LINEAR ALGEBRA
UNIVERSITY OF BRISTOL

Lecturer: Dr. Charles Cox

These Lecture Notes[1] are for the second half of the course MATH10015, on *Linear Algebra*.

Contents

We have used $\mathbb{R}^n$ as a key example so far. This is an example of a vector space. We will not see the formal definition of this object until later, and so an important point must be made. When we said $\mathbb{R}^n$, we viewed this as a set of vectors of length n with real entries equipped with an addition '$+$' (defined as component-wise addition of real numbers) and a scalar multiplication '$\cdot$' (defined as multiplication of every entry by some specific scalar $\lambda \in \mathbb{R}$). In Section 6 we will formalise both of these ideas, with the additional layer of abstraction allowing our results to apply to many more general settings. But to start with, we will only need the examples of $\mathbb{R}^n$ and $\mathbb{C}^n$. The only difference when working with $\mathbb{C}^n$ is that both the entries and our scalars will be from $\mathbb{C}$ rather than just $\mathbb{R}$. The results we prove for finite dimensional vector spaces will apply to these two examples, as well as others. But we will also see a formal definition of dimension and encounter *infinite dimensional* vector spaces, for which some ideas generalise but other results break down. For this reason it is important to have our examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ in mind, but also that our statements can be applied more generally.

[1]Modified November 4, 2024 by Charles Garnet Cox

1. The different forms of a linear function

In this section we will see how to convert between three representations of a given linear function.

Definition 1.1. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called* $\mathbb{R}$**-linear** *if it satisfies two properties:*

- $f(v + w) = f(v) + f(w)$ *for all* $v, w \in \mathbb{R}^n$*; and*
- $f(\lambda v) = \lambda f(v)$ *for all* $v \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$*.*

Definition 1.2. *A function $f : \mathbb{C}^n \to \mathbb{C}^m$ is called* $\mathbb{C}$**-linear** *if it satisfies two properties:*

- $f(v + w) = f(v) + f(w)$ *for all* $v, w \in \mathbb{C}^n$*; and*
- $f(\lambda v) = \lambda f(v)$ *for all* $v \in \mathbb{C}^n$ *and* $\lambda \in \mathbb{C}$*.*

Both $\mathbb{R}$ and $\mathbb{C}$ are examples of *fields*. Another example is $\mathbb{Q}$. The exact nature of a field is not important for now, but we can now provide a definition of linear that allows us to deal with the examples of $\mathbb{Q}^n$, $\mathbb{R}^n$, and $\mathbb{C}^n$ simultaneously.

Definition 1.3. *Let V and W be vector spaces (technically over a field $\mathbb{F}$). Then a function $f : V \to W$ is called* $\mathbb{F}$**-linear** *if it satisfies two properties:*

- $f(v + w) = f(v) + f(w)$ *for all* $v, w \in V$*; and*
- $f(\lambda v) = \lambda f(v)$ *for all* $v \in V$ *and* $\lambda \in \mathbb{F}$*.*

If, from the context, the field $\mathbb{F}$ is clear, then we will talk about the function f being linear and omit the notation of $\mathbb{F}$.

Our focus in this section is to see three different ways to represent a linear map, and understand how to convert between these different forms. These forms generally are not named in the literature, and so we give informal names to them (and since these are not standard names, you may find other authors refer to them differently).

Example 1.4. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $f(x, y, z) = (3x, x + y + z)$. We can check this is $\mathbb{R}$-linear in two steps. First,

$$\begin{aligned} f((x_1, y_1, z_1) + (x_2, y_2, z_2)) &= f((x_1 + x_2, y_1 + y_2, z_1 + z_2)) \\ &= (3(x_1 + x_2), x_1 + x_2 + y_1 + y_2 + z_1 + z_2) \\ &= (3x_1, x_1 + y_1 + z_1) + (3x_2, x_2 + y_2 + z_2) \\ &= f((x_1, y_1, z_1)) + f((x_2, y_2, z_2)) \end{aligned}$$

and similarly

$$\begin{aligned} f(\lambda(x_1, y_1, z_1)) &= f((\lambda x_1, \lambda y_1, \lambda z_1)) \\ &= (3(\lambda x_1), \lambda x_1 + \lambda y_1 + \lambda z_1) \\ &= (\lambda 3(x_1), \lambda(x_1 + y_1 + z_1)) \\ &= \lambda(3(x_1), (x_1 + y_1 + z_1)) \\ &= \lambda f((x_1, y_1, z_1)) \end{aligned}$$

which clearly apply to any $x_1, x_2, y_1, y_2, z_1, z_2, \lambda \in \mathbb{R}$.

Example 1.5. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(x, y) = (x, -y, 1)$. Then this is not $\mathbb{R}$-linear. We can see this by producing a counterexample. This should either be a choice of $v, w \in \mathbb{R}^2$ such that $f(v + w) \neq f(v) + f(w)$ or a choice of $v \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$ such that $f(\lambda v) \neq \lambda f(v)$. Below we give one option for each such kind of counterexample.

- Let $v = (0, 0) = 0$ and $w = 0$. Then $f(v + w) = f(0 + 0) = f(0) = (0, 0, 1)$ whereas $f(v) + f(w) = f(0) + f(0) = (0, 0, 1) + (0, 0, 1) = (0, 0, 2)$ and so $f(v + w) \neq f(v) + f(w)$.
- Let $v = (0, 0)$ and $\lambda = 2$. Then $f(\lambda v) = f(0, 0) = (0, 0, 1)$ whereas $\lambda f(v) = 2f(0, 0) = (0, 0, 2)$ and so $f(\lambda v) \neq \lambda f(v)$.

Remark 1.6. From these examples, we can see that this definition is useful since it is easy to compute with, but has the drawback that it might not be immediately obvious that it is a linear

map[2]. This first form might be sensible to call the **algebraic form** of a linear map. Our other two forms always provide us with a linear map.

The following vectors are so ubiquitous that we give them their own notation.

Definition 1.7. *Fix an $n \in \{1, 2, 3, \ldots\}$ and choose an $i \in \{1, \ldots, n\}$. Then e_i is the vector in $\mathbb{R}^n$ with ith entry 1 and all other entries zero. Another way to phrase this is that the jth entry of e_i is given by the Kronecker delta function δ_{ij}.*

Example 1.8. We first look at the above vectors in $\mathbb{R}^n$ for various n.

- In $\mathbb{R}$, the only such vector is e_1, with $e_1 = (1)$.
- In $\mathbb{R}^2$, there are two such vectors, $e_1 = (1, 0)$ and $e_2 = (0, 1)$.
- In $\mathbb{R}^3$ we have $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$.

How this generalises to $\mathbb{R}^n$ is now hopefully clear.

Example 1.9. We consider how to write particular vectors using these e_i.

- We note that $(1, 1, 1) = e_1 + e_2 + e_3$.
- Similarly $(-1, 5, 0, \sqrt{2}) = -e_1 + 5e_2 + \sqrt{2}e_4$.
- Given $e_1 + 3e_2$ we know the first two entries of the vector, but do not know which space we are in. If we were told that this is a vector in $\mathbb{R}^5$, we would have that $e_1 + 3e_2 = (1, 3, 0, 0, 0)$.

It is now easier to introduce our second form for a linear map by first proving a lemma.

Lemma 1.10. *Let V and W be vector spaces over $\mathbb{F}$ and f and g be $\mathbb{F}$-linear functions from V to W such that $f(e_i) = g(e_i)$ for all valid i. Then $f = g$.*

Note: with the following proof, consider how you would justify each line.

Proof. Pick a vector $x = \sum_{i=1}^{k} x_i e_i$ where $x_1, \ldots, x_k \in \mathbb{F}$. Then

$$\begin{aligned} f(x) &= f(x_1 e_1 + \ldots + x_k e_k) \\ &= f(x_1 e_1) + \ldots + f(x_k e_k) \\ &= x_1 f(e_1) + \ldots + x_k f(e_k) \\ &= x_1 g(e_1) + \ldots + x_k g(e_k) \\ &= g(x_1 e_1) + \ldots + g(x_k e_k) \\ &= g(x_1 e_1 + \ldots + x_k e_k) \\ &= g(x). \end{aligned}$$

Since x was arbitrary, we have that f and g agree on all inputs, and hence $f = g$. □

Remark 1.11. The previous lemma says that a linear function is uniquely determined by its outputs on the vectors e_i. Also, for any choice of outputs for the e_i there is one, and only one, linear function with those images.

Example 1.12. Let $f : e_1 \mapsto e_1 + e_3$, $f : e_2 \mapsto -e_2 + e_3$ and be a linear map from $\mathbb{R}^2$ to $\mathbb{R}^3$. We will write f in the first form that we saw. The idea is really that from the previous proof. For ease of following the steps, we break them into bullet points.

- We wish to work out the image of (x, y) where $x, y \in \mathbb{R}$ are arbitrary. Let $a, b \in \mathbb{R}$.
- First, note that $f(ae_1) = af(e_1)$ because f is linear. Thus $f(ae_1) = ae_1 + ae_3$.
- Secondly, $f(be_2) = bf(e_2)$ and so $f(be_2) = -be_2 + be_3$.
- Putting these together, $f(ae_1 + be_2) = f(ae_1) + f(be_2) = ae_1 + ae_3 - be_2 + be_3$.
- Hence $f(a, b) = (a, -b, a + b)$. But we have not used any property of our $a, b \in \mathbb{R}$ and so $f(x, y) = (x, -y, x + y)$ for every $x, y \in \mathbb{R}$. This is our linear function in algebraic form.

[2]Any linear map f satisfies $f(0) = 0$, but show that $g : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (0, \sin(x))$ is not linear in order to deduce that $g(0) = 0$ does not guarantee that a map is linear.

Remark 1.13. Note that $\{e_1, \ldots, e_n\}$ is known as the standard basis of $\mathbb{R}^n$. We could therefore name this second form the **standard basis form** of the linear map or more informally perhaps the e_i **form** of a linear map.

The previous example showed how to change a linear operator in standard basis form into one of algebraic form. The next example shows us how to convert the other way.

Example 1.14 (Example 1.4 continued)**.** We will see how to convert the linear map above to standard basis form. The function was $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by $f(x, y, z) = (3x, x + y + z)$.

- Our first question is 'What size set must we use to determine f?'. The answer is 3.
- From the function, we immediately know the image of e_1, e_2, and e_3. Consider why this should be our aim, and how we the image of these vectors.
- The correct answer is therefore that $f : \mathbb{R}^3 \to \mathbb{R}^2$ is given by $f(e_1) = 3e_1 + e_2$, $f(e_2) = e_2$, and $f(e_3) = e_2$.

Our final form for a linear map is that given by a matrix. It is for their relation to linear maps that matrices are so scrutinised in courses on linear algebra. We have already justified the following, but will be so useful to us that we will refer to it as 'the fact' from now on.

FACT. *Let A be a matrix consisting of n columns $c_1, \ldots, c_n$ each of length m. Then multiplication by A provides a function taking vectors of length n and outputs vectors of length m. Furthermore:*

i) *multiplication by A is a linear map; and*
ii) *we have that $Ae_i = c_i$ for each $i = 1, \ldots, n$.*

Remark 1.15. An immediate consequence of (i) is that, given $A \in M_{m\times n}(\mathbb{R})$, the function $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f : x \mapsto Ax$ is a linear function. We will therefore call it the **matrix form** of a linear map.

Part (ii) of the fact is immensely useful. Our first application is to convert between the standard basis form and the matrix form of a linear map.

Example 1.16. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $f : \mathbb{R}^2 \to \mathbb{R}^2$, $x \mapsto Ax$. Our aim is to write f in standard basis form. The information can just be read off from the matrix A. From the fact, we know that A sends e_1 to c_1 and e_2 to c_2, where c_1 is the first column of A and c_2 is the second column of A. Thus $A(e_1) = e_1$ and $A(e_2) = e_1 + e_2$. Hence the linear function f is uniquely determined by knowing that $f(e_1) = e_1$ and $f(e_2) = e_1 + e_2$.

Example 1.17 (Example 1.4 continued)**.** Recall that we know the function f in standard basis form. We had that $f : \mathbb{R}^3 \to \mathbb{R}^2$ is given by $f(e_1) = 3e_1 + e_2$, $f(e_2) = e_2$, and $f(e_3) = e_2$. We will use this knowledge to write f in matrix form.

- First, we wish to determine the dimensions of A. We know that it should represent the given function f, and that $f : \mathbb{R}^3 \to \mathbb{R}^2$. Thus A takes inputs of length 3 and outputs vectors of length 2. This makes A a matrix with 3 columns, each of length 2. In other words A is a real 2×3 matrix, or an element of $M_{2,3}(\mathbb{R})$. This means we wish to find the columns c_1, c_2, and c_3 of A, each of length 2.
- From part (ii) of the fact, the column c_i should be the result of Ae_i. But we want this to be $f(e_i)$ so that f agrees with multiplication by A. Thus $c_1 = f(e_1)$, $c_2 = f(e_2)$, and $c_3 = f(e_3)$. This information is given above, and so in this case we have that

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Note that we can easily check our answer by directly computing Ae_i for $i = 1, 2, 3$.

At this stage we can now convert freely between our three forms of a linear map. For completeness, we give examples converting directly between the algebraic form and the matrix form of a linear map.

Example 1.18 (Example 1.4 continued)**.** Recall the function $f : \mathbb{R}^3 \to \mathbb{R}^2$ given in algebraic form by $f(x, y, z) = (3x, x + y + z)$. We will now write f in matrix form.

- It suffices to find the columns of a matrix A so that $f(x) = Ax$ for all $x \in \mathbb{R}^3$.
- Using part (ii) of the fact, we can find the ith column of A by computing $f(e_i)$.
- In this case we have $f(e_1) = (3, 1)$, $f(e_2) = (0, 1)$, and $f(e_3) = (0, 1)$. Thus we know the columns of A, and hence know A.

Example 1.19. Let $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$. We will write the map $f(x) = Ax$ in algebraic form.

- We first decide on the domain and codomain of f. Since A takes vectors of length 2 and outputs vectors of length 3, we will say[3] that $f : \mathbb{R}^2 \to \mathbb{R}^3$.
- We now wish to compute, given a, $b \in \mathbb{R}$, where A sends (a, b). This can be computed directly from A.
- In this case we have $A(a, b) = (0, b, a)$. Because this computation applies generally, we have that $A(x, y) = (0, y, x)$ for all $x, y \in \mathbb{R}$.

We therefore have, in algebraic form, the function $f : \mathbb{R}^2 \to \mathbb{R}^3$, $(x, y) \mapsto (0, y, x)$.

It is now natural to question what happens if we try and find the matrix or standard basis form for a map that is not linear. We first apply the steps above in a naive way, and then see what goes wrong.

Example 1.20 (Example 1.5 continued)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(x, y) = (x, -y, 1)$. We will use this information to **try to** write f in matrix form **but this will go wrong**. Before continuing, consider why it will go wrong.

- First, we wish to determine the dimensions of A. We know that it should represent the given function f, and that $f : \mathbb{R}^2 \to \mathbb{R}^3$. Thus A takes inputs of length 2 and outputs vectors of length 3. This makes A a matrix with 2 columns, each of length 3. In other words A is a real 3×2 matrix, or an element of $M_{3,2}(\mathbb{R})$. This means we wish to find the columns c_1 and c_2 of A, each of length 3.
- From part (ii) of the fact, the column c_i should be the result of Ae_i. But we want this to be $f(e_i)$ so that f agrees with multiplication by A. Thus $c_1 = f(e_1)$ and $c_2 = f(e_2)$. This information can be computed directly from the function, and so in this case we have that

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

At this point we may feel either like toasting our success, or be confused that we have found a matrix. We might even check our answer by considering where A sends e_1 and e_2. We will have that $Ae_i = f(e_i)$. Importantly, though, we do not have that $A(x) = f(x)$ for all $x \in \mathbb{R}^2$. To see this we could consider the image of $(1, -1)$ under A and under f. We get $A(e_1 - e_2) = e_1 + e_2$, but know this cannot be the image of $e_1 - e_2$ under f as all images under f have final coordinate equal to 1. Checking, we see that $f(1, -1) = (1, 1, 1)$.

Remark 1.21. The above example shows us that we must proceed with caution when applying a method. The method we have produces a matrix that, by construction, agrees with a given function on the basis vectors $\{e_1, \ldots, e_n\}$. If the function we were given is linear, then (by Remark 1.11) we will have found the unique linear function which sends e_i to $f(e_i)$ for $i = 1, \ldots, n$. But if the given function is not linear, then we have just found a linear function that agrees with f on the set $\{e_1, \ldots, e_n\}$ but that will not agree with f for all $x \in \mathbb{R}^n$. Note, also, that A and f might agree on more than just the set $\{e_1, \ldots, e_n\}$, but there will be an $x \in \mathbb{R}^n$ where they disagree.

This example actually gives us an alternative way to see whether a given function, in algebraic form, is linear.

[3] We are assuming this is a real function, but will discuss this later.

Example 1.22. In a footnote above we came across the function $g : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (0, \sin(x))$. We will now see that it is not linear. Such an exercise can help us build intuition for what linear functions 'look like'.

- We begin by building a matrix A which agrees with g on the vectors e_1 and e_2. This is
$$A = \begin{pmatrix} 0 & 0 \\ \sin(1) & 0 \end{pmatrix}.$$
- Next find a new function, f, so that $f(x) = Ax$ for all $x \in \mathbb{R}^2$. We can compute f in algebraic form by applying A to a general vector (x, y). In our case we get that
$$f(x) = \begin{pmatrix} 0 & 0 \\ \sin(1) & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ x\sin(1) \end{pmatrix}.$$
- Note that g is linear if and only if $g = f$. In this case we note that $x\sin(1)$ and $\sin(x)$ cannot be the same function since $x\sin(1)$ is unbounded whereas $\sin(x)$ is bounded. Thus g is not linear.

We end with a comment on how the above applies to maps that are over $\mathbb{C}$. The terminology can be a bit loose when using the duality of matrices with linear maps. The issue is that the matrix appears a concrete object, with entries coming from some field (think: $\mathbb{C}$ or $\mathbb{R}$) but when considering a matrix as a linear map, we should really be giving all of the information for a function. This means stating the domain, the codomain, and the 'rule' (which is almost always that x is sent to Ax).

Example 1.23 (Example 1.19 continued)**.** In the previous example we were given the matrix but not told the domain or codomain. We had the matrix
$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$
and, with it being a real matrix, we considered this as a function $\mathbb{R}^2 \to \mathbb{R}^3$. We could have equally well considered this as a function from $\mathbb{C}^2$ to $\mathbb{C}^3$. We can approach this with the same ideas as above. Note that it sends e_i to c_i, where c_i denotes the ith column of A. Also, since it is a $\mathbb{C}$-linear map, we then know the image of any $(x, y) \in \mathbb{C}^2$. This approach matches with what we would obtain if we used the usual matrix multiplication on vectors from $\mathbb{C}^2$.

We end with examples of matrices working over different fields.

Example 1.24. Let $A = \begin{pmatrix} 1 & i \\ 0 & 1+i \end{pmatrix}$, so that we have the $\mathbb{C}$-linear map $f : \mathbb{C}^2 \to \mathbb{C}^2$, $x \mapsto Ax$. We can compute where this sends a general $(x, y) \in \mathbb{C}^2$ by using the usual matrix multiplication:
$$\begin{pmatrix} 1 & i \\ 0 & 1+i \end{pmatrix} \begin{pmatrix} a+bi \\ c+di \end{pmatrix} = \begin{pmatrix} (a+bi)+i(c+di) \\ (1+i)(c+di) \end{pmatrix} = \begin{pmatrix} (a-d)+(b+c)i \\ (c-d)+(c+d)i \end{pmatrix}.$$

Example 1.25. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, so that we have the $\mathbb{Q}$-linear map $f : \mathbb{Q}^2 \to \mathbb{Q}^2$, $x \mapsto Ax$. This matrix can also be used to define an $\mathbb{R}$-linear map and a $\mathbb{C}$-linear map. (It turns out that $\mathbb{Z}$ is not a field, and so we will **not** look at $\mathbb{Z}$-linear maps.) We'll see this matrix again soon.

2. Eigenvalues and Eigenvectors

In this section we introduce, for an $\mathbb{F}$-linear operator T, the notions of an eigenvector and eigenvalue. The 'eigen' part comes from the German prefix meaning 'same'. If we think of a vector as a direction (technically together with a length) then an eigenvector is one where applying T leaves the direction unchanged.

Definition 2.1. *Let V be a vector space over $\mathbb{F}$ and $T : V \to V$ a linear operator. Then a vector $v \in V$ with $v \neq 0$ is called an* **eigenvector** *of T if there exists a $\lambda \in \mathbb{F}$ such that*

$$T(v) = \lambda v \ .$$

The number λ is the called an **eigenvalue** *of T.*

This might look like a rather strange concept, and it is not clear if and why such vectors should exist. In this section we will learn how to compute eigenvalues and eigenvectors of matrices and explore the limits of our approach.

Example 2.2. Let $V = \mathbb{C}^2$ and $T : \mathbb{C}^2 \to \mathbb{C}^2$ be given by $T(e_1) = 2e_1$ and $T(e_2) = -3e_2$. Then e_1 and e_2 are eigenvector with eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -3$, respectively.

Example 2.3. A less obvious example is $T(e_1) = e_2$ and $T(e_2) = e_1$. Then one can check that $v_1 = e_1 + e_2$ is an eigenvector with eigenvalue $\lambda_1 = 1$ and $v_2 = e_1 - e_2$ is an eigenvector with eigenvalue $\lambda_2 = -1$.

It is helpful to have a name for the set of eigenvalues for a given linear map.

Definition 2.4. *For an $\mathbb{F}$-linear function T, we call the set of eigenvalues of T the spectrum of T, denoted* $\operatorname{spec} T$.

At present these eigenvectors appear to have little structure. Our first lemma is a nice observation about the eigenvectors relating to a specific eigenvalue.

Lemma 2.5. *Let $T : V \to V$ be a linear operator and λ be an eigenvalue for T. Then the set of eigenvectors for λ, together with the zero vector, form a subspace of V.*

Proof. If v is an eigenvector of T with eigenvalue λ, then for any non-zero $\alpha \in \mathbb{F}$, we have that αv is an eigenvector of T with eigenvalue λ since

$$T(\alpha v) = \alpha T(v) = \alpha \lambda v = \lambda(\alpha v).$$

In a similar way, if v and w are eigenvectors of T with the same eigenvalue λ then $v + w$ is either zero or an eigenvector with eigenvalue λ since

$$T(v + w) = T(v) + T(w) = \lambda v + \lambda w = \lambda(v + w).$$

Therefore the set of eigenvectors with the same eigenvalue, together with $v = 0$, form a subspace of V. □

We next look at how to find the eigenvalues for any given linear operator.

2.1. The characteristic polynomial of a matrix.

The following neat observation motivates our definitions for this section.

Lemma 2.6. *Let $T : V \to V$ be a linear operator and $\dim V < \infty$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $\det(T - \lambda I) = 0$.*

Proof. We perform a series of steps to the eigenvalue equation[4]

$$\begin{aligned} Tv &= \lambda v \\ \Rightarrow Tv - \lambda v &= 0 \\ \Rightarrow Tv - (\lambda I)v &= 0 \\ \Rightarrow (T - (\lambda I))v &= 0 \end{aligned}$$

[4]It is helpful to consider why multiplication by λ can be considered as multiplication by a matrix, and what this matrix is.

which are actually if and only if statements. This reduces our eigenvalue equation to a matrix equation of the form $Av = 0$, where $A = (T - (\lambda I))$. Now, consider that $\det(T - (\lambda I)) \neq 0$. This would imply that A^{-1} exists, and so our equation would become $v = A^{-1}0$. But $A^{-1}0 = 0$ by definition, and so $v = 0$ is the only solution in this case. On the other hand, if $\det(T - (\lambda I)) = 0$, then the matrix A has non-trivial kernel, and so there is a $v \neq 0$ such that $Av = 0$. Thus there is a $v \neq 0$ if and only if $\det(T - (\lambda I)) = 0$, and $Tv = \lambda v$ has a solution if and only if $\det(T - (\lambda I)) = 0$. □

From the above proof, we are now interested in finding all λ such that $\det(T - (\lambda I)) = 0$, and for each given λ the eigenvectors are then the non-zero elements of $\ker(T - \lambda I)$.

Definition 2.7. *Let V be vector space over $\mathbb{F}$ and $T : V \to V$ be a linear operator,*

- *if $\dim V < \infty$ then the* **characteristic polynomial**[5] *of T is defined as*

$$p_T(x) := \det(T - xI).$$

- *if $\lambda \in \mathbb{F}$ is an eigenvalue of T the corresponding* **eigenspace** *is defined as*

$$E(\lambda) := \ker(T - \lambda I).$$

By now we are likely very happy with how to apply a matrix to a vector. But note that, on an eigenspace $E(\lambda)$, the action of T is extremely simple as it is just multiplication by the eigenvalue λ. This can be stated algebraically, for $v \in E(\lambda)$, as $Tv = \lambda v$. (To express this one uses the notation $T|_{E(\lambda)} = \lambda I$, where $T|_{E(\lambda)}$ means restricting the domain of T from V to the subspace $E(\lambda)$.)

Lemma 2.6 allows us to compute the eigenvalues of a operator T first, and then we can solve the system of linear equations $(T - \lambda I)v = 0$ to find the corresponding eigenvectors. Let us look at a few simple examples, using $V = \mathbb{C}^2$.

Example 2.8. For $T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, we have that

$$p_T(\lambda) = \det\left(\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right) = \det \begin{pmatrix} 1-\lambda & 0 \\ 0 & 2-\lambda \end{pmatrix} = (1-\lambda)(2-\lambda),$$

and so we see that the condition $p_T(\lambda) = 0$ gives $\lambda_1 = 1$ and $\lambda_2 = 2$ as eigenvalues of T. To find an eigenvector $v_1 = (x, y)$ with eigenvalue $\lambda_1 = 1$ we have to find a solution to $(T - \lambda_1 I)v = (T - I)v = 0$ and this gives

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

This gives the condition $y = 0$, hence any vector $v_1 = (x, 0)$ with $x \neq 0$ is an eigenvector, so we can choose for instance $x = 1$. Similarly for $\lambda_2 = 2$ we want to find $v_2 = (x, y)$ with $(T - 2I)v_2 = 0$ which gives

$$\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Thus $x = 0$, and so any vector $v_2 = (0, y)$ with any $y \neq 0$ is an eigenvector and to pick one we can choose for instance $y = 1$. So we found that T has two eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$ with corresponding eigenvectors $v_1 = (1, 0)$ and $v_2 = (0, 1)$. The eigenvalues are uniquely determined, but the eigenvectors are only determined up to a multiplicative constant, the corresponding eigenspaces are $E(1) = \{(x, 0)\,,\, x \in \mathbb{F}\}$ and $E(2) = \{(0, y)\,,\, y \in \mathbb{F}\}$.

We now come back to the example we saw at the end of Section 2.

[5]The Cayley-Hamilton Theorem is an important result relating to the characteristic polynomial and will be covered in Linear Algebra 2 and also ODEs 2.

Example 2.9. For $T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, we find[6]

$$p_T(\lambda) = \det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1.$$

Therefore the characteristic polynomial has the two roots $\lambda_1 = \mathrm{i}$ and $\lambda_2 = -\mathrm{i}$. So if $\mathbb{F} = \mathbb{R}$, then this operator has no eigenvalues in $\mathbb{F}$, but if $\mathbb{F}$ contains i, for instance if $\mathbb{F} = \mathbb{C}$, then we have two eigenvalues. To find an eigenvector $v_1 = (x, y)$ with eigenvalue $\lambda_1 = \mathrm{i}$ we have to solve $(T - \mathrm{i})v = 0$ which is

$$\begin{pmatrix} -\mathrm{i} & -1 \\ 1 & -\mathrm{i} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Thus $-\mathrm{i}x - y = 0$ and $x - \mathrm{i}y = 0$. But the second equation is just $-\mathrm{i}$ times the first equation, so what we find is that $y = -\mathrm{i}x$, so any $(x, -\mathrm{i}x)$ is an eigenvector, and we can choose for instance $x = 1$ to obtain $v_1 = (1, -\mathrm{i})$. Similarly we get for $\lambda_2 = -\mathrm{i}$ that

$$\begin{pmatrix} \mathrm{i} & -1 \\ 1 & \mathrm{i} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

has the solutions $(x, \mathrm{i}x)$, and so choosing $x = 1$ gives $v_2 = (1, \mathrm{i})$.

Remark 2.10. Even when the matrix elements are real, the eigenvalues need not be real. This means that a operator can have no eigenvalues when we look at it as a function over $\mathbb{R}$, but it will have eigenvalues over $\mathbb{C}$. This is why we often work over $\mathbb{C}$ when dealing with eigenvalues.

Example 2.11. If $T = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$, then $p_T(\lambda) = \lambda^2 - 1$ and so there are two eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$. The eigenvectors corresponding to $\lambda_1 = 1$ are determined by

$$\begin{pmatrix} -1 & -\mathrm{i} \\ \mathrm{i} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

which gives $-x - \mathrm{i}y = 0$ and $\mathrm{i}x - y = 0$ and so $y = \mathrm{i}x$. Choosing $x = 1$ gives us $v_1 = (1, \mathrm{i})$, and similarly we find for $\lambda_2 = -1$ that $v_2 = (\mathrm{i}, 1)$ is an eigenvector.

Remark 2.12. A matrix with complex entries can still have all of its eigenvalues as real numbers, but then the eigenvectors must be complex.

Example 2.13. If $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then

$$p_T(\lambda) = \det \begin{pmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{pmatrix} = (\lambda - 1)^2$$

and so we have one eigenvalue $\lambda_1 = 1$. The corresponding eigenvectors are determined by

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

which gives the one condition $y = 0$. Hence any vector $(x, 0)$ (with $x \neq 0$) is an eigenvector and we can choose for instance $v_1 = (1, 0)$. In this example, contrary to the previous ones, we found only one eigenvalue and a one-dimensional eigenspace.

Example 2.14. With $T = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, we get $p_T(\lambda) = (2 - \lambda)^2$, so $\lambda_1 = 2$ is the only eigenvalue. But now we have two linearly independent eigenvectors $v_1 = e_1$ and $v_2 = e_2$, since $T - 2I = 0$.

Remark 2.15. In all the cases where we had two eigenvalues, the eigenvectors actually formed a basis. In these last two examples, we only found one eigenvalue and then in the first case we found only a one-dimensional eigenspace, so there is no basis of eigenvectors, whereas in the second case we found two linearly independent eigenvectors which then formed a basis of $\mathbb{C}^2$.

[6] We can also approach this without as much theory. Note that this function sends (a, b) to $(-b, a)$. Thus an eigenvalue λ would satisfy $(-b, a) = \lambda(a, b)$, i.e. the equations $-b = \lambda a$ and $a = \lambda b$. Putting these together give us that $a = -\lambda^2 a$, which for $\lambda \in \mathbb{R}$ gives us only the solution $a = b = 0$.

2.2. **Algebraic and geometric multiplicity.** In order to gain a more systematic understanding of eigenvalues and eigenvectors, we need to know more about the roots of polynomials. The following list of properties of polynomials will be proved in courses on complex analysis and algebra: we only quote them here.

Definition 2.16. *A polynomial of degree n over $\mathbb{C}$ is an expression of the form*

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

with $a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{C}$ and $a_n \neq 0$. Often we think of such expressions as functions, since for each x (in $\mathbb{R}$ or in $\mathbb{C}$) we can evaluate $p(x)$.

Definition 2.17. *For a polynomial $p(x)$ we say that $\lambda_1 \in \mathbb{C}$ is a* **root** *of p if $p(\lambda_1) = 0$. We can also talk about the* **multiplicity** *of a root. Specifically, if there exists a polynomial $q(x)$ of degree $n - m_1$ with $q(\lambda_1) \neq 0$ where*

$$p(x) = (x - \lambda_1)^{m_1} q(x),$$

then λ_1 is a root of multiplicity m_1 Note that the multiplicity of any root is a natural number.

Theorem 2.18. *Every polynomial of degree n has exactly n roots in $\mathbb{C}$, counted with multiplicity. In other words, for every polynomial of degree n there exist $\alpha, \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{C}$ and $m_1, m_2, \ldots, m_k \in \mathbb{N}$ with*

$$p(x) = \alpha(x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \ldots (x - \lambda_k)^{m_k}$$

where $m_1 + m_2 + \ldots + m_k = n$.

Remark 2.19. These results rely on working over $\mathbb{C}$ rather than $\mathbb{R}$, and follow from the crucial fact that every polynomial has at least one root in $\mathbb{C}$ (called the Fundamental Theorem of Algebra which was proved by Gauss in his PhD thesis, published in 1799).

The above results allow us to draw some conclusions about eigenvalues.

Lemma 2.20. *Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear operator. Then T has at least one and at most n different eigenvalues.*

Proof. One can check that the characteristic polynomial of T is exactly of order $n = \dim V$. (We do this later, in the proof of Lemma 2.33.) Thus it has at most n different roots. The Fundamental Theorem of Algebra then states that this polynomial has at least one root in $\mathbb{C}$. □

Definition 2.21. *Let T be a linear operator. For each $\lambda \in \operatorname{spec} T$ we say that*

- *the* **geometric multiplicity** *of λ, denoted $m_g(\lambda)$, is $\dim E(\lambda)$; and*
- *the* **algebraic multiplicity** *of λ, denoted $m_a(\lambda)$, is the multiplicity of the root λ in $p_T(x)$.*

Note, for any $\lambda \in \operatorname{spec} T$, that $m_g(\lambda), m_a(\lambda) \in \mathbb{N}$ (do consider why these are non-zero).

The algebraic multiplicity can be found by determining all the roots of the characteristic polynomial. The geometric multiplicity can be equivalently given as

$$m_g(\lambda) = \text{nullity}(A - \lambda I).$$

A natural question is to query how these two multiplicities relate to one another.

Theorem 2.22. *Let T be a linear operator and $\lambda \in \operatorname{spec} T$. Then $m_g(\lambda) \leq m_a(\lambda)$.*

We will not prove this, but note that in our examples above, in all but one of the cases we had $m_g(\lambda) = m_a(\lambda)$ for every $\lambda \in \operatorname{spec} T$. The exception was Example 2.13 which sent e_1 to e_1 and e_2 to $e_1 + e_2$. We then had that $\lambda = 1$ was the only eigenvalue and $m_g(1) = 1$ but $m_a(1) = 2$.

Before we dive into the next theorem, on the linear independence of eigenvectors, let us consider a simple case. Take a linear operator T with two distinct eigenvalues λ_1 and λ_2 with corresponding eigenvectors v_1 and v_2. We will show that v_1, v_2 are linearly independent. If v_1, v_2 are linearly dependent, they are proportional to each other, so they both lie in the same one-dimensional subspace. That means that if two eigenvectors are linearly dependent, then the intersection of the corresponding subspaces is at least one-dimensional, $E(\lambda_1) \cap E(\lambda_2) \neq \{0\}$. But if $v \neq 0$ is in $E(\lambda_1) \cap E(\lambda_2)$, then $\lambda_1 v = T(v) = \lambda_2 v$ and this can only happen if $\lambda_1 = \lambda_2$. We now see how to generalise this argument.

Proposition 2.23. *Let $T : V \to V$ be a linear operator, $\dim V = n$ and $\{v_1, v_2, \ldots, v_k\}$ a set of eigenvectors with different eigenvalues. Then the set $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.*

Proof. We first fix some notation. From our hypotheses, we have that $T(v_i) = \lambda_i v_i$ for $i = 1, \ldots, k$ and that $\lambda_i = \lambda_j \Leftrightarrow i = j$. Let us assume that we can find constants $\alpha_1, \ldots, \alpha_k$ such that

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = 0$$

where $\alpha_i \neq 0$ for some $i \in \{1, \ldots, k\}$. Thus our assumption is that $\{v_1, \ldots, v_k\}$ is linearly dependent. But if the set is linearly dependent, then there is a smallest $m \in \mathbb{N}$ where we can find $\{w_1, \ldots, w_m\} \subseteq \{v_1, \ldots, v_k\}$ with $\{w_1, \ldots, w_m\}$ also linearly dependent. Thus we have

$$a_1 w_1 + \ldots + a_m w_m = 0 \tag{2.1}$$

where $a_1, \ldots, a_m \in \mathbb{F} \setminus \{0\}$. (If any a_i were zero, then we could remove w_i from our set.) We now show that $\{w_2, \ldots, w_m\}$ is also linearly dependent, providing us with a contradiction. Directly computing, we apply T to equation (2.1) and multiply equation (2.1) by λ_1 to obtain

$$\lambda_1 a_1 v_1 + \ldots + \lambda_m a_m v_m = 0 \text{ and } \lambda_1 a_1 v_1 + \ldots + \lambda_1 a_m v_m = 0$$

respectively. The difference of these equations is therefore

$$(\lambda_2 - \lambda_1) a_2 v_2 + \ldots + (\lambda_m - \lambda_1) a_m v_m = 0. \tag{2.2}$$

By assumption, $\lambda_i - \lambda_1 \neq 0$ for $i = 2, \ldots, m$. Hence equation (2.2) is a linear combination with coefficients $(\lambda_i - \lambda_1)a_i$ for $i \in \{2, \ldots, m\}$. Moreover, $a_i \neq 0 \Rightarrow (\lambda_i - \lambda_1)a_i \neq 0$ for $i \in \{2, \ldots, m\}$. This means we have reduced to have a linear combination

$$b_2 w_2 + \ldots + b_m w_m = 0$$

where $b_2, \ldots, b_m \neq 0$. Hence $\{w_2, \ldots, w_m\}$ is also a linearly dependent set, contradicting the minimality of m. □

A more geometric formulation of our goal to find a basis of eigenvectors is to try to decompose the vector space into eigenspaces, and on each eigenspace the operator T is then just multiplication by an eigenvalue. Thus, in what follows, the function $T|_{E(\lambda_i)}$ is the restriction of the map T to the subspace $E(\lambda_i)$, so that $T|_{E(\lambda_i)} = \lambda_i I$. This means that $T|_{E(\lambda_i)}$ is a new function with domain $E(\lambda_i)$ which behaves like the map $\lambda_i I$ on $E(\lambda_i)$. This is just an alternative way of saying $E(\lambda_i)$ is the eigenspace corresponding to λ_i.

Proposition 2.24. *A linear operator $T : V \to V$ has a basis of eigenvectors if and only if V can be decomposed into a direct sum of eigenspaces*

$$V = E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_k) \ ,$$

where $T|_{E(\lambda_i)} = \lambda_i I$ for $i = 1, \ldots, k$.

Proof. If we have such a decomposition, then we can choose a basis $\mathcal{B}_i$ of each eigenspace and the union of these bases $\mathcal{B} = \bigcup \mathcal{B}_i$ will be a basis of V which consists of eigenvectors. (This requires some checking: take any $v \in V$, write it as $\sum w_i$ where $w_i \in E(\lambda_i)$, and then write each w_i as a linear combination in $\mathcal{B}_i$.)

On the other hand, if we have a basis of eigenvectors then group these by eigenvalue:

$$v_1^{(i)}, \ldots, v_{d_i}^{(i)} \in E(\lambda_i) \text{ for each eigenvalue } \lambda_i.$$

Fix an i, and let $d := d_i$ and $v_j := v_j^{(i)}$. Our claim is that $U_i := \text{span}\{v_1, \ldots, v_d\}$ equals $E(\lambda_i)$. Assume not, so that there exists $w \in E(\lambda_i) \setminus U_i$. But then $w \in V$ and so can be written as a linear combination in our basis. Let w' be the element in U_i that has the same coefficients as w in front of $v_1, \ldots, v_d$. We note that $w - w' = w_i \in E(\lambda_i)$, since it is a subspace, but also $w - w'$ is expressed in our basis as $\sum_{j \neq i} w_j$ where $w_j \in E(\lambda_j)$ for $j \neq i$. Take $J := \{j \neq i : w_j \neq 0\}$. Our assumption states that $w - w' \neq 0$, and so $J \neq \emptyset$. Hence

$$(w - w') - (w - w') = 0 \Rightarrow w_i - \sum_{j \in J} w_j = 0$$

which contradicts Proposition 2.23. □

This gives us one important criterium to decide when a operator has a basis of eigenvectors.

Lemma 2.25. *Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear operator with n different eigenvalues. Then T has a basis of eigenvectors.*

Proof. If T has n different eigenvalues, then by Proposition 2.23 the corresponding eigenvectors are linearly independent. Then these n linearly independent vectors must form a basis for $\mathbb{C}^n$. □

From this lemma, we see that the possible obstruction to the existence of enough linearly independent eigenvectors is that the characteristic polynomial can have roots of multiplicity larger then 1. Then the condition for the existence of a basis of eigenvectors becomes

$$m_a(\lambda) = m_g(\lambda) \ , \quad \text{for all} \quad \lambda \in \operatorname{spec} T.$$

Unfortunately in general this condition can only be checked after one has computed all the eigenvectors. Fortunately, if there are not enough eigenvectors for a basis, there is a natural generalisation of the concept of an eigenvector, called a *root vector*. These are dealt with in Linear Algebra 2. There are precisely as many linearly independent root vectors as on needs to make up for the difference between the algebraic and geometric multiplicity of an eigenvalue.

2.3. **Using eigenvectors to aid computations: diagonalisation.** If we have found a basis of eigenvectors $\{v_1, v_2, \ldots, v_n\}$ for $A \in M_n(\mathbb{C})$, then we can **diagonalise** the matrix A. This means that we can find an invertible matrix C such that $C^{-1}AC = D$, where D is a diagonal matrix, i.e., only has nonzero entries on the diagonal. To do so, let

$$C := (v_1 \cdots v_n)$$

which has the eigenvectors as columns and (since these form a basis) is invertible. We wish to determine the form of the matrix $C^{-1}AC$, and can do so by applying the ideas of Section 1. We note, for each $i = 1, \ldots, n$, that

$$e_i \xrightarrow{C} v_i \xrightarrow{A} \lambda_i v_i \xrightarrow{C^{-1}} \lambda_i e_i.$$

From the FACT, we see that the ith column of $C^{-1}AC$ must therefore be $\lambda_i e_i$. Hence the matrix $C^{-1}AC$ is diagonal.

Remark 2.26. One can reverse the above argument to show that if A is diagonalisable, then the column vectors of the matrix C must be eigenvectors and the elements of the diagonal matrix are the eigenvalues. Since the eigenvalues are uniquely determined, the diagonal matrix is unique up to reordering of the elements on the diagonal. But the matrix C is not unique, since one can for instance multiply any column by an arbitrary non-zero number, and still get an eigenvector.

Let us now summarise the method of how to compute eigenvalues and eigenvectors for operators on finite dimensional vector spaces. We will almost always work over $\mathbb{R}$ or $\mathbb{C}$.

(i) We begin by computing the characteristic polynomial $p_T(x) = \det(T - xI)$.
(ii) Then we have to find all roots of $p_T(x)$ with multiplicity. We know that there are n of them in $\mathbb{C}$. There are now 3 possibilities.
   – We have n distinct roots and they all lie in the field $\mathbb{F} \subset \mathbb{C}$. Then we immediately know that we can find a basis of eigenvectors.
   – There are less than n roots, counted with multiplicity, in the field $\mathbb{F}$. Then we cannot find a basis of eigenvectors, and T is not diagonalisable.
   – All roots are in $\mathbb{F}$ (which is always the case if $\mathbb{F} = \mathbb{C}$), but some have higher multiplicity than 1. Then we cannot decide yet whether there is a basis of eigenvectors.
(iii) To find the eigenvectors we have to solve for each eigenvalue λ the system of n linear equations

$$(T - \lambda I)v = 0.$$

We can do this, for example, by using Gaussian elimination. In order to find a basis of eigenvectors, we must find $m_a(\lambda)$ linearly independent solutions for each $\lambda \in \operatorname{spec} T$.

We look at two examples of 3×3 matrices to see how this works.

Example 2.27. Our first example is given by the following matrix A, which we consider as a function from $\mathbb{R}^3$ to $\mathbb{R}^3$.

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$$

We begin by computing the characteristic polynomial and its roots.

$$\begin{aligned} p_A(x) = \det(A - xI) &= \det \begin{pmatrix} 4-x & 1 & -1 \\ 2 & 5-x & -2 \\ 1 & 1 & 2-x \end{pmatrix} \\ &= (4-x)\det \begin{pmatrix} 5-x & -2 \\ 1 & 2-x \end{pmatrix} - \det \begin{pmatrix} 2 & -2 \\ 1 & 2-x \end{pmatrix} - \det \begin{pmatrix} 2 & 5-x \\ 1 & 1 \end{pmatrix} \\ &= (4-x)[(5-x)(2-x)+2] - 2(2-x) - 2 - 2 + (5-x) \\ &= (4-x)(5-x)(2-x) + 2(4-x) - 8 + 2x + (5-x) \\ &= (5-x)[(4-x)(2-x)+1] \\ &= (5-x)[x^2 - 6x + 9] = (5-x)(x-3)^2 \end{aligned}$$

These computations may appear confusing at first, but two parts of our approach are common. First, that we found the determinant by expanding along the first row. Secondly, that we didn't multiply out all terms immediately, but instead kept relevant factors which appear in the final factorisation. This avoids the need to factorise a cubic. We then see that the eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 3$, and that 5 has algebraic multiplicity 1 and 3 has algebraic multiplicity 2. So we can't yet say if the matrix is diagonalisable; we have to see if there are two linearly independent eigenvectors with eigenvalue 3.

We now look at finding an eigenvector $v_1 = (x, y, z)$ with eigenvalue $\lambda_1 = 5$. Thus v_1 is a solution to the system of 3 linear equations $(A - 5I)v_1 = 0$, and

$$A - 5I = \begin{pmatrix} -1 & 1 & -1 \\ 2 & 0 & -2 \\ 1 & 1 & -3 \end{pmatrix} \equiv \begin{pmatrix} -1 & 1 & -1 \\ 0 & 2 & -4 \\ 0 & 2 & -4 \end{pmatrix} \equiv \begin{pmatrix} -1 & 1 & -1 \\ 0 & 2 & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

where the '$\equiv$' sign means that we have simplified the matrix using elementary row operations. In the first step we added the first row to the third and added 2 times the first row to the second. In the second step we just subtracted the second row from the third. So the system of equations is now $-x + y - z = 0$ and $2y - 4z = 0$, which can be rewritten as

$$y = 2z \text{ and } x = y - z = z.$$

This gives us a one parameter family of solutions, which is what we expect, since eigenvectors are only defined up to a multiplicative factor. To pick one particularly simple eigenvector we can choose for instance $z = 1$, and so obtain

$$E(5) = \text{span}\{v_1\} \text{ where } v_1 = (1, 2, 1).$$

To find the eigenvectors for $\lambda_2 = 3$ we proceed along the same lines, with our aim to solve $(A - 3I)v = 0$. This gives

$$A - 3I = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 2 & -2 \\ 1 & 1 & -1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where we have subtracted row one from row three and two times row one from row two. So this gives just the one equation

$$x = z - y,$$

which means that we have two free parameters in the solution, and any vector of the form

$$v = (z - y, y, z)$$

for arbitrary $(y, z) \neq 0$ is an eigenvector. Therefore they form a two dimensional space, and we just have to pick two that form a basis. One option would be to choose $y = 1, z = 0$ and then $y = 0, z = 1$, so that

$$v_2 = (-1, 1, 0) \text{ and } v_3 = (1, 0, 1)$$

form a basis of the eigenspace $E(3)$.

We have found three linearly independent eigenvectors, and therefore A is diagonalisable with

$$C = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \text{ and } C^{-1}AC = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Notice that C depends on the choices we made for the eigenvectors. If we had chosen different eigenvectors, the matrix C would look different but would still diagonalise A. A good final computation to try is to compute $C^{-1}AC$ with

$$C = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}.$$

Example 2.28. For our second example we will consider

$$B = \begin{pmatrix} 3 & -1 & 1 \\ 7 & -5 & 1 \\ 6 & -6 & 2 \end{pmatrix}.$$

We first compute $p_B(x)$.

$$\begin{aligned} \det(B - xI) &= (3 - x) \det \begin{pmatrix} -5 - x & 1 \\ -6 & 2 - x \end{pmatrix} + \det \begin{pmatrix} 7 & 1 \\ 6 & 2 - x \end{pmatrix} + \det \begin{pmatrix} 7 & -5 - x \\ 6 & -6 \end{pmatrix} \\ &= (3 - x)[-(5 + x)(2 - x) + 6] + 7(2 - x) - 6 - 42 + 6(5 + x) \\ &= -(3 - x)(5 + x)(2 - x) + 7(2 - x) \\ &= (2 - x)[7 - (3 - x)(5 + x)] \\ &= (2 - x)[x^2 + 2x - 8] \\ &= -(x - 2)(x - 2)(x + 4) = -(x - 2)^2(x + 4) \end{aligned}$$

Hence the eigenvalues are $\lambda_1 = -4$ with multiplicity 1 and $\lambda_2 = 2$ with algebraic multiplicity 2. We then find the eigenvectors for each eigenvalue in turn. For $\lambda_1 = -4$, we solve $(B + 4I)v = 0$. Note

$$B + 4I = \begin{pmatrix} 7 & -1 & 1 \\ 7 & -1 & 1 \\ 6 & -6 & 6 \end{pmatrix} \equiv \begin{pmatrix} 7 & -1 & 1 \\ 0 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \equiv \begin{pmatrix} 0 & 6 & -6 \\ 0 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

which gives the two equations $y = z$ and $x = y - z = 0$, so any vector $(0, z, z)$ with $z \neq 0$ is an eigenvector. Choosing $z = 1$ gives us $v_1 = (0, 1, 1)$. Now for $\lambda_2 = 2$, we get

$$B - 2I = \begin{pmatrix} 1 & -1 & 1 \\ 7 & -7 & 1 \\ 6 & -6 & 0 \end{pmatrix} \equiv \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & -6 \\ 0 & 0 & -6 \end{pmatrix} \equiv \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

which gives the equations $y = 0$ and $x + z = 0$. These give us only a one parameter family, i.e., y is fixed, and once we have chosen z, the value of x is fixed, too. So the eigenspace $E(2)$ is one-dimensional and spanned by

$$v_2 = (1, 0, -1).$$

Hence the geometric multiplicity of $\lambda_2 = 2$ is 1. This means B does not have a basis of eigenvectors, and cannot be diagonalised.

The second matrix B gave us an example which cannot be diagonalised. The drawback of our approach is that only at the very end of our computation we actually found out that the matrix is not diagonalisable. It would be much more efficient if we had some criteria to tell us in advance if

a matrix is diagonalisable. Such criteria can be given if we introduce additional structure, namely an inner product. This will be the subject of Section 5.

2.4. **Properties of similar matrices.**

Definition 2.29. *We say that matrices A and B are* **similar** *if there exists an invertible matrix C such that $B = C^{-1}AC$. We can then also say that A and B are* **conjugate** *(by C).*

One can check that the notion of being similar is an equivalence relation for square matrices of a fixed dimension. Many properties are preserved across similar matrices.

Lemma 2.30. *Let $A, B \in M_n(\mathbb{F})$ be similar matrices. Then $p_A(x) = p_B(x)$.*

Proof. Our definition of similar says that $B = C^{-1}AC$ for some invertible $C \in M_n(\mathbb{F})$. We have

$$\begin{aligned} p_B(x) = \det(B - xI) = \det(C^{-1}AC - xI) = \det(C^{-1}(A - xI)C) \\ = \det C^{-1} \det(A - xI) \det C = \det(A - xI) = p_A(x). \end{aligned}$$

□

An almost identical proof tells us that similar matrices have the same determinant. We now consider more carefully the coefficients of the characteristic polynomial to gain further insights.

Lemma 2.31. *Let $A \in M_n(\mathbb{F})$. Then the constant coefficient of $p_A(x)$ equals $\det(A)$. Furthermore, if A is similar to B, then $\det(A) = \det(B)$.*

Proof. We note from the explicit form of the matrix $A - xI$ that

$$p_A(x) = (-1)^n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \text{ where } a_{n-1}, \ldots, a_0 \in \mathbb{F}.$$

Moreover, from the above expression, $a_0 = p_A(0) = \det(A - 0I) = \det A$ and so the term a_0 in the characteristic polynomial equals $\det A$. The claim for similar matrices then immediately follows from Lemma 2.30. □

Definition 2.32. *Let $A \in M_n(\mathbb{F})$. Then the* **trace** *of A, denoted $\operatorname{tr} A$, is equal to $\sum_{i=1}^n a_{ii}$.*

It is possible to directly show that the similar matrices have the same trace,[7] but we will do this using the coefficients of the characteristic polynomial.

Lemma 2.33. *Let $A \in M_n(\mathbb{F})$. Then the x^{n-1} coefficient of $p_A(x)$ equals $(-1)^{n-1}\operatorname{tr}(A)$. Furthermore, if A is similar to B, then $\operatorname{tr}(A) = \operatorname{tr}(B)$.*

Proof. We call upon the Leibniz formula for the determinant. Let $B := A - xI$, so that

$$\begin{aligned} \det(A - xI) &= \sum_{\sigma \in S_n} \operatorname{sign}\sigma \prod_{j=1}^n b_{\sigma(j)j} \\ &= \prod_{i=1}^n (a_{ii} - x) + \sum_{\sigma' \in S_n \setminus \{1\}} \operatorname{sign}\sigma' \prod_{j=1}^n b_{\sigma'(j)j} \end{aligned}$$

where the second line corresponds to taking out the identity permutation. We will now see that the sum in the second line does not contribute to the coefficients of the x^n or x^{n-1} terms.

We observe that any permutation $\sigma \in S_n$ is a bijection, and so knowing the values of σ on $n-1$ values of $j \in \{1, 2, \ldots, n\}$ uniquely determines the remaining one. Hence, a nonidentity permutation has at least two indices j which it alters, that is $\sigma'(j) \neq j$. It follows that no product in the sum in the latter formula can involve more than $n-2$ diagonal elements of $A - xI$, and hence the whole sum cannot contribute higher powers of x to $p_A(x)$ than x^{n-2}. Now

$$\prod_{i=1}^n (a_{ii} - x) = (-1)^n x^n + (-1)^{n-1}(a_{11} + \ldots + a_{nn})x^{n-1} + \ldots,$$

from which our statement follows. Lemma 2.30 then implies our second statement. □

[7] A nice approach is to show that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ by directly computing the relevant entries of the matrices AB and BA, and then use this result to show that $\operatorname{tr}(C^{-1}AC) = \operatorname{tr} A$. Hint: make a clever choice for B.

Let us summarise the above.

Proposition 2.34. *Let $T : V \to V$ be a linear operator and $\dim V = n$. Then*

$$p_T(x) = (-1)^n x^n + (-1)^{n-1} \operatorname{tr} T\, x^{n-1} + \ldots + \det T,$$

where $\ldots$ stands for terms with powers of x between $n-2$ and 1, and $\operatorname{tr} T$ denotes the trace of T.

In particular, for $n = 2$ Proposition 2.34 describes every term in the characteristic polynomial $p_T(x)$, without omissions.

Suppose we are working with $\mathbb{R}^n$ or $\mathbb{C}^n$. Then the Fundamental Theorem of Algebra guarantees the existence of the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic polynomial $p_T(x)$, not necessarily distinct. Then we can factor $p_T(x)$ as

$$p_T(x) = (\lambda_1 - x)(\lambda_2 - x) \ldots (\lambda_n - x).$$

Comparing this with the claim of Proposition 2.34 we draw the following conclusion.

Lemma 2.35. *Let T be a linear operator on $\mathbb{C}^n$ or $\mathbb{R}^n$, with (not necessarily distinct) eigenvalues $\lambda_1, \ldots, \lambda_n$. Then*

$$\det T = \prod_{i=1}^{n} \lambda_i, \qquad \operatorname{tr} T = \sum_{i=1}^{n} \lambda_i.$$

We have therefore expressed the determinant and trace of a linear operator in terms of its eigenvalues. We revisit an earlier example to see illustrate this result.

Example 2.36 (Example 2.9 continued)**.** We had the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and found that the eigenvalues were i and $-i$. We also found the eigenvectors $(1, -i)$ and $(1, i)$. Note that $\operatorname{tr}(A) = i + (-i) = 0$. We could also now diagonalise A and obtain

$$D = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

where $\operatorname{tr}(D) = 0$ and $\det(D) = 1 = \det(A)$ as expected.

3. Linear combinations and change of basis

We continue working with $\mathbb{R}^n$ and $\mathbb{C}^n$, but note that the theory in this chapter applies to a vector space over any field $\mathbb{F}$. To begin, we consider the idea of writing one basis in terms of another.

3.1. **Linear combinations.** Let us start in $\mathbb{R}^2$. This is an object where we naturally view vectors (a, b) as the sum of an x-direction and a y-direction. Another way to say this is that $(a, b) = ae_1 + be_2$. But, of course, we could also work with another basis. Given the basis $S = \{(2, 0), (0, 1)\}$, we note that $0.5a(2, 0)+b(0, 1)$ is the linear combination that gives the element (a, b). We now see some other examples of doing this for bases of $\mathbb{R}^2$.

Example 3.1. We write a general vector $v = (a, b)$ for the following bases S.

i) $S = \{(1, 1), (0, 1)\}$.
ii) $S = \{(2, 1), (0, 1)\}$.
iii) $S = \{(1, 0), (0, 0.5)\}$.
iv) $S = \{(2, 0), (1, 1)\}$.

It is recommended that you attempt these on your own. The solutions are then below[8].

The actual 'skill' here is solving simultaneous linear equations. But it is worth us dwelling on this a little longer, to build some intuition. In the following, it is best to think of our examples $\mathbb{R}^n$ and $\mathbb{C}^n$.

Definition 3.2. *Let $V = \mathbb{F}^n$. Then the* **standard basis** *for V is $\{e_1, e_2, \ldots, e_n\}$, denoted $\mathcal{E}$.*

Thus our above example is really writing an element of the standard basis in terms of some new basis. This inspires a new approach.

Example 3.3. Let $V = \mathbb{R}^2$, $s_1 = (1, 1)$, $s_2 = (1, -1)$, and $S = \{s_1, s_2\}$. We will write a general element (a, b) in terms of the basis S.

- We begin by writing the element e_1 in terms of S. We could do this by solving linear equations, or note that $(1, 1) + (1, -1) = 2e_1$. This gives us that $e_1 = \frac{1}{2}(1, 1) + \frac{1}{2}(1, -1)$.
- Note that $e_2 = (1, 1) - (1, 0)$, and so we can use our solution in the above line to obtain $e_2 = (1, 1) - (\frac{1}{2}(1, 1) + \frac{1}{2}(1, -1)) = \frac{1}{2}(1, 1) - \frac{1}{2}(1, -1)$.
- Thus $ae_1 + be_2 = a(\frac{1}{2}(1, 1) + \frac{1}{2}(1, -1)) + b(\frac{1}{2}(1, 1) - \frac{1}{2}(1, -1)) = (\frac{a}{2} + \frac{b}{2})s_1 + (\frac{a}{2} - \frac{b}{2})s_2$.

We could also work in the other direction, writing the elements of a basis S in terms of $\mathcal{E}$.

Example 3.4. Let $V = \mathbb{R}^2$, $s_1 = (1, 1)$, $s_2 = (1, -1)$, and $S = \{s_1, s_2\}$. We will write a general element $as_1 + bs_2$ in terms of the standard basis $\{e_1, e_2\}$.

- We have that $s_1 = (1, 1) = e_1 + e_2$ and $s_2 = (1, -1) = e_1 - e_2$.
- Thus $as_1 + bs_2 = a(e_1 + e_2) + b(e_1 - e_2) = (a + b)e_1 + (a - b)e_2$.

Note, from LA1a, that having a basis means each vector can be expressed uniquely as a linear combination in the basis. This means our solutions are the *only* solutions in each case. Our final aim is to simplify the mechanical process in the above examples. The following definition is not standard, but helps us see what is going on above.

Definition 3.5. *Let $S = \{s_1, \ldots, s_n\}$ be an ordered[9] basis for V. Then denote the linear combination $\sum_{i=1}^{n} a_i s_i$ by the vector $(a_1, a_2, \ldots, a_n)_S$.*

We now use this definition with respect to our previous examples.

---

[8] We solve each in turn.
$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (b - a) \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2}a \begin{pmatrix} 2 \\ 1 \end{pmatrix} + (b - \frac{1}{2}a) \begin{pmatrix} 0 \\ 1 \end{pmatrix}$,
$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2b \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}$, and $\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2}(a - b) \begin{pmatrix} 2 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

[9] This is very important from now on: the bases $\{v_1, v_2\}$ and $\{v_2, v_1\}$ are distinct for our purposes.

Example 3.6. Let $V = \mathbb{R}^2$, $\mathcal{A} = \{(2,1),(0,1)\}$, $\mathcal{B} = \{(1,1),(1,-1)\}$, and $\mathcal{E} = \{e_1, e_2\}$. We can therefore summarise our work above, using our new notation:

$$\begin{pmatrix} a \\ b \end{pmatrix}_{\mathcal{E}} = \frac{1}{2}a \begin{pmatrix} 2 \\ 1 \end{pmatrix}_{\mathcal{E}} + [b - \frac{1}{2}a] \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} \frac{1}{2}a \\ b - \frac{1}{2}a \end{pmatrix}_{\mathcal{A}} \text{ and } \begin{pmatrix} a \\ b \end{pmatrix}_{\mathcal{A}} = a \begin{pmatrix} 2 \\ 1 \end{pmatrix}_{\mathcal{E}} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} 2a \\ a+b \end{pmatrix}_{\mathcal{E}};$$

$$\begin{pmatrix} a \\ b \end{pmatrix}_{\mathcal{E}} = [\frac{a}{2} + \frac{b}{2}] \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{E}} + [\frac{a}{2} - \frac{b}{2}] \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} \frac{a}{2} + \frac{b}{2} \\ \frac{a}{2} - \frac{b}{2} \end{pmatrix}_{\mathcal{B}} \text{ and } \begin{pmatrix} a \\ b \end{pmatrix}_{\mathcal{B}} = a \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{E}} + b \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} a+b \\ a-b \end{pmatrix}_{\mathcal{E}}.$$

Let us review Example 3.4. We first worked out that s_1 was a linear combination of e_1 and e_2. Note that once we know this linear combination, we also know as_1 in terms of $\mathcal{E}$. Similarly, once we know s_2 in terms of $\mathcal{E}$, then we know bs_2 in terms of $\mathcal{E}$. And if we know as_1 and bs_2 in terms of $\mathcal{E}$, then we know $as_1 + bs_2$ in terms of $\mathcal{E}$. This should remind us of something.

Example 3.7 (Example 3.4 revisited)**.** In our new notation, we showed that

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{E}} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{B}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{\mathcal{E}}$$

from which we could obtain the form of a general vector $(a,b)_{\mathcal{B}}$. The key observation is then that changing between linear combinations is a linear map. That is, we can define the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

which sends $(1,0) \to (1,1)$, $(0,1) \to (1,-1)$ and, because matrix multiplication is linear, sends (a,b) to $(a+b, a-b)$. That is, multiplication by this matrix changes linear combinations in $\mathcal{B}$ to linear combinations in $\mathcal{E}$. We could also compute that

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}_{\mathcal{B}} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}_{\mathcal{B}} \text{ to obtain the matrix } \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$$

which changes linear combinations in $\mathcal{E}$ to linear combinations in $\mathcal{B}$.

We now phrase our findings as a definition.

Definition 3.8. *Let $\mathcal{A}$ and $\mathcal{B}$ be bases of a vector space V. Then $C_{\mathcal{BA}}$ denotes the* **change of basis matrix** *taking a linear combination in $\mathcal{A}$ to the equivalent linear combination in $\mathcal{B}$.*

Before covering some theory, we give one more example.

Example 3.9 (Example 3.6 continued)**.** We find $C_{\mathcal{BA}}$ directly. We note

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{A}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}_{\mathcal{E}} = \frac{3}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{E}} + \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{\mathcal{E}} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{A}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{E}} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{\mathcal{E}} - \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{\mathcal{E}}$$

so that

$$C_{\mathcal{BA}} = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

which takes a linear combination in $\mathcal{A}$ and outputs the corresponding linear combination in $\mathcal{B}$.

These ideas naturally generalise.

Proposition 3.10. *Let $\mathcal{A} = \{v_1, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, \ldots w_n\}$ be bases for $V = \mathbb{F}^n$. Then there exist $c_{ij} \in \mathbb{F}$ such that*

$$\begin{aligned} v_1 &= c_{11}w_1 + c_{21}w_2 + \ldots + c_{n1}w_n \\ v_2 &= c_{12}w_1 + c_{22}w_2 + \ldots + c_{n2}w_n \\ &\vdots \\ v_n &= c_{1n}w_1 + c_{2n}w_2 + \ldots + c_{nn}w_n. \end{aligned}$$

Moreover, given any $a = (a_1, \ldots, a_n)_{\mathcal{A}}$ and $(b_1, \ldots, b_n)_{\mathcal{B}} = b$ with $(a)_{\mathcal{A}} = (b)_{\mathcal{B}}$, we have that $b = C_{\mathcal{BA}}a$ where $(C_{\mathcal{BA}})_{ij} = c_{ij}$.

Proof. The equations for $v_1, \ldots, v_n$ follow because $\mathcal{B}$ is a basis for V. Let $v = (a)_{\mathcal{A}}$. Then

$$\begin{aligned} v &= a_1 v_1 + a_2 v_2 + \ldots + a_n v_n \\ &= a_1 \left(\sum_{i=1}^{n} c_{i1} w_i \right) + a_2 \left(\sum_{i=1}^{n} c_{i2} w_i \right) + \ldots + a_n \left(\sum_{i=1}^{n} c_{in} w_i \right) \\ &= (a_1 c_{11} + \ldots + a_n c_{1n}) w_1 + (a_1 c_{21} + \ldots + a_n c_{2n}) w_2 + \ldots + (a_1 c_{n1} + \ldots + a_n c_{nn}) w_n \\ &= b_1 w_1 + b_2 w_2 + \ldots + b_n w_n \end{aligned}$$

and so $b = C_{\mathcal{BA}} a$ as required. □

We now make some observations about these change of basis matrices.

Lemma 3.11. *Let $\mathcal{A} = \{v_1, \ldots, v_n\}$ be a basis for $\mathbb{F}^n$. Then $C_{\mathcal{AA}} = I_n$, the $n \times n$ identity matrix.*

Proof. With $w_i = v_i$ for each $i = 1, \ldots, n$ in the above proof, we get the required constants. □

Lemma 3.12. *Let $\mathcal{A} = \{v_1, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, \ldots w_n\}$ be bases for $\mathbb{F}^n$. Then $C_{\mathcal{AB}} = C_{\mathcal{BA}}^{-1}$.*

Proof. Note, since $\mathcal{A}$ and $\mathcal{B}$ span $\mathbb{F}^n$, that for any $a = (a_1, \ldots, a_n)_{\mathcal{A}}$ there exists $b = (b_1, \ldots, b_n)_{\mathcal{B}}$ such that $(a)_{\mathcal{A}} = v = (b)_{\mathcal{B}}$. Then $b = C_{\mathcal{BA}} a$ but also reversing the roles of a and b we have $C_{\mathcal{AB}} b = a$. Thus $C_{\mathcal{AB}} C_{\mathcal{BA}} a = a$ for all $a \in \mathbb{F}^n$, i.e. $C_{\mathcal{BA}}^{-1} = C_{\mathcal{AB}}$. □

Remark 3.13. This lemma has a powerful application. Assume $\{w_1, \ldots, w_n\}$ is basis and $\{v_1, \ldots, v_n\}$ are defined by $v_i = \sum_j c_{ji} w_j$. Then $\{v_1, \ldots, v_n\}$ is a basis if and only if $C = (c_{ij})$ is invertible, i.e. $\det(C) \neq 0$. The most concrete approach to prove this is to consider the form of a matrix $C_{\mathcal{EA}}$ and then note what the image and kernel tell us about the columns of this matrix[10].

We illustrate this remark with two examples.

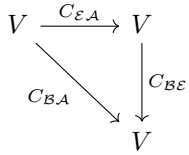
Example 3.14. We use examples involving $\mathbb{R}^3$ and $\mathbb{C}^3$.

- Consider the vectors $(x_1, 0, 0)$, $(y_1, y_2, 0)$, and $(z_1, z_2, z_3) \in \mathbb{R}^3$. Then these form a basis for $\mathbb{R}^3$ if and only if $x_1 y_2 z_3 \neq 0$, which occurs exactly when all of x_1, y_2, z_3 are nonzero.
- We can compute that $\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 0$, and so the columns of this matrix are linearly dependent in $\mathbb{C}^3$ and in $\mathbb{R}^3$. Note also that $\det(A) = \det(A^t)$ for any $A \in M_n(\mathbb{F})$, and so the rows of the given matrix are also linearly dependent in $\mathbb{C}^3$ and $\mathbb{R}^3$.

Lemma 3.15. *Let $\mathcal{A}$ and $\mathcal{B}$ be bases of $\mathbb{F}^n$. Then $C_{\mathcal{BE}} C_{\mathcal{EA}} = C_{\mathcal{BA}}$.*

Proof. Let $a \in \mathbb{F}^n$. Then $(a)_{\mathcal{A}} = (x)_{\mathcal{E}} = (b)_{\mathcal{B}}$. We will show that $C_{\mathcal{BE}} C_{\mathcal{EA}} a = C_{\mathcal{BA}} a$. Since a is arbitrary, we will then have that $C_{\mathcal{BE}} C_{\mathcal{EA}} = C_{\mathcal{BA}}$ as functions and so also as matrices. Calling on Proposition 3.10, we see that $C_{\mathcal{EA}} : a \to x$, $C_{\mathcal{BE}} : x \to b$ meaning $C_{\mathcal{BE}} C_{\mathcal{EA}} a = b$; we then note that also, by definition, $C_{\mathcal{BA}} a = b$. □

The idea of the above proof can be nicely demonstrated with what is known as a **commutative diagram**. This consists of arrows (representing maps) going between letters (representing sets). The diagram is said to be commutative since it does not matter which order we follow sequential arrows in: the resulting output for any element of a set should be the same. We will also use such diagrams in the next section.



[10]The image tells us whether the columns span $\mathbb{F}^n$, and a non-trivial kernel says that they are linearly dependent; the relationship between these is captured by the Rank-Nullity Theorem.

The following result is a consequence of the more general Proposition 3.26, which we will prove in the next section.

Proposition 3.16. *Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be three bases of $\mathbb{F}^n$. Then $C_{\mathcal{C}\mathcal{B}} C_{\mathcal{B}\mathcal{A}} = C_{\mathcal{C}\mathcal{A}}$.*

Proof. We put together our work from the above smaller results, noting that

$$\begin{aligned} C_{\mathcal{C}\mathcal{B}} C_{\mathcal{B}\mathcal{A}} &= (C_{\mathcal{C}\mathcal{E}} C_{\mathcal{E}\mathcal{B}})(C_{\mathcal{B}\mathcal{E}} C_{\mathcal{E}\mathcal{A}}) \\ &= C_{\mathcal{C}\mathcal{E}} (C_{\mathcal{B}\mathcal{E}}^{-1} C_{\mathcal{B}\mathcal{E}}) C_{\mathcal{E}\mathcal{A}} \\ &= C_{\mathcal{C}\mathcal{E}} C_{\mathcal{E}\mathcal{A}} \\ &= C_{\mathcal{C}\mathcal{A}}. \end{aligned}$$

□

This indicates another approach for finding $C_{\mathcal{B}\mathcal{A}}$, which is often simpler.

Example 3.17 (Example 3.6 in another way)**.** We find $C_{\mathcal{B}\mathcal{A}}$ by finding $C_{\mathcal{B}\mathcal{E}} C_{\mathcal{E}\mathcal{A}}$. Recall

$$C_{\mathcal{B}\mathcal{E}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \text{ and } C_{\mathcal{E}\mathcal{A}} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

so that

$$C_{\mathcal{B}\mathcal{A}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

which matches with the answer we found above.

In Section 1 we saw how to find a matrix to represent a linear map. But really this was only half the story, because we can represent any linear map with respect to any choices of bases $\mathcal{A}$ and $\mathcal{B}$.

Definition 3.18. *Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, w_2, \ldots, w_m\}$ be bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively, and $T : \mathbb{F}^n \to \mathbb{F}^m$ be an $\mathbb{F}$-linear map. Then*

$$M_{\mathcal{B}\mathcal{A}}(T) = (a_{ij}) \in M_{m,n}(\mathbb{F})$$

denotes the **matrix representing T with respect to the bases $\mathcal{A}$ and $\mathcal{B}$,** *where the elements $a_{ij} \in \mathbb{F}$ are defined by*

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i, \quad \text{for } i = 1, 2, \ldots, n.$$

We emphasise again that the existence and uniqueness of the matrix elements a_{ij} follows from the fact that $\mathcal{B} = \{w_1, \ldots, w_m\}$ is a basis, but the computation of these numbers requires usually some work and will in general lead to a system of nm linear equations.We will now see that the computations can be rather involved.

Example 3.19. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(e_1) = 2e_1 - e_2$ and $T(e_2) = e_2$. Then

$$M_{\mathcal{E}\mathcal{E}}(T) = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix}.$$

We will now find $M_{\mathcal{B}\mathcal{E}}(T)$ where $\mathcal{B} = \{w_1, w_2\}$ with $w_1 = (1, 2)$ and $w_2 = (1, -1)$. To do so, we find constants $a_{ij} \in \mathbb{R}$ such that $T(e_1) = a_{11} w_1 + a_{21} w_2$ and $T(e_2) = a_{12} w_1 + a_{22} w_2$. We could solve this directly (by solving the simultaneous linear equations). Alternatively we could write these as a matrix equation, i.e.,

$$\begin{aligned} T(e_1) = \begin{pmatrix} 2 \\ -1 \end{pmatrix} &= a_{11} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + a_{21} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} \text{ and} \\ T(e_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= a_{12} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + a_{22} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}. \end{aligned}$$

These two equations can then be realised as the single matrix equation

$$\begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

where the first equation above corresponds to the first column, and the second equation to the second column of this matrix equation. Thus

$$M_{\mathcal{BE}}(T) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} = \frac{-1}{3} \begin{pmatrix} -1 & -1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 5 & -1 \end{pmatrix}.$$

The next example illustrates the importance of the bases being used[11].

Example 3.20. Let id denote the identity map from $\mathbb{R}^2$ to $\mathbb{R}^2$, and let $\mathcal{A} = \{(a, c), (b, d)\}$ be a basis for $\mathbb{R}^2$. We will find $M_{\mathcal{EA}}(\mathrm{id})$. Note that $\mathrm{id} : (a, c) \to (a, c)$ and $(b, d) \to (b, d)$. Thus

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{A}} \to \begin{pmatrix} a \\ c \end{pmatrix}_{\mathcal{E}} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{A}} \to \begin{pmatrix} b \\ d \end{pmatrix}_{\mathcal{E}}$$

which, from our understanding of forming the matrix of a linear map, means

$$M_{\mathcal{EA}}(\mathrm{id}) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Note that this is far from being $I_2 = M_{\mathcal{EE}}(\mathrm{id})$.

Let us now understand the above examples in complete generality.

Proposition 3.21. *Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, w_2, \ldots, w_m\}$ be bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively, and $T : \mathbb{F}^n \to \mathbb{F}^m$ be an $\mathbb{F}$-linear map. Given any $v = (x_1, \ldots, x_n)_{\mathcal{A}} = (x)_{\mathcal{A}}$, we then have that $T(v) = (y_1, \ldots, y_m)_{\mathcal{B}} = (y)_{\mathcal{B}}$ where $y = M_{\mathcal{BA}}(T)x$.*

Proof. We know that T is $\mathbb{F}$-linear and $T(v_j) = \sum_{i=1}^m a_{ij} w_i$ for $j = 1, \ldots, n$. Hence

$$\begin{aligned} T(v) &= T(x_1 v_1 + \ldots + x_n v_n) \\ &= x_1 T(v_1) + \ldots + x_n T(v_n) \\ &= x_1 \Big(\sum_{i=1}^n a_{i1} w_i \Big) + \ldots + x_n \Big(\sum_{i=1}^n a_{in} w_i \Big) \\ &= \Big(\sum_{j=1}^n x_j a_{1j} \Big) w_1 + \ldots + \Big(\sum_{j=1}^n x_j a_{nj} \Big) w_n \end{aligned}$$

and so $y_i = \sum_{j=1}^n a_{ij} x_j$ for $i = 1, \ldots, m$. Hence $y = M_{\mathcal{BA}}(T)x$. □

This proof is really using that a double summand commutes, i.e. that $\sum_i \sum_j x_i y_j = \sum_j \sum_i x_i y_j$. We'll soon use this idea again in order to prove Proposition 3.26.

Example 3.22. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be an $\mathbb{R}$ linear map with $T(e_1) = e_1 + 3e_2$ and $T(e_2) = -2e_1$. With $\mathcal{B} = \{(1, 3), (-2, 0)\}$, we find $M_{\mathcal{BE}}(T)$. In this case, our computations are simple:

$$T(e_1) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_{\mathcal{B}} \text{; and } T(e_2) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}_{\mathcal{E}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_{\mathcal{B}}$$

which gives us $a_{11} = 1$, $a_{12} = 0$, $a_{21} = 0$, and $a_{22} = 1$. Hence $M_{\mathcal{BE}}(T) = I_2$.

Lemma 3.23. *Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a bijective linear map and $\mathcal{A} = \{v_1, \ldots, v_n\}$ be a basis of $\mathbb{F}^n$. If we set $\mathcal{B} = \{T(v_1), \ldots, T(v_n)\}$, then $M_{\mathcal{BA}}(T) = I_n$.*

Proof. If $w_i = T(v_i)$, then the matrix coefficients a_{ij} are 1 if $i = j$ and 0 otherwise. □

Remark 3.24. In the above lemma it is vital that T is bijective. If it were not, then the given set $\mathcal{B}$ would not be a basis for $\mathbb{F}^n$.

Lemma 3.25. *Let $\mathcal{A}$ and $\mathcal{B}$ be bases of $\mathbb{F}^n$. Then $M_{\mathcal{BA}}(\mathrm{id}) = C_{\mathcal{BA}}$.*

Proof. The matrix $M_{\mathcal{BA}}(\mathrm{id})$ sends the jth vector of $\mathcal{A}$ to the jth vector of $\mathcal{B}$. □

[11] Also, it is very important to note here that our FACT can be useful in computing such matrices, but does not immediately tell us the columns of a matrix $M_{\mathcal{BA}}(T)$.

We may query what matrix multiplication relates to in our framework. With the following theorem, we see that it is actually composition of maps. Note that, in light of the previous lemma, the following is a generalisation of Lemma 3.15.

Proposition 3.26. *Let $S : \mathbb{F}^m \to \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^p$ be $\mathbb{F}$-linear maps and $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be bases for $\mathbb{F}^m$, $\mathbb{F}^n$, and $\mathbb{F}^p$ respectively. Then*

$$M_{\mathcal{CA}}(T \circ S) = M_{\mathcal{CB}}(T)M_{\mathcal{BA}}(S).$$

Proof. Let $\mathcal{A} = \{u_1, \ldots, u_m\}$, $\mathcal{B} = \{v_1, \ldots, v_n\}$, and $\mathcal{C} = \{w_1, \ldots, w_p\}$. Then

$$S(u_j) = \sum_{i=1}^{n} a_{ij} v_i \quad \text{and} \quad T(v_i) = \sum_{k=1}^{p} b_{ki} w_k$$

where $M_{\mathcal{BA}}(S) = (a_{ij})$ and $M_{\mathcal{CB}}(T) = (b_{ki})$. Applying T to $S(u_j)$ (and using $\mathbb{F}$-linearity) we get

$$\begin{aligned}(T \circ S)(u_j) &= T(S(u_j)) = \sum_{i=1}^{n} a_{ij} T(v_i) = \sum_{i=1}^{n} a_{ij} \left(\sum_{k=1}^{p} b_{ki} w_k \right) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{p} a_{ij} b_{ki} w_k = \sum_{k=1}^{p} \sum_{i=1}^{n} a_{ij} b_{ki} w_k = \sum_{k=1}^{p} \underbrace{\left(\sum_{i=1}^{n} b_{ki} a_{ij} \right)}_{c_{kj}} w_k.\end{aligned}$$

Note that this is the rule for matrix multiplication, i.e. writing $M_{\mathcal{CA}}(T \circ S) = (c_{kj})$, we have $(c_{kj}) = (b_{ki})(a_{ij})$, as required. □

Remark 3.27. Suppose that a linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ is invertible. It follows from the above theorem that the matrices $M_{\mathcal{BA}}(T)$ and $M_{\mathcal{AB}}(T^{-1})$ (for any choice of bases $\mathcal{A}$ and $\mathcal{B}$) are inverse to each other. (Their composition is the identity operator, which has matrix I_n relative to any ordered basis.) In particular, the matrix of T is invertible.

Example 3.28. We apply Proposition 3.26 to some examples.

- Let $\mathcal{A} = \{(1,1), (1,-1)\}$, and S and T be $\mathbb{R}$-linear functions from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by $S : e_1 \mapsto \frac{1}{2}(e_1 + e_2)$, $S : e_2 \mapsto \frac{1}{2}(-e_1 + e_2)$, $T : e_1 \mapsto 2e_1$, and $T : e_2 \mapsto 0$. We will find $M_{\mathcal{AA}}(T \circ S)$ by computing $M_{\mathcal{AE}}(T)$ and $M_{\mathcal{EA}}(S)$. Using that S is linear, we have $S((1,1)) = (0,1)_{\mathcal{E}}$ and $S((1,-1)) = (1,0)_{\mathcal{E}}$. Then $T((1,0)) = (2,0)_{\mathcal{E}} = (1,1)_{\mathcal{A}}$ and $T((0,1)) = (0,0)_{\mathcal{E}} = (0,0)_{\mathcal{A}}$. Therefore

  $$M_{\mathcal{EA}}(S) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } M_{\mathcal{AE}}(T) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

  which gives us that

  $$M_{\mathcal{AA}}(T \circ S) = M_{\mathcal{AE}}(T)M_{\mathcal{EA}}(S) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

- Let $S : \mathbb{R}^3 \to \mathbb{R}^2$ and $T : \mathbb{R}^2 \to \mathbb{R}^4$ be $\mathbb{R}$-linear maps where

  $$S : \begin{array}{l} e_1 \mapsto e_1 + e_2 \\ e_2 \mapsto e_1 \\ e_3 \mapsto e_2 \end{array} \qquad T : \begin{array}{l} e_1 \mapsto e_2 \\ e_2 \mapsto e_3 \end{array}$$

  We find $M_{\mathcal{EE}}(T \circ S)$. First, $M_{\mathcal{EE}}(S)$ has 3 columns, $e_1 + e_2$, e_1, and e_2. Then $M_{\mathcal{EE}}(T)$ has 2 columns, e_2 and e_3. Putting these together we see that

  $$M_{\mathcal{EE}}(T \circ S) = M_{\mathcal{EE}}(T)M_{\mathcal{EE}}(S) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

  which means that $T \circ S$ sends $e_1 \mapsto e_2 + e_3$, $e_2 \mapsto e_2$, and $e_3 \mapsto e_3$. Note that we can compose these maps because of the correct codomain of S and domain of T, and that these correspond to the correct dimensions of matrices for us to be able to multiply them.

The following gives us new understanding of how $M_{\mathcal{BA}}(T)$ and $M_{\mathcal{EE}}(T)$ are related.

Lemma 3.29. *Let $T : \mathbb{F}^m \to \mathbb{F}^n$ be an $\mathbb{F}$-linear map and $\mathcal{A}$ and $\mathcal{B}$ be bases of $\mathbb{F}^m$ and $\mathbb{F}^n$ respectively. Then*

$$M_{\mathcal{BA}}(T) = C_{\mathcal{BE}} M_{\mathcal{EE}}(T) C_{\mathcal{EA}} = C_{\mathcal{EB}}^{-1} M_{\mathcal{EE}}(T) C_{\mathcal{EA}}.$$

Proof. We use Lemma 3.25 (that $M_{\mathcal{BB}'}(\mathrm{id}) = C_{\mathcal{BB}'}$) and apply Proposition 3.26 twice:

$$\begin{aligned} M_{\mathcal{BA}}(T) &= M_{\mathcal{BA}}(T \circ \mathrm{id}) \\ &= M_{\mathcal{BE}}(T) M_{\mathcal{EA}}(\mathrm{id}) \\ &= M_{\mathcal{BE}}(\mathrm{id} \circ T) M_{\mathcal{EA}}(\mathrm{id}) \\ &= M_{\mathcal{BE}}(\mathrm{id}) M_{\mathcal{EE}}(T) M_{\mathcal{EA}}(\mathrm{id}) = C_{\mathcal{BE}} M_{\mathcal{EE}}(T) C_{\mathcal{EA}}. \end{aligned}$$

The final statement follows from the fact that $C_{\mathcal{BE}}^{-1} = C_{\mathcal{EB}}$, which is Lemma 3.12. □

A corollary to this theorem is that $M_{\mathcal{B}'\mathcal{A}'}(T) = C_{\mathcal{B}'\mathcal{B}} M_{\mathcal{BA}}(T) C_{\mathcal{AA}'} = C_{\mathcal{BB}'}^{-1} M_{\mathcal{BA}}(T) C_{\mathcal{AA}'}$, which is left as an exercise. Really, this theorem is a useful tool for how we can compute $M_{\mathcal{BA}}(T)$ when given $T : \mathbb{F}^m \to \mathbb{F}^n$ and the bases $\mathcal{A}$, $\mathcal{B}$ of $\mathbb{F}^m$ and $\mathbb{F}^n$. It states that we only need find $M_{\mathcal{EE}}(T)$ and can then use the matrices $C_{\mathcal{BE}}$ and $C_{\mathcal{EA}}$ of the previous section in order to find $M_{\mathcal{BA}}(T)$. This is nicely summarised by the following diagram.

$$\begin{array}{ccc} \mathbb{F}^m & \xrightarrow{M_{\mathcal{BA}}(T)} & \mathbb{F}^n \\ {\scriptstyle C_{\mathcal{EA}}}\downarrow & & \uparrow{\scriptstyle C_{\mathcal{BE}}} \\ \mathbb{F}^m & \xrightarrow[M_{\mathcal{EE}}(T)]{} & \mathbb{F}^n \end{array}$$

Particular attention is given to the case where $m = n$, which means that $M_{\mathcal{BA}}(T)$ is a square matrix. Generally we then also impose that $\mathcal{A} = \mathcal{B}$, using the notation $M_{\mathcal{A}}(T)$ or sometimes just M_T if this is unambiguous (for example when using the standard basis in $\mathbb{R}^n$). We now see the above result in these restricted circumstances.

Corollary 3.30. *Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be an $\mathbb{F}$-linear operator, and $\mathcal{B}, \mathcal{B}'$ be bases of $\mathbb{F}^n$. Set $M := M_{\mathcal{BB}}(T)$ and $M' := M_{\mathcal{B}'\mathcal{B}'}(T)$. Then*

$$M' = C^{-1} M C,$$

where $C = C_{\mathcal{BB}'}$ is the matrix of change of coordinates from $\mathcal{B}'$ to $\mathcal{B}$.

Remark 3.31. This result means that a change of coordinates relates to the matrices being similar, which we saw in Section 2.4. Since the choice of basis does not impact on the determinant, the trace, or the characteristic polynomial of a matrix, we can therefore define these for an abstract linear map as the answer we obtain for **any** choice of basis.

This remark naturally raises whether we can find a basis $\mathcal{B}$ such that $M_{\mathcal{BB}}(T)$ is particularly simple[12]. We now recall our ideas from Section 2 in terms of a choice of basis.

Theorem 3.32. *Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be an $\mathbb{F}$-linear operator. If we can find a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of eigenvectors of T, i.e., $T(v_i) = \lambda_i v_i$, then*

$$M_{\mathcal{BB}}(T) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

where $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the diagonal matrix with elements $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the diagonal. Furthermore, if $\mathcal{B} = \{v_1, \ldots, v_n\}$ is basis such that $M_{\mathcal{BB}}(T) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ for some numbers $\lambda_i \in \mathbb{F}$, then each vector v_i is an eigenvector of T with corresponding eigenvalue λ_i. Hence a square matrix is diagonalisable if and only if it has a basis of eigenvectors.

In the next section we look at special kinds of bases which make the computations of the matrices in this section far simpler.

[12] "Particularly simple" ideally means diagonal. But we saw in general that this is not always possible, and the full answer is given by the so-called Jordan normal form, which we leave to Linear Algebra 2.

4. Inner products

This section only relates to $\mathbb{R}^n$ and $\mathbb{C}^n$. Recall the dot product and norm

$$x \cdot y = \sum_{i=1}^{n} x_i y_i \ , \quad \|x\| = \sqrt{x \cdot x} = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}},$$

which allowed to measure length of vectors and angles between them, and in particular gave us the notion of orthogonality. Our aim is to generalise the notion of a dot product. We do this with an *inner product*. The key motivation for it comes from beyond this course, in particular quantum mechanics. Note that this is intrinsically complex, rather than real.

First, we extend the notion of dot product and norm to $\mathbb{C}^n$. In this case the answers will be in $\mathbb{C}$. In order to give rise to a norm, we want that the inner product of a vector with itself to output a positive real number. For this reason, let us consider

$$\bar{x} \cdot y := \sum_{i=1}^{n} \bar{x}_i y_i$$

where $\bar{x}$ denotes complex conjugation. All the generalisations of the dot product share some key features which we take now to define the general notion of an inner product.

Definition 4.1. *For a vector space V over $\mathbb{C}$ (think: $\mathbb{C}^n$), an* **inner product** *on V is a map $\langle\, , \rangle : V \times V \to \mathbb{C}$ which has the following properties for all $u, v, w \in V$ and $\lambda \in \mathbb{C}$:*

(i) $\langle v, v \rangle \in \mathbb{R}_{\geqslant 0}$ *and* $\langle v, v \rangle = 0$ *if and only if* $v = 0$*;*
(ii) $\langle v, u + w \rangle = \langle v, u \rangle + \langle v, w \rangle$ *and* $\langle v, \lambda w \rangle = \lambda \langle v, w \rangle$*;*
(iii) $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

We will see that $\langle \lambda v, w \rangle \neq \lambda \langle v, w \rangle$ for $\lambda \in \mathbb{C} \setminus \mathbb{R}$, and instead get $\langle \lambda v, w \rangle = \bar{\lambda} \langle v, w \rangle$. Informally, we can "pull out" the scalar λ from the second position in $\langle , \rangle$ but not the first[13]. The reasoning for this is to ensure that $\langle v, v \rangle =: \|v\|^2$ is in $\mathbb{R}$. In the case that V is a vector space over $\mathbb{R}$ we have the same definition, but condition (iii) can be more simply written as $\langle v, w \rangle = \langle w, v \rangle$.

Definition 4.2. *For a vector space V over $\mathbb{R}$ (think: $\mathbb{R}^n$), an* **inner product** *on V is a map $\langle\, , \rangle : V \times V \to \mathbb{R}$ which has the following properties for all $u, v, w \in V$ and $\lambda \in \mathbb{R}$:*

(i) $\langle v, v \rangle \geq 0$ *and* $\langle v, v \rangle = 0$ *if and only if* $v = 0$*;*
(ii) $\langle v, u + w \rangle = \langle v, u \rangle + \langle v, w \rangle$ *and* $\langle v, \lambda w \rangle = \lambda \langle v, w \rangle$*;*
(iii) $\langle v, w \rangle = \langle w, v \rangle$.

Example 4.3. We see two inner products, which will be our main examples for now.

(i) For $V = \mathbb{C}^n$ we have the (standard) inner product given by

$$\langle x, y \rangle := \sum_{i=1}^{n} \bar{x}_i y_i = \bar{x} \cdot y.$$

(ii) Let $A \in M_n(\mathbb{R})$ be a matrix which is symmetric ($A^t = A$) and positive definite ($x \cdot Ax > 0$ for all $x \in \mathbb{R}^n \backslash \{0\}$). Then we obtain an inner product on $V = \mathbb{R}^n$ from

$$\langle x, y \rangle_A := x \cdot Ay = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} y_j.$$

Definition 4.4. *A vector space V with an inner product $\langle\, , \rangle$ defined on it is called an* **inner product space** *$(V, \langle\, , \rangle)$. If we wish to further specify the field as $\mathbb{C}$ or $\mathbb{R}$, then we say V is a complex inner product space or a real inner product space respectively.*

Remark 4.5. For many of our proofs we will imagine that V is a complex inner product space, since this is a greater restriction on V. Our proofs therefore hold for any inner product space.

[13] This is a matter of convention: some authors insist that the inner product is linear in the first variable rather than the second.

Let us note now a few simple consequences of our definition.

Lemma 4.6. *Let $(V, \langle , \rangle)$ be an inner product space over $\mathbb{C}$. Then*

$$\langle u + w, v\rangle = \langle u, v\rangle + \langle w, v\rangle \quad \textit{and} \quad \langle \lambda w, v\rangle = \bar{\lambda}\langle w, v\rangle$$

for all $u, v, w \in V$ and $\lambda \in \mathbb{C}$.

Proof. This follows from combining (ii) and (iii) in the definition of $\langle , \rangle$. Let us show the second assertion: $\langle \lambda w, v\rangle = \overline{\langle v, \lambda w\rangle} = \overline{\lambda\langle v, w\rangle} = \bar{\lambda}\overline{\langle v, w\rangle} = \bar{\lambda}\langle w, v\rangle$. The other is similar. □

If $(V, \langle , \rangle)$ is a real inner product space then we have instead $\langle \lambda v, w\rangle = \lambda\langle v, w\rangle$. Also, the above properties can be extended to linear combinations of vectors. We have

$$\langle \sum_{i=1}^{k} \lambda_i v_i, w\rangle = \sum_{i=1}^{k} \bar{\lambda}_i \langle v_i, w\rangle \quad \text{and} \quad \langle v, \sum_{i=1}^{k} \lambda_i w_i\rangle = \sum_{i=1}^{k} \lambda_i \langle v, w_i\rangle.$$

Having an inner product enables us to define a norm.

Definition 4.7. *Let $(V, \langle , \rangle)$ be an inner product space. Then we define an associated norm by*

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

Example 4.8 (Example 4.3 continued)**.** We find the norm for our examples of inner products.

(i) $\|x\| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$

(ii) $\|x\|_A = \left(\sum_{i,j=1}^{n} a_{ij} x_i x_j \right)^{\frac{1}{2}}$

We used the dot product previously to define as well the angle between vectors. But on a complex vector space the inner product gives usually a complex number, so we can't easily define an angle. The notion of orthogonality, however, can be extended directly.

Definition 4.9. *Let $(V, \langle , \rangle)$ be an inner product space, then*

(i) *$v, w \in V$ are* **orthogonal***, denoted $v \perp w$, if $\langle v, w\rangle = 0$;*

(ii) *two subspaces $U, W \subset V$ are called* **orthogonal***, denoted $U \perp W$, if $u \perp w$ for all $u \in U$ and $w \in W$.*

Example 4.10. We work with $\mathbb{C}^2$ with the standard inner product $\langle x, y\rangle = \bar{x} \cdot y$.

i) For $v_1 = (\mathrm{i}, 1)$ and $v_2 = (1, \mathrm{i})$, we have $v_1 \perp v_2$.

ii) With v_1 and v_2 as above, $U = \mathrm{span}\{v_1\}$ and $W = \mathrm{span}\{v_2\}$ are orthogonal.

The following is probably most natural when the subset is a subspace.

Definition 4.11. *Let V be an inner product space and S a subset of V. Then the* **orthogonal complement** *of S, denoted $S^\perp$, is*

$$S^\perp := \{v \in V,\ v \perp s \textit{ for all } s \in S\}.$$

Example 4.12. We look at some examples of orthogonal complements.

i) For $W = \mathrm{span}\{(\mathrm{i}, 1)\}$, we have $W^\perp = \mathrm{span}\{(1, \mathrm{i})\}$.

ii) For $\mathbb{R}^n$ and $k < n$, if $W = \mathrm{span}\{e_1, \ldots, e_k\}$ then $W^\perp = \mathrm{span}\{e_{k+1}, \ldots, e_n\}$.

iii) For $\mathbb{R}^n$ and $k < n$, if $S = \{e_1, \ldots, e_k\}$ then $S^\perp = \mathrm{span}\{e_{k+1}, \ldots, e_n\}$.

The following lemma is helpful in determining orthogonal complements.

Lemma 4.13. *Let V be an inner product space and S a subset. Then $S^\perp$ is a subspace of V.*

Proof. We use the subspace test. We first note that $0 \in S^\perp$. Secondly, that for $v_1, v_2 \in S^\perp$ and any $s \in S$ we have $\langle v_1 + v_2, s\rangle = \langle v_1, s\rangle + \langle v_2, s\rangle = 0 + 0$. Finally, that for $\lambda \in \mathbb{C}$, $v \in V$, and $s \in S$, that $\langle \lambda v, s\rangle = \bar{\lambda}\langle v, s\rangle = \bar{\lambda} 0 = 0$. □

We have as well a Pythagoras theorem for orthogonal vectors.

Theorem 4.14. *Let V be an inner product space and $v, w \in V$. If $v \perp w$, then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Proof. We have $\|v + w\|^2 = \langle v + w, v + w\rangle = \langle v, v\rangle + \langle w, w\rangle + \langle v, w\rangle + \langle w, v\rangle$ which is then $\|v\|^2 + \|w\|^2 + \langle v, w\rangle + \overline{\langle v, w\rangle}$. Our assumption that $v \perp w$ gives us $\|v + w\|^2 = \|v\|^2 + \|w\|^2$. □

Note in the above proof that $\langle v, w\rangle + \overline{\langle v, w\rangle} = 0$ would imply that $\|v+w\|^2 = \|v\|^2 + \|w\|^2$. For a complex inner product we could have that $\langle v, w\rangle = iy$ for some $y \in \mathbb{R} \setminus \{0\}$ and so our statement is not an if and only if. For a real inner product we get that $\langle v, w\rangle + \overline{\langle v, w\rangle} = 2\langle v, w\rangle$ and so the only way for this to be zero is if $\langle v, w\rangle = 0$, i.e., $v \perp w$ (making the statement an if and only if).

One of the advantages of having an inner product on a vector space is that we can introduce the notion of an orthonormal basis.

Definition 4.15. *Let $(V, \langle , \rangle)$ be an inner product space. Then a basis $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ is called an* **orthonormal basis** *(often abbreviated as ONB) if*

$$\langle v_i, v_j\rangle = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Example 4.16. We give some examples of orthonormal bases.

i) For $V = \mathbb{C}^n$ with the standard inner product, $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ is an ONB.
ii) For (i) with $n = 2$, the vectors $v_1 = \frac{1}{\sqrt{2}}(\mathrm{i}, 1)$ and $v_1 = \frac{1}{\sqrt{2}}(1, \mathrm{i})$ form an ONB.
iii) On $V = \mathbb{R}^n$ with $\langle , \rangle_A$, where $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ the set $\mathcal{B} = \{v_1, \ldots, v_n\}$ with $v_i = (\alpha_i)^{-1/2} e_i$, $i = 1, 2, \ldots, n$ is an ONB.

Working with the standard basis has had many advantages. For one, given $x \in \mathbb{R}^n$, we had $x = (x \cdot e_1)e_1 + (x \cdot e_2)e_2 + \ldots + (x \cdot e_n)e_n$. This actually follows from $\mathcal{E}$ being an ONB.

Proposition 4.17. *Let $(V, \langle , \rangle)$ be an inner product space and $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ an orthonormal basis. Then for any $v, w \in V$ we have each of the following.*

i) $v = \sum_{i=1}^n \langle v_i, v\rangle v_i$
ii) $\langle v, w\rangle = \sum_{i=1}^n \overline{\langle v_i, v\rangle}\langle v_i, w\rangle$
iii) $\|v\| = \left(\sum_{i=1}^n |\langle v_i, v\rangle|^2\right)^{\frac{1}{2}}$

Proof. Since $\mathcal{B}$ is a basis, there are $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ such that $v = \sum_{j=1}^n \lambda_j v_j$. Then

$$\langle v_i, v\rangle = \langle v_i, \sum_{j=1}^n \lambda_j v_j\rangle = \sum_{j=1}^n \lambda_j \langle v_i, v_j\rangle = \sum_{j=1}^n \lambda_j \delta_{ij} = \lambda_i$$

from which (i) follows. For (ii), we use that $v = \sum_{k=1}^n \langle v_k, v\rangle v_k$ and get

$$\langle v, w\rangle = \langle \sum_{k=1}^n \langle v_k, v\rangle v_k, w\rangle = \sum_{k=1}^n \langle \langle v_k, v\rangle v_k, w\rangle = \sum_{k=1}^n \overline{\langle v_k, v\rangle}\langle v_k, w\rangle.$$

Finally, (iii) follows from (ii) by setting $v = w$. □

Remark 4.18. The result (ii) above gives us a way to compute the coefficients for a linear combination in an ONB $\mathcal{B} = \{v_1, \ldots, v_n\}$. That is, given $v \in V$, we have $v = x_1 v_1 + \ldots + x_n v_n$ where $x_i = \langle v_i, v\rangle$ for $i = 1, \ldots, n$. Moreover, in these coordinates, the inner product becomes the standard inner product on $\mathbb{C}^n$. That is, for $(a)_{\mathcal{E}} = (x)_{\mathcal{B}}$ and $(b)_{\mathcal{E}} = (y)_{\mathcal{B}}$, we have

$$\langle a, b\rangle = \sum_{i=1}^n \bar{x}_i y_i = \bar{x} \cdot y.$$

At this point, we may wonder whether we can always find an ONB.

Theorem 4.19. *Let V be an inner product space with basis $\mathcal{A} = \{u_1, \ldots, u_n\}$. Then there exists an ONB for V.*

Proof. We will turn the basis $\mathcal{A} = \{u_1, \ldots, u_n\}$ of V into an orthonormal one in the following way (generally known as the Gram–Schmidt process). We set

$$\begin{aligned}
v_1 &:= \frac{1}{\|u_1\|}\, u_1 \\
v_2 &:= \frac{1}{\|u_2 - \langle v_1, u_2\rangle v_1\|}(u_2 - \langle v_1, u_2\rangle v_1) \\
v_3 &:= \frac{1}{\|u_3 - \langle v_2, u_3\rangle v_2 - \langle v_1, u_3\rangle v_1\|}(u_3 - \langle v_2, u_3\rangle v_2 - \langle v_1, u_3\rangle v_1) \\
&\vdots \\
v_n &:= \frac{1}{\|u_n - \langle v_{n-1}, u_n\rangle v_{n-1} - \ldots - \langle v_1, u_n\rangle v_1\|}(u_n - \langle v_{n-1}, u_n\rangle v_{n-1} - \ldots - \langle v_1, u_n\rangle v_1)
\end{aligned}$$

and this defines a set of n orthonormal vectors, hence an orthonormal basis. □

An advantage of having an inner product on a space V is that, for any subspace $U \subset V$, we can find a unique complementary subspace consisting of all orthogonal vectors.

Proposition 4.20. *Let V be an inner product space and U a finite dimensional subspace. Then*

$$V = U \oplus U^{\perp}.$$

Proof. The previous theorem says that there is an ONB $\{v_1, \ldots, v_k\}$ for U. We begin by showing that $V = U + U^{\perp}$. Let $v \in V$. Set

$$u := \langle v_1, v\rangle v_1 + \langle v_2, v\rangle v_2 + \ldots + \langle v_k, v\rangle v_k$$

which is an element of U since $\langle v_j, v\rangle \in \mathbb{C}$ for $j = 1, \ldots, k$. Now let

$$w := v - \langle v_1, v\rangle v_1 - \langle v_2, v\rangle v_2 - \ldots - \langle v_k, v\rangle v_k$$

which is actually in $U^{\perp}$. To see this we note, for any $j = 1, \ldots, k$, that $\langle v_j, w\rangle = \langle v_j, v\rangle - \langle v_j, v\rangle$. Thus $v = u + w$ where $u \in U$ and $w \in U^{\perp}$. Finally we show this is a direct sum. Consider if $v \in U \cap U^{\perp}$. But then $\langle v, v\rangle = 0$, and so $v = 0$. □

We can check that the above Theorem does indeed hold for the spaces in Example 4.12.

Proposition 4.21. *Let V be an inner product space and U a finite dimensional subspace. Then*

$$(U^{\perp})^{\perp} = U.$$

Proof. Note that $\langle u, v\rangle = 0$ if and only if $\langle v, u\rangle = 0$. We show that $U \subseteq (U^{\perp})^{\perp}$ and $(U^{\perp})^{\perp} \subseteq U$.

- Take $u \in U$. Then $\langle w, u\rangle = 0$ for every $w \in U^{\perp}$. But then $\langle u, w\rangle = 0$ and so $u \in (U^{\perp})^{\perp}$.
- Take $v \in (U^{\perp})^{\perp}$. Then, by Proposition 4.20, $v = u + w$ where $u \in U$ and $w \in U^{\perp}$. Thus $v - u = w \in U^{\perp}$. But also $v, u \in (U^{\perp})^{\perp}$, and so $v - u \in (U^{\perp})^{\perp}$. Thus $v - u \in U^{\perp} \cap (U^{\perp})^{\perp}$, and so $\langle v - u, v - u\rangle = 0$. Hence $v - u = 0$, and $v = u$, i.e. $v \in U$. □

We have two applications of Remark 4.18 that simplify the computations of Section 3.

Example 4.22. We will express a general vector as a linear combination in an ONB.

i) For $\mathbb{C}^2$ we know that $v_1 = \frac{1}{\sqrt{2}}(\mathrm{i}, 1)$ and $v_1 = \frac{1}{\sqrt{2}}(1, \mathrm{i})$ form an ONB. To expand an arbitrary vector $v = (z_1, z_2)$ in $\{v_1, v_2\}$, we can simply compute $\langle v_1, v\rangle = \frac{-\mathrm{i}z_1 + z_2}{\sqrt{2}}$ and $\langle v_2, v\rangle = \frac{z_1 - \mathrm{i}z_2}{\sqrt{2}}$ to obtain that

$$v = \frac{-\mathrm{i}z_1 + z_2}{\sqrt{2}}\, v_1 + \frac{z_1 - \mathrm{i}z_2}{\sqrt{2}}\, v_2$$

which allows us to avoid solving a system of two linear equations. Note also that this allows us to easily compute $C_{\mathcal{BA}}$ for any basis $\mathcal{A}$ of $\mathbb{C}^2$.

ii) One can check that

$$v_1 = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\2\end{pmatrix}\ , \quad v_2 = \frac{1}{\sqrt{5}}\begin{pmatrix}-2\\0\\1\end{pmatrix}\ , \quad v_3 = \frac{1}{\sqrt{30}}\begin{pmatrix}1\\-5\\2\end{pmatrix}$$

is an ONB for $\mathbb{R}^3$ by computing $\langle v_i, v_j\rangle$ for $i, j = 1, 2, 3$. To expand a vector $v = (x, y, z)$ in this basis, we would have previously had to solve a system of three equations for three unknowns. Instead, compute $\langle v_1, v\rangle = \frac{x+y+2z}{\sqrt{6}}$, $\langle v_2, v\rangle = \frac{-2x+z}{\sqrt{5}}$ and $\langle v_3, v\rangle = \frac{x-5y+2z}{\sqrt{30}}$ to immediately obtain that

$$v = \frac{x+y+2z}{\sqrt{6}}\, v_1 + \frac{-2x+z}{\sqrt{5}}\, v_2 + \frac{x-5y+2z}{\sqrt{30}}\, v_3.$$

iii) For $V = \mathbb{C}^n$ with $\langle x, y\rangle = \bar{x}\cdot y$ and the standard basis $\mathcal{E}$, we have for $x = (x_1, \ldots, x_n)$ that $\langle e_i, x\rangle = x_i$. Hence the expansion formula just gives (as we would expect)

$$x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n.$$

The above examples may also remind us of how we defined a matrix to represent a linear transformation T. The entries of the matrix were given by $a_{ij} = e_i \cdot T(e_j)$. A similar approach can be taken whenever we have an ONB.

Proposition 4.23. *Let $(V, \langle,\rangle)$ be an inner product space and $\mathcal{B} = \{v_1 \ldots, v_n\}$ an orthonormal basis. Given a linear operator $T : V \to V$, we have that*

$$M_{\mathcal{B}\mathcal{B}}(T) = (\langle v_i, Tv_j\rangle).$$

Proof. The entries of $M_{\mathcal{B}\mathcal{B}}(T)$ are defined by the equation $Tv_j = \sum_k a_{kj}v_k$ for $j = 1, \ldots, n$. Let $i \in \{1, \ldots, n\}$. Then we have $\langle v_i, Tv_j\rangle = \langle v_i, \sum_k a_{kj}v_k\rangle = a_{ij}$ because $\mathcal{B}$ is an ONB. □

This result is an immensely practical one: rather than solving a system of n linear systems of equations to define, one by one, the columns of a matrix representing T, we merely need calculate n^2 inner products.

Example 4.24 (Example 4.22(ii) continued)**.** Recall that we found an ONB

$$v_1 = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\2\end{pmatrix}\ , \quad v_2 = \frac{1}{\sqrt{5}}\begin{pmatrix}-2\\0\\1\end{pmatrix}\ , \quad v_3 = \frac{1}{\sqrt{30}}\begin{pmatrix}1\\-5\\2\end{pmatrix}.$$

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map defined by $M_{\mathcal{E}\mathcal{E}}(T) := \begin{pmatrix}1 & 3 & 0\\0 & 1 & 0\\0 & 2 & -1\end{pmatrix}$. We will find $M_{\mathcal{B}\mathcal{B}}(T)$. We begin by finding that

$$T(v_1) = \frac{1}{\sqrt{6}}\begin{pmatrix}4\\1\\0\end{pmatrix} \qquad T(v_2) = \frac{1}{\sqrt{5}}\begin{pmatrix}-2\\0\\-1\end{pmatrix}\ , \quad T(v_3) = \frac{1}{\sqrt{30}}\begin{pmatrix}-14\\-5\\-12\end{pmatrix}.$$

We can then obtain the matrix elements $a_{ij} = v_i \cdot T(v_j)$ as

$$M_{\mathcal{B}\mathcal{B}}(T) = \begin{pmatrix}\frac{5}{6} & \frac{-4}{\sqrt{30}} & \frac{-43}{6\sqrt{5}}\\ \frac{-8}{\sqrt{30}} & \frac{3}{5} & \frac{16}{5\sqrt{6}}\\ \frac{-1}{6\sqrt{5}} & \frac{-4}{5\sqrt{6}} & \frac{-13}{30}\end{pmatrix}. \tag{4.1}$$

Thus, by computing 9 inner products, we have found $M_{\mathcal{B}\mathcal{B}}(T)$. This is far less work than solving the relevant simultaneous linear equations.

5. The adjoint and important classes of matrices

In this section we restrict ourselves to complex inner product spaces of finite dimension. We do this as then any linear operator has at least one eigenvalue, since any polynomial has at least one root over $\mathbb{C}$.

Definition 5.1. *Let $(V, \langle , \rangle)$ be an inner product space and $T : V \to V$ a linear operator. Then the* **adjoint operator** *$T^* : V \to V$ is defined by the relation*

$$\langle T^* v, w\rangle = \langle v, Tw\rangle \text{ for all } v, w \in V.$$

Example 5.2. We start simply. Let $\lambda \in \mathbb{C}$ and $T(v) := \lambda v$ for all $v \in V$. We have $T^* = \bar{\lambda} I$, since

$$\langle v, Tw\rangle = \langle v, \lambda w\rangle = \lambda\langle v, w\rangle = \langle \bar{\lambda} v, w\rangle.$$

Upon first seeing this definition, we may be concerned whether in general T^* exists and if it is unique. One can develop some general arguments answering both questions affirmatively, but the best way to get some understanding of the adjoint is to look at matrices.

Lemma 5.3. *Let V be an inner product space and $T : V \to V$ be a linear operator. If $\mathcal{B}$ is an orthonormal basis and T has the matrix $M_{\mathcal{BB}}(T) = (a_{ij})$ in that basis, then*

$$M_{\mathcal{BB}}(T^*) = (\bar{a}_{ji}).$$

Proof. By definition, the entries of $M_{\mathcal{BB}}(T)$ are given by $a_{ij} = \langle v_i, Tv_j\rangle$. To find the entries b_{ij} of $M_{\mathcal{BB}}(T^*)$ we note that $b_{ij} = \langle v_i, T^* v_j\rangle = \overline{\langle T^* v_j, v_i\rangle} = \overline{\langle v_j, Tv_i\rangle}$ which equals $\bar{a}_{ji}$. □

Thus we can obtain $M_{\mathcal{BB}}(T^*)$ from $M_{\mathcal{BB}}(T)$ by taking the complex conjugate of the entries and then take the transpose. It is worthwhile to give this operation on matrices an extra definition.

Definition 5.4. *Let $A = (a_{ij}) \in M_{n,m}(\mathbb{C})$ be an $m \times n$ matrix with complex elements, then the matrix $A^* = (\bar{a}_{ji}) \in M_{m,n}(\mathbb{C})$ is called the* **adjoint matrix**.

Example 5.5. Let us find the adjoint of particular matrices. Let

$$A = \begin{pmatrix} 2-\mathrm{i} & 1+3\mathrm{i} \\ -\mathrm{i} & 2 \end{pmatrix} \quad B = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \quad C = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i} & 1 \\ 1 & \mathrm{i} \end{pmatrix} \quad D = \begin{pmatrix} 3 & 2-\mathrm{i} & \mathrm{e}^{2\mathrm{i}} \\ 0 & \mathrm{i} & 3 \\ 11\mathrm{i}-1 & 12 & \pi \end{pmatrix}.$$

Then

$$A^* = \begin{pmatrix} 2+\mathrm{i} & \mathrm{i} \\ 1-3\mathrm{i} & 2 \end{pmatrix} \quad B^* = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \quad C^* = \frac{1}{\sqrt{2}} \begin{pmatrix} -\mathrm{i} & 1 \\ 1 & -\mathrm{i} \end{pmatrix} \quad D^* = \begin{pmatrix} 3 & 0 & -11\mathrm{i}-1 \\ 2+\mathrm{i} & -\mathrm{i} & 12 \\ \mathrm{e}^{-2\mathrm{i}} & 3 & \pi \end{pmatrix}.$$

Notice that $B^* = B$, and by direct computation we have that $C^* C = I$.

The following lemma will be helpful for understanding properties of the adjoint.

Lemma 5.6. *Let V be an inner product space, and T and T' be linear maps. If $\langle v, Tw\rangle = \langle v, T'w\rangle$ for all $v, w \in V$, then $T = T'$.*

Proof. One approach is to fix an ONB $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V and note that $M_{\mathcal{BB}}(T) = M_{\mathcal{BB}}(T')$ since $\langle v_i, T(v_j)\rangle = \langle v_i, T'(v_j)\rangle$ for all $i, j \in \{1, \ldots, n\}$. Another approach, where we do not choose a finite basis, is to note that

$$\langle v, Tw\rangle = \langle v, T'w\rangle \Leftrightarrow \langle v, Tw\rangle - \langle v, T'w\rangle = 0 \Leftrightarrow \langle v, Tw - T'w\rangle = 0$$

and so if $Tw - T'w = 0$ for all $w \in V$ then we are done (since then $Tw = T'w$ for all $w \in V$). Now, imagine that there is a $x \in V$ such that $Tx - T'x = u \neq 0$. But then choose $v := u$ in the above computation. This provides a contradiction since

$$\langle u, Tx - T'x\rangle = \langle u, u\rangle = 0$$

only occurs when $u = 0$ (from our definition of an inner product). □

Similarly if $\langle Tv, w\rangle = \langle T'v, w\rangle$ for all $v, w \in V$, then $T = T'$. We are now ready to observe a few consequences of the definition of the adjoint.

Lemma 5.7. *Let V be an inner product space and $S, T : V \to V$ be linear operators. Then*

i) $(T^*)^* = T$.
ii) $(S+T)^* = S^* + T^*$.
iii) $(TS)^* = S^*T^*$.
iv) *if T is invertible, then* $(T^{-1})^* = (T^*)^{-1}$.

Proof. We lean heavily on the previous lemma, and deal with each statement in turn.

i) We have $\langle v, Tw\rangle = \langle T^*v, w\rangle = \overline{\langle w, T^*v\rangle} = \overline{\langle (T^*)^*w, v\rangle} = \langle v, (T^*)^*w\rangle$.
ii) Here $\langle (S+T)^*v, w\rangle = \langle v, (S+T)w\rangle = \langle v, Sw + Tw\rangle = \langle v, Sw\rangle + \langle v, Tw\rangle$ and then
$$\langle v, Sw\rangle + \langle v, Tw\rangle = \langle S^*v, w\rangle + \langle T^*v, w\rangle = \langle S^*v + T^*v, w\rangle = \langle (S^*+T^*)v, w\rangle.$$
iii) Now $\langle (TS)^*v, w\rangle = \langle v, T(S(w))\rangle = \langle T^*v, Sw\rangle = \langle S^*(T^*(v)), w\rangle = \langle S^*T^*v, w\rangle$.
iv) We note that $\langle (T^{-1})^*T^*v, w\rangle = \langle T^*v, T^{-1}w\rangle = \langle v, T(T^{-1}(w))\rangle = \langle v, w\rangle$ and so $(T^{-1})^*T^*$ is the identity, as required. □

Definition 5.8. *Let V be an inner product space and $T : V \to V$ a linear operator. Then we say*

i) *T is* **hermitian**, *or* **self-adjoint**, *if $T^* = T$.*
ii) *T is* **unitary** *if $T^*T = I$ and $TT^* = I$.*
iii) *T is* **normal** *if $T^*T = TT^*$.*

Example 5.9 (Example 5.5 continued)**.** The same definitions hold for matrices in general.

- The matrix B is hermitian.
- The matrix C, from our computation, is unitary.

To see how some of these properties interact, it is worthwhile proving each of the following.

Lemma 5.10. *If T is hermitian, then T is normal.*

Lemma 5.11. *If T is unitary, then T is normal.*

Example 5.12. We find an example of a matrix which is not normal (and so is also not hermitian or unitary). Let
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ so that } A^* = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$
from which we see, by direct computation, that $AA^* \neq A^*A$.

We will return now to the study of eigenvalues and eigenvectors and look at consequences of the above definitions for them. We start with hermitian operators.

Lemma 5.13. *Let $(V, \langle,\rangle)$ be an inner product space and $T : V \to V$ a hermitian linear operator. Then all of the eigenvalues of T are real valued.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of T, and let $v \in V_\lambda$ be an eigenvector with $\|v\| = 1$. Then $\lambda = \langle v, Tv\rangle$ because
$$\langle v, Tv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle = \lambda.$$
Now, using $T = T^*$, we see that $\lambda = \langle v, Tv\rangle = \langle T^*v, v\rangle = \langle Tv, v\rangle = \langle \lambda v, v\rangle = \bar{\lambda}\langle v, v\rangle = \bar{\lambda}$. □

We can also say something about the eigenvectors of a hermitian operator.

Proposition 5.14. *Let $(V, \langle,\rangle)$ be an inner product space and $T : V \to V$ a hermitian linear operator. Then eigenvectors with different eigenvalues are orthogonal, i.e., if $\lambda_1 \neq \lambda_2$, then*
$$E(\lambda_1) \perp E(\lambda_2).$$

Proof. Let $v_1 \in E(\lambda_1)$ and $v_2 \in E(\lambda_2)$, i.e,, $Tv_1 = \lambda_1 v_1$ and $Tv_2 = \lambda_2 v_2$. We consider $\langle v_1, Tv_2\rangle$. On the one hand, we have
$$\langle v_1, Tv_2\rangle = \langle v_1, \lambda_2 v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$
On the other hand, since $T^* = T$, we have
$$\langle v_1, Tv_2\rangle = \langle Tv_1, v_2\rangle = \langle \lambda_1 v_1, v_2\rangle = \bar{\lambda}_1\langle v_1, v_2\rangle.$$

But Lemma 5.13 states that $\bar{\lambda}_1 = \lambda_1$. So $\lambda_2\langle v_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle$ or

$$(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0,$$

and if $\lambda_1 \neq \lambda_2$ we must conclude that $\langle v_1, v_2\rangle = 0$. □

Note that the above shows, for hermitian matrices, a stronger property for the eigenvectors than the one we saw in Section 2 (that in general we only have that eigenvectors for different eigenvalues are linearly independent). We will soon show that this stronger property applies to all normal operators, and so also all unitary ones.

Example 5.15 (Example 5.5 continued)**.** We return to our matrix B, which was hermitian. This was also the matrix of Example 2.11, where we found the eigenvalues 1 and -1 and corresponding eigenvectors $v_1 = (1, i)$ and $v_2 = (i, 1)$. Hence $E(1) \perp E(-1)$, as expected.

Example 5.16. We will see some examples and properties of **orthogonal projections**. These are projections (meaning $P^2 = P$) that are also hermitian.

(a) Let $V = \mathbb{C}^2$ with the standard inner product. Then P_1 and P_2 are orthogonal projections.

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad P_2 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

(b) Let $w_0 \in V$ be a vector with $\|w_0\| = 1$. Then

$$P(v) := \langle w_0, v\rangle w_0$$

is an orthogonal projection. Both previous examples are special cases of this construction: if we choose $w_0 = (1, 0)$ then we get P_1, and if we choose $w_0 = \frac{1}{\sqrt{2}}(1, 1)$ then we get P_2.

We now investigate eigenvalues and eigenvectors for orthogonal projections. Given $Pv = \lambda v$, then by $P^2 = P$ we obtain $\lambda^2 v = \lambda v$ which gives $(\lambda^2 - \lambda)v = 0$. Hence

$$\lambda^2 = \lambda.$$

Therefore P can have as eigenvalues only 1 or 0. For the eigenspaces $E(0)$ and $E(1)$, we can also make progress. If $v \in E(0)$, then $Pv = 0$ and so $E(0) = \ker P$. If $v \in E(1)$, then $v = Pv$ and this means $v \in \operatorname{Im} P$. On the other hand, if $v \in \operatorname{Im} P$ then $v = Pv$ since $v = Pw$ for some $w \in V$ and then $Pv = P^2 w = Pw = v$. Hence $E(1) = \operatorname{Im} P$. We can apply the Rank Nullity Theorem to obtain that

$$V = E(0) \oplus E(1).$$

We now turn our attention to unitary operators.

Lemma 5.17. *Let V be an inner product space and $S, T : V \to V$ unitary operators. Then S^{-1}, ST, and S^* are all unitary.*

Proof. We use properties of the adjoint found in Lemma 5.7.

- From $SS^* = I$, we have $S^* = S^{-1} \Rightarrow I = (S^*)^{-1}S^{-1} \Rightarrow I = (S^{-1})^* S^{-1}$.
- Using $S^{**} = S$, we see S^* then satisfies $S^*(S^*)^* = I$.
- We compute that $ST(ST)^* = STT^*S^* = SS^* = I$. □

Lemma 5.18. *Let V be an inner product space and T be unitary. Then $\|Tv\| = \|v\|$ for any $v \in V$ and if λ is an eigenvalue of T, then $|\lambda| = 1$.*

Proof. We first note that $\|Tv\| = \langle Tv, Tv\rangle^{\frac{1}{2}} = \langle T^*Tv, v\rangle^{\frac{1}{2}} = \langle v, v\rangle^{\frac{1}{2}} = \|v\|$. Now, given an eigenvector $w \in V$ with corresponding eigenvalue λ, we see that $\|w\| = \|Tw\| = \|\lambda w\| = |\lambda|\|w\|$ and since $w \neq 0$, we must have that $|\lambda| = 1$. □

Remark 5.19. Combining Lemma 5.13 and Lemma 5.18 mean a matrix that is both unitary and hermitian can only have eigenvalues from $\{1, -1\}$.

Lemma 5.20. *Let $A \in M_n(\mathbb{C})$ be unitary, with $A = (c_1 \cdots c_n)$. Then $\{c_1, \ldots, c_n\}$ form an ONB for $\mathbb{C}^n$, with respect to the standard inner product.*

Proof. Since A is unitary, we have that $A^*A = I$, where I denotes the identity matrix. Then the ith row of A^* is given by $\bar{c}_1$, and the matrix equation states that $\bar{c}_i \cdot c_j = \delta_{ij}$, i.e., that $\{c_1, \ldots, c_n\}$ is an ONB for $\mathbb{C}^n$. □

Finally we move onto the most general case, of normal operators.

Lemma 5.21. *Let S be a normal operator and λ an eigenvalue of S. Then $S' := S - \lambda I$ is also normal.*

Proof. We start with just S and T as normal operators, and compute that

$$\begin{aligned}(S+T)(S+T)^*v &= (S+T)[S^*v + T^*v] \\ &= SS^*v + ST^*v + TS^*v + TT^*v \\ &= (SS^* + ST^* + TS^* + TT^*)v \text{ and} \\ (S+T)^*(S+T)v &= (S^*S + S^*T + T^*S + T^*T)v.\end{aligned}$$

If we set $T := -\lambda I$, we note that $T^* = -\bar{\lambda} I$. But then $S^*T = TS^*$ and $T^*S = ST^*$. Hence $(S+T)(S+T)^* = (S+T)^*(S+T)$ in this case. □

Proposition 5.22. *Let $(V, \langle , \rangle)$ be an inner product space and $T : V \to V$ be a normal operator. If v is an eigenvector of T with eigenvalue λ, then v is an eigenvector of T^* with eigenvalue $\bar{\lambda}$.*

Proof. Our aim is to show that $T^*v = \bar{\lambda}v$ for our eigenvector v corresponding to our eigenvalue λ. From Lemma 5.21 we know that

$$S := T - \lambda I$$

is normal. Using $SS^* = S^*S$ we find for an arbitrary $v \in V$

$$\|Sv\|^2 = \langle Sv, Sv\rangle = \langle v, S^*Sv\rangle = \langle v, SS^*v\rangle = \langle S^*v, S^*v\rangle = \|S^*v\|^2$$

and now if v is an eigenvector of T with eigenvalue λ, then $\|Sv\| = 0$ and so $\|S^*v\| = 0$ which means $S^*v = 0$. But since $S^* = T^* - \bar{\lambda} I$ this implies $T^*v = \bar{\lambda}v$. □

Let us show that Proposition 5.14 also applies to normal operators.

Proposition 5.23. *Let $(V, \langle , \rangle)$ be an inner product space and $T : V \to V$ a normal operator. If λ_1, λ_2 are eigenvalues of T with $\lambda_1 \neq \lambda_2$, then we have*

$$E(\lambda_1) \perp E(\lambda_2).$$

Proof. The proof is almost identical to the one for the hermitian case above, but now we use $T^*v_1 = \bar{\lambda}_1 v_1$. We consider $\langle v_1, Tv_2\rangle$, with $v_1 \in E(\lambda_1)$ and $v_2 \in E(\lambda_2)$. On the one hand,

$$\langle v_1, Tv_2\rangle = \langle v_1, \lambda_2 v_2\rangle = \lambda_2 \langle v_1, v_2\rangle.$$

On the other hand,

$$\langle v_1, Tv_2\rangle = \langle T^*v_1, v_2\rangle = \langle \bar{\lambda}_1 v_1, v_2\rangle = \lambda_1 \langle v_1, v_2\rangle.$$

Thus $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0$, and so $\langle v_1, v_2\rangle = 0$. □

We come now to the central result about normal operators, which will imply that they can be diagonalised. The proof below is a generalisation of the standard proof for hermitian operators, which we leave as a challenge to produce as a particular case of the following.

Theorem 5.24. *Let $(V, \langle , \rangle)$ be a finite dimensional complex inner product space and $T : V \to V$ a normal operator with* $\operatorname{spec} T = \{\lambda_1, \ldots, \lambda_k\}$. *Then*

$$V = E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_k).$$

Proof. We set $U = E(\lambda_1) + E(\lambda_2) + \ldots + E(\lambda_k)$, and note that these are direct sums by Proposition 2.23. We now wish to show that $V = U$, i.e., that V can be completely decomposed into eigenspaces of T (so that there is nothing left). Since $V = U \oplus U^\perp$ by Proposition 4.20, we will do this by showing that $U^\perp = \{0\}$.

Since eigenvectors of T are eigenvectors of T^*, too, we know that U is invariant under T^*, i.e., $T^*(U) \subset U$. But then $U^\perp$ is also invariant under T. To see this, consider $u \in U$ and $w \in U^\perp$.

Then $\langle Tw, u\rangle = \langle w, T^*u\rangle = 0$, because $T^*u \in U$, and since this is true for any $u \in U$ and $w \in U^\perp$ we get $T(U^\perp) \subset U^\perp$.

So if $U^\perp \neq \{0\}$, then the operator $T : U^\perp \to U^\perp$ must have at least one eigenvalue[14]. But then we have an eigenspace of T in $U^\perp$, when by assumption all the eigenspaces are in U. This is a contradiction, and hence $U^\perp = \{0\}$. □

We obtain the following as an immediate consequence of the previous two results.

Corollary 5.25. *Let $(V, \langle,\rangle)$ be a finite dimensional complex inner product space and $T : V \to V$ a normal operator. Then V has an orthonormal basis of eigenvectors of T.*

Proof. From the previous theorem we can write V as a direct sum of spaces $E(\lambda_i)$ where $i = 1, \ldots, k$. For each such space we choose an orthonormal basis. Now, since $E(\lambda_i) \perp E(\lambda_j)$ if $i \neq j$, the union of all these bases is an orthonormal basis of V consisting of eigenvectors of T. □

We have therefore determined a criteria for a linear operator to have of a basis of eigenvectors. Any normal operator, or in particular any hermitian and any unitary operator, has a basis of eigenvectors, and hence is diagonalisable. We can actually say a little more for matrices.

Theorem 5.26. *Let $A \in M_n(\mathbb{C})$ be a normal matrix. Then there exists a matrix $U \in M_n(\mathbb{C})$ such that*

$$U^{-1}AU = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, counted with multiplicity. Furthermore, we can choose U to be a unitary matrix (so that $U^{-1} = U^*$) where its columns form an orthonormal basis consisting of eigenvectors of A.

The advantages from this result are that we know that the given matrix A is diagonalisable and we only need to find U^* rather than U^{-1}.

Example 5.27 (Example 5.5 continued)**.** We return to our hermitian matrix

$$B = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

We already discussed in Example 2.11 that B has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$ with corresponding eigenvectors $v_1 = (1, \mathrm{i})$ and $v_2 = (\mathrm{i}, 1)$. If we now choose the normalised eigenvectors $\tilde{v}_1 = \frac{1}{\sqrt{2}}(1, \mathrm{i})$ and $\tilde{v}_2 = \frac{1}{\sqrt{2}}(\mathrm{i}, 1)$ for the columns of U, then the corresponding matrix

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix}$$

is unitary and diagonalises A:

$$U^*AU = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that the actual process of finding eigenvalues and eigenvectors for hermitian, unitary or in general normal matrices is identical to the examples discussed in Section 2. The only difference is that the eigenvectors are orthogonal, and if we choose *normalised* eigenvectors, then the change of basis matrix U we obtain, is unitary. The additional theory we developed does not really help us with the computational aspect, but it tells us in advance if it is worth starting the computation.

5.1. **Real matrices.** We have some new vocabulary for real matrices.

Definition 5.28. *Let $A \in M_n(\mathbb{R})$.*

- *If A is hermitian, then we call it* **symmetric**. *In particular, we have $A = A^t$.*
- *If A is unitary, then we call it* **orthogonal**. *In particular, we have $A^t = A^{-1}$.*

[14]Here we are using that our space is complex, so that the characteristic polynomial has at least one root in $\mathbb{C}$.

We have focused on complex matrices so far, because if we work over $\mathbb{C}$ then we always have n eigenvalues, including multiplicity. But many applications involve only real valued quantities and so we should like to work with real matrices. We now want to give one result about diagonalisation in that context. We look at hermitian matrices since we have seen these have only real eigenvalues.

Theorem 5.29. *Let $A \in M_n(\mathbb{R})$ be symmetric. Then there exists a matrix $O \in M_n(\mathbb{R})$ such that*

$$O^{-1}AO = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are the eigenvalues of A. Furthermore, we can choose O to be an orthogonal matrix where its columns form an orthonormal basis consisting of eigenvectors of A.

Proof. A real symmetric $n \times n$ matrix A has n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, counted with multiplicity. The corresponding eigenvectors are solutions to

$$(A - \lambda_i I)v_i = 0$$

but since this is a system of linear equations with real coefficients, the number of linearly independent solutions over $\mathbb{R}$ is the same as over $\mathbb{C}$. Thus we can choose $\dim E(\lambda_i)$ orthogonal eigenvectors with real components, and can find a orthonormal basis of real eigenvectors $v_1, \ldots v_n \in \mathbb{R}^n$ of A. Hence the matrix $O = [v_1 \ldots v_n]$ will diagonalise A. □

Lemma 5.30. *Let $O \in M_n(\mathbb{R})$ be an orthogonal matrix. Then the column vectors $v_1, \ldots, v_n$ of O satisfy $v_i \cdot v_j = \delta_{ij}$, i.e., they form an orthonormal basis.*

Lemma 5.31. *If O_1, O_2 are orthogonal matrices, then O_1O_2 and $O_1^{-1} = O_1^t$ are orthogonal, too.*

In other words, orthogonal matrices of a given size form a group with respect to matrix multiplication. We leave the proof as an exercise.

Example 5.32. We work with

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

This matrix has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$ and normalised eigenvectors $v_1 = \frac{1}{\sqrt{2}}(1, 1)$ and $v_2 = \frac{1}{\sqrt{2}}(1, -1)$. We therefore have that

$$O = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } O^tAO = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

An application of this result is the classification of quadratic forms. A function $g : \mathbb{R}^n \to \mathbb{R}$ is called a **quadratic form** if

$$g(x) = \frac{1}{2} x \cdot Qx$$

where $Q \in M_n(\mathbb{R})$ is a symmetric matrix. We want to find a simple representation of this function which allows us for instance to determine if $x = 0$ is a maximum or a minimum of $g(x)$, or neither of the two. By Theorem 5.29 there exist an orthogonal matrix O such that $O^tQO = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the operator $Q : \mathbb{R}^n \to \mathbb{R}^n$. So if we introduce new coordinates y by $y = O^t x$, or $x = Oy$, then

$$G(y) := g(Oy) = \frac{1}{2} y \cdot O^tQOy = \frac{1}{2}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2).$$

Hence the behaviour of the quadratic form is completely determined by the eigenvalues[15].

This is used for instance in the study of critical points of functions of several variables in multivariable calculus, because the so-called Hessian matrix composed of second partial derivatives of a sufficiently differentiable function of several variables is symmetric, hence diagonalisable.

[15] $x = 0$ is: a minimum if all positive, a maximum if all negative, and a generalised saddle point otherwise.

6. Vector spaces and subspaces

We now see the notion of a vector space, which generalises the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ that we have worked with so far. This constitutes the foundation for Linear Algebra within more advanced areas of modern mathematics. Our abstract exposition has this in mind.

6.1. **Some abstract algebra.** Mathematics is often a discipline of abstraction. This is a powerful tool in the mathematician's arsenal, since it allows us to prove results with one example in mind ($\mathbb{R}^n$, for instance) whilst actually proving results about a much more general class of examples (vector spaces, in our case). For linear algebra, we first need the notion of a *field* of numbers, often denoted by $\mathbb{F}$. Before defining a field, let us define what a *group* is. Groups arose in the study of transformations and symmetry and constitute the vast subject of group theory.

Definition 6.1. *A* ***group*** $(G, *)$ *consists of a nonempty set G and the binary operation $*$, called group multiplication, which assigns to each ordered pair $(g, h) \in G \times G$ an element $g * h \in G$. Furthermore, the following properties must hold.*

- *Closedness under group operation: if $f, g \in G$, then $f * g \in G$.*
- *Associativity: if $f, g, h \in G$, then $(f * g) * h = f * (g * h)$.*
- *Existence of a unique identity element, $e \in G$: if $g \in G$, then $e * g = g * e = g$.*
- *Existence of an inverse: for each $g \in G$, there exists an $h \in G$ such that $g * h = h * g = e$ (and in this case one denotes h as g^{-1}).*

A group where $g * h = h * g$ for every g, h is called **commutative** *or* **abelian**. *When a group G is abelian, one usually uses the notation $+$ instead of $*$, 0 instead of e, and $-f$ instead of f^{-1}.*

When the binary operation is understood, we often just write G for a group rather than $(G, *)$, and write gh rather than $g * h$.

Example 6.2. We see some examples of groups. In each case, consider which element is the identity and, for each element of the set, what the inverse might be.

(1) With the binary operation of addition, $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are abelian groups.
(2) With the binary operation of multiplication, $\mathbb{Q} \setminus \{0\}, \mathbb{R} \setminus \{0\}, \mathbb{C} \setminus \{0\}$ are abelian groups.
(3) The unit circle in $\mathbb{C}$, that is $\{z \in \mathbb{C} : |z| = 1\}$, with multiplication, is an abelian group.
(4) An important example of a non-commutative group is denoted as $GL(n, \mathbb{R})$, where GL stands for "general linear". This is the set of all $n \times n$ real invertible matrices, with $*$ being matrix multiplication.
(5) Given any $n \in \mathbb{N}$, the set of bijections of $\{1, \ldots, n\}$ with composition of functions forms a group, denoted S_n, the symmetric group. Consider why this group has $n!$ elements.

Many would argue that it is not natural to work with just one binary operation. With integers, matrices, functions, etc. we have two operations: addition and multiplication. This motivates dealing with abstract structures which involve two algebraic operations. Perhaps unsurprisingly, these operations are referred to as abstract addition and abstract multiplication. Of these structures, we skip the most general one, called a *ring* (motivated by polynomials with integer coefficients) and pass on to the most restrictive one, called a *field*. A field $\mathbb{F}$ is a set of numbers for which the *binary* operations of addition, subtraction, multiplication and division are defined and satisfy the usual rules. For our purposes, a list of examples is probably sufficient, but the axioms are also listed below. Such algebraic objects are studied in Algebra 2.

Definition 6.3. *A* ***field*** $(\mathbb{F}, +, \times)$ *consists of a set $\mathbb{F}$ of size at least two, and commutative binary operations $+$ and $\times$ (both from $\mathbb{F}$ to $\mathbb{F}$) such that*

- *$(\mathbb{F}, +)$ is an abelian group with identity element 0;*
- *$(\mathbb{F} \setminus \{0\}, \times)$ is an abelian group; and*
- *$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$ for all $\alpha, \beta, \gamma \in \mathbb{F}$.*

This final condition is often referred to as 'multiplication is distributive over addition'.

Note that the second condition precludes division by zero. From now on we will write $\mathbb{F}$ to denote a field, but you can think of it as just being $\mathbb{R}$ or $\mathbb{C}$ (the two most important cases for us).

Example 6.4. Let us see some examples of fields.

(1) The standard infinite fields are $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{Q}$, the set of complex, real, or rational numbers, with standard addition and multiplication.
(2) The sets $\mathbb{N}$ and $\mathbb{Z}$ are not fields, since in $\mathbb{N}$ one cannot subtract arbitrary numbers, and in $\mathbb{Z}$ one cannot divide by arbitrary numbers.
(3) Sets of the form $\mathbb{Q}[\mathrm{i}] := \{a + \mathrm{i}b \ , \ a, b \in \mathbb{Q}\}$ or $\mathbb{Q}[\sqrt{2}] := \{a + \sqrt{2}\, b \ , \ a, b \in \mathbb{Q}\}$ are examples of fields, and there many fields of this type which one obtains by extending the rational numbers by certain complex or real numbers. These are important in Number Theory.
(4) The key example of a *finite* field is the set $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$, where p is a prime number and addition and multiplication are done modulo p, called the *prime residue field*.[16] Finite fields are widely used, e.g., in Number Theory and Cryptography.

From these examples we may notice that $0 \times a = 0$ for every $a \in \mathbb{F}$. Let us show that this is in fact always the case.

Lemma 6.5. *Let $(\mathbb{F}, +, \cdot)$ be a field with additive identity 0. Then $a \cdot 0 = 0 \cdot a = 0$ for every $a \in \mathbb{F}$.*

Proof. Note that our definition of 0 is that $a + 0 = 0 + a = a$ for every $a \in \mathbb{F}$. Calling on the axiom that multiplication is distributive over addition, for any $a \in \mathbb{F}$ we have

$$a \cdot (0 + 0) = a \cdot 0 + a \cdot 0 \text{ and } a \cdot (0 + 0) = a \cdot (0) = a \cdot 0$$

from which $a \cdot 0 + a \cdot 0 = a \cdot 0$ becomes $a \cdot 0 = 0$ (we can add $-(a \cdot 0)$ to both sides since $a \cdot 0 \in \mathbb{F}$). □

The notion of a field generalises $\mathbb{R}$ and $\mathbb{C}$, and we note that the properties we have used for these sets are properties of arbitrary fields. Therefore almost all the results we have developed remain true if we replace $\mathbb{R}$ with a general field $\mathbb{F}$. In particular, we can define

$$\mathbb{F}^n := \{(x_1, x_2, \ldots, x_n) \ : \ x_1, x_2, \ldots, x_n \in \mathbb{F}\}$$

i.e., the space of vectors with n components given by elements of $\mathbb{F}$. We can also define

$$M_{m,n}(\mathbb{F}) := \{A = (a_{ij}) \ : \ a_{ij} \in \mathbb{F} \ \}$$

the set of matrices with elements in $\mathbb{F}$. Matrix multiplication only relies on addition and multiplication, and so applies equally well in this new setting. More precisely, the equation $Ax = y$ corresponds to

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \text{ for } i = 1, 2, \ldots, m$$

and the theory of systems of linear equations we developed before remains valid if we replace $\mathbb{R}$ by a general field $\mathbb{F}$. Thus we can consider the equation $Ax = b$ where the coefficients of A and entries of b are in $\mathbb{F}$, and the entries of x are sought in $\mathbb{F}$. Hence if $\mathbb{F} = \mathbb{Q}$, then we have rational coefficients and look for rational solutions only, whereas if $\mathbb{F} = \mathbb{C}$ then we allow everything to be complex. Since elementary row operations use only operations which are defined in every field $\mathbb{F}$, we can use the same methods for solving systems of linear equations. For completeness, we restate our results in this new language.

Theorem 6.6. *Let $Ax = b$ where $A \in M_{m,n}(\mathbb{F})$ and $b \in \mathbb{F}^m$ are known and $x \in \mathbb{F}^n$ is not. Let M be the row echelon form of the associated augmented matrix. Then*

(i) *the system has no solutions if and only if the last column of M contains a leading 1;*
(ii) *the system has a unique solution if every column except the last one of M contains a leading 1;*
(iii) *if, in addition, $\mathbb{F}$ has infinitely many elements, then the system has infinitely many solutions if the last column of M does not contain a leading 1 and there are less than n leading 1's. Then there are $n-k$ unknowns which can be chosen arbitrarily, where k is the number of leading 1 ' s of M.*

This leads to an immediate, useful, result.

[16] Exercise: show that $\mathbb{F}_p$ is, indeed, a field, and explain why p must be prime.

Corollary 6.7. *Let $A \in M_{m,n}(\mathbb{F})$ and assume that the only solution to $Ax = 0$ is $x = 0$. Then $m \geq n$, i.e., we need at least as many equations as unknowns to determine a unique solution.*

We end our discussion on solutions to equations with the following. We do not give the proof here, since it is identical to the case $\mathbb{F} = \mathbb{R}$.

Theorem 6.8. *Let $A \in M_n(\mathbb{F})$. Then the following are equivalent:*

(i) *A is invertible;*
(ii) *$\det A \neq 0$ in $\mathbb{F}$;*
(iii) *The rows of A are linearly independent over $\mathbb{F}$;*
(iv) *The columns of A are linearly independent over $\mathbb{F}$;*
(v) *The reduced row-echelon form of A is the identity matrix;*
(vi) *For any $b \in \mathbb{F}^n$, the a system of equations $Ax = b$ has a unique solution.*

6.2. **Formal definition of a vector space.** We are now ready to give a proper definition of a vector space. Intuitively this definition says that we have a set of objects that we can add together and can multiply by elements from $\mathbb{F}$, but let us see a formal definition.

Definition 6.9. *A **vector space over** $\mathbb{F}$ consists of an abelian group $(V, +)$, a field $(\mathbb{F}, +_{\mathbb{F}}, \cdot_{\mathbb{F}})$, and the notion of multiplication by $\mathbb{F}$ on V, denoted by placing an element of $\mathbb{F}$ on the left of an element of V. For all $v, w \in V$ and $\lambda, \mu \in \mathbb{F}$, this scalar multiplication must satisfy*

(1) $\lambda v \in V$
(2) $\lambda(v + w) = \lambda v + \lambda w$
(3) $(\lambda +_{\mathbb{F}} \mu)v = \lambda v + \mu v$
(4) $(\lambda \cdot_{\mathbb{F}} \mu)v = \lambda(\mu v)$
(5) $1_{\mathbb{F}} v = v$
(6) $0_{\mathbb{F}} v = 0_V$

where $0_{\mathbb{F}}$ and $1_{\mathbb{F}}$ denote the identity elements of the groups $(\mathbb{F}, +_{\mathbb{F}})$ and $(\mathbb{F} \setminus \{0_{\mathbb{F}}\}, \cdot_{\mathbb{F}})$, respectively. As with groups, we omit the subscripts (for $+_{\mathbb{F}}, \cdot_{\mathbb{F}}, 1_{\mathbb{F}}, 0_{\mathbb{F}}$, and 0_V) if the meaning is clear. Thus '$+$' denotes the binary operation for V and for $\mathbb{F}$: we did this when working with $\mathbb{R}^n$ over $\mathbb{R}$.

Elements of V are called *vectors* and elements of $\mathbb{F}$ are called *scalars*. We will now try to refrain using boldface for vectors, mostly assuming the convention that vectors are Latin u, v, w, etc., while scalars are Greek such as $\lambda, \mu, \epsilon, \upsilon, \rho, \eta, \kappa, \alpha$. With this convention, 0 denotes both the zero vector in V and the zero scalar in $\mathbb{F}$ (which are distinct things). One can find different but equivalent sets of axioms. In particular, the last property follows from the others.

Lemma 6.10. *Axiom (6) in Definition 6.9 is redundant (it follows from the other axioms).*

Proof. We have $v = 1v = (1+0)v = 1v + 0v = v + 0v$, and so $v = v + 0v$. Now add $-v$, the inverse to v in $(V, +)$, to both sides. This gives $0 = 0 + 0v = 0v$. □

Lemma 6.11. *Let V be a vector space over $\mathbb{F}$. Then $0 \in V$ is unique, and for each $v \in V$ the inverse element $-v \in V$ is also unique.*

Proof. (The proof is exactly that used for any group.) Assume there is another zero $0' \neq 0$. Then $0' + 0$ is equal simultaneously to $0'$ and 0, a contradiction. Now let $v \in V$ and assume there are distinct $u, w \in V$ with $v + w = v + u = 0$. Adding the element $-v$ to both sides, we get $u = w = -v$. □

Example 6.12. We note that a vector space can be finite. As an example, we will carefully check the axioms are satisfied for the vector space $V = (\mathbb{F}_2)^2$ over the field $\mathbb{F}_2$ of 2 elements. Here we visualise V as consisting of vectors of length two, where the entries are in $\mathbb{F}_2$ (and so are either 0 or 1).

$$(\mathbb{F}_2)^2 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Then the addition $+$ for V is component-wise, that is

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

where $a+c$ and $b+d$ are understood by the addition in $\mathbb{F}_2$: $0+0=0$, $0+1=1+0=1$, $1+1=0$. We can check that $(V,+)$ is an abelian group. This binary operation is closed, the identity is $(0,0)^t \in V$, each element is its own inverse (check this), and the addition is associative since it is associative component-wise in $\mathbb{F}_2$. We could actually check associativity directly, by checking for any $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{F}_2$ that

$$\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right) + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \left(\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \right).$$

The scalar multiplication is also thought of component-wise. In particular, for any $\lambda \in \mathbb{F}_2$ we have

$$\lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \end{pmatrix}$$

where λa_1 and λa_2 are computed using the multiplication in $\mathbb{F}_2$: $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$ and $1 \cdot 1 = 1$. We now go through each axiom of Definition 6.9 in turn. In general we would compute each side separately and see that they are equal. Let us say that $v = (a_1, a_2)^t$ and $w = (b_1, b_2)^t$ where $a_1, a_2, b_1, b_2 \in \mathbb{F}_2$. Note that checking (6) and (5) early on is often helpful.

(1) $\lambda v \in V$. Since multiplication in $\mathbb{F}_2$ is a binary operation, this means that λa_1 and λa_2 are in $\mathbb{F}_2$ and so $\lambda v \in V$ for any $\lambda \in \mathbb{F}_2$ and $v \in V$.

(6) $0_{\mathbb{F}} v = 0_V$. Although not technically needed in light of Lemma 6.10, we show this to simplify our later computations. We note that

$$0 \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \cdot a_1 \\ 0 \cdot a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

since $0 \cdot a = 0$ for every $a \in \mathbb{F}_2$.

(5) $1_{\mathbb{F}} v = v$. In a way akin to (6), we see that

$$1 \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \cdot a_1 \\ 1 \cdot a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

since $1 \cdot a = a$ for every $a \in \mathbb{F}_2$.

(2) $\lambda(v+w) = \lambda v + \lambda w$. The standard approach would be to find each side and note they are equal. In our case V contains only 4 elements and our scalar λ can only be from $\mathbb{F}_2 = \{0,1\}$. If $\lambda = 0$, then

$$(v+w) = \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right) = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \text{ where } a_1 + b_1, a_2 + b_2 \in \mathbb{F}_2$$

and so $\lambda(v+w) = (0,0)^t$. The same computation shows that $\lambda v = (0,0)^t$ and $\lambda w = (0,0)^t$ meaning $\lambda v + \lambda w = (0,0)^t + (0,0)^t = (0,0)^t = \lambda(v+w)$. If $\lambda = 1$ then the above computation shows that $\lambda(v+w) = (a_1 + b_1, a_2 + b_2)^t$. We then see that $\lambda v = (a_1, a_2)^t$ and $\lambda w = (b_1, b_2)^t$. Hence

$$\lambda v + \lambda w = (a_1, a_2)^t + (b_1, b_2)^t = (a_1 + b_1, a_2 + b_2)^t$$

which is the same expression we had for $\lambda(v+w)$.

(3) $(\lambda +_{\mathbb{F}} \mu)v = \lambda v + \mu v$. Again it is best to find each side, but to more concrete we work with $(\mathbb{F}_2)^2$ we find each possibility. Note $\lambda, \mu \in \{0,1\}$. Then the possibilities for the left hand side are

$$(0 +_{\mathbb{F}} 0)v = 0v = 0, (0 +_{\mathbb{F}} 1)v = 1v = v, (1 +_{\mathbb{F}} 0)v = 1v = v, \text{ and } (1 +_{\mathbb{F}} 1)v = (0)v = 0$$

and the corresponding possibilities for the right hand side are

$$(0)v + (0)v = 0 + 0 = 0, (0)v + (1)v = 0 + v = v, (1)v + (0)v = v + 0 = v, \text{ and } (1)v + (1)v = v + v = 0$$

where the last equation holds since $a + a = 0$ for any $a \in \mathbb{F}_2$.

(4) $(\lambda \cdot_{\mathbb{F}} \mu)v = \lambda(\mu v)$. We again consider all of the possibilities (rather than the general approach of finding an algebraic expression for each side). We note that if $0 \in \{\lambda, \mu\}$, then $\lambda \cdot_{\mathbb{F}} \mu = 0$. We

then note that the same behaviour occurs for the expression $\lambda(\mu v)$. Finally, $(1 \cdot_{\mathbb{F}} 1)v = 1v = v$ and $1(1v) = 1(v) = v$. Hence $(\lambda \cdot_{\mathbb{F}} \mu)v$ and $\lambda(\mu v)$ are equal.

Example 6.13. We see some more examples of vector spaces. It is worth writing down the addition and scalar multiplication in each case, and checking that the vector space axioms are satisfied.

(1) If we choose $V := \mathbb{R}^n$ and $\mathbb{F} := \mathbb{R}$ or $V := \mathbb{C}^n$ and $\mathbb{F} := \mathbb{C}$ then we obtain the objects $\mathbb{R}^n$ and $\mathbb{C}^n$ that we have been used to so far.
(2) The above generalises in the natural way. With $V := \mathbb{F}^n$, we obtain a vector space over $\mathbb{F}$. This is with component-wise addition of vectors and our usual scalar multiplication.
(3) Take $V = \mathbb{C}$ but now choose $\mathbb{F} = \mathbb{R}$. Then V is a vector space over $\mathbb{R}$.
(4) For $M_{m,n}(\mathbb{F}) := \{A = (a_{ij}) : a_{ij} \in \mathbb{F}\}$, we can define addition by $(a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij})$ and scalar multiplication by $\lambda(a_{ij}) := (\lambda a_{ij})$.

6.3. **Subspaces.** As in $\mathbb{R}^n$, we can look at subspaces of general vector spaces.

Definition 6.14. *Let V be a vector space over $\mathbb{F}$. A subset $U \subset V$ is called a* **subspace** *if U is a vector space over $\mathbb{F}$ with the addition and scalar multiplication induced by V.*

This is a natural definition, let us look at some examples.

Example 6.15 (Example 6.13 continued)**.** We will mostly continue with $\mathbb{R}^n$ and $\mathbb{C}^n$.

(1) For $V := \mathbb{R}^n$ over $\mathbb{R}$, the examples of linear subspaces are all subspaces. These were often expressed as spans, which we will see again soon.
(2) For $V := \mathbb{C}^2$ over $\mathbb{C}$, the set $\{\lambda(1, 0) \ : \ \lambda \in \mathbb{C}\}$ is a subspace of V.
(3) For $V := \mathbb{C}^2$ over $\mathbb{C}$, the set $\{\lambda(1, 0) \ : \ \lambda \in \mathbb{R}\}$ is **not** a subspace of V.
(4) For $V := \mathbb{C}^2$ over $\mathbb{R}$, the set $\{\lambda(1, 0) \ : \ \lambda \in \mathbb{R}\}$ is a subspace of V.
(5) For $V := M_2(\mathbb{C})$ over $\mathbb{C}$, we have that $\{\lambda I_2 \ : \ \lambda \in \mathbb{C}\}$ is a subspace of V.[17]

The drawback of this definition is that in order to check it we have to go through all the axioms for a vector space. Luckily, there's a simpler criterion, which comes about by noting which axioms automatically hold for an arbitrary subset.

Theorem 6.16 (The subspace test)**.** *Let V be a vector space over $\mathbb{F}$. A subset $U \subset V$ is a subspace if the following three conditions hold.*

(i) *That U is not empty, i.e. $U \neq \emptyset$.*
(ii) *That U is closed under addition: $u + u' \in U$ for all $u, u' \in U$.*
(iii) *That U is closed under multiplication by scalars: $\lambda u \in U$ for all $\lambda \in \mathbb{F}$ and $u \in U$.*

Proof. We need to show that U with the addition and scalar multiplication induced by V is a vector space, *i.e.*, it satisfies all the axioms in Definition 6.9. Since $U \neq \emptyset$, the first condition is fulfilled and there exists a $u \in U$. Axioms (2),(3), (4) and (5) are simply inherited from V. Next we check that $(U, +)$ is an abelian group. Closedness is followed from condition (ii), and associativity and commutativity are inherited from the abelian group $(V, +)$. For any $u \in U$, the element $(-1)u = -u$ is also in U, by (iii). To see that $-u$ is the inverse for u in $(U, +)$ we write $(-1)u + u = (-1 + 1)u = 0u$, since axiom (3) is fulfilled. We need to observe that $0u$ is the identity element of $(U, +)$, then existence of inverse and identity element is proved. As V is a vector space, by the last axiom in Definition 6.9 we have $0u = 0_V$, where 0_V denotes the identity element of $(V, +)$. Now, any element $u \in U$ is also an element in $(V, +)$, and we have $u + 0_V = 0_V + u = u$. So 0_V is also playing the role of the identity element in $(U, +)$. Now we have checked all the axioms of Definition 6.9 and the proof is complete. □

The above theorem gives us a simple test to check whether a given subset is a subspace. The conditions are also generally easy to check. We will see many uses of this test in the remainder of our course. The first applications are given below.

Lemma 6.17. *Let V be a vector space over $\mathbb{F}$ and $U = \{0\}$. Then U is a subspace of V.*

[17] Here, I_2 denotes the 2×2 identity matrix.

Lemma 6.18. *Let V be a vector space over $\mathbb{F}$ and $U, W \subset V$ be subspaces. Then $U \cap W$ is also a subspace of V.*

Proof. We use the subspace test.

- Non-empty: 0 is in both U and W, and hence in $U \cap W$.
- Closed under addition: take $v, v' \in U \cap W$, then $v, v' \in U$ and $v, v' \in W$. Hence $v + v' \in U$ and $v + v' \in W$ (because each of these are subspaces). Hence $v + v' \in U \cap W$.
- Closed under scalar multiplication: take $v \in U \cap W$ and $\lambda \in \mathbb{F}$. Then $\lambda v \in U$ and $\lambda v \in W$ (again because each of these are subspaces). Hence $\lambda v \in U \cap W$.

Hence $U \cap W$ satisfies all 3 conditions for the subspace test, and is a subspace of V over $\mathbb{F}$. □

We end this section with more examples of vector spaces and subspaces, and start by applying the subspace test to some specific subsets.

Example 6.19. Before looking at the below, all of the subsets given in the previous example can be checked using the subspace test.

(1) Let $V = \mathbb{F}^n$ over $\mathbb{F}$ and $U = \{(a_1, a_2, \ldots, a_n) \ : \ a_1 = 0\}$. Then U is non-empty, and closed under addition and scalar multiplication. Hence U is a subspace of V.
(2) Let $V = \mathbb{C}^n$ over $\mathbb{C}$ and $U = \{(a_1, a_2, \ldots, a_n) \ : \ a_1 = 1\}$. Then U is non-empty but fails to be closed under addition or scalar multiplication. Is this true if $\mathbb{C}$ is replaced by $\mathbb{F}_2$?
(3) Let $V = \mathbb{Q}^2$ over $\mathbb{Q}$ and $U = \mathbb{Z}^2 = \{(x, y) \in \mathbb{Q}^2 \ : \ x, y \in \mathbb{Z}\}$. Then U is non-empty and closed under addition, but not under scalar multiplication.
(4) Let $V = \mathbb{R}^2$ over $\mathbb{R}$ and $U = \{(x, 0) \in \mathbb{R}^2\} \cup \{(0, y) \in \mathbb{R}^2\}$. Then U is non-empty and closed under scalar multiplication, but not closed under addition.

Example 6.20. Consider the set of functions from the set $S := \{2, 4, 5\}$ to $\mathbb{R}$. This is a vector space V over $\mathbb{R}$, with addition given by $(f + g)(x) := f(x) + g(x)$ for all $x \in S$ and scalar multiplication given by $(\lambda f)(x) := \lambda f(x)$ for all $x \in S$. Consider the following:

- What is the zero element, i.e. 0, of V?
- Given a function $f \in V$, what is the inverse of f in $(V, +)$?
- What does a subspace consist of in this case? Can you find examples/non-examples?

Example 6.21. We end with a slightly more elaborate example. Let $V = \mathbb{P}(\mathbb{R})$ denote the set of all polynomials with real coefficients. This is a vector space over $\mathbb{R}$, with addition being the usual addition of polynomials, and scalar multiplication being defined as $\lambda \sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n} \lambda a_i x^i$. We now consider some subsets of V, and determine which of these are subspaces.

(1) Real polynomials of degree at most 3.
(2) Real polynomials of degree exactly 2.
(3) Real polynomials p with $p(1) = 0$.
(4) The union of $S_1 := \{p(x) = bx + a \ : \ a, b \in \mathbb{R}\}$ and $S_2 := \{p(x) = bx^2 + ax \ : \ a, b \in \mathbb{R}\}$.
(5) The intersection of S_1 and S_2. What does this look like?
(6) Polynomials in S_1 with the property that $p(1) = 0$. What do these look like?
(7) Real polynomials with cubic coefficient equal to their quadratic coefficient.
(8) Polynomials with coefficients in $\mathbb{Z}$. (How about coefficients in $\mathbb{Q}$?)
(9) Real polynomials with constant coefficient double that of their quadratic coefficient.
(10) Real polynomials whose coefficients sum to 0. (How about those which sum to 1?)
(11) Real polynomials p with at least 2 values a_1, $a_2 \in \mathbb{R}$ such that $p(a_1) = p(a_2) = 0$.
(12) Real monic polynomials (those with leading coefficient 1).

Example 6.22. Another vector space is given by $V = \{(a_j)_{j \in \mathbb{N}}, , a_j \in \mathbb{F}\}$, the set of infinite sequences of elements from $\mathbb{F}$, i.e., $(a_j)_{j \in \mathbb{N}}$ is a shorthand for the sequence $(a_1, a_2, a_3, a_4, \ldots)$ where the numbers a_j are chosen from $\mathbb{F}$. On V we can define

- addition: $(a_j)_{j \in \mathbb{N}} + (b_j)_{j \in \mathbb{N}} := (a_j + b_j)_{j \in \mathbb{N}}$
- scalar multiplication: $\lambda (a_j)_{j \in \mathbb{N}} := (\lambda a_j)_{j \in \mathbb{N}}$

which is similar to the case $\mathbb{F}^n$ and so V is often denoted as $\mathbb{F}^\infty$.

7. Spans, linear independence, and dimension

Another common way in which subspaces occur is by taking all the linear combinations of a given set of elements from V. The definition below is similar to the one we have seen previously for $\mathbb{R}^n$.

Definition 7.1. *Let V be a vector space over $\mathbb{F}$ and $S \subset V$ a subset.*

(i) *Then $v \in V$ is a* **linear combination** *of elements from S if $v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_k v_k$ for some $v_1, v_2, \ldots, v_k \in S$ and $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$.*

(ii) *The* **span** *of S, denoted* $\operatorname{span}(S)$*, is the set of all linear combinations of elements from S.*

The integer k that appears in part (i) can be an arbitrary number. If S is finite and has n elements, than it is natural to choose $k = n$. In the case where S contains infinitely many elements, we may be tempted to alter our definition. We will not! A linear combination *always* contains only a finite number of terms, and the span is *always* defined as the set of linear combinations with finitely many elements from S. The reason for this restriction is that for a general vector space we have no notion of convergence of infinite sums, so we simply cannot say what the meaning of an infinite sum would be[18]. The span of a subset is actually a subspace.

Lemma 7.2. *Let V be a vector space over $\mathbb{F}$ and $S \subset V$ a subset with $S \neq \emptyset$. Then* $\operatorname{span} S$ *is a subspace of V.*

Proof. We again use the subspace test.

- Since S is nonempty, there exists a $v \in S$ and so $v = 1v \in \operatorname{span} S$, and $\operatorname{span} S \neq \emptyset$.
- The sum of two linear combinations is again a linear combination, and so the set $\operatorname{span} S$ is closed under addition.
- Any multiple of a linear combination is again a linear combination, and so $\operatorname{span} S$ is closed under scalar multiplication.

Hence Theorem 6.16 states that $\operatorname{span} S$ is subspace. □

Example 7.3. We consider the examples from the last section, but expressed as spans.

(1) From Example 6.13, the vector space $M_2(\mathbb{C})$ can be expressed as

$$\operatorname{span}\left\{\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right\}.$$

(2) The subspace of $M_2(\mathbb{C})$ given in Example 6.15 was produced by considering a span:

$$\left\{\lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \;:\; \lambda \in \mathbb{C}\right\} = \operatorname{span}\left\{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right\}.$$

(3) The set of polynomials of degree at most n is naturally expressed as a span. Take

$$S_n := \{1, x, x^2, \ldots, x^n\},$$

of all simple powers ranging from 0 to n. Then $S_n \subset \mathbb{P}(\mathbb{R})$ and $\operatorname{span} S_n$ is a subspace of $\mathbb{P}(\mathbb{R})$ consisting of polynomials of degree at most n. Defining $\operatorname{span} S_n =: \mathbb{P}_n(\mathbb{R})$ appears logical in this instance.

(4) One may now wonder about a set that spans $\mathbb{P}(\mathbb{R})$, and whether this set even exists. We need an infinite set in this case (a finite set will be contained within $\operatorname{span} S_n$ for some $n \in \mathbb{N}$). A reasonable choice would be

$$S_\infty := \{x^n : n = 0, 1, 2, \ldots\} \subset \mathbb{P}(\mathbb{R}).$$

Notice that $P_\infty := \operatorname{span} S_\infty$ consists only of *finite* linear combinations of powers, i.e., $p(x) \in P_\infty$ if there exists a $k \in \mathbb{N}$ and $n_1, \ldots, n_k \in \mathbb{N}$, $p_1, \ldots, p_k \in \mathbb{F}$ such that

$$p(x) = \sum_{i=1}^{k} p_i x^{n_i}$$

[18] Ok, this is not quite true. When we have a norm on our vector space, we can drop this restriction and allow infinite linear combinations. We will mention an example of this later, but you need not dwell on this idea.

and so $P_\infty = \mathbb{P}(\mathbb{R})$ and we have that $\mathbb{P}(\mathbb{R})$ is therefore a vector space.

(5) Let $S := \{e_i \ : \ i \in \mathbb{N}\}$ where e_i denotes a 'vector' having entries indexed by $\mathbb{N}$ with $e_i(j) := \delta_{ij}$. We can think of these as elements of $\mathbb{R}^\infty$ from Example 6.22, that is, sequences consisting of real entries. Then $\operatorname{span} S$ is **not** equal to $\mathbb{R}^\infty$ but is rather a subspace, most easily stated as the set of sequences in $\mathbb{R}^\infty$ with only a *finite* number of nonzero entries. That is, for each element in $\operatorname{span} S$, there is some $N \in \mathbb{N}$ such that $a_n = 0$ for every $n \geq N$. This may actually remind us of the vector space $\mathbb{P}(\mathbb{R})$.

Following the same strategy as for subspaces in $\mathbb{R}^n$, we want to see if we can pick 'nice' subsets $\mathcal{B} \subset V$ such that $V = \operatorname{span} \mathcal{B}$ and $\mathcal{B}$ is in some sense optimal (is of minimal size). Such a set will be called a basis, and the size of the set will be called the dimension of V. This endevour of 'smallest possible' leads naturally to the notions of linear dependence and independence.

Definition 7.4. *Let V be a vector space over $\mathbb{F}$ and $S \subset V$.*

(a) *We say that S is* **linearly dependent** *if there exist distinct $v_1, v_2, \ldots, v_k \in S$ and $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ with $\{\lambda_1, \ldots, \lambda_k\} \neq \{0\}$ such that*

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_k v_k = 0.$$

(b) *We say that S is* **linearly independent** *if for any distinct $v_1, \ldots v_k \in S$ the equation*

$$\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$$

has only the trivial solution in $\mathbb{F}$, namely $\lambda_1 = \cdots = \lambda_k = 0$.

Linear dependence means that we can find a collection of vectors $v_1, \ldots, v_k$ in S and coefficients $\lambda_1, \ldots, \lambda_k \in \mathbb{F} \setminus \{0\}$ such that $\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$. This means in particular that

$$v_1 = \frac{-1}{\lambda_1}(\lambda_2 v_2 + \ldots + \lambda_k v_k). \tag{7.1}$$

Thus if $S' := S \backslash \{v_1\}$, then $\operatorname{span} S = \operatorname{span} S'$. So if S is linearly dependent, then one can find a smaller set which has the same span as S. This useful observation can be expressed as a lemma.

Lemma 7.5. *Let V be a vector space over $\mathbb{F}$ and $S \subset V$. Then S is linearly dependent if and only if there exists a $v \in S$ such that* $\operatorname{span} S = \operatorname{span}(S \backslash \{v\})$.

Proof. If $\operatorname{span} S = \operatorname{span}(S \backslash \{v\})$, then $v \in \operatorname{span}(S \backslash \{v\})$ meaning v is a linear combination in $S \backslash \{v\}$ and so S is linearly dependent. Now assume S is linearly dependent. Then, by (7.1), there is an element $v := v_1 \in S$ which can be written as a linear combination $v_1 = \mu_2 v_2 + \ldots + \mu_k v_k$ for some $v_2, \ldots v_k \in S \setminus \{v_1\}$. Now assume $y \in \operatorname{span} S$. Then y can be written as a linear combination of elements from S. If v_1 is not contained in this linear combination, then $y \in \operatorname{span}(S \backslash \{v_1\})$. If v_1 is contained in this linear combination, then

$$\begin{aligned} y &= \lambda_1 v_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n \\ &= \lambda_1(\mu_2 v_2 + \ldots + \mu_k v_k) + \lambda_2 w_2 + \ldots + \lambda_n w_n \\ &= (\lambda_1 \mu_2) v_2 + \ldots + (\lambda_1 \mu_k) v_k + \lambda_2 w_2 + \ldots + \lambda_n w_n \end{aligned}$$

where $(\lambda_1 \mu_2) v_2 + \ldots + (\lambda_1 \mu_k) v_k, \lambda_2 w_2 + \ldots + \lambda_n w_n \in \operatorname{span}(S \backslash \{v_1\})$. Hence $y \in \operatorname{span}(S \backslash \{v\})$ and $\operatorname{span} S = \operatorname{span}(S \backslash \{v\})$. □

Example 7.6. We consider some examples of linear dependence and independence.

(i) Let $V = \mathbb{C}^2$ over $\mathbb{C}$ and $v_1 = (1, 1)$ and $v_2 = (\mathrm{i}, \mathrm{i})$. Then $v_1 + \mathrm{i} v_2 = 0$, and so the set $S = \{v_1, v_2\}$ is linearly dependent. Note that the previous lemma therefore applies.

(ii) If we view $V = \mathbb{C}^2$ as a vector space over $\mathbb{R}$, then $v_1 = (1, 1)$ and $v_2 = (\mathrm{i}, \mathrm{i})$ are linearly independent, since in order that $\lambda_1 v_1 + \lambda_2 v_2 = 0$ we must have $\lambda_1 = -\mathrm{i}\lambda_2$ which is impossible for nonzero $\lambda_1, \lambda_2 \in \mathbb{R}$.[19]

(iii) The set $S_n = \{1, x, x^2, \ldots, x^n\}$ is linearly independent. We will show this in the exercises.

[19] The take home message here is that linear dependence or independence can depend on the field $\mathbb{F}$ we choose.

Other than determining linear dependence by observation, it can be computationally frustrating to determine whether a set is linearly independent (by solving simultaneous linear equations). Remark 3.13 is of great help here, since it says that we can merely compute the determinant of the change of basis matrix.

Example 7.7. For each set we decide on linear independence by computing the determinant.

(i) For $v_1 = (1, 2\mathrm{i}), v_2 = (-\mathrm{i}, 3) \in \mathbb{C}^2$ we find $\det \begin{pmatrix} 1 & -\mathrm{i} \\ 2\mathrm{i} & 3 \end{pmatrix} = 1$. Hence the vectors are linearly independent over $\mathbb{C}$.

(ii) For $v_1 = (1, -1, 3), v_2 = (2, 0, -1), v_3 = (-1, -2, 0) \in \mathbb{C}^3$ we find

$$\det \begin{pmatrix} 1 & 2 & -1 \\ -1 & 0 & -2 \\ 3 & -1 & 0 \end{pmatrix} = -15,$$

and so $\{v_1, v_2, v_3\}$ are linearly independent in $V = \mathbb{C}^3$ over $\mathbb{C}$.

(iii) Given $S \subset \mathbb{R}^n$ that is linearly independent, we automatically have that S is linearly independent in $V = \mathbb{C}^n$ over $\mathbb{C}$ (and also linearly independent in $V = \mathbb{C}^n$ over $\mathbb{R}$).

We are now ready for the definition of a basis.

Definition 7.8. *Let V be a vector space over $\mathbb{F}$. A subset $\mathcal{B} \subset V$ is called a **basis** of V if*

(i) *$\mathcal{B}$ spans V, i.e., $V = \operatorname{span} \mathcal{B}$; and*

(ii) *$\mathcal{B}$ is linearly independent.*

Example 7.9. Many of the sets we have seen are actually bases.

(i) A field $\mathbb{F}$ as a vector space over itself requires at least one element to span. In fact, any non-zero element of $\mathbb{F}$ is a basis.

(ii) Let $V = \mathbb{F}^n$, then $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$, with the jth entry of e_i being given by the Kronecker delta function δ_{ij}, is a basis. It is known as the **standard basis** of $\mathbb{F}^n$.

(iii) The set $S_n = \{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathbb{P}_n(\mathbb{R})$.

The proof of the following is almost identical to the one for $\mathbb{R}^n$ we saw previously.

Proposition 7.10. *Let V be a vector space over $\mathbb{F}$ and $\mathcal{B} \subset V$ a basis of V. Given $v \in V \setminus \{0\}$, there exist unique $v_1, v_2, \ldots, v_k \in \mathcal{B}$ and $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F} \setminus \{0\}$ such that*

$$v = \lambda_1 v_1 + \ldots + \lambda_k v_k.$$

Proof. The existence of such $v_1, \ldots, v_k$ and $\lambda_1, \ldots, \lambda_k$ follows from the fact that $\mathcal{B}$ spans V. We now wish to show that the combination is unique. For the specific $v \in V$ above, imagine that there exist $w_1, \ldots, w_n \in \mathcal{B}$ and $\mu_1, \ldots, \mu_n \in \mathbb{F} \setminus \{0\}$ such that $v = \mu_1 w_1 + \ldots + \mu_n w_n$. If $\{v_1, \ldots, v_k\} \cap \{w_1, \ldots, w_n\} = \emptyset$, then

$$\lambda_1 v_1 + \ldots + \lambda_k v_k = \mu_1 w_1 + \ldots + \mu_n w_n \Rightarrow \lambda_1 v_1 + \ldots + \lambda_k v_k + (-\mu_1) w_1 + \ldots + (-\mu_n) w_n = 0$$

which, because $\mathcal{B}$ is linearly independent, means $\{\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_n\} = \{0\}$. The same logic means that, if there is $w_i \notin \{v_1, \ldots, v_k\}$ or $v_j \notin \{w_1, \ldots, w_n\}$, then μ_i or λ_j must be zero. Hence $\{v_1, \ldots, v_k\} = \{w_1, \ldots, w_n\}$ and so $k = n$ and, after potentially reordering, we have that $w_i = v_i$ for $i = 1, \ldots, k$. Then

$$\lambda_1 v_1 + \ldots + \lambda_k v_k = \mu_1 v_1 + \ldots + \mu_n v_k \Rightarrow (\lambda_1 - \mu_1) v_1 + \ldots + (\lambda_k - \mu_k) v_k = 0$$

which is a linear combination equal to zero. Now, because $\mathcal{B}$ is a linearly independent set, we must have that $\lambda_i = \mu_i$ for $i = 1, \ldots, k$. □

Our main goal is now to show that if V has a basis $\mathcal{B}$ with finitely many elements, then any other basis of V will have the same number of elements. Hence the number of elements a basis contains is well defined and can be called the dimension of V. For this reason we restrict ourselves to the case of vector spaces which can be spanned by finite sets. We have seen vector spaces without this property: $\mathbb{P}(\mathbb{R})$ and $\mathbb{F}^\infty$.

Definition 7.11. *We call a vector space V over a field $\mathbb{F}$* **finite dimensional** *if there exists a set $S \subset V$ with $V = \operatorname{span} S$ and $|S| < \infty$. Otherwise, V is infinite dimensional.*

The following is not true in the infinite dimensional setting, unless we assume something known as the axiom of choice (which we will leave for now).

Proposition 7.12. *Let V be a vector space over $\mathbb{F}$ and $S \subset V$ a set with $|S| < \infty$ and $\operatorname{span} S = V$. Then S contains a basis of V. In particular, every finite dimensional vector space has a basis.*

Proof. We iterate Lemma 7.5 until we obtain a linearly independent set. If S is linearly independent, then S is already a basis and we are done. If S is linearly dependent, then by Lemma 7.5 there exists a $v_1 \in S$ such that $S_1 := S \backslash \{v_1\}$ spans V. Now, if S_1 is linearly independent, then it forms a basis. If S_1 is not linearly independent, we apply Lemma 7.5 again to obtain a smaller set S_2 which still spans V. Continuing this process we get a sequence of sets $S, S_1, S_2, \dots$ with $|S_{i+1}| = |S_i| - 1$, i.e., with strictly decreasing size, and since we started with a finite set S this sequence must stop. Hence at some step k the corresponding set S_k will be linearly independent and span V (and therefore be a basis of V). □

The next result shows that a linearly independent set cannot contain more element than a basis, and is our main tool to show that any two bases have the same number of elements.

Theorem 7.13. *Let V be a vector space over $\mathbb{F}$, $\mathcal{B} \subset V$ a basis with $|\mathcal{B}| = n \in \mathbb{N}$, and $S \subset V$ a linearly independent subset. Then $|S| \leq |\mathcal{B}|$.*

Proof. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ and assume we can choose distinct $w_1, \dots, w_{n+1} \in S$. Our approach will be to replace v_i with w_i for $i = 1, \dots, n$ so that w_{n+1} is a linear combination of $w_1, \dots, w_n$.[20]

Note that $w_1 \in V$. Hence $w_1 \in \operatorname{span} \mathcal{B}$, and so $w_1 = \sum \lambda_i v_i$. After reordering the v_i, we can assume that $\lambda_1 \neq 0$, and so rearrange $w_1 = \sum \lambda_i v_i$ to express v_1 as a linear combination of $\{w_1, v_2, \dots, v_n\} =: \mathcal{B}^{(1)}$. Thus $v_1 \in \operatorname{span} \mathcal{B}^{(1)}$, and $\mathcal{B}^{(1)}$ spans V. Thus $w_2 \in \mathcal{B}^{(1)}$, and so $w_2 = \lambda_1' w_1 + \sum_{i=2}^{n} \lambda_i' v_i$. If $\lambda_2', \dots, \lambda_n'$ are all zero, then w_2 is a multiple of w_1, contradicting that S is linearly independent. Hence we can again reorder the v_i to assume that $\lambda_2' \neq 0$. Thus $\mathcal{B}^{(2)} := \{w_1, w_2, v_3, \dots, v_n\}$ spans V and continuing in this way $\mathcal{B}^{(n)} := \{w_1, w_2, \dots, w_n\}$ spans V. Hence $w_{n+1} \in \operatorname{span} \mathcal{B}^{(n)}$, contradicting that S is linearly independent. □

Corollary 7.14. *Let V be a vector space over $\mathbb{F}$. If V has a basis with finitely many elements, then any other basis of V has the same number of elements.*

Proof. Let $\mathcal{B}, \mathcal{B}' \subset V$ be two bases of V. Since $\mathcal{B}'$ is linearly independent, we get $|\mathcal{B}'| \leq |\mathcal{B}|$. But reversing the roles of $\mathcal{B}$ and $\mathcal{B}'$ we get as well $|\mathcal{B}| \leq |\mathcal{B}'|$, and hence $|\mathcal{B}| = |\mathcal{B}'|$. □

An immediate consequence of the previous corollary is that our naive notion of dimension (being the size of a basis) is well defined.

Definition 7.15. *Let V be vector space and assume that V has a basis $\mathcal{B}$ with finitely many elements. Then we define the* ***dimension*** *of V as $|\mathcal{B}|$, and denote this by $\dim V$.*

Assuming the ultrafilter lemma[21], this definition can be used for infinite bases as well, i.e., we can define the dimension of a vector space to be the cardinality of a basis. We will not dwell on this, but rather mention, for interest, that the approach in the proof of Theorem 7.13 can be extended to the infinite dimensional setting, and then Corollary 7.14 immediately applies.

Example 7.16 (Example 7.9 continued.)**.** In fact, many sets we have seen are bases.

(i) We have that $\dim \mathbb{F}^n = n$, since $\mathcal{E} = \{e_1, e_2, \dots, e_n\}$ is a basis of $\mathbb{F}^n$.
(ii) We have $\dim M_{m,n}(\mathbb{F}) = mn$, since if we set $E_{kl} \in M_{m,n}(\mathbb{F})$ to have ijth entry equal to $\delta_{ik}\delta_{jl}$, then the set of E_{kl}, $k = 1, \dots, m$, $l = 1, \dots, n$, form a basis of $M_{m,n}(\mathbb{F})$.
(iii) With the basis $S_n = \{1, x, x^2, \dots, x^n\}$, we see that $\dim P_n(\mathbb{R}) = n + 1$.

[20] For this reason, the result is sometimes referred to as the Steinitz–Mac Lane exchange lemma (where exchange refers to how the result is proved, and Steinitz and Mac Lane proved the result in two different instances).

[21] A logically weaker assumption than the axiom of choice, which we leave for discussion by the logic/set theorists.

Lemma 7.17. *Assume $S \subset V$ is linearly independent and* $\operatorname{span} S \neq V$. *Given $v \in V \setminus \operatorname{span} S$, the set $S \cup \{v\}$ is linearly independent.*

Proof. Consider the linear combination

$$\lambda_1 v_1 + \ldots + \lambda_k v_k + \lambda v = 0$$

where $v_1, \ldots v_k \in S$. If $\lambda \neq 0$, then $v = -1/\lambda(\lambda_1 v_1 + \ldots + \lambda_k v_k) \in \operatorname{span} S$, which is a contradiction. Thus $\lambda = 0$. But the remaining vectors are in S and, since S is linearly independent, we must have that $\lambda_1 = \ldots = \lambda_k = 0$. □

Proposition 7.18. *Let V be a vector space over $\mathbb{F}$ and assume V is finite dimensional. Then any linearly independent subset $S \subset V$ can be extended to a basis $\mathcal{B}$.*

Proof. If $\operatorname{span} S = V$, then S is already a basis. If $\operatorname{span} S \neq V$, apply the previous lemma to extend S to $S^{(1)} := S \cup \{v\}$ where $v \in V \setminus \operatorname{span} S$. Continuing in this way, either $S^{(1)}$ spans V and so is a basis of V, or otherwise we can extend $S^{(1)}$ to $S^{(2)} := S^{(1)} \cup \{v'\}$ where $v' \in V \setminus \operatorname{span} S^{(1)}$. But note that the sets keep strictly increasing in size and are still linearly independent. Hence, by Theorem 7.13, the process has to stop. □

Let us summarise key properties for bases of finite dimensional vector spaces.

Corollary 7.19. *Let V be a vector space of finite dimension and let $S \subset V$.*

(i) *If S is linearly independent, then S has at most* $\dim V$ *elements.*
(ii) *If S spans V, then S has at least* $\dim V$ *elements.*
(iii) *If S is linearly independent and has* $\dim V$ *elements, then S is a basis of V.*
(iv) *If S spans V and has* $\dim V$ *elements, then S is a basis of V.*

Proof. We show each statement in turn.

(i) This is a rephrasing, with our new terminology of dimension, of Theorem 7.13.
(ii) Assume that $|S| < \dim V$. Applying Proposition 7.12, S contains $\mathcal{B}$, a basis for V. But then $|\mathcal{B}| \leq |S| < \dim V$, a contradiction.
(iii) Take S as linearly independent but not a basis for V. Then S cannot span V. Applying the previous proposition, we can extend S to a basis $\mathcal{B}$, where $|\mathcal{B}| > |S| = n$. This contradicts $\dim V = n$.
(iv) Since S spans V, Proposition 7.12 states that S contains a basis for V. Let us denote this specific basis by $\mathcal{B}$. Assuming $S \neq \mathcal{B}$ would mean $|\mathcal{B}| < n$, a contradiction. □

This corollary gives a simpler criterion to detect a basis than the original definition. If we know the dimension of V, then any set which has $\dim V$ elements and is either linearly independent or spans V must be a basis. This means, in the finite dimensional case, we only have to check one of the two conditions in the definition of a basis. We see some examples of this below.

Example 7.20 (Example 7.7 continued.)**.** We decide on linear independence to work out which sets are bases.

(i) Since $v_1 = (1, 2\mathrm{i})$ and $v_2 = (-\mathrm{i}, 3)$ are linearly independent, they form a basis of $\mathbb{C}^2$.
(ii) Similarly $v_1 = (1, -1, 3), v_2 = (2, 0, -1), v_3 = (-1, -2, 0) \in \mathbb{C}^3$ were linearly independent, and so they form a basis of $\mathbb{C}^3$.
(iii) For $V = \mathbb{P}_3(\mathbb{R})$ we had the basis $\mathcal{B} = \{1, x, x^2, x^3\}$. Consider the polynomials[22]

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x.$$

For the change of basis matrix we find the constants c_{ij} satisfying $T_j = \sum_i c_{ij} x^i$. Then

$$C = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

and since $\det C = 8$, the polynomials $\{T_0, T_1, T_2, T_3\}$ form a basis of $\mathbb{P}_3(\mathbb{R})$.

[22] These are the first four so-called Chebycheff polynomials.

Finally let us look at subspaces; the following appears quite natural.

Proposition 7.21. *Let V be a vector space over $\mathbb{F}$ with $\dim V = n$ and U a subspace of V.*

(i) *Then U is finite dimensional, and furthermore $\dim U \leq \dim V$.*
(ii) *If $\dim U = \dim V$, then $U = V$.*

Proof. We start with (i). Assume, for a contradiction, that $\dim U > n$. Then we can find

$$u_1 \in U \setminus \{0\}, u_2 \in U \setminus \operatorname{span}\{u_1\}, \ldots, u_{n+1} \in U \setminus \operatorname{span}\{u_1, \ldots, u_n\}.$$

Lemma 7.17 states that $u_1, \ldots, u_{n+1}$ are linearly independent in V, contradicting Corollary 7.19(i). For (ii), consider if we had a subspace U of V with $\dim U = \dim V = n$ but $U \neq V$. Hence we have a basis $\mathcal{B}_U := \{u_1, \ldots, u_n\}$ for U where $\mathcal{B}_U$ does not span V. Choose $v \in V \setminus \operatorname{span} \mathcal{B}_U$. By Lemma 7.17, $\mathcal{B}_U \cup \{v\}$ is a linear independent subset of V. This contradicts Corollary 7.19(i). □

Note that the definition of dimension depends on the field we consider. For $V = \mathbb{C}^2$ over $\mathbb{C}$, we have the standard basis e_1, e_2, meaning $\dim \mathbb{C}^2 = 2$. But we can view $\mathbb{C}^2$ as well as a vector space over $\mathbb{R}$. Then e_1, e_2 are now longer a basis, since linear combinations of e_1, e_2 with real coefficients give us only vectors in $\mathbb{R}^2 \subset \mathbb{C}^2$. In this case $e_1, \mathrm{i}e_1, e_2, \mathrm{i}e_2$ do form a basis, so as a vector space over $\mathbb{R}$ we have $\dim \mathbb{C}^2 = 4$. This dependence on the field $\mathbb{F}$ is sometimes emphasised by putting $\mathbb{F}$ as a subscript, i.e., $\dim_{\mathbb{F}} V$ is the dimension of V over $\mathbb{F}$. In our example we found

$$\dim_{\mathbb{C}} \mathbb{C}^2 = 2 \text{ and } \dim_{\mathbb{R}} \mathbb{C}^2 = 4.$$

The difference can be even more dramatic: viewing $\mathbb{R}$ as a vector space over $\mathbb{R}$ and over $\mathbb{Q}$ we get $\dim_{\mathbb{R}} \mathbb{R} = 1$ but $\dim_{\mathbb{Q}} \mathbb{R} = \infty$.

7.1. **Direct sums.** We have seen the notion of a direct sum for $\mathbb{R}^n$. An example to have in mind is $\mathbb{R}^3$ decomposed as the direct sum of the xy-plane and the z-axis[23]. As in other recent sections, we wish to formalise this idea for arbitrary vector spaces.

Definition 7.22. *Let V be a vector space over $\mathbb{F}$ and $U, W \subset V$ be subspaces. Then*

$$U + W := \{u + w \,;\, u \in U \,,\, w \in W\}$$

is the **sum** *of U and W. If we have $U \cap W = \{0\}$, then we use the notation $U \oplus W$ for the sum of U and W, and call this the* **direct sum** *of U and W.*

We see that the sum and direct sum are both subspaces in the exercises.

Lemma 7.23. *Let V be a vector space over $\mathbb{F}$ and $U, W \subset V$ be subspaces satisfying $U \cap W = \{0\}$. Then any $v \in U \oplus W$ has a unique decomposition $v = u + w$ with $u \in U$ and $w \in W$.*

Proof. By the definition of the sum of vector spaces there exists $u \in U$ and $w \in W$ such that $v = u + w$. To show that they are unique let us assume that $v = u' + w'$ with $u' \in U$ and $w' \in W$. Then $u + w = u' + w'$ and this gives $u - u' = w' - w$. But $u - u' \in U$ and $w - w' \in W$, and since $U \cap W = \{0\}$ we must have $u - u' = 0$ and $w - w' = 0$. Hence $u = u'$ and $w = w'$. □

Proposition 7.24. *Let V be a vector space over $\mathbb{F}$ and $U, W \subset V$ be finite dimensional subspaces satisfying $U \cap W = \{0\}$. Then*

$$\dim(U \oplus W) = \dim U + \dim W.$$

Proof. Let $\mathcal{B}_U$ and $\mathcal{B}_W$ be bases of U and W respectively. We consider the set $\mathcal{B}_U \cup \mathcal{B}_W$.

- We first see that $\operatorname{span} \mathcal{B}_U \cup \mathcal{B}_W = U \oplus W$. This follows since any $v \in U \oplus W$ can be written as $v = u + w$ and $u \in \operatorname{span} \mathcal{B}_U$ and $w \in \operatorname{span} \mathcal{B}_W$.
- Next, $\mathcal{B}_U \cup \mathcal{B}_W$ is linearly independent. Lemma 7.23 states, by the uniqueness of decomposition, that the only solution to $0 = u + w$ is if $u = 0$ and $w = 0$. Since $\mathcal{B}_U$ and $\mathcal{B}_W$ are linearly independent, the only way to get 0 as a linear combination is to choose all of the coefficients to be 0.

[23] In this case every vector in $\mathbb{R}^3$ can be uniquely represented as the sum of a horizontal vector in the xy-plane and a vector parallel to the z-axis.

Hence $\mathcal{B}_U \cup \mathcal{B}_W$ is a basis for $U \oplus W$. Using that $\mathcal{B}_U \cap \mathcal{B}_W = \emptyset$, we get $|\mathcal{B}_U \cup \mathcal{B}_W| = |\mathcal{B}_U| + |\mathcal{B}_W|$ and so $\dim U \oplus W = \dim U + \dim W$. □

We can generalise the above to obtain $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$.

Definition 7.25. *Let V be a vector space over $\mathbb{F}$, and U a subspace of V. Then a subspace W of V is called a* **complement** *of U in V if $V = U \oplus W$.*

An immediate question is whether a complement always exists. We answer this in the finite dimensional case (with the assumption that every finite dimensional vector space has a basis; we'll show this later).

Proposition 7.26. *Let V be a finite dimensional vector space over $\mathbb{F}$ and $U \subset V$ a proper subspace. Then there exists a subspace W which is a complement of U in V.*

Proof. Let $\mathcal{B}_U$ be a basis of U and let $\mathcal{B}_V$ be a basis of V with $\mathcal{B}_U \subset \mathcal{B}_V$. We claim that

$$W = \operatorname{span}\left(\mathcal{B}_V \setminus \mathcal{B}_U\right)$$

is a complement of U in V. By construction $V = U + W$, since $U + W$ contains a basis of V. Let $v \in U \cap W$. Then $v \in \operatorname{span} \mathcal{B}_U$ and also $v \in \operatorname{span}(\mathcal{B}_V \setminus \mathcal{B}_U)$. But $\mathcal{B}_V$ is a basis and so linearly independent, meaning $v = 0$. □

Example 7.27. We see some examples of complements. Example 7.20 is helpful for this.

(i) Let $V = \mathbb{R}^2$ and $U = \operatorname{span}\{v\}$ for some $v \in V \setminus \{0\}$. Then U is a line, and for any $v' \in V$ such that $\{v, v'\}$ form a basis of $\mathbb{R}^2$ we have $\mathbb{R}^2 = \operatorname{span}\{v\} \oplus \operatorname{span}\{v'\}$.

(ii) For $U = \operatorname{span}\{(\mathrm{i}, 1, \mathrm{i}), (0, \mathrm{i}, 1)\} \subset \mathbb{C}^3$ then $W = \operatorname{span}\{(1, 0, 0)\}$ is a complement since

$$\det \begin{pmatrix} \mathrm{i} & 0 & 1 \\ 1 & \mathrm{i} & 0 \\ \mathrm{i} & 1 & 0 \end{pmatrix} = 2$$

and therefore the vectors $\{(\mathrm{i}, 1, \mathrm{i}), (0, \mathrm{i}, 1), e_1\}$ form a basis of $\mathbb{C}^3$ over $\mathbb{C}$.

(iii) We can generalise (ii). Let $U \subset \mathbb{F}^n$ have the basis $v_1, v_2, \ldots, v_k$. Then to find a complement of U, we must find $v_{k+1}, \ldots, v_n \in \mathbb{F}^n$ such that $v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n$ form a basis of $\mathbb{F}^n$. With this setup, $W := \operatorname{span}\{v_{k+1}, \ldots, v_n\}$ is the complement of U in $\mathbb{F}^n$.

8. The Rank-Nullity Theorem and Isomorphisms

We first make some observations about linear maps.

Definition 8.1. *Let V and W be vector spaces over $\mathbb{F}$. A function $T : V \to W$ is called* **linear** *if*

(i) $T(u+v) = T(u) + T(v)$ *for all* $u, v \in V$.
(ii) $T(\lambda v) = \lambda T(v)$ *for all* $\lambda \in \mathbb{F}$ *and* $v \in V$.

Example 8.2. Let us look at some examples involving vector spaces over $\mathbb{R}$.

(i) Let $\lambda \in \mathbb{R}$. We can check that $T_\lambda : \mathbb{R}^n \to \mathbb{R}^n$, $v \mapsto \lambda v$ is linear.
(ii) Let $V = \mathbb{P}_n(\mathbb{R})$ and $W = \mathbb{P}_{n-1}(\mathbb{R})$, the spaces of polynomials of degree at most n and $n-1$ respectively. Let $D : V \to W$ be $D(p(x)) = p'(x)$, the derivative. Then D reduces the order by 1, and so the domain and codomain are suitable. It is also linear.
(iii) Let $q(x) := x^3 - x^2$. Then $M_q : \mathbb{P}_n(\mathbb{R}) \to \mathbb{P}_{n+3}(\mathbb{R})$, $p(x) \mapsto q(x)p(x)$ defines a linear map.
(iv) Let $\alpha \in \mathbb{R}$. Then $T_\alpha : \mathbb{P}_n(\mathbb{R}) \to \mathbb{P}_n(\mathbb{R})$, $p(x) \mapsto p(x+\alpha)$ is linear (called the *shift map*).
(v) Let $a \in \mathbb{R}$. Then $\delta_a : \mathbb{P}(\mathbb{R}) \to \mathbb{R}$, $p(x) \mapsto p(a)$ is linear (called the *evaluation map*).

We make some initial observations for Definition 8.1.

Lemma 8.3. *Let V be a vector space over $\mathbb{F}$ and $T : V \to W$ a linear map. Then*

(i) $T(0) = 0$.
(ii) $T(-v) = -T(v)$ *for all* $v \in V$.
(iii) *If* $v = \sum_{i=1}^k \lambda_i v_i$, *with* $v_i \in V$ *and* $\lambda_i \in \mathbb{F}$ *for* $i = 1, 2, \ldots, k$, *then* $T(v) = \sum_{i=1}^k \lambda_i T(v_i)$.

Proof. We prove each part separately, using the properties of a linear map.

(i) We use that $T(\lambda v) = \lambda T(v)$ for all $\lambda \in \mathbb{F}$ and $v \in V$. Setting $\lambda = 0$ gives the result.
(ii) Recall that $T(v+w) = T(v) + T(w)$ for all $v, w \in V$. Setting $w = -v$, we get
$$T(0) = T(v - v) = T(v) + T(-v) \Rightarrow 0 = T(v) + T(-v) \Rightarrow -T(v) = T(-v).$$
(iii) We have
$$T(v) = T\left(\sum_{i=1}^k \lambda_i v_i\right) = \sum_{i=1}^k T(\lambda_i v_i) = \sum_{i=1}^k \lambda_i T(v_i)$$
where properties of linear maps are applied at each stage. □

We now investigate the image of a subspace under a linear map.

Lemma 8.4. *Let V, W be vector spaces over $\mathbb{F}$, $T : V \to W$ a linear map, and $U \subset V$ a subspace. Then $T(U) = \{T(u) \,:\, u \in U\} \subset W$ is a subspace, too.*

Proof. We use the subspace test.

- We know that $0 \in U$. Then, by Lemma 8.3(i), we have that $0 \in T(U)$.
- Take $w, w' \in T(U)$. Then there exist $u, u' \in U$ such that $T(u) = w$ and $T(u') = w'$. Hence $w + w' = T(u) + T(u') = T(u + u')$ and so $w + w' \in T(u)$.
- Take $w \in T(U)$ and $\lambda \in \mathbb{F}$. Then $\lambda w = \lambda T(u) = T(\lambda u)$ and so $\lambda w \in T(U)$. □

Two important examples of subspaces are the following.

Definition 8.5. *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map.*

(i) *the* **kernel** *of T is defined as*
$$\ker T := \{v \in V \,:\, T(v) = 0\}.$$
(ii) *the* **image** *of T is defined as*
$$\operatorname{Im} T := \{w \in W \,:\, \text{there exists } v \in V \text{ with } T(v) = w\}.$$

Lemma 8.6. *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map. Then* $\ker T \subset V$ *and* $\operatorname{Im} T \subset W$ *are subspaces.*

Proof. Apply the subspace test (as we did previously with the specific case of $V = \mathbb{R}^n$ over $\mathbb{R}$). Note that $\operatorname{Im} T = T(V)$ and so for $\operatorname{Im} T$ we could also just apply the previous lemma. □

Example 8.7 (Example 8.2 continued)**.** We find the image and kernel in each case.

(i) We note $\ker(T_0) = V$ and $\mathrm{Im}(T_0) = \{0\}$. If $\lambda \neq 0$, then $\mathrm{Im}(T_\lambda) = V$ and $\ker(T_\lambda) = \{0\}$.
(ii) Here $\ker D = \mathbb{P}_0(\mathbb{R})$, the space of polynomial of degree 0 and $\mathrm{Im}\, D = \mathbb{P}_{n-1}(\mathbb{R})$.
(iii) Let $p(x) \in \mathbb{P}_n(\mathbb{R})$. To determine whether $q(x)p(x) = 0$ for all $x \in \mathbb{R}$, we look at the coefficients of $q(x)p(x)$. The only way for these to all be zero is if $p = 0$ (check this). So $\ker M_q = \{0\}$. But $\mathrm{Im}\, M_q$ is harder to compute; we will find the dimension of $\mathrm{Im}\, M_q$ later.
(iv) For the shift map we have $\ker(T_\alpha) = \{0\}$ and $\mathrm{Im}(T_\alpha) = \mathbb{P}_n(\mathbb{R})$ for every $\alpha \in \mathbb{R}$.
(v) For the evaluation map we have $\ker \delta_a = \{f : f(a) = 0\}$ and $\mathrm{Im}\, \delta_a = \mathbb{R}$ for every $a \in \mathbb{R}$.

We have used bases to put the notion of dimension on a firm ground. This allows us to state the Rank-Nullity Theorem for general vector spaces.

Definition 8.8. *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map. Then we define*

(i) *the* **rank** *of T as* $\mathrm{rank}\, T := \dim \mathrm{Im}\, T$
(ii) *the* **nullity** *of T as* $\mathrm{nullity}\, T := \dim \ker T$.

Example 8.9. We work over $\mathbb{R}$ and over $\mathbb{C}$.

(i) Let $T : \mathbb{C}^2 \to \mathbb{C}$ be defined by $T(z_1, z_2) = z_1 - z_2$. Then $T(z_1, z_2) = 0$ if $z_1 = z_2$, i.e., the kernel of T consists of multiples of $(1, 1)$, so $\mathrm{nullity}\, T = 1$. Since $T(z, 0) = z$ we have $\mathrm{Im}\, T = \mathbb{C}$ and so $\mathrm{rank}\, T = 1$. (Note that this changes if we work with $V = \mathbb{C}^2$ over $\mathbb{R}$.)
(ii) Let $T : \mathbb{C}^2 \to \mathbb{C}^3$ be defined by $T(z_1, z_2) = (z_1, z_2, z_1 - z_2)$. Then $T(z_1, z_2) = 0$ implies $z_1 = z_2 = 0$, so $\mathrm{nullity}\, T = 0$. Now $\mathrm{Im}\, T$ is spanned by $w_1 = (1, 0, 1)$ and $w_2 = (0, 1, -1)$, since $T(z_1, z_2) = z_1 w_1 + z_2 w_2$. Since w_1, w_2 are linearly independent, we find $\mathrm{rank}\, T = 2$.
(iii) For the derivative $D : \mathbb{P}_n(\mathbb{R}) \to \mathbb{P}_{n-1}(\mathbb{R})$, we get $\mathrm{nullity}\, D = 1$ and $\mathrm{rank}\, D = n$.

The rank and nullity determine some crucial properties of T. These are left as an exercise.

Proposition 8.10. *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map. Then*

(i) *T is injective if, and only if,* $\mathrm{nullity}\, T = 0$.
(ii) *T is surjective if, and only if,* $\mathrm{rank}\, T = \dim W$.
(iii) *T is bijective if, and only if,* $\mathrm{nullity}\, T = 0$ *and* $\mathrm{rank}\, T = \dim W$.

If we know a linear map is injective or surjective, this gives us information about the image of sets that are linearly independent or spanning.

Proposition 8.11. *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$ be a linear map.*

(i) *Assume $S \subset V$ is linearly independent and T is injective. Then $T(S) \subset W$ is linearly independent.*
(ii) *Assume $S \subset V$ spans V and T is surjective. Then $T(S)$ spans W.*

Proof. We prove each part in turn.

(i) Any element in $T(S)$ is of the form $w = T(v)$ for some $v \in S$. Hence to test linear independence, we have to see if we can find $v_1, \ldots, v_k \in S$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ such that $\sum_{i=1}^k \lambda_i T(v_i) = 0$. By Lemma 8.3(iii), $\sum_{i=1}^k \lambda_i T(v_i) = T(\sum_{i=1}^k \lambda_i v_i)$ meaning our condition $\sum_{i=1}^k \lambda_i T(v_i) = 0$ is satisfied if and only if $\sum_{i=1}^k \lambda_i v_i \in \ker T$. Calling on our assumption that T is injective, $\ker T = \{0\}$ means $\sum_{i=1}^k \lambda_i v_i = 0$, and since S is linear independent we must have $\lambda_1 = \cdots = \lambda_k = 0$. Hence $T(S)$ is linearly independent.
(ii) As T is surjective, given any $w \in W$ there exists a $v \in V$ such that $T(v) = w$. Since S spans V, we can find $v_1, \ldots, v_k \in S$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ such that $v = \sum_{i=1}^k \lambda_i v_i$. Hence $w = T(v) = \sum_{i=1}^k \lambda_i T(v_i) \in \mathrm{span}\{T(S)\}$ and $\mathrm{span}\{T(S)\} = W$. □

Corollary 8.12. *Let V, W be vector spaces over $\mathbb{F}$, $T : V \to W$ a linear map, and* $\dim V < \infty$. *If* $\mathrm{nullity}\, T = 0$, *then*

$$\mathrm{rank}\, T = \dim V.$$

Proof. Let $\mathcal{B}_V$ is a basis of V. The condition $\mathrm{nullity}\, T = 0$ means that T is injective. Then, from the previous result, $T(\mathcal{B}_V)$ is linearly independent and by construction $T(\mathcal{B}_V)$ spans $\mathrm{Im}\, T$. Therefore $T(\mathcal{B}_V)$ is a basis of $\mathrm{Im}\, T$ and $\mathrm{rank}\, T = \dim \mathrm{Im}\, T = |T(\mathcal{B}_V)| = |\mathcal{B}_V| = \dim V$. □

Example 8.13 (Example 8.2(iii) continued)**.** We had $M_q : \mathbb{P}_n(\mathbb{R}) \to \mathbb{P}_{n+3}(\mathbb{R})$, $p(x) \mapsto q(x)p(x)$ where $q(x) = x^3 - x^2$. We found $\ker M_q = \{0\}$ but $\operatorname{Im} M_q$ was harder to describe explicitly. We can gain information by applying the above result. Note $\operatorname{nullity} M_q = 0$ and $\dim \mathbb{P}_n(\mathbb{R}) = n + 1$, meaning $\operatorname{rank} M_q = n + 1$. Hence $\dim \operatorname{Im} M_q = n + 1$.

We can strengthen Corollary 8.12 to what is known as the Rank-Nullity Theorem.

Theorem 8.14. *Let V, W be vector spaces over $\mathbb{F}$, $T : V \to W$ a linear map, and $\dim V < \infty$. Then*

$$\operatorname{rank} T + \operatorname{nullity} T = \dim V.$$

Proof. We present two approaches, which are worthwhile filling in the details for.

Approach 1. Take a basis[24] $\{u_1, \ldots, u_k\}$ for $\ker T$, extend this to a basis for V by introducing the vectors $\{v_1, \ldots, v_m\}$, and then check that $\{T(v_1), \ldots, T(v_m)\}$ is a basis of size m for $\operatorname{Im} T$.

Approach 2. Let U be a complement of $\ker T$ in V. Hence $V = \ker T \oplus U$ and, by Proposition 7.24, we have $\dim V = \operatorname{nullity} T + \dim U$. Now any $v \in V$ can be written as $v = \tilde{v} + u$ with $\tilde{v} \in \ker T$ and $u \in U$. Hence $T(v) = T(u)$ and so $\operatorname{Im} T = T(U)$. But the restriction of T to U, $T|_U$, has $\operatorname{nullity} T|_U = 0$ and $\operatorname{rank} T|_U = \dim T(U) = \operatorname{rank} T$, and so by applying Corollary 8.12 to $T|_U$ we get $\dim U = \operatorname{rank} T|_U = \operatorname{rank} T$. □

The following is a useful application of the Rank-Nullity Theorem.

Corollary 8.15. *Let V, W be finite dimensional vector spaces over $\mathbb{F}$ and $T : V \to W$ a linear map.*

(i) *If $\dim W > \dim V$, then T is not surjective.*
(ii) *If $\dim W < \dim V$, then T is not injective.*
(iii) *If $\dim V = \dim W$, then T is surjective if and only if T is injective.*

Proof. Left as an exercise. (Proposition 8.10 is helpful here). □

Example 8.16. We will see some applications of the previous result.

(i) A linear function $f : \mathbb{R}^2 \to \mathbb{R}^3$ cannot be surjective.
(ii) A linear function $g : \mathbb{R}^3 \to \mathbb{R}^2$ cannot be injective, and so has a non-trivial kernel.
(iii) A linear function $h : \mathbb{R}^3 \to \mathbb{R}^3$ with $h(e_1), h(e_2), h(e_3) \in \operatorname{span}\{e_1, e_2\}$ is clearly not surjective (no vector is sent to e_3) and so cannot be injective (meaning it has a non-trivial kernel). Thus Corollary 8.15(iii) means for a linear function $T : V \to V$ that checking one of injectivity or surjectivity determines the other one.

The word *isomorphism* in mathematics usually refers to *a map that preserves the desired structure*. In this course, linearity is significant, and isomorphic linear spaces are somehow "equal" in linear algebra. Similar words will occur in other areas of mathematics, for example, a homeomorphism preserves the topology, a diffeomorphism preserves the differential structure, a group isomorphism preserves the structure of a group. In all of these cases, an "isomorphism" is a special kind of bijective map.

Definition 8.17. *Let V, W be vector spaces over $\mathbb{F}$. A linear map $T : V \to W$ that is bijective is called an* **isomorphism***. Two vector spaces V, W over $\mathbb{F}$ are then called* **isomorphic***, denoted $V \cong W$, if there exists an isomorphism $T : V \to W$.*

Example 8.18. We work with vector spaces over $\mathbb{R}$.

(i) Let $V = \mathbb{R}^2$, $W = \mathbb{C}$, and $T(x, y) := x + \mathrm{i}y$. Then T is linear and bijective, and so an isomorphism. Hence $\mathbb{C}$ and $\mathbb{R}^2$ are isomorphic as vector spaces over $\mathbb{R}$.
(ii) Let $V = \mathbb{R}^{n+1}$ and $W = \mathbb{P}_n(\mathbb{R})$. Define $T : \mathbb{R}^{n+1} \to \mathbb{P}_n(\mathbb{R})$ by

$$T(a_n, a_{n-1}, \ldots, a_1, a_0) \mapsto a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0.$$

This is an isomorphism, and so $\mathbb{P}_n(\mathbb{R})$ is isomorphic to $\mathbb{R}^{n+1}$.

[24] A finite basis for $\ker T$ exists by Proposition 7.21.

Note that $\mathbb{P}_n(\mathbb{R})$ and $\mathbb{R}^{n+1}$ 'feel' like distinct structures. We can think of them as being equal only if we strip them from all other properties except the ones related to addition and scalar multiplication. With many mathematical objects, determining whether they are isomorphic can be a real challenge. But isomorphic vector spaces must have the same dimension.

Proposition 8.19. *Let V, W be vector spaces over $\mathbb{F}$ which are isomorphic and let $\dim V = n \in \mathbb{N}$. Then $\dim V = \dim W$.*

Proof. Let $T : V \to W$ be an isomorphism. We show two approaches.

Approach 1. Apply the Rank-Nullity Theorem so that $\operatorname{rank} T + \operatorname{nullity} T = \dim V$. But T is bijective, and so injective and surjective. Using Proposition 8.10, we see that $\operatorname{rank} T = \dim W$ and $\operatorname{nullity} T = 0$. Hence $\dim W = \dim V$.

Approach 2. Let $\mathcal{A}$ be a basis of V, and set $\mathcal{B} := T(\mathcal{A}) \subset W$. We now apply Proposition 8.11. Since T is injective, $\mathcal{B}$ is linearly independent. Since T is surjective, $\operatorname{span} \mathcal{B} = W$. Hence $\mathcal{B}$ is a basis of W. But $\mathcal{B}$ has the same number of elements as $\mathcal{A}$, and therefore $\dim V = \dim W$. □

Remark 8.20. When reading these proofs, we may feel that the first is more straightforward. The reason for the second approach is that it does not appeal to the fact that V is finite dimensional. It does, however, ask that we have a basis for V. The assumption that every vector space has a basis is equivalent to *The Axiom of Choice*; courses on logic and set theory will discuss this.

The above theorem means that, in order to be isomorphic, the vector spaces must have the same dimension. Perhaps surprisingly, the inverse of this result is as well true: whenever two vector spaces have the same dimension, over the same field, then they are isomorphic. To prove this we introduce a specific linear map.

Definition 8.21. *Let V and W be vector spaces over $\mathbb{F}$ and $\dim V = \dim W = n$. Given bases $\mathcal{A} = \{v_1, \ldots, v_n\} \subset V$ and $\mathcal{B} = \{w_1, \ldots, w_n\} \subset W$, let*

$$T_{\mathcal{BA}}(x_1 v_1 + \ldots + x_n v_n) := x_1 w_1 + \ldots + x_n w_n$$

for all $x_1, \ldots, x_n \in \mathbb{F}$. Note that $T_{\mathcal{BA}}$ is well defined and linear since $\mathcal{A}$ is a basis for V.

Lemma 8.22. *Let V and W be vector spaces over $\mathbb{F}$ with bases $\mathcal{A}$ and $\mathcal{B}$ respectively. Then $T_{\mathcal{BA}}$ is an isomorphism.*

Proof. From the definition we see that $\operatorname{Im} T_{\mathcal{BA}} = \operatorname{span} \mathcal{B} = W$, since on the right hand side all linear combinations of vectors from the basis $\mathcal{B}$ appear if we vary $x_1, \ldots, x_n$. Hence $T_{\mathcal{BA}}$ is surjective and so Corollary 8.15(iii) states $T_{\mathcal{BA}}$ is bijective (and therefore an isomorphism). □

Theorem 8.23. *Let V, W be vector spaces over $\mathbb{F}$ with $\dim V = \dim W = n \in \mathbb{N}$. Then $V \cong W$.*

Proof. We note that V and W have bases $\mathcal{A} = \{v_1, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, \ldots, w_n\}$ respectively. Hence the function $T_{\mathcal{BA}}$ provides an isomorphism from V to W. □

Assuming the Axiom of Choice, Theorem 8.23 generalises to vector spaces of arbitrary dimension.

Example 8.24. We recall some observations from Example 7.16.

(i) Since $\dim M_{m,n}(\mathbb{F}) = mn$, we have that $M_{m,n}(\mathbb{F}) \cong \mathbb{F}^{mn}$.
(ii) The space $\mathbb{P}(\mathbb{R})$ is not isomorphic to $\mathbb{R}^n$ for any $n \in \mathbb{N}$, since $\mathbb{P}(\mathbb{R})$ is not finite dimensional over $\mathbb{R}$ whereas $\dim(\mathbb{R}^n) = n$.
(iii) Recall the subspace of real sequences from Example 7.3, given by $V = \operatorname{span}\{e_i \,:\, i \in \mathbb{N}\}$ where each e_i is a vector indexed by $\mathbb{N}$ with $e_i(j) := \delta_{ij}$. The set $\{e_i \,:\, i \in \mathbb{N}\}$ is linearly independent, and so a basis, meaning $\dim V$ is countably infinite. We also found that $\mathbb{P}(\mathbb{R}) = \operatorname{span}\{x^i \,:\, i \in \mathbb{N} \cup \{0\}\}$ was of countably infinite dimension. Hence $V \cong \mathbb{P}(\mathbb{R})$.

We provide a convenient condition to check whether a linear function is an isomorphism. Note that $\det T$ is well defined, since it is independent from our choice of finite basis.

Lemma 8.25. *Let V and W be vector spaces over $\mathbb{F}$ and $\dim V = \dim W = n$. Then a linear operator $T : V \to W$ is an isomorphism if, and only if, $\det T \neq 0$.*

9. Spaces of functions

We now consider vector spaces of functions, and see what isomorphisms can tell us about these.

Definition 9.1. *Let V, W be vector spaces over $\mathbb{F}$. Then $L(V, W)$ denotes the set of linear maps from V to W.*

Proposition 9.2. *Let V, W be vector spaces over $\mathbb{F}$. If $f \in L(V, W)$ is a bijection, then f^{-1} exists and is in $L(W, V)$.*

Proof. That f^{-1} exists follows immediately from f being a bijection. Take $w, w' \in W$. Thus there exist $v, v' \in V$ such that $f(v) = w$ and $f(v') = w'$. Hence

$$f^{-1}(w + w') = f^{-1}(f(v) + f(v')) = f^{-1}(f(v + v')) = v + v' = f^{-1}(w) + f^{-1}(w')$$

as required. □

The following construction is the source of many examples of vector spaces.

Definition 9.3. *Let V be a vector space over $\mathbb{F}$ and S a non-empty set. Then $F(S, V)$ denotes the set of all functions from S to V. On $F(S, V)$ we have a natural addition and scalar multiplication defined, for every $f, g \in F(S, V)$ and $\lambda \in \mathbb{F}$, by*

- $(f + g)(s) := f(s) + g(s)$*; and*
- $(\lambda f)(s) := \lambda f(s)$ *for all $s \in S$.*

We have that $f(s), g(s) \in V$, and hence they can be added together and also multiplied by elements from $\mathbb{F}$. We say $f = g$ for $f, g \in F(S, V)$ if $f(s) = g(s)$ for all $s \in S$.

It is useful to highlight a subtlety of the notion that $f = g$ for functions $f, g \in F(S, V)$.

Example 9.4. We now see why we have not looked at $\mathbb{P}(\mathbb{F})$ for any finite field $\mathbb{F}$.

(i) Given $p, q \in \mathbb{P}(\mathbb{R})$, we have that distinct coefficients yield distinct functions[25].
(ii) In $\P(\mathbb{F}_2)$, we have that $p(x) := x$ and $q(x) := x^2$ are the same function.

Proposition 9.5. *Let V be a vector space over the field $\mathbb{F}$ and S a non-empty set. Then $F(S, V)$ is a vector space over $\mathbb{F}$.*

Proof. We go through each vector space axiom (from Definition 6.9) in turn.

- Closed under addition: given $f, g \in F(S, V)$ we have $f + g \in F(S, V)$, since the sum of two functions is a function from S to V.
- Commutativity of addition: note $(f+g)(s) = f(s)+_V g(s) = g(s)+_V f(s)$ as $f(s), g(s) \in V$ and V is a vector space by assumption. Hence $f + g = g + f$ as elements in $F(S, V)$.
- The identity element of $F(S, V)$ is the zero function which maps all of S to 0_V.
- The inverse of f is the function $-f$ defined by $(-f)(s) := -f(s)$ for all $s \in U$, where we use that each $v \in V$ has an additive inverse.
- One has $(f+(g+h))(u) = f(u)+(g+h)(u) = f(u)+(g(u)+h(u))$ which using associativity in V gives $f(u)+(g(u)+h(u)) = (f(u)+g(u))+h(u) = (f+g)(u)+h(u) = ((f+g)+h)(u)$.

Hence $F(S, V)$ is an abelian group with respect to usual addition of functions. Then

- $(\lambda(f + g))(s) = \lambda((f + g)(s)) = \lambda(f(s) + g(s)) = \lambda f(s) + \lambda g(s) = (\lambda f)(s) + (\lambda g)(s)$.
- $((\lambda + \mu)f)(s) = (\lambda + \mu)f(s) = \lambda f(s) + \mu f(s) = (\lambda f)(s) + (\mu f)(s)$.
- $((\lambda\mu)f)(s) = (\lambda\mu)f(s) = \lambda(\mu f(s)) = \lambda(\mu f)(s) = (\lambda(\mu f))(s)$.
- $(1f)(s) = 1f(s) = f(s)$. (Also $(0f)(s) = 0f(s) = 0$, but Lemma 6.10 states this.) □

Example 9.6. We see some specific cases for the above construction.

(i) We have that $F(\mathbb{R}, \mathbb{R})$, the set of real valued functions, is a vector space over $\mathbb{R}$.
(ii) With $\mathbb{C}$ as a vector space over $\mathbb{C}$ we have that $F(\mathbb{R}, \mathbb{C})$, the set of complex valued functions, is a vector space over $\mathbb{C}$.
(iii) Considering $\mathbb{C}$ as a vector space over $\mathbb{R}$ gives us $F(\mathbb{R}, \mathbb{C})$ as a vector space over $\mathbb{R}$.

[25] One approach is to note that $p, q \in \mathbb{P}_n(\mathbb{R})$ for some $n \in \mathbb{N}$, and so $p(x) - q(x)$ can have at most n roots.

Example 9.7. We can also construct vector spaces that we have already seen.

(i) Let $S = \{1, 2, \ldots, n\}$ and $V = \mathbb{F}$. Then $F(S, \mathbb{F})$ consists of functions $f : \{1, 2, \ldots, n\} \to \mathbb{F}$. Such a function is completely determined by the values it takes on the first n integers, i.e., by the list $(f(1), f(2), \ldots, f(n))$. But this is an element in $\mathbb{F}^n$, and since the functions can take arbitrary values we find $F(S, \mathbb{F}) \cong \mathbb{F}^n$. Another approach is that both are vector spaces over $\mathbb{F}$ of dimension n.

(ii) If $S = \mathbb{N}$ and $V = \mathbb{F}$, then an element in $F(\mathbb{N}, \mathbb{F})$ is a function $f : \mathbb{N} \to \mathbb{F}$ which is defined by the list of values it takes on all of the positive integers

$$(f(1), f(2), f(3), \ldots, f(k), \ldots)$$

which is nothing but an infinite sequence. Hence $F(\mathbb{N}, \mathbb{F}) \cong \mathbb{F}^\infty$.

(iii) Let $S = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} = \{(i, j)\,;\, i = 1, 2, \ldots, m\,,\, j = 1, 2, \ldots, n\}$ and $V = \mathbb{F}$. Then $F(S, \mathbb{F}) \cong M_{m,n}(\mathbb{F})$, the set of $m \times n$ matrices with elements in $\mathbb{F}$.

Lemma 9.8. *Let V, W be vector spaces over $\mathbb{F}$. Then $L(V, W)$ is a vector space over $\mathbb{F}$.*

Proof. We have $L(V, W) \subset F(V, W)$, and so can apply the subspace test.

(i) Let $f_0(v) := 0$ for all $v \in V$. Then f_0 is a linear map, and so $L(V, W) \neq \emptyset$.

(ii) We wish to show that the sum of linear functions is a linear function. Take $f, g \in L(V, W)$. Then $(f+g)(u+v) = f(u+v)+g(u+v) = f(u)+f(v)+g(u)+g(v) = (f+g)(u)+(f+g)(v)$ and $(f + g)(\lambda v) = f(\lambda v) + g(\lambda v) = \lambda f(v) + \lambda g(v) = \lambda(f + g)(v)$.

(iii) We wish to show, for any $\mu \in \mathbb{F}$ and $f \in L(V, W)$, that (μf) is a linear function. We have that $(\mu f)(u + v) = \mu(f(u + v)) = \mu(f(u) + f(v)) = \mu f(u) + \mu f(v) = (\mu f)(u) + (\mu f)(v)$. Similarly $(\mu f)(\lambda v) = \mu f(\lambda v) = \mu\lambda f(v) = \lambda\mu f(v) = \lambda(\mu f)(v)$. □

With what we have seen with isomorphisms, we may keenly ask whether $L(V, W)$ is a vector space that we have already seen. If it is finite dimensional, then Theorem 8.23, it will be isomorphic to $\mathbb{F}^k$ for some $k \in \mathbb{N}$. Using the connection between linear maps and matrices allows us to completely answer this question in the case where V and W are finite dimensional.

Proposition 9.9. *Let V and W be vector spaces over $\mathbb{F}$ and $\dim V = n$, $\dim W = m$. Then $\dim L(V, W) = m \times n$. We therefore have that $L(V, W) \cong \mathbb{F}^{m\times n}$.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be bases for V and W respectively. From our earlier work, we have that for each $f \in L(V, W)$ there exists a unique matrix $A_f := M_{\mathcal{B}\mathcal{A}}(f) \in M_{m,n}(\mathbb{F})$ that represents f. This allows us to define a bijection $\Psi : L(V, W) \to M_{m,n}(\mathbb{F})$, $f \mapsto A_f$. From direct computation[26], we have that $A_{f+g} = A_f + A_g$ and $A_{\lambda f} = \lambda A_f$. Thus

$$\Psi(f + g) = \Psi(f) + \Psi(g) \text{ and } \Psi(\lambda f) = \lambda\Psi(f)$$

meaning Ψ is linear, and so an isomorphism. Hence $L(V, W) \cong M_{m,n}(\mathbb{F})$. In Example 8.24(i) we showed that $\dim M_{m,n}(\mathbb{F}) = m \times n$. Another (somewhat similar) approach would be to find the dimension of $L(V, W)$ in order to conclude the isomorphism. □

We end by extending some of the ideas from examples introduced in this section.

Example 9.10. Examples (iii) and (iv) are not examinable, but included for interest.

(i) Let U be a subset of $M_{m,n}(\mathbb{F})$ that is closed under addition and also scalar multiplication by $\mathbb{F}$. Then U is a subspace of $M_{m,n}(\mathbb{F})$, and $\dim U \in \{0, \ldots, m \times n\}$.

(ii) Let $S = \{1, \ldots, n\}$ and W be a vector space over $\mathbb{F}$ of dimension m. Then, extending Example 9.7(i), $\dim F(S, W) = n \times m$. Thus, given a vector space V over $\mathbb{F}$ of dimension n, we have that $L(V, W) \cong F(S, W)$.

(iii) As a more involved example, we find the cardinality of $F(\mathbb{R}, \mathbb{R})$. A function f from $\mathbb{R}$ to $\mathbb{R}$ can be interpretted as the set $\{(r, f(r)) \,:\, r \in \mathbb{R}\}$. Each of these are elements of $P(\mathbb{R} \times \mathbb{R})$, where $P(S)$ denotes the *power set* of S. This observation, together with $|\mathbb{R} \times \mathbb{R}| = |\mathbb{R}|$, implies that $|F(\mathbb{R}, \mathbb{R})| \leq |P(\mathbb{R} \times \mathbb{R})| = |P(\mathbb{R})| = 2^{|\mathbb{R}|}$. For a lower bound, consider the

[26] That $(A_f + A_g)(x) = A_f(x) + A_g(x) = f(x) + g(x) = (f + g)(x)$ and $\lambda A_f(x) = \lambda f(x) = (\lambda f)(x)$ for all $x \in V$.

functions $\{f_S \ : \ S \subseteq \mathbb{R}\}$ defined by $f_S(x) = 1$ if $x \in S$ and $f_S(x) = 0$ if $x \notin S$. Thus $|F(\mathbb{R},\mathbb{R})| \geq |\{f_S \ : \ S \subseteq \mathbb{R}\}| = |P(\mathbb{R})|$. Hence $|F(\mathbb{R},\mathbb{R})| = 2^{|\mathbb{R}|}$. An extra argument[27] shows that $\dim(F(\mathbb{R},\mathbb{R})) = 2^{|\mathbb{R}|}$.

(iv) Finally we compare $C^0(\mathbb{R},\mathbb{R})$, the vector space of continuous functions from $\mathbb{R} \to \mathbb{R}$, with $F(\mathbb{Q},\mathbb{R})$. There is a neat way to see that these are isomorphic. Take $f \in F(\mathbb{Q},\mathbb{R})$. Then f is uniquely determined by its outputs $f(q)$ across all $q \in \mathbb{Q}$. In the same way, once a function $g \in C^0(\mathbb{R},\mathbb{R})$ is defined on $\mathbb{Q}$, its outputs are known for all $r \in \mathbb{R}$ (which is not an obvious step, but follows from the definition of continuity). Furthermore, for any $f \in F(\mathbb{Q},\mathbb{R})$ we can find a $g \in C^0(\mathbb{R},\mathbb{R})$ such that $f(x) = g(x)$ for all $x \in \mathbb{Q}$. We can use this to define a bijection between $F(\mathbb{Q},\mathbb{R})$ and $C^0(\mathbb{R},\mathbb{R})$, and check this is a linear function. In this case $|F(\mathbb{Q},\mathbb{R})| = 2^{|\mathbb{N}|} = |\mathbb{R}|$, meaning the argument in the footnote cannot be used to determine $\dim F(\mathbb{Q},\mathbb{R})$. We can apply this argument less directly, however. Take $F(\mathbb{Q},\mathbb{Q})$ as a vector space over $\mathbb{Q}$, and let S be a basis. Then $|S| > |\mathbb{N}|$ and S is linearly independent over $\mathbb{Q}$. Thus, for any $n \in \mathbb{N}$ and $s_1, \ldots, s_n \in S$, we have that $a_1 s_1 + \ldots + a_n s_n = 0$ has only the solution $a_1 = \ldots = a_n = 0$ when $a_1, \ldots, a_n$ can be chosen from $\mathbb{Q}$. This can be written as an equation involving M, a matrix with columns $s_1, \ldots, s_n$, applied to $(a_1, \ldots, a_n) \in \mathbb{Q}^n$ with output 0. By Remark 3.13, we have $\det(M) \neq 0$. Thus, with $b_1, \ldots, b_n \in \mathbb{R}$, the only solution to $b_1 s_1 + \ldots + b_n s_n = 0$ is $b_1 = \ldots = b_n = 0$. Hence S is linearly independent over $\mathbb{R}$, and we have a set of cardinality $|S| > |\mathbb{N}|$. Applying Theorem 7.13 in the infinite dimensional setting, we have that $|\mathbb{N}| < \dim(F(\mathbb{Q},\mathbb{R})) \leq 2^{|\mathbb{N}|}$. If we believe the Continuum Hypothesis, then we have uniquely determined $\dim(F(\mathbb{Q},\mathbb{R}))$.

9.1. **Revisiting concepts using $F(\mathbb{R},\mathbb{R})$ and $F(\mathbb{R},\mathbb{C})$.** We look back at key concepts using these two examples. Unless stated, we will consider $F(\mathbb{R},\mathbb{R})$ as a vector space over $\mathbb{R}$ and $F(\mathbb{R},\mathbb{C})$ as over $\mathbb{C}$. Recall that given $f, g \in F(\mathbb{R},\mathbb{C})$, we have $f = g$ if and only if $f(x) = g(x)$ for all $x \in \mathbb{R}$.

Example 9.11. We start with linear dependence/independence.

(i) Let $S = \{\cos x, \sin x, \mathrm{e}^{\mathrm{i}x}\} \subset F(\mathbb{R},\mathbb{C})$. Then by $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$, the set S is linearly dependent.

(ii) The smaller set $S = \{\cos x, \sin x\}$ is linearly independent: if $\lambda_1 \cos x + \lambda_2 \sin x = 0$ for all $x \in \mathbb{R}$, then for $x = 0$ we get $\lambda_1 = 0$ and for $x = \pi/2$ we get $\lambda_2 = 0$.

Example 9.12. We can check a subset is a subspace with the subspace test.

(i) If $F(\mathbb{R},\mathbb{C})$ is a vector space over $\mathbb{R}$, then $F(\mathbb{R},\mathbb{R})$ is a subspace of $F(\mathbb{R},\mathbb{C})$.

(ii) The sets $\mathbb{P}(\mathbb{R})$ and $\mathbb{P}(\mathbb{C})$ of polynomials with real or complex coefficients are subspaces of $F(\mathbb{R},\mathbb{R})$ and $F(\mathbb{R},\mathbb{C})$ respectively. and it is closed under addition and scalar multiplication, hence it is a vector space. We further often just write P_N, with some $\mathbb{F}$ in mind.

(iii) We have that $U = \{f : \mathbb{R} \to \mathbb{C} \ ; \ f(0) = 0\} \subset F(\mathbb{R},\mathbb{C})$ is a subspace.

(iv) The set $PF(\mathbb{R},\mathbb{C}) := \{f \in F(\mathbb{R},\mathbb{C}) \ ; \ f(x+1) = f(x) \text{ for all } x \in \mathbb{R}\}$ is the set of all periodic functions with period 1 on $\mathbb{R}$. This set is closed under addition and multiplication by scalars, and hence is a vector space.

(v) The set $C^0(\mathbb{R},\mathbb{R})$, defined as the set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a subset of $F(\mathbb{R},\mathbb{R})$ which is closed under addition and multiplication by scalars. Similarly

$$C^k(\mathbb{R},\mathbb{R}) := \left\{ f \in C^0(\mathbb{R},\mathbb{R}) : \frac{d^m f}{dx^m} \in C^0(\mathbb{R},\mathbb{R}) \text{ for } 1 \leq m \leq k \right\}$$

is a vector space. Indeed, this follows from well-known theorems about the limit and derivative of the sum of two functions or a scalar multiple of a function.

(vi) The set of bounded real functions $C_b(\mathbb{R},\mathbb{R}) \subset C(\mathbb{R},\mathbb{R})$, defined by $f \in C_b(\mathbb{R},\mathbb{R})$ if there exists a $C_f > 0$ such that $|f(x)| \leq C_f$ for all $x \in \mathbb{R}$, is a vector space.

[27] For any subset S, linear combinations in S with coefficients from $\mathbb{R}$ can be thought of as finite sequences in $S \times \mathbb{R}$, i.e., in bijection with a subset of $(S \times \mathbb{R}) \times \mathbb{N}$. Thus no set S with $|S| < 2^{|\mathbb{R}|}$ can span $F(\mathbb{R},\mathbb{R})$, since $|\operatorname{span} S| \leq |(S \times \mathbb{R}) \times \mathbb{N}| = |S \times \mathbb{R}| < 2^{|\mathbb{R}|} = |F(\mathbb{R},\mathbb{R})|$.

Example 9.13. We now see examples similar to the above and important in Fourier Analysis.

(i) Almost periodic functions $AP := \text{span}(S_{\mathbb{R}})$, where $S_{\mathbb{R}} := \{e^{i\omega x} \ : \ \omega \in \mathbb{R}\} \subset F(\mathbb{R}, \mathbb{C})$.
(ii) The subspace of AP given by $\text{span}(S_{\mathbb{Z}})$, where $S_{\mathbb{Z}} := \{e^{2\pi n x} \ : \ n \in \mathbb{Z}\} \subset F(\mathbb{R}, \mathbb{C})$.
(iii) The subspace of AP and of $\text{span}(S_{\mathbb{Z}})$ consisting of trigonometric polynomials, for $N \in \mathbb{N}$, given by $\text{span}(T_N)$ where $T_N := \{e^{2\pi i n x}; \, n = -N, -N+1, \ldots, -1, 0, 1, \ldots, N-1, N\}$.

Each of the sets T_N, $S_{\mathbb{Z}}$, and $S_{\mathbb{R}}$ can also be shown to be linearly independent. Hence T_N is a subspace of dimension $2N+1$, and both AP and $\text{span}(S_{\mathbb{Z}})$ are infinite dimensional.

We can also determine whether a set is a basis in the same way we did for subsets of $\mathbb{C}^n$.

Example 9.14. Let $\text{span}\, T_2 := \{\sum_{|n|\leq 2} a_n e^{2\pi i n x} \, ; a_n \in \mathbb{C}\}$ be the space (over $\mathbb{C}$) of trigonometric polynomials of order 2. Then the set $\mathcal{A} = \{e^{-2\pi i 2x}, e^{-2\pi i x}, 1, e^{2\pi i x}, e^{2\pi i 2x}\}$ is a basis of T_2. Now we can expand $e^{2\pi i n x} = \cos(2\pi n x) + i \sin(2\pi n x)$, and so we expect that

$$\mathcal{B} = \{\cos(2\pi 2x), \sin(2\pi 2x), \cos(2\pi x), \sin(2\pi x), 1\}$$

is as well a basis for T_2. The corresponding matrix is given by

$$C_{\mathcal{BA}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ -i & 0 & 0 & 0 & i \\ 0 & 1 & 0 & 1 & 0 \\ 0 & -i & 0 & i & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

for which we compute that $\det C_{\mathcal{BA}} = -4$. Thus $C_{\mathcal{BA}}$ is nonsingular and $\mathcal{B}$ is indeed a basis.

Example 9.15. The notions of eigenvalue and eigenvector also apply to spaces of functions.

(i) Let D denote the derivative map on $C^1(\mathbb{R}, \mathbb{R})$, so that $D(f) := f'$. Then $f(x) = e^{\lambda x}$ is an eigenvector for D with corresponding eigenvalue λ.
(ii) Similarly, for the same map D as above, every element in $S_{\mathbb{R}}$ is an eigenvector in AP.
(iii) For $\mathbb{P}(\mathbb{R})$ over $\mathbb{R}$, we can consider the derivative function $D : \mathbb{P}(\mathbb{R}) \to \mathbb{P}(\mathbb{R})$ and also $A : \mathbb{P}(\mathbb{R}) \to \mathbb{P}(\mathbb{R})$, $p(x) \mapsto xp(x)$. Setting $T := A \circ D$ leads to each element from $\{x, x^2, x^3, \ldots\}$ being an eigenvector (each with a different eigenvalue).

Example 9.16. The image and kernel are defined for any linear map.

(i) For D above, on $C^1(\mathbb{R}, \mathbb{R})$, we have that $\ker D = \mathbb{P}_0$ and $\text{Im}\, D = C^0(\mathbb{R}, \mathbb{R})$.
(ii) For the shift map $T_1 : F(\mathbb{R}, \mathbb{C}) \to F(\mathbb{R}, \mathbb{C})$, $f(x) \mapsto f(x+1)$ and $I : f(x) \mapsto f(x)$, we have that $\ker(T_1 - I) = PF(\mathbb{R}, \mathbb{C})$ the space of periodic functions with period 1.
(iii) For the evaluation map $\delta_1 : F(\mathbb{R}, \mathbb{C}) \to \mathbb{C}$, $f(x) \mapsto f(1)$, we have $\ker \delta_1 = \{f \ : \ f(1) = 0\}$ and $\text{Im}\, \delta_1 = \mathbb{C}$.

Example 9.17. We also have new examples of inner products.

(i) Let $V = M_n(\mathbb{R})$ over $\mathbb{R}$. Then $\langle A, B \rangle := \text{tr}(A^t B)$ defines an inner product on V. In one way this is not new: we sum, over j, the dot products of rows of A and columns of B, and so we could equally well view A, B as vectors in $\mathbb{R}^{n^2}$, listing, column after column, their entries as a one-dimensional array.
(ii) On $C[a, b] := \{f : [a, b] \to \mathbb{C} \ : \ f \text{ is continuous}\}$ we have an inner product given by $\langle f, g \rangle = \int_a^b \bar{f}(x) g(x) \, dx$. We also have an associated norm. The key motivation for these comes from quantum mechanics.

Example 9.18. With an inner product we can then find an orthonormal basis.

(i) For our inner product above on $M_n(\mathbb{R})$, an ONB is really just one in $\mathbb{R}^{n^2}$. Thus the basis seen before consisting of matrices with exactly one (i, j) such that $a_{ij} = 1$ and all other entries zero defines an ONB (and corresponds to the standard basis $\{e_1, \ldots, e_{n^2}\}$).
(ii) Let $V = C[0, 1]$ and $e_k(x) := e^{2\pi i x}$ for $k \in \mathbb{Z}$. Then for $k \neq l$ and $k, l \in \mathbb{Z}$ we get

$$\langle e_k, e_l \rangle = \int_0^1 e^{2\pi i (l-k) x} \, dx = \left[\frac{1}{2\pi i (l-k)} e^{2\pi i (l-k) x} \right]_0^1 = 0$$

so $e_k \perp e_l$ if $k \neq l$. Hence the above sets $S_{\mathbb{Z}}$ and T_N, where $N \in \mathbb{N}$, are actually ONB.

(iii) One also has that $\{1, \sqrt{2}\cos x, \sqrt{2}\sin x, \ldots, \sqrt{2}\cos kx, \sqrt{2}\sin kx\}$ is an ONB for $\operatorname{span} T_N$.

When we have an ONB, it is then natural to consider orthogonal projections. Recall these are linear functions P which are projections (meaning $P^2 = P$) that are hermitian (meaning $P^* = P$).

Theorem 9.19. *Let V be an inner product space, $P : V \to V$ an orthogonal projection and $W = \operatorname{Im} P$. Then, for each $w \in W$, we have that*

$$\|v - w\| \geq \|v - Pv\|.$$

Proof. We apply Theorem 4.14 by noting $Pv \in W$ and $v - Pv \in W^\perp$:

$$\|v - Pv\|^2 \leq \|v - Pv\|^2 + \|Pv - w\|^2 = \|v - Pv + Pv - w\|^2 = \|v - w\|^2.$$

Thus, because this applies to every $w \in W$, the result holds. □

Thus $Pv \in W$ is the vector in W closest to v and the distance from v to W is actually $\|v - Pv\|$. We end with a motivation for an ONB in the infinite dimensional setting. Let $V = C[0, 1]$, the space of continuous functions on $[0, 1]$, and recall the inner product from Example 9.18(ii). With $W_N = \operatorname{span} T_N$ and $e_k(x) := \mathrm{e}^{2\pi \mathrm{i} x}$ for $k \in \mathbb{Z}$, we have an orthogonal projection onto W_N given by

$$P_N(f)(x) := \sum_{k=-N}^{N} \langle \mathrm{e}_k, f \rangle \mathrm{e}_k(x).$$

Theorem 9.19 tells us that for any function $f(x) \in F(\mathbb{R}, \mathbb{C})$ the trigonometric polynomial

$$f_N(x) := P_N(f)(x) = \sum_{k=-N}^{N} \langle \mathrm{e}_k, f \rangle \mathrm{e}_k(x), \quad \text{with} \quad \langle \mathrm{e}_k, f \rangle = \int_0^1 f(x) \mathrm{e}^{-2\pi \mathrm{i} k x}\, dx$$

gives the best approximation of f in the sense that $\|f - f_N\| \leq \|f - g\|$ for all $g \in W_N$. This is called a finite Fourier series of f. In Fourier Analysis one shows that if $N \to \infty$, then $\|f - f_N\| \to 0$. Let us now touch on the subject of Functional Analysis. When we introduced the general notion of a basis we required that every vector can be written as a linear combination of a *finite* number of basis vectors. The reason for this was that, for a general vector space, we cannot define an infinite sum of vectors as we have no notion of convergence. But with an inner product and the associated norm $\|v\|$, the situation is different. Given infinite sequences $(v_n)_{n \in \mathbb{N}}$ and $(\lambda_n)_{n \in \mathbb{N}}$, we say that the sum $\sum_{i=1}^{\infty} \lambda_i v_i$ converges to v, denoted $v = \sum_{i=1}^{\infty} \lambda_i v_i$, if

$$\lim_{N \to \infty} \left\| v - \sum_{i=1}^{N} \lambda_i v_i \right\| = 0.$$

We can then introduce a different notion of basis, a *Hilbert space basis*, which is an orthonormal set of vectors $\{v_1, v_2, \ldots\}$ such that every vector can be written as

$$v = \sum_{i=1}^{\infty} \langle v_i, v \rangle v_i.$$

Example 9.20. The set $S_{\mathbb{Z}} = \{e_n(x) \ : \ n \in \mathbb{Z}\}$ is a Hilbert space basis of $C[0, 1]$[28]. In this case the sum $f(x) = \sum_{k \in \mathbb{Z}} \langle e_k, f \rangle e_k(x)$ is called the *Fourier series* of f.

[28]... well, almost. We should take the completion of $C[0, 1]$, which is $L^2[0, 1]$, but leave this for another day.