} & B_{22} & \cdots & B_{2m} \\
\vdots & & \ddots & \vdots \\
B_{k1} & B_{k2} & \cdots & B_{km}
\end{bmatrix} =
$$

$$
\begin{bmatrix}
(AB)_{11} & (AB)_{12} & \cdots & (AB_{1m} \\
(AB)_{21} & (AB)_{22} & \cdots & (AB)_{2m} \\
\vdots & & \ddots & \vdots \\
(AB)_{k1} & (AB)_{k2} & \cdots & (AB)_{nm}
\end{bmatrix}
$$

by defining the entries of AB to be

$$
(AB)_{ij} := A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ik}B_{kj}.
$$

Proposition 69. *The above operations on matrices satisfy the following rules whenever the sums and products involved are defined:*

1. $(A + B)C = AC + BC$

2. $A(B + C) = AB + AC$

3. $(kA)B = A(kB) = k(AB)$

4. $(AB)C = A(BC)$

Proof. The proofs of (1) - (3) are all routine checks which you have hopefully done before. Let us prove (4), to practice using Σ-notation! Suppose A, B and C have sizes $n \times k$, $k \times r$ and $r \times m$ respectively, so that the matrix

products make sense. Then:

$$((\mathsf{AB})\mathsf{C})_{ij} = \sum_{p=1}^{r} (\mathsf{AB})_{ip}\mathsf{C}_{pj}$$

$$= \sum_{p=1}^{r} \left(\sum_{q=1}^{k} \mathsf{A}_{iq}\mathsf{B}_{qp} \right) \mathsf{C}_{pj}$$

$$= \sum_{p,q} \mathsf{A}_{iq}\mathsf{B}_{qp}\mathsf{C}_{pj}$$

$$= \sum_{q=1}^{k} \mathsf{A}_{iq} \left(\sum_{p=1}^{r} \mathsf{B}_{qp}\mathsf{C}_{pj} \right)$$

$$= \sum_{q=1}^{k} \mathsf{A}_{iq}(\mathsf{BC})_{qj}$$

$$= (\mathsf{A}(\mathsf{BC}))_{ij}.$$

$\square$

I hope the Σ-notation does not confuse you in the above proof! Let me write out the exact same proof *without* Σ-notation, in the simple case where A, B and C are all 2×2 matrices, and we are trying to calculate, say, the entry at position 11.

$$((\mathsf{AB})\mathsf{C})_{11} = (\mathsf{AB})_{11}\mathsf{C}_{11} + (\mathsf{AB})_{12}\mathsf{C}_{21}$$
$$= (\mathsf{A}_{11}\mathsf{B}_{11} + \mathsf{A}_{12}\mathsf{B}_{21})\mathsf{C}_{11} + (\mathsf{A}_{11}\mathsf{B}_{12} + \mathsf{A}_{12}\mathsf{B}_{22})\mathsf{C}_{21}$$
$$= \mathsf{A}_{11}\mathsf{B}_{11}\mathsf{C}_{11} + \mathsf{A}_{12}\mathsf{B}_{21}\mathsf{C}_{11} + \mathsf{A}_{11}\mathsf{B}_{12}\mathsf{C}_{21} + \mathsf{A}_{12}\mathsf{B}_{22}\mathsf{C}_{21}$$
$$= \mathsf{A}_{11}(\mathsf{B}_{11}\mathsf{C}_{11} + \mathsf{B}_{12}\mathsf{C}_{21}) + \mathsf{A}_{12}(\mathsf{B}_{21}\mathsf{C}_{11} + \mathsf{B}_{22}\mathsf{C}_{21})$$
$$= \mathsf{A}_{11}(\mathsf{BC})_{11} + \mathsf{A}_{12}(\mathsf{BC})_{21}$$
$$= (\mathsf{A}(\mathsf{BC}))_{11}.$$

Do you understand the Σ-notation proof now? The crucial step (going from the second to the fourth lines) is called *exchanging the order of summation*.

The *transpose* of an $n \times m$ matrix A is the $m \times n$ matrix A^T whose entries are given by

$$(\mathsf{A}^T)_{ij} := \mathsf{A}_{ji}.$$

Appendix B

Hypersurfaces

Definition 70. Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a function. The *zero set* of f is the subset $M \subset \mathbb{R}^{n+1}$ given by

$$M := \{\mathbf{x} \in \mathbb{R}^{n+1} : f(\mathbf{x}) = 0\}.$$

Definition 71. We say that a subset $M \subset \mathbb{R}^{n+1}$ is an *n-dimensional hypersurface* if there exists a function $f : \mathbb{R}^{n+1} \to \mathbb{R}$ such that M is the zero set of f, and if for all $\mathbf{p} \in M$,

$$\nabla_{\mathbf{p}} f \text{ exists and is not equal to } \mathbf{0}.$$

Example B.1 (Hyperbola)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function
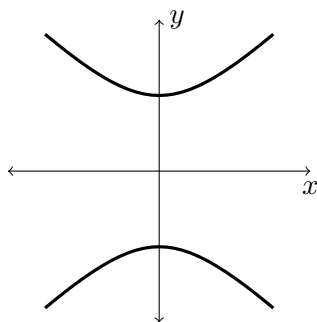
$$f(x, y) = y^2 - x^2 - 1$$

Draw a picture of the zero set M of f, and show that it is a hypersurface.

Solution. The zero level set of f is

$$M = \{(x, y) \in \mathbb{R}^2 : y^2 - x^2 - 1 = 0\}$$

which is a hyperbola: tikzalign

To show that M is a hypersurface, we first compute the gradient of f at a general point $\mathbf{r} = (x, y) \in \mathbb{R}^2$:

$$\nabla_{\mathbf{r}} f = (2y, -2x)$$

The gradient clearly exists everywhere. Moreover, for a point $\mathbf{p} = (a, b)$ on M, we have

$$\nabla_{\mathbf{p}} f = (2b, -2a) \quad \text{where } b^2 = 1 + a^2$$

Suppose $\nabla_{\mathbf{p}} f = (0, 0)$. Then in particular, $2b = 0$ so that $b = 0$, which means that $0 = 1 + a^2$ using the equation for M. This is impossible. Hence $\nabla_{\mathbf{p}} f \neq (0, 0)$ for all $\mathbf{p} \in M$, so that M is a hypersurface.

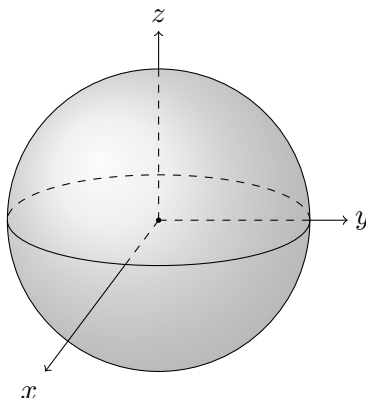**Example B.2** (Two-sphere). Let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function

$$f(x, y, z) = x^2 + y^2 + z^2 - 1.$$

Draw a picture of the level set M of f, and show that it is a differentiable hypersurface.

Solution. The level set of f is

$$M = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 - 1 = 0\}$$

This is called the *2-sphere*, and is pictured below: tikzalign

To show it is a hypersurface, we first compute ∇f at a general point $\mathbf{r} = (x, y, z) \in \mathbb{R}^3$:

$$\nabla_{\mathbf{r}} f = (2x, 2y, 2z)$$

This clearly exists everywhere. Moreover, for a point $\mathbf{p} = (a, b, c)$ on M, we have
$$\nabla_{\mathbf{p}} f = (2a, 2b, 2c) \neq (0, 0, 0)$$
since $a^2 + b^2 + c^2 = 1$. Hence M is a hypersurface.

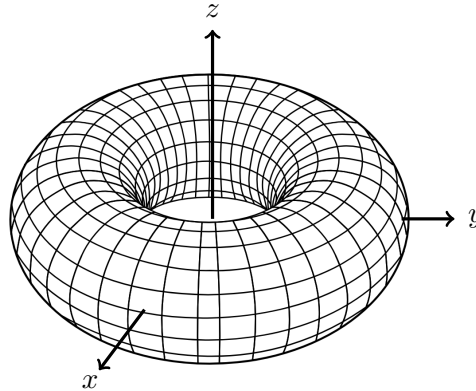**Example B.3.** Let $f : \mathbb{R}^3 \to \mathbb{R}$ be the function
$$f(x, y, z) = (2 - \sqrt{x^2 + y^2})^2 + z^2 - 1.$$

Draw a picture of the level set M of f, and show that it is a hypersurface.

Solution. The zero set of f is
$$M = \{(x, y, z) \in \mathbb{R}^3 : (2 - \sqrt{x^2 + y^2})^2 + z^2 - 1 = 0\}$$

This is a *torus*, and is pictured below: tikzalign

To show M is a hypersurface, we first compute ∇f at a general point $\mathbf{r} = (x, y, z) \in \mathbb{R}^3$:
$$\nabla_{\mathbf{r}} f = \left(\frac{2x(2 - \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}, \frac{2y(2 - \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}, 2z \right)$$

Note that ∇f doesn't exist *everywhere* in $\mathbb{R}^3$, since there is a problem when $x^2 + y^2 = 0$ (i.e. on the z-axis). But for a point $\mathbf{p} = (a, b, c)$ on M, we cannot have $a^2 + b^2 = 0$ since that implies, using the equation for M, that
$$(2 - \sqrt{0})^2 + c^2 - 1 = 0,$$

in other words, $c^2 = -3$ which is impossible. So $\nabla_{\mathbf{p}} f$ exists everywhere on M. Moreover, suppose $\nabla_{\mathbf{p}} f = (0, 0, 0)$ and $\mathbf{p}$ lies on M. Then in particular $2c = 0$ so $c = 0$, and hence, using the equation for M,

$$(2 - \sqrt{a^2 + b^2})^2 = 1$$

which implies that a and b cannot both be zero. This is a contradiction, hence $\nabla_{\mathbf{p}} \neq \mathbf{0}$ everywhere on M. So M is a hypersurface.

Definition 72. Let $M \subset \mathbb{R}^{n+1}$ be a hypersurface associated to a function $f : \mathbb{R}^{n+1} \to \mathbb{R}$. The *tangent space* of M at $\mathbf{p} \in M$ is:

$$T_{\mathbf{p}} M := \{\mathbf{v} \in \mathbb{R}^{n+1} : \nabla_{\mathbf{p}} f \cdot \mathbf{v} = 0\}$$

From Multivariable Calculus, you know that we can equivalently express the tangent space to M at p as

$$T_{\mathbf{p}} M = \{\mathbf{v} \in \mathbb{R}^{n+1} : D_{\mathbf{v}} f(\mathbf{p}) = 0\}$$
$$= \left\{\mathbf{v} \in \mathbb{R}^{n+1} : \frac{d}{dt}\Big|_{t=0} f(\mathbf{p} + t\mathbf{v}) = 0\right\}$$

Lemma 73. *If M*

Example B.4. Compute a basis for the tangent space to the hyperbola M from Example B.1 at $\mathbf{p} = (1, \sqrt{2})$. Draw a picture to illustrate the result.

Solution. We have already computed the gradient of f at a point $\mathbf{p} = (a, b) \in M$:
$$\nabla_{\mathbf{p}} = (2b, -2a) \quad \text{where } b^2 = 1 + a^2.$$
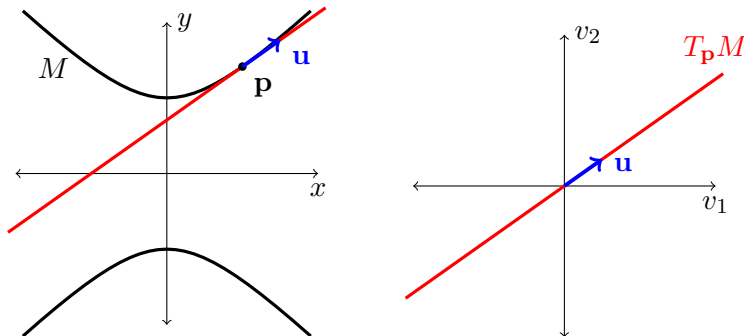Hence at $\mathbf{p} = (1, \sqrt{2})$, we have

$$\nabla_{\mathbf{p}} f = (2\sqrt{2}, -2).$$

Thus

$$T_{\mathbf{p}} M = \{(v_1, v_2) \in \mathbb{R}^2 : \nabla_{\mathbf{p}} f \cdot (v_1, v_2) = 0\}$$
$$= \{(v_1, v_2) \in \mathbb{R}^2 : 2\sqrt{2} v_1 - 2v_2 = 0\}$$

A basis for $T_{\mathbf{p}} M$ is $\mathbf{u} = (1, \frac{1}{\sqrt{2}})$. The picture is: tikzalign

The red line in the first picture, drawn for the sake of illustration, is the tangent space shifted so that it passes through **p**. The actual tangent space $T_{\mathbf{p}}M$ goes through the origin as in the second picture.