MATH 341
Linear Algebra

Antoine Remond-Tiedrez

University of Wisconsin – Madison
Fall 2021
Figure 1: Two vectors in the plane can be added together to produce a third vector.

Introduction

This course is a first course in linear algebra with an added emphasis on how to read and write mathematical arguments, i.e. proofs. Before diving into the course let us say a few words about what linear algebra is for, as well as a few words on what proofs are about.

Linear algebra

Fundamentally, linear algebra is the study of linear transformations which act on so-called vector spaces. These vectors may be the ones you are familiar with, such as those encountered in physics and geometry, depicted in Figure 1, and which are fully characterized by a finite set of coordinates.

But these vectors can also be more fancy-looking: we will soon see that functions are vectors too! The only, but significant, difference is that functions sometimes require us to specify infinitely-many “coordinates” (see Figure 2).

Figure 2: A function is a vector fully specified by its value at every point – i.e. fully specified by possibly infinitely many “coordinates”.
However, as mentioned above, linear algebra only cares about vector spaces insofar as these spaces are those acted upon by linear transformations. When dealing with familiar two or three-dimensional vectors, linear transformations can be thought of as rotations, shears, or other geometrically meaningful transformations (this is depicted in Figure 3).

![Figure 3: Rotations and shears, along with some other geometric transformations, are linear.](image)

When dealing with functions, the linear transformations at play may not always have such clear geometric interpretations, but they are nonetheless crucial. For example: the transformation which takes a differentiable function to its derivative is linear! (See Figure 4.)

\[ f \mapsto f' \]

![Figure 4: A linear transformation on the vector space of differentiable functions.](image)

Proofs

To make sure that we are all on the same page when we use the word “proof” we provide a definition of a mathematical proof.

**Definition 0.1** (Mathematical proof). A mathematical proof is an argument, written in complete sentences using English words and mathematical symbols, which convinces a reasonable reader that a particular statement is true, without requiring the reader to fill out any details. In other words a proof is:

1. readable,
2. complete, and
3. correct.
A key point to keep in mind is that a proof is a form of communication between the reader and the writer. In particular this means that a proof cannot be correct “but the reader did not understand it”. If a proof cannot be understood then it is wrong.

We are now ready to start talking about linear algebra, and to start proving things about linear algebra.
1 Vector spaces

We begin our study of linear algebra by discussing vector spaces. Why start there?

Ultimately linear algebra is the study of linear processes, which are encoded mathematically as linear transformations. Linear transformations are nothing more than special functions between special sets, sets on which a notion of linearity can be defined (i.e. we want to be able to add elements to one another - this idea will be made precise soon). These special sets are precisely vector spaces. That is why we start our journey into linear algebra by discussing vector spaces.

1.1 Vector spaces

It turns out that vector spaces themselves rely on the notion of a field, so we start there.

Definition 1.1 (Field). A field \( \mathbb{F} \) is a set equipped with two operations, called addition and multiplication, which satisfy the following properties.

1. Commutativity: \( a + b = b + a \) and \( ab = ba \) for every \( a, b \in \mathbb{F} \).
2. Associativity: \( (a+b)+c = a+(b+c) \) and \( (ab)c = a(bc) \) for every \( a, b, c \in \mathbb{F} \).
3. Identities: There exist two special elements in \( \mathbb{F} \), denoted by 0 and 1, for which \( 0 + a = a \) and \( 1a = a \) for every \( a \in \mathbb{F} \).
4. Inverses: For every \( a, b \in \mathbb{F} \) with \( b \neq 0 \) there exist \( c, d \in \mathbb{F} \) such that \( a + c = 0 \) and \( bd = 1 \).
5. Distributivity: \( a(b + c) = ab + ac \) for every \( a, b, c \in \mathbb{F} \).

Example 1.2. The set of real numbers \( \mathbb{R} \), the set of complex numbers \( \mathbb{C} \), and the set of rational numbers \( \mathbb{Q} \) are all examples of fields. The set of integers \( \mathbb{Z} \) is not a field since it lacks multiplicative inverses (there is no integer which can multiply 2 in order to produce 1).

With the notion of a field in hand we can now introduce the main objects of interest of this section, namely vector spaces.

Definition 1.3 (Vector space). A vector space \( V \) over a field \( \mathbb{F} \) is a set equipped with two operations, called vector addition and scalar multiplication, such that the following hold.

1. Addition is commutative: for every \( x, y \in V \), \( x + y = y + x \).
2. Addition is associative: for every \( x, y, z \in V \), \( x + (y + z) = (x + y) + z \).

1Words that appear highlighted in blue are hyperlinks that can be used to go to the definition of the word itself. This is particularly handy when you are reading a complicated statement and may have forgotten the precise definition of one of the terms employed.
3. There exists an element in $V$, denoted by 0 and called the zero vector, such that $x + 0 = x$ for every $x \in V$.

4. For every $x \in V$ there exists an additive inverse $y \in V$ such that $x + y = 0$.

5. For every $x \in V$, $1x = x$ (here 1 denotes the multiplicative identity of the field $F$).

6. Multiplication is associative for every $a, b \in F$ and every $x \in V$, $(ab)x = a(bx)$.

7. Multiplication distributes over addition: for every $a \in F$ and every $x, y \in V$, $a(x + y) = ax + ay$.

8. Addition distributes over multiplication: for every $a, b \in F$ and every $x \in V$, $(a + b)x = ax + bx$.

Elements of $V$ are called vectors and elements of $F$ are called scalars.

**Example 1.4.** Here are some example of vector spaces. In some cases you will already be used to calling the elements of these spaces by the name “vector”, in other cases you will know them by different names. The beauty of linear algebra is that it allows to treat these seemingly disparate objects in a unified manner.

1. The set of $n$-tuples with entries from a field $F$ is a vector space denoted by $F^n$. For every $v = (v_1, v_2, \ldots, v_n), w = (w_1, w_2, \ldots, w_n) \in V$ and $a \in F$, $v + w$ and $av$ are defined via

   \[
   v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n) \quad \text{and} \quad av = (av_1, av_2, \ldots, av_n).
   \]

   For example: $\mathbb{R}$ and $\mathbb{C}$ are vector spaces (so are $\mathbb{R}$ and $\mathbb{C}$).

2. For any non-empty set $S$ and any field $F$ the set $\mathcal{F}(S, F)$ of all functions from $S$ to $F$ is a vector space. For every $f, g \in \mathcal{F}(S, F)$ and $a \in F$, $f + g$ and $af$ are defined via

   \[
   (f + g)(s) = f(s) + g(s) \quad \text{and} \quad (af)(s) = af(s).
   \]

   For example: $\mathcal{F}(\mathbb{R}, \mathbb{R})$, the set of all real-valued functions on the real line, and $\mathcal{F}(S^2, \mathbb{R})$, the set of all real-valued functions on the 2-dimensional sphere $S^2$ (this set of function can be used, for example, to describe the set of possible profiles of topographical elevation on the globe), are vector spaces.

3. In any field $F$ a polynomial is an expression of the form

   \[a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0\]

   where $n \geq 0$ is an integer and $a_k \in F$ for every $k$. The set of all polynomials over $F$, denoted by $P(F)$, is a vector space.
Note that each polynomial over $\mathbb{F}$ gives rise to a function from the field $\mathbb{F}$ to itself. In other words, the vector space $P(\mathbb{F})$ in the third example above is contained in the vector space $\mathcal{F}(\mathbb{F}, \mathbb{F})$ in the second example above. It is often the case that vector spaces are “nested” in interesting ways and so we will, in the next section, consider this idea more generally.

1.2 Subspaces

Motivating example 1.5. Any polynomial $p$ over $\mathbb{F}$ gives rise to a function $p : \mathbb{F} \to \mathbb{F}$, i.e. the two vector spaces $P(\mathbb{F})$ and $\mathcal{F}(\mathbb{F}, \mathbb{F})$ are nested within one another such that $P(\mathbb{F}) \subseteq \mathcal{F}(\mathbb{F}, \mathbb{F})$. We often care about how this nesting takes place and therefore introduce the notion of a subspace.

Definition 1.6 (Subspace). A subset $W$ of a vector space $V$ over a field $\mathbb{F}$ is called a subspace of $V$ if $W$ is a vector space over $\mathbb{F}$.

Example 1.7. Here are examples (and counter-examples!) of subspaces.

1. As shown in Figure 5 the line $x = y$ is a subspace of $\mathbb{R}^2$ while the line $x = y + 1$ is not (since it fails to contain the zero vector).

![Figure 5: Two subsets of the plane $\mathbb{R}^2$. One is a subspace, the other is not.](image)

2. As shown in Figure 6 the line $y = 2z$ is a subspace of the plane $x = 0$, which itself is a subspace of $\mathbb{R}^3$. This means that the line $y = 2z$ is also a subspace of $\mathbb{R}^3$.

The good news about the definition of a subspace is that it is remarkably simple. The downside of that simplicity is that, if we only use that definition, then proving that a given subset $S$ of a vector space $V$ is itself a vector space can be quite cumbersome (after all we need to verify that eight different properties hold in order to establish that something is a vector space!) Thankfully the subset $S$ automatically inherits many of the properties of the vector space $V$ itself and we are actually left with only three properties to verify. This is recorded precisely in the theorem below.

Theorem 1.8 (Characterisation of subspaces). Let $V$ be a vector space over a field $\mathbb{F}$ and let $W$ be a subset of $V$. $W$ is a subspace of $V$ if and only if the following hold:
(a) $0 \in W$ (the zero vector belongs to $W$),

(b) $x + y \in W$ for every $x, y \in W$ (closure under vector addition), and

(c) $ax \in W$ for every $x \in W$ and every $a \in F$ (closure under scalar multiplication).

Proof. Suppose that $W$ is a subspace of $V$. Then items (b) and (c) hold and it remains to establish (a). Since $W$ is a vector space there exists $0 \in W$ such that, for every $x \in W$, $x + 0 = x$. In particular $x + 0 = x = x + 0$ (note that 0 is the zero vector in $W$ whereas 0 is the zero vector in $V$ – we do not yet know that they are one and the same, this is precisely what we are trying to show). Therefore, by the Cancellation Law for Vector Addition (proved in Assignment 1), $\hat{0} = 0$ and indeed (a) holds.

Suppose now that (a)–(c) hold. Since $W \subseteq V$, properties (1), (2), and (5)–(8) of a vector space (as defined in Definition 1.3) also hold in $W$. Since, by (a), $0 \in W$, it therefore suffices to show that additive inverses of elements in $W$ lie in $W$. The key observation to that effect is that, for any vector $x \in W$, its additive inverse is precisely $(-1)x$ since $x + (-1)x = (1 + (-1))x = 0x = 0$ and hence $(-1)x \in W$ by (c).

Example 1.9. Here are some more examples and counter-examples of subspaces.

1. The space $P(F)$ of polynomials over a field $F$ is a subspace of $\mathcal{F}(F,F)$, the space of functions $F$ to $F$, since the three properties of Theorem 1.8 can be verified. Indeed: (a) holds since $f(x) = 0$ is a polynomial, (b) holds since the sum of two polynomials is a polynomial, and (c) holds since the scalar multiple of a polynomial remains a polynomial.

Similarly: the set $P_n(F)$ of all polynomials over a field $F$ of degree at most $n$ is a subspace of $P(F)$. 7
2. As shown in Figure 7 the curve $y = x^2$ is *not* a subspace of $\mathbb{R}^2$. Although the zero vector does belong to that curve, it is not closed under scalar multiplication: the vector $(1, 1)$ belongs to the curve but its scalar multiple $(2, 2)$ does not.

![Figure 7: This parabola is not a subspace of the plane since it is neither closed under scalar multiplication nor scalar addition.](image)

### 1.3 Linear combinations and span

**Motivating example 1.10.** To describe a line or a plane (in $\mathbb{R}^2$ or $\mathbb{R}^3$) *parametrically* we only need to use one or two vectors, respectively. Can we use this idea to obtain a succinct description of subspaces in general? Well yes, we can, using the notion of a *span*. First, however, we must discuss *linear combinations*.

**Definition 1.11 (Linear combination).** Let $V$ be a vector space over a field $\mathbb{F}$. We say that $v \in V$ is a *linear combination* of $u_1, u_2, \ldots, u_n \in V$ if there exist scalars $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n.$$

**Example 1.12.** Here are examples and counter-examples of linear combinations.

1. $v = (2, -5, 6)$ is a linear combination of $u_1 = (1, 0, 2)$, $u_2 = (0, 3, -1)$, and $u_3 = (1, 1, 2)$ since $v = u_1 - 2u_2 + u_3$.

2. $v = (6,0)$ is a linear combination of $u_1 = (1,0)$ and $u_2 = (2,0)$ since $v = 6u_1$. Note that $v$ can be written as a *different* linear combination of these two vectors since we could write $v = 3u_2$, $v = 10u_1 - 2u_2$, etc.
3. The polynomial \( p(x) = 3x^3 - 2x^2 + 7x + 8 \) is not a linear combination of
\[
\begin{align*}
q_1(x) &= x^3 - 2x^2 - 5x - 3 \\
q_2(x) &= 3x^3 - 5x^2 - 4x - 9.
\end{align*}
\]
Indeed: if that were the case, i.e. if \( p = aq_1 + bq_2 \) for some scalars \( a \) and \( b \) then equating the coefficients of each power tells us that
\[
\begin{align*}
3 &= a + 3b, \\
-2 &= -2a - 5b, \\
7 &= -5a - 4b, \text{ and} \\
8 &= -3a - 9b.
\end{align*}
\]
Eliminating \( a \) yields
\[
\begin{align*}
3 &= a + 3b, \\
4 &= b, \\
22 &= 11b, \text{ i.e. } b = 2, \text{ and} \\
17 &= 0.
\end{align*}
\]
Since it is impossible that \( 17 = 0 \) (or that both \( b = 4 \) and \( b = 2 \) simultaneously) we deduce that indeed \( p \) cannot be a linear combination of \( q_1 \) and \( q_2 \).

Linear combinations are central concepts to linear algebra but for now they are most helpful it allowing us to define the span. The span will allow us to answer the question opening this questions, namely “How can we fully describe a subspace very concisely?”

**Definition 1.13 (Span).** Let \( S \) be a non-empty subset of a vector space \( V \). The span of \( S \), denoted \( \text{span}(S) \), is the set of all linear combinations of the vectors in \( S \). For convenience we define \( \text{span}() = \{0\} \), where recall that \( \emptyset \) denotes the empty set.

**Example 1.14.** Here are examples of spans.

1. \( \text{span} \left( \{(1,0,0), (0,1,0)\} \right) \) is the \( z = 0 \) plane in \( \mathbb{R}^3 \), as depicted in Figure 8,

2. \( \text{span} \left( \{1, x, x^2\} \right) = \mathcal{P}_2 (\mathbb{F}) \).

A particularly useful characterization of the span is the following.

**Theorem 1.15.** The span of any subset \( S \) of a vector space \( V \) is the smallest subspace of \( V \) containing \( S \), i.e.

1. \( \text{span}(S) \) is a subspace of \( V \) and

2. \( \text{span}(S) \subseteq W \) for any subspace \( W \) of \( V \) containing \( S \).
Figure 8: The span of the two vectors depicted here is a plane.

Proof. If $S = \emptyset$ then the result is immediate since $\text{span}(\emptyset) = \{0\}$ which is a subspace of all subspaces of $V$.

Now suppose that $S$ is non-empty. We first verify (2): let $W$ be a subspace of $V$ containing $S$. Then, by closure of $W$ under vector addition and scalar multiplication, any linear combination of vectors in $S$ belongs to $W$, i.e. indeed $\text{span}(S) \subseteq W$.

Now we verify (1) using the theorem on the characterisation of subspaces (Theorem 1.8). Since $S \neq \emptyset$ there exists $z \in S$ such that $0 = 0z \in \text{span}(S)$. Now let $x, y \in \text{span}(S)$ and $a \in F$, i.e.

\[
\begin{aligned}
x &= a_1u_1 + \cdots + a_nu_n \text{ for } a_i \in F \text{ and } u_i \in S \\
y &= b_1v_1 + \cdots + b_nv_n \text{ for } b_i \in F \text{ and } v_i \in S.
\end{aligned}
\]

Then $x + y$ and $ax$ are themselves linear combination of elements of $S$, i.e. $x + y, ax \in \text{span}(S)$ such that indeed $\text{span}(S)$ is a subspace of $V$.  

1.4 Linear dependence and independence

Motivating example 1.16. We can write the plane $\mathbb{R}^2$ as the span of the three vectors $v_1$, $v_2$, and $v_3$ depicted in Figure 9 but that is somewhat inefficient. Indeed: we can also describe the plane $\mathbb{R}^2$ as the span of two of these vectors – the third one is redundant!

In other words: introducing linear combinations and, more importantly, the span allowed us to describe subspaces in a concise manner. The next order of business is therefore to find the most concise way to describe any given subspace by describing it as the span of the fewest number of vectors possible.

This is where the notions of linear dependence and independence come in.

Definition 1.17 (Linear dependence and independence). A subset $S$ of a vector space $V$ is called linearly dependent if there exist vectors $u_1, \ldots, u_n$ in $S$ and scalars $a_1, \ldots, a_n$ not all equal to zero such that

\[a_1u_1 + \cdots + a_nu_n = 0.\]
Figure 9: The plane is spanned by any two of these three vectors.

A subset which is not linearly dependent is called **linearly independent**.

**Example 1.18.** Here are examples of linearly dependent and independent sets.

1. The set \{x^2, x^2 - x, x - 1, 1\} is linearly dependent because
   \((-1)x^2 + (x^2 - x) + (x - 1) + 1 = 0\).

2. The set \{(1, 1, 0), (1, 0, 1)\} is linear independent because if
   \[
   a \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0
   \]
   then
   \[
   \begin{cases} 
   a + b = 0, \\
   a = 0, \text{ and} \\
   b = 0,
   \end{cases}
   \]
i.e. \(a = b = 0\).

Remember: we care about linear independence because it tells us when a given set of vectors has, in some sense, redundant information. This will be very important in the next section when we talk about **bases**.

For now however we turn our attention to a result which allows us to combine some of the concepts we have discussed so far rather nicely.

**Lemma 1.19** (Relation between linear independence and the span). Let \(S\) be a linearly independent subset of a vector space \(V\) and let \(v\) be a vector in \(V\) but not in \(S\). \(S \cup \{v\}\) is linearly dependent if and only if \(v \in \text{span}(S)\).

**Proof.** Suppose that \(S \cup \{v\}\) is linearly dependent. This means that there are vectors \(u_1, \ldots, u_n \in S \cup \{v\}\) and scalars \(a_1, \ldots, a_n\) not all equal to zero such that \(a_1u_1 + \cdots + a_nu_n = 0\). Crucially: since \(S\) is linearly independent one of the \(u_i\)'s must be equal to \(v\) and the corresponding scalar \(a_i\) must be nonzero. Up to re-indexing we may take \(i = 1\) such that
   \[
a_1v + a_2u_2 + \cdots + a_nu_n = 0.
   \]
Rearranging yields 
\[ v = -\frac{a_2}{a_1}u_2 - \cdots - \frac{a_n}{a_1}u_n, \]
i.e. indeed \( v \in \text{span}(S) \), as desired.

Conversely, suppose that \( v \in \text{span}(S) \). Then \( v = a_1u_1 + \cdots + a_nu_n \) for some scalars \( a_1, \ldots, a_n \) and \( u_1, \ldots, u_n \in S \) and hence 
\[ a_1u_1 + \cdots + a_nu_n + (-1)v = 0, \]
i.e. indeed \( S \cup \{v\} \) is linearly dependent. \( \square \)

### 1.5 Bases and dimension

We have described in the previous two sections how the span can be used to describe any subspace using only small number of vectors and how the notion of linear dependence allowed us to say precisely when some of these vectors were, in some sense, “redundant”. Naturally we now combine these two notions: when a set of spanning vectors are linearly independent, such that they fully describe the subspace they span without any redundancies, we call them a **basis**.

**Definition 1.20 (Basis).** Let \( V \) be a vector space. A subset \( S \) of \( V \) is called a **basis** of \( V \) if it is a linearly independent subset which spans \( V \) (i.e. \( \text{span}(S) = V \)).

**Example 1.21.** Here are some examples of bases for vector spaces we have encountered before.

1. The vectors \( e_1 = (1, 0, 0), e_2 = (0, 1, 0), \) and \( e_3 = (0, 0, 1) \) form a basis of \( F^3 \) for any field \( F \).
2. The set \( \{1, x, x^2, \ldots, x^n\} \) forms a basis of \( P_n(F) \) for any field \( F \).
3. The vectors \( (1, 0, 0), (1, 1, 0), \) and \( (1, 1, 1) \) form a basis of \( \mathbb{R}^3 \). They are linearly independent since, if 
   \[
   a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
   \]
   then 
   \[
   \begin{cases}
   a + b + c = 0, \\
   b + c = 0, \\
   c = 0
   \end{cases}
   \]
   from which we conclude that \( a = b = c = 0 \). These vectors also span \( \mathbb{R}^3 \) since any vector \( (x, y, z) \in \mathbb{R}^3 \) may be written as
   \[
   \begin{pmatrix} x \\ y \\ z \end{pmatrix} = z \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (y - z) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + (x - y) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.
   \]
So far we have been somewhat vague when motivating why one should care
about combining the notion of a span with the notion of linear independence,
mentioning only that it provides an efficient description of subspaces. The
theorem below makes this idea precise.

**Theorem 1.22.** Let \( V \) be a vector space. The vectors \( v_1, \ldots, v_n \) form a basis
of \( V \) if and only if every \( v \in V \) can be written uniquely as
\[
v = a_1 v_1 + \cdots + a_n v_n,
\]
i.e. the coefficients are uniquely determined by the vector \( v \).

**Proof.** Suppose that \( v_1, \ldots, v_n \) form a basis of \( V \). Since they span \( V \), every
\( v \in V \) can be written as \( v = a_1 v_1 + \cdots + a_n v_n \). Suppose there is another such
representation, say \( v = b_1 v_1 + \cdots + b_n v_n \). Then

\[
0 = v - v = (a_1 - b_1)v_1 + \cdots + (a_n - b_n)v_n
\]
such that, by linear independence of the \( v_i \)'s,
\[
a_1 - b_1 = 0, \ldots, a_n - b_n = 0
\]
which proves that indeed the \( a_i \)'s are uniquely determined by \( v \).

Conversely, suppose that every \( v \in V \) can be written uniquely in the form \( v = a_1 v_1 + \cdots + a_n v_n \). Then clearly the \( v_i \)'s span \( V \). Moreover: if \( a_1 v_1 + \cdots + a_n v_n = 0 \) then
\[
a_1 v_1 + \cdots + a_n v_n = 0 v_1 + \cdots + 0 v_n
\]
and so, by the uniqueness of the scalars \( a_i \), we know that \( a_1 = \cdots = a_n = 0 \),
i.e. indeed \( v_1, \ldots, v_n \) are linear independent. \( \square \)

Another reason why bases are so important is that they allow us to precisely
define the *dimension* of a vector space. However, in order to do so, we must
first state the following result.

**Theorem 1.23.** All finite bases of a vector space \( V \) contain the same number
of vectors.

**Proof.** See assignment 4. \( \square \)

The reason **Theorem 1.23** is so important is that it tells us that the number
of vectors in a basis does not depend on how we construct that basis, but rather
is an intrinsic property of the vector space itself. We give that property a name:
dimension.

**Definition 1.24** (Dimension). A vector space \( V \) which possesses a finite basis
is called *finite-dimensional* and its *dimension*, denoted \( \dim(V) \), is the number
of vectors in a basis of \( V \).

**Example 1.25.** We now record the dimensions of some familiar spaces.
1. \( \dim(F^n) = n. \)

2. \( \dim(P_n(F)) = n + 1. \)

3. Over the field \( \mathbb{C} \), \( \dim(\mathbb{C}) = 1 \) since a basis is \( \{1\} \) but, over the field \( \mathbb{R} \), \( \dim(\mathbb{C}) = 2 \) since a basis is \( \{1, i\} \).
2 Linear maps and matrices

We know enter the meat of the subject. Remember: linear algebra is ultimately the study of linear transformations, such as for example rotations and shears in $\mathbb{R}^2$ or $\mathbb{R}^3$ or, when acting on functions, operations like differentiation and integration. The only reason we discussed vector spaces in the first place is because those are the spaces between which linear transformations act. Having discussed vector spaces we are therefore now ready to introduce linear transformations, or linear maps, and discuss some of their properties.

2.1 Linear maps, kernels, and images

**Definition 2.1** (Linear map). Let $V$ and $W$ be vector spaces over the same field $\mathbb{F}$. A linear map (or linear transformation) is a function $T : V \rightarrow W$ satisfying

1. $T(u + v) = T(u) + T(v)$ for every $u, v \in V$, and
2. $T(au) = aT(u)$ for every $u \in V$ and $a \in \mathbb{F}$,

i.e. $T$ preserves vector addition and scalar multiplication.

**Example 2.2.** Here are some examples of linear maps.

1. The projection $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by $T(x, y, z) = (x, y, 0)$.
2. Differentiation $T : P_n(\mathbb{R}) \rightarrow P_{n-1}(\mathbb{R})$ defined by $T(p) = p'$.
3. For any $a \in \mathbb{F}$ the shift $T_a : \mathcal{F}(\mathbb{F}, \mathbb{F}) \rightarrow \mathcal{F}(\mathbb{F}, \mathbb{F})$ defined by
   $[T_a(f)](x) = f(x - a)$.

There are many interesting properties of linear maps to discuss – this is what we will spend most of the course doing! We start with two fundamental objects associated to linear maps, namely kernels and images.

**Definition 2.3** (Kernel and image). Let $T : V \rightarrow W$ be a linear map. The kernel and image of $T$ are defined respectively as

$\ker T = \{v \in V : T(v) = 0\}$ and
$\text{im } T = \{w \in W : w = T(v) \text{ for some } v \in V\}$.

**Example 2.4.** We record the kernels and images of the linear transformations discussed in Example 2.2 above.

1. For the projection $T(x, y, z) = (x, y, 0)$ we have that
   $\ker T = \{(0, 0, z) : z \in \mathbb{R}\}$ = \{z-axis\} and
   $\text{im } T = \{(x, y, 0) : x, y \in \mathbb{R}^2\}$ = \{plane $z = 0$\}.

This can also be seen in Figure 10.
Figure 10: The kernel and image of the linear map $T(x, y, z) = (x, y, 0)$.

2. For the differentiation map $T: P_n(\mathbb{R}) \rightarrow P_{n-1}(\mathbb{R})$ given by $T(p) = p'$ we have that
   \[
   \begin{align*}
   \ker T &= \{ c : c \in \mathbb{R} \}, \text{ i.e. constants, and} \\
   \im T &= P_{n-1}(\mathbb{R}).
   \end{align*}
   \]
   Indeed, every polynomial of degree $n-1$ is the derivative of a polynomial of degree $n$ since
   \[
   a_{n-1}x^{n-1} + \cdots + a_0 = (\frac{a_{n-1}}{n}x^n + \cdots + a_0x)'.
   \]

3. For any $a \in \mathbb{F}$ consider the shift $T_a(f) = f(\cdot - a)$. Then
   \[
   \ker T = \{ 0 \} \text{ and } \im T = \mathcal{F}(\mathbb{F}, \mathbb{F}).
   \]
   Indeed, for any function $f : \mathbb{F} \rightarrow \mathbb{F}$ we may find $g$ such that $T_a(g) = f$.
   To do so is suffices to define $g = T_{-a}f$ as then $T_a(T_{-a}f) = f$.

A fundamental observation about the kernel and image of a linear map is that they are examples of subspaces.

**Theorem 2.5.** Let $V : V \rightarrow W$ be a linear map. Its kernel and image are subspaces of $V$ and $W$, respectively.

**Proof.** We only prove that the kernel of $T$ is a subspace of $V$, see assignment 3 for the proof that the image of $T$ is a subspace of $W$.

Clearly $0 \in \ker T$ since $T(0) = T(0v) = 0T(v) = 0$. Now let $u, v \in V$ and $a$ be a scalar. Then:
   \[
   T(u + v) = T(u) + T(v) = 0 + 0 = 0 \text{ and } T(au) = aT(u) = a0 = 0
   \]
   i.e. indeed $u + v, au \in \ker T$, which proves that $\ker T$ is a subspace of $V$. \hfill \Box

The kernel and image of a linear map are such fundamental objects that their dimensions are given their own special name.
**Definition 2.6** (Rank and nullity). Let $T : V \to W$ be a linear map between finite-dimensional vector spaces. Its **rank** and **nullity** are defined to be

$$\text{rank}(T) = \dim(\text{im}(T)) \text{ and } \text{nullity}(T) = \dim(\ker(T)).$$

The rank and nullity of a linear map are very neatly related to each other, as recorded in the theorem below.

**Theorem 2.7** (Rank-Nullity Theorem). Let $T : V \to W$ be a linear map between finite-dimensional vector spaces. The following holds:

$$\dim V = \text{rank} T + \text{nullity} T.$$

What follows was not covered in lecture and is only provided if you are curious as to how one would prove the rank-nullity theorem. Note that this proof was not covered in class due to its length but that it is not too challenging otherwise.

**Proof.** Since $V$ is finite-dimensional and $\ker(T)$ is a subspace of $V$, $\ker(T)$ itself is finite-dimensional (any basis of $V$ will necessarily span $\ker(T)$ and so, by a result proved in a problem session, can be reduced to a basis). So let $\text{nullity}(T) = s$ and let $u_1, \ldots, u_s$ be a basis of $\ker(T)$. Since this set of vectors is linearly independent it can, by a result proved in assignment 3, be extended to a basis $u_1, \ldots, u_s, v_1, \ldots, v_r$ of $V$. Then $\dim(V) = s + r$ and so in order to prove the theorem it suffices to show that $\text{rank} T = r$, i.e. that the dimension of $\text{im}(T)$ is equal to $r$.

Since $u_1, \ldots, u_s, v_1, \ldots, v_r$ is a basis of $V$ it spans $V$ and hence

$$T(u_1), \ldots, T(u_s), T(v_1), \ldots, T(v_r)$$

must also span $\text{im}(T)$. Moreover, since

$$T(u_1) = \cdots = T(u_s) = 0$$

we know that it actually suffices to consider $T(v_1), \ldots, T(v_r)$ in order to span $\text{im}(V)$. To conclude the proof it therefore suffices to prove that $T(v_1), \ldots, T(v_r)$ are linearly independent, since they would then constitute a basis of $\text{im}(T)$.

Well suppose that

$$a_1 T(v_1) + \cdots + a_r T(v_r) = 0.$$

Then, by linearity of $T$, we know that

$$T(a_1 v_1 + \cdots + a_r v_r) = 0,$$

i.e. $a_1 v_1 + \cdots + a_r v_r \in \ker(T)$. Since $u_1, \ldots, u_s$ is a basis of $\ker(T)$ this tells us that there exist scalars $b_1, \ldots, b_s$ such that

$$a_1 v_1 + \cdots + a_r v_r = b_1 u_1 + \cdots + b_s u_s.$$
and hence we may rearrange this equation into

\[ a_1 v_1 + \cdots + a_r v_r - b_1 u_1 - \cdots - b_s u_s = 0. \]

Crucially: the vectors \( v_1, \ldots, v_r, u_1, \ldots, u_r \) form a basis of \( V \), and in particular this means that they are linearly independent. This allows us to deduce that all of the scalars above, the \( a_i \)'s and the \( b_i \)'s, must be equal to zero. This establishes the linear independence of \( v_1, \ldots, v_r \) and concludes the proof.

This is the end of the portion of the notes covering extra material not discussed in lectures.

2.2 The matrix representation of a linear map

When introducing bases, in Section 1.5 we noted that they allowed us to find coordinates for any vector space. Indeed: if \( v_1, \ldots, v_n \) is a basis of \( V \) then any vector \( v \in V \) has a unique representation of the form

\[ v = a_1 v_1 + \cdots + a_n v_n \]

and so we may think of \( a_1, \ldots, a_n \) as the coordinates of \( v \) along the basis \( v_1, \ldots, v_n \). In this section we will see how to take this idea one step further: we will see that linear transformations can be, once bases are chosen, characterized by arrays of numbers in a similar fashion to how vectors can be characterized by a list of coordinates.

These arrays that will be used to characterize linear maps are called matrices.

Before introducing matrices and their most fundamental operation, namely how they multiply vectors, we motivate the appearance of matrices through linear systems.

Motivating example 2.8 (Matrices come from linear systems.). Consider the linear system

\[
\begin{align*}
    x_1 - x_2 + x_3 + 2x_4 &= 5 \\
    -2x_1 + x_2 - x_3 - x_4 &= -1.
\end{align*}
\]

This linear system is characterized by its coefficients so it can be written using matrix-vector multiplication as

\[
\begin{pmatrix}
    1 & -1 & 1 & 2 \\
    -2 & 1 & -1 & -1
\end{pmatrix}
\begin{pmatrix}
    x_1 \\
    x_2 \\
    x_3 \\
    x_4
\end{pmatrix}
= 
\begin{pmatrix}
    5 \\
    -1
\end{pmatrix}
\]

where each row of the matrix correspond to one of the equations from the linear system.
Definition 2.9 (Matrix and matrix-vector multiplication). Let \( \mathbb{F} \) be a field. A \( n \)-by-\( m \) matrix \( A \) is an array of the form

\[
A = \begin{pmatrix}
a_{11} & \cdots & a_{1m} \\
\vdots & \ddots & \vdots \\
a_{n1} & \cdots & a_{nm}
\end{pmatrix}
\]

where \( a_{ij} \in \mathbb{F} \) for every \( i, j \).

Note that \( A \) has \( n \) rows and \( m \) columns and that, in \( a_{ij} \), \( i \) is the row index and \( j \) is the column index.

Given a vector \( v \in \mathbb{F}^m \), the matrix-vector product \( Av \in \mathbb{F}^n \) is defined by

\[
(Av)_i = \sum_{j=1}^{m} a_{ij}v_j,
\]

i.e. the \( i \)-th coordinate of \( Av \) comes from the \( i \)-th row of \( A \).

Example 2.10. Observe that, according to the definition above,

\[
\begin{pmatrix}
1 & -1 & 1 & 2 \\
-2 & 1 & -1 & -1
\end{pmatrix}
\begin{pmatrix}
-1 \\
1 \\
3
\end{pmatrix}
= \begin{pmatrix}
1 \cdot (-1) + (-1) \cdot 1 + 1 \cdot 1 + 2 \cdot 3 \\
(-2) \cdot (-1) + 1 \cdot 1 + (-1) \cdot 1 + (-1) \cdot 3
\end{pmatrix}
= \begin{pmatrix}
5 \\
-1
\end{pmatrix},
\]

i.e. we have found a solution

\[
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix}
= \begin{pmatrix}
-1 \\
1 \\
1 \\
3
\end{pmatrix}
\]

of the linear system discussed in Example 2.8 above.

Now that we have introduced matrices we may focus on the main purpose of this section: determine how to represent a given linear map as a matrix. We begin with an example.

Motivating example 2.11 (Representing a linear map as a matrix). Consider the linear map \( T : P_2(\mathbb{R}) \to P_2(\mathbb{R}) \) defined by

\[
[T(p)](x) = p(x) - xp'(x) - p''(x).
\]

Then

\[
T(ax^2 + bx + c) = (ax^2 + bx + c) - x(2ax + b) - 2a = -ax^2 + (c - 2a).
\]

In other words, if we write \( T(ax^2 + bx + c) \) in the usual form \( dx^2 + ex + f \), we
see that
\[ T(ax^2 + bx + c) = dx^2 + ex + f \iff \begin{cases} -a = d \\ 0 = e \\ c - 2a = f \end{cases} \iff \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} d \\ e \\ f \end{pmatrix}. \]

The matrix \( A \) what we will call the matrix representation of \( T \).

Note in particular that \( T(x^2) = -x^2 - 2 \) and that, expressed in the basis \( \{x^2, x, 1\} \) of \( P_2(\mathbb{R}) \), the coordinates of
\[ T(x^2) = (-1) \cdot x^2 + 0 \cdot x + (-2) \cdot 1 \]
are precisely the numbers appearing in the first column of \( A \).

We now generalize the idea discussed above to an arbitrary linear map.

**Definition 2.12** (Matrix representation of linear maps). Let \( T : V \to W \) be a linear map where \( \dim V = m \) and \( \dim W = n \) and suppose that \( v_1, \ldots, v_m \) and \( w_1, \ldots, w_n \) are bases of \( V \) and \( W \), respectively. For every \( j \), \( T(v_j) \) can be uniquely written as a linear combination of \( w_1, \ldots, w_n \) so there exist scalars \( a_{ij} \) such that
\[ T(v_j) = a_{1j}w_1 + \cdots + a_{nj}w_n = \sum_{i=1}^{n} a_{ij}w_i. \]
The \( n \)-by-\( m \) matrix \( A = (a_{ij})_{i=1,j=1}^{n,m} \) is called the matrix of \( T \) with respect to the bases of \( V \) and \( W \) chosen above.

**Example 2.13.** The vectors \( v_1 = (1, 0, 0) \), \( v_2 = (1, 1, 0) \), and \( v_3 = (1, 1, 1) \) form a basis of \( \mathbb{R}^3 \). What is the matrix \( A \) of \( T(x, y, z) = (x - z, y, 0) \) with respect to this basis? Observe that
\[
\begin{align*}
T(v_1) &= v_1 = 1 \cdot v_1 + 0 \cdot v_2 + 0 \cdot v_3, \\
T(v_2) &= v_2 = 0 \cdot v_1 + 1 \cdot v_2 + 0 \cdot v_3, \quad \text{and} \\
T(v_3) &= (0, 1, 0) = v_2 - v_1 = (-1) \cdot v_1 + 1 \cdot v_2 + 0 \cdot v_3
\end{align*}
\]
such that
\[
A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.
\]

**Theorem 2.14** (Correspondence between linear maps and matrices). Let \( V \) and \( W \) be vector spaces over a field \( \mathbb{F} \) with bases \( v_1, \ldots, v_m \) and \( w_1, \ldots, w_n \), respectively. There is a one-to-one correspondence between linear maps from \( V \) to \( W \) and \( n \)-by-\( m \) matrices over \( \mathbb{F} \).

**Proof.** See assignment 4. \( \square \)
2.3 Composition of linear maps & matrix multiplication

Now that we know how to associate matrices to linear maps we seek to answer the following question: two linear maps \( T \) and \( S \) will have matrices \( A \) and \( B \) associated with them, and their composition \( S \circ T \) will also have some matrix \( C \) associated to it – is there a simple way to write \( C \) in terms of \( A \) and \( B \)? Before we can answer this question we must take a closer look at a detail we overlooked when formulating our question above: are we even sure that the composition \( S \circ T \) is itself a linear map? Thankfully the answer is yes.

**Lemma 2.15.** The composition of two linear maps is itself a linear map.

**Proof.** Let \( T : U \to V \) and \( S : V \to W \) be linear maps. Then, for every \( x, y \in U \) and every scalar \( a \), it follows from the linearity of \( T \) and \( S \) that

\[
(S \circ T)(x + y) = S(T(x + y)) = S(T(x) + T(y)) = S(T(x)) + S(T(y)) = (S \circ T)(x) + (S \circ T)(y)
\]

and

\[
(S \circ T)(ax) = S(T(ax)) = S(aT(x)) = aS(T(x)) = a(S \circ T)(x)
\]

such that indeed the composition \( S \circ T \) is linear. \( \square \)

Now that we now that the composition of two linear maps is always linear we return to our original question: given the matrices of two linear maps, how may we determine the matrix associated with their composition? As we work our way to an answer below we will see that we need to introduce the product of two matrices in a very natural way.

**Motivating example 2.16 (The matrix of a composition of linear maps).** Consider two linear maps \( T : U \to V \) and \( S : V \to W \), suppose that

- \( u_1, \ldots, u_m \) is a basis of \( U \),
- \( v_1, \ldots, v_n \) is a basis of \( V \), and
- \( w_1, \ldots, w_p \) is a basis of \( W \),

and suppose that \( A \) and \( B \) are the matrices of \( T \) and \( S \), respectively, with respect to these bases. Then, recalling that the coordinates of \( T(u_k) \) are encoded in the \( k \)-th column of \( A \) and that the coordinates of \( S(v_j) \) are encoded in the \( j \)-th
column of \( B \), we see that

\[
(S \circ T)(u_k) = S(T(u_k))
\]

\[
= S \left( \sum_j a_{jk} v_j \right)
\]

\[
= \sum_j a_{jk} S(v_j)
\]

\[
= \sum_j a_{jk} \left( \sum_i b_{ij} w_i \right)
\]

\[
= \sum_i \left( \sum_j b_{ij} a_{ij} \right) w_i
d_{ik}
\]

where \( C = (c_{ik}) \) is the matrix of \( S \circ T \). Crucially, we note that

\[
c_{ik} = (i\text{-th row of } B) \cdot (k\text{-th column of } A).
\]

This motivates our definition of matrix-matrix multiplication below.

**Definition 2.17** (Matrix-matrix multiplication). Let \( \mathbb{F} \) be a field, let \( A \) be an \( n \times m \) matrix over \( \mathbb{F} \), and let \( B \) be a \( p \times n \) matrix over \( \mathbb{F} \). Then \( C = BA \) is given by \( C = (c_{ik})_{i=1}^p_{k=1} \) where

\[
c_{ik} = \sum_{j=1}^n b_{ij} a_{jk}.
\]

**Example 2.18.** Here are some examples of matrix multiplication going right, and going awry.

1. We may compute

\[
\begin{pmatrix}
  2 & 1 & 0 & -1 \\
  1 & -2 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
  0 & -2 & 1 \\
  2 & 1 & 0 \\
  3 & 0 & 1 \\
  -1 & 0 & 2
\end{pmatrix}
= \begin{pmatrix}
  2 + 1 & -4 + 1 & 2 - 2 \\
  -4 + 3 & -2 - 2 & 1 + 1 \\
  3 & -3 & 0 \\
  -1 & -4 & 2
\end{pmatrix}
\]

Note that we can relate the size of each factor and of the product in the following way:

\((2\text{-by-}4) \cdot (4\text{-by-}3) = 2\text{-by-}3\).
The product
\[
\begin{pmatrix}
0 & -2 & 1 \\
2 & 1 & 0 \\
3 & 0 & 1 \\
-1 & 0 & 2
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 \\
1 & -2 & 1 \\
0 & -1 & 2
\end{pmatrix}
\]
is not defined since the first matrix has three columns but the second matrix has only two rows!

Having defined matrix-matrix multiplication we may now verify that, indeed, the matrix of a composition is given by the product of the matrices corresponding to the linear maps being composed. Note that the proof is essentially already done in Example 2.16 above!

**Theorem 2.19** (The matrix of a composition). Let \( T : U \rightarrow V \) and \( S : V \rightarrow W \) be linear maps, let \( \{u_1, \ldots, u_m\} \), \( \{v_1, \ldots, v_n\} \), and \( \{w_1, \ldots, w_p\} \) be bases of \( U \), \( V \), and \( W \) respectively, and let \( A \) and \( B \) be the respective matrices of \( T \) and \( S \) with respect to these bases. The matrix of \( S \circ T \) with respect to these same bases is \( BA \).

**Proof.** For any \( k \) we have that
\[
(S \circ T)(u_k) = S \left( \sum_j a_{jk} v_j \right) = \sum_j a_{jk} \left( \sum_i b_{ij} w_i \right) = \sum_i \left( \sum_j b_{ij} a_{jk} \right) w_i
\]
and so indeed that matrix of \( S \circ T \) is \( BA \).

### 2.4 Inverses

In the previous section we established that the composition of linear maps translates to the product of matrices. In this section and the following we will use this idea to continue building a dictionary between linear maps and matrices. More precisely: in this section we will discuss how to determine the matrix of the inverse of a linear map while in the following section we will discuss how to relate two matrices of the same linear map computed using different bases. Once again, an important thing to keep in mind here is that the translation between composition of linear maps and multiplication of matrices will play a fundamental role in both of the two further topics discussed above.

For now we therefore turn our attention to invertible linear maps. Before saying anything fancy above those we really ought to define them precisely.

**Definition 2.20** (Identities, invertible linear maps, and invertible matrices). Let \( V \) and \( W \) be vector spaces and let \( F \) be a field.

1. The identity map on \( V \), denoted \( \text{id} \) or \( \text{id}_V \), is the map \( \text{id} : V \rightarrow V \) defined by \( \text{id}(v) = v \) for every \( v \in V \).
2. A linear map \( T : V \to W \) is called \textit{invertible} if it has an \textit{inverse} \( T^{-1} : W \to V \) such that \( T \circ T^{-1} = \text{id}_W \) and \( T^{-1} \circ T = \text{id}_V \).

3. For any integer \( n \geq 0 \) the \( n \)-by-\( n \) \textit{identity matrix} \( I \) over \( \mathbb{F} \) has entries

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j
\end{cases}
\]

where the symbol \( \delta_{ij} \) is known as the \textit{Kronecker delta}.

4. A matrix \( A \) is called \textit{invertible} if it has an \textit{inverse} matrix \( A^{-1} \) such that \( AA^{-1} = I \) and \( A^{-1}A = I \).

**Remark 2.21.** As we will prove in a problem session: invertible matrices must be \textit{square}.

**Example 2.22.** Here are some examples and counter-examples of invertible linear maps and matrices.

1. Consider \( T : P_2(\mathbb{R}) \to P_2(\mathbb{R}) \) defined by \( T(p) = p(\cdot - 2) \). This matrix is invertible and its inverse is given by \( T^{-1}(p) = p(\cdot + 2) \).

2. The rotation by an angle \( \theta \) in \( \mathbb{R}^2 \) given by

\[
A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}
\]

is invertible with

\[
A^{-1} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.
\]

Indeed: we may compute that

\[
AA^{-1} = A^{-1}A = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & 0 \\ 0 & \cos^2 \theta + \sin^2 \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.
\]

3. The matrix

\[
A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}
\]

is \textit{not} invertible because \( A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \) but \( A^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \) is impossible!

4. The map \( T : P_1(\mathbb{R}) \to P_2(\mathbb{R}) \) given by \( T(p) = p^2 \) is \textit{not} invertible since, for \( p(x) = x \), we cannot find a polynomial \( q \in P_1(\mathbb{R}) \) such that \( T(q) = p \) (we would need \( q(x) = \sqrt{x} \)). As dutifully pointed out by folks in class: this map also fails to be linear (since the sum of squares is not equal to the square of the sum)!
5. To make up for the example above, here is another one of a truly linear map which fails to be invertible. Consider \( T : P(\mathbb{R}) \to P(\mathbb{R}) \) given by \([T(p)](x) = xp(x)\). This map is not invertible since we cannot find a polynomial \( p \) such that \( T(p) = 1 \) (we would need \( p(x) = 1/x \), which is not a polynomial).

Using the identification of the composition of linear maps with the product of matrices we can deduce, as recorded in the theorem below, that the matrix of an inverse is the inverse of that matrix.

**Theorem 2.23.** Let \( T : V \to W \) be a linear map between finite-dimensional vector spaces and let \( A \) be the matrix of \( T \) with respect to some bases of \( V \) and \( W \). \( T \) is invertible if and only if \( A \) is invertible and the matrix of \( T^{-1} \) is \( A^{-1} \).

**Proof.** Suppose that \( T \) is invertible and let \( B \) be the matrix of \( T^{-1} \). Since \( T \circ T^{-1} = \text{id}_W \) and \( T^{-1} \circ T = \text{id}_V \) and the matrix of \( \text{id}_V \) or \( \text{id}_W \) is \( I \) we deduce that \( AB = I \) and \( BA = I \), i.e. \( A \) is invertible and \( B = A^{-1} /\)

Conversely: if \( A \) is invertible then let \( S \) be the linear map with matrix \( A^{-1} \). Then, since \( AA^{-1} = I \) and \( A^{-1}A = I \), we deduce that \( T \circ S = \text{id}_W \) and \( S \circ T = \text{id}_V \) such that indeed \( S = T^{-1} \). \(\square\)

### 2.5 Change of basis matrix and similar matrices

In the previous section we used the identification of the composition of linear maps with the multiplication of matrices to deduce that the matrix of an inverse is the inverse of a matrix.

In this section we use this same identification once more, now to determine how to relate two matrices who represent the same linear map with respect to two different bases.

**Motivating example 2.24** (Change of basis). We consider the linear map \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) given by \( T \left( \begin{array}{c} x \\ y \end{array} \right) = \left( \begin{array}{c} x + 3y \\ 3x + y \end{array} \right) \). We want to consider this same linear map but with respect to a *different* basis, namely the basis consisting of the vectors \( \left( \begin{array}{c} 1 \\ 1 \end{array} \right) \) and \( \left( \begin{array}{c} 1 \\ -1 \end{array} \right) \). We therefore compute

\[
T \left[ \alpha \left( \begin{array}{c} 1 \\ 1 \end{array} \right) + \beta \left( \begin{array}{c} 1 \\ -1 \end{array} \right) \right] = T \left( \begin{array}{c} \alpha + \beta \\ \alpha - \beta \end{array} \right) = \left( \begin{array}{c} 4\alpha - 2\beta \\ 4\alpha + 2\beta \end{array} \right) = 4 \left( \begin{array}{c} 1 \\ 1 \end{array} \right) + (-2) \left( \begin{array}{c} 1 \\ -1 \end{array} \right).
\]

In other words, in terms of this *new* basis \( \left\{ \left( \begin{array}{c} 1 \\ 1 \end{array} \right), \left( \begin{array}{c} 1 \\ -1 \end{array} \right) \right\} \), the linear map \( T \) can be written as

\[
\tilde{T} : \left( \begin{array}{c} \alpha \\ \beta \end{array} \right) \to \left( \begin{array}{c} 4\alpha - 2\beta \\ 4\alpha + 2\beta \end{array} \right).
\]

All of this is also depicted in Figure 11.
As Example 2.24 above shows, changing bases can sometimes make linear transformations much easier to analyze. The question is now: can we devise a general recipe to change bases? That is, given a linear map and its matrix in some given basis, can we devise a recipe that, given a new linear map, will produce the matrix of that same linear map with respect to that new basis? The answer is that yes, we can and will do that – the key component of this recipe is a so-called change of basis matrix.

Definition 2.25 (Change of basis matrix). Let \( u_1, \ldots, u_n \) and \( v_1, \ldots, v_n \) be bases of a vector space \( V \). The change of basis matrix from \( u_1, \ldots, u_n \to v_1, \ldots, v_n \) is the matrix of \( \text{id} : V \to V \) where the first basis is used for the domain and the second basis is used for the codomain / target.

Example 2.26. Consider the bases \( u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ u_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \) and \( v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \). Then

\[
\text{id}(u_1) = u_1 = (v_1 + v_2)/2 \quad \text{and} \quad \text{id}(u_2) = u_2 = 1 \cdot v_1 + 0 \cdot v_2
\]

so the change of basis matrix from \( u_1, u_2 \to v_1, v_2 \) is \( Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 0 \end{pmatrix} \).

We now come to the main result of this section which answers the following question: given a linear map and its matrix in some basis and given a new basis, how do we compute the matrix of that same linear map in the new basis?
Figure 12: On the left: writing $T = \text{id} \circ T \circ \text{id}^{-1}$. On the right: writing $B = QAQ^{-1}$. Note how each of the linear maps on the left has a corresponding matrix on the right.

**Theorem 2.27** (Change of basis). Let $T : V \to V$ be a linear map, let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be bases of $V$, let $A$ and $B$ be the respective matrices of $T$ with respect to these bases, and let $Q$ be the change of basis matrix from $u_1, \ldots, u_n$ to $v_1, \ldots, v_n$. Then

$$B = QAQ^{-1}.$$  

**Proof.** This follows from Theorem 2.19 (which tells us that the matrix of a composition is the product of the matrices) since $T = \text{id} \circ T \circ \text{id}^{-1}$ and so we deduce that $B = QAQ^{-1}$, as depicted in Figure 12. 

Unsurprisingly, two matrices which represent the same linear map, but with respect to two different bases, will share many properties in common. Such matrices are therefore given a name and called *similar*.

**Definition 2.28** (Similar matrices). Let $A$ and $B$ be square matrices of the same size. We say that $A$ and $B$ are similar if there exists an invertible matrix $Q$ such that $B = QAQ^{-1}$.
3 Elementary row operations and systems of linear equations

As we enter a new section of the course it is worth taking a step back and taking stock of where we've been and where we are going. Remember: the goal of linear algebra is to provide tools that can be used to better understand linear transformations. In that regard, a key take-away of Section 2 is that, once bases are chosen, every linear map can be identified with a matrix.

We have then seen a few examples of how properties (such as invertibility) of the linear map could then be deduced by only considering its matrix. We will push this idea further here. Essentially we will see that purely arithmetic computations in matrix-world can be translated back into meaningful statements about the linear maps they represent.

A set of particularly useful computations to carry out are so-called elementary row operations, which we now introduce and discuss.

3.1 Elementary row operations

Motivating example 3.1 (Elementary row operations arise when solving linear systems by elimination). We will use so-called elementary row operations to deduce useful information about matrices. First, however, we discuss where these operations come from, namely from solving linear systems by elimination.

Consider the linear system

\[
\begin{align*}
    x + 2y + 3z & = -1, \\
    x + 2y + 2z & = -2, \quad \text{and} \\
    x + 3y + 4z & = -2.
\end{align*}
\]

Subtracting the first equations from the second and third equations yields

\[
\begin{align*}
    x + 2y + 3z & = -1, \\
    -z & = -1, \quad \text{and} \\
    y + z & = -1.
\end{align*}
\]

Swapping the second and third rows tells us that

\[
\begin{align*}
    x + 2y + 3z & = -1, \\
    y + z & = -1, \quad \text{and} \\
    -z & = -1.
\end{align*}
\]

Finally we multiply the last equation by $-1$, which tells us that

\[
z = 1.
\]

We are now in a great position to solve the linear system: plugging the equation above into the second equation of the system tells us that

\[
y = -1 - z = -2
\]
which we may then plug into the first equation of the system to deduce that

\[ x = -1 - 2y - 3z = 0. \]

We may also write the procedure above in matrix notation, making it clearer that the operations we carried out to simplify the system are all row operations. We begin with the so-called augmented matrix

\[
\begin{pmatrix}
1 & 2 & 3 & -1 \\
1 & 2 & 2 & -2 \\
1 & 3 & 4 & -3
\end{pmatrix}
\]

whose entries fully describe the linear system considered above. The row operations carried out above can then be viewed as:

\[
\begin{pmatrix}
1 & 2 & 3 & -1 \\
1 & 2 & 2 & -2 \\
1 & 3 & 4 & -3
\end{pmatrix}
\xrightarrow{R_3 \rightarrow R_3 - R_1 - R_2}
\begin{pmatrix}
1 & 2 & 3 & -1 \\
0 & 0 & -1 & -1 \\
0 & 1 & 1 & -1
\end{pmatrix}
\xrightarrow{R_2 \leftrightarrow R_3}
\begin{pmatrix}
1 & 2 & 3 & -1 \\
0 & 1 & 1 & -1 \\
0 & 0 & 1 & 1
\end{pmatrix}
\xrightarrow{R_3 \rightarrow -R_3}
\begin{pmatrix}
1 & 2 & 3 & -1 \\
0 & 1 & 1 & -1 \\
0 & 0 & 1 & 1
\end{pmatrix}
\]

Note that, in the example above, we have only used three types of row operations, either swapping rows, multiplying a row by a (nonzero) scalar, or linearly combining some rows together. These are precisely the so-called elementary row operations.

**Definition 3.2 (Elementary row operations).** Given a matrix \( A \) the following row operations are called elementary:

(a) swapping two rows,

(b) multiplying a row by a nonzero scalar, and

(c) adding a scalar multiple of a row to another row.

It turns out that elementary row operations can be represented as left-multiplication by invertible matrices. This will be tremendously useful whenever we seek to prove anything about a matrix obtained after a sequence of row operations.

**Theorem 3.3 (Elementary row operations as left-multiplication).** Let \( A \) be an \( n \)-by-\( m \) matrix and let \( B \) be the matrix obtained by performing an elementary row operation on \( A \). If \( E \) denotes the matrix obtained by performing the same row operation on the identity matrix \( I_n \) then \( B = EA \). Moreover \( E \) is invertible.
Proof. We omit the proof of the first part since it comes down to a particularly tedious mess with indices, although the principle is clear by looking at a few examples (see below). The second part then follows immediately since every elementary row operation can be undone by another elementary row operation of the same type.

The matrices $E$ appearing in Theorem 3.3 as the left-multiplication representative of elementary row operations are so useful that they are given their own name.

Definition 3.4 (Elementary matrix). A matrix obtained by performing an elementary row operation on the identity matrix is called an elementary matrix.

Now let us rephrase the motivating example above in terms of left-multiplication by these elementary matrices.

Example 3.5 (Elementary matrices in action). Each of the steps of Example 3.1 can be written as left-multiplication by the appropriate elementary matrix. For instance, subtracting the first row from the second and third rows may be written as

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
-1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
0 & 1 & 1
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 \\
-1 & 2 & 2 \\
0 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
0 & 0 & -1 \\
0 & 1 & 1
\end{pmatrix}
\]

while swapping the last two rows takes the form

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
0 & 0 & -1 \\
0 & 1 & 1
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & -1
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & -1
\end{pmatrix}
\]

and finally multiplying the last row by $-1$ looks like

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & -1
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{pmatrix}
\]

Putting it all together we see that we may write the simplified form of the augmented matrix as the initial augmented matrix left-multiplied by a sequence of augmented matrices:

\[
\begin{pmatrix}
1 & 2 & 3 \\
0 & 1 & 1 \\
0 & 0 & 1
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
-1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 \\
1 & 2 & 2 \\
1 & 3 & 4
\end{pmatrix}
\]

3.2 Computing the rank via elementary row operations

Remember: we introduced elementary row operations in order to be able to deduce various properties of linear maps from computations involving their associated matrices. As an example of this procedure we will, in this section,
discuss how to compute the rank of a linear map by performing elementary row operations on its matrix.

In a nutshell, we will show that for a linear map $T$ whose matrix is $A$, the following holds: (where $E_i$ denotes elementary matrices corresponding to elementary row operations)

$$\text{rank } T = \text{rank } A = \text{rank } (E_n \ldots E_1 A).$$

Note that, throughout this and the following section we will write things like $\text{rank } A$, $\text{im } A$, $\text{ker } A$, etc. which imply that we think of $A$ as a linear map. This is perfectly legitimate since $A$ can be viewed as linear map acting from $\mathbb{F}^m$ to $\mathbb{F}^n$.

To get started we record a useful characterisation of the rank of a linear map.

**Lemma 3.6.** Let $T : V \rightarrow W$ be a linear map and let $v_1, \ldots, v_n$ be a basis of $V$. The rank of $T$ is equal to the size of the largest linearly independent subset of $\{T(v_1), \ldots, T(v_n)\}$.

**Proof.** This follows from the observation that the vectors $T(v_1), \ldots, T(v_n)$ span $\text{im } T$ and so may be reduced to a basis of $\text{im } T$. $\square$

Since we will ultimately prove that $\text{rank } T = \text{rank } A$ if $A$ is a matrix associated with a linear map $T$, and since we have a useful characterization of the rank of a linear map thanks to the lemma above, we now seek a useful description of the rank of a matrix. We begin with an elementary observation which we can build on.

**Lemma 3.7.** Let $A$ be an $n$-by-$m$ matrix over a field $\mathbb{F}$ and let $x \in \mathbb{F}^m$. Then

$$Ax = x_1 C_1(A) + \cdots + x_m C_m(A)$$

where $C_i(A) \in \mathbb{F}^n$ denotes the $i$-th column of $A$.

**Proof.** This follows from a simple computation:

$$\begin{pmatrix} a_{11} & \ldots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_{11} x_1 + \cdots + a_{1m} x_m \\ \vdots \\ a_{n1} x_1 + \cdots + a_{nm} x_m \end{pmatrix} = x_1 C_1(A) + \cdots + x_m C_m(A).$$

$\square$

The lemma above allows us to better understand the image of $A$, which will naturally lead us to a useful description of the rank of $A$.

**Corollary 3.8.** The image of a matrix $A$ is the span of the columns of $A$.

**Proof.** Since $\text{im } A = \{Ax : x \in \mathbb{F}^m\}$ the claim follows immediately from Lemma 3.7 immediately above. $\square$
We are now ready to prove that the rank of a linear map and the rank of its matrix are one and the same.

**Theorem 3.9.** Let $V$ and $W$ be vector spaces with respective bases $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$, let $T : V \to W$ be a linear map, and let $A$ be its corresponding matrix with respect to these bases. Then $\text{rank } T = \text{rank } A$.

**Proof.** The $j$-th column of $A$ corresponds to the coordinates of $T(v_j)$ in the basis $w_1, \ldots, w_n$, so, for any scalars $a_1, \ldots, a_m$,

$$a_1 T(v_1) + \cdots + a_m T(v_m) = 0 \iff a_1 C_1(A) + \cdots + a_m C_m(A) = 0.$$ 

Therefore a subset of $\{T(v_1), \ldots, T(v_m)\}$ is linearly independent if and only if the corresponding subset of $\{C_1(A), \ldots, C_m(A)\}$ is linear independent. So finally, by Lemma 3.6 and Corollary 3.8,

$$\text{rank } T = \text{size of the largest linearly independent subset of } \{T(v_1), \ldots, T(v_m)\} = \text{size of the largest linearly independent subset of } \{C_1(A), \ldots, C_m(A)\} = \text{rank } A. \quad \Box$$

Now that we know that the rank of a linear map is equal to the rank of its matrix, it remains to make sure that we can indeed compute the rank of a matrix easily. The way to do that is recorded below: we can perform elementary row operations, since they preserve the rank. It will then suffice to read off the rank of the matrix once elementary row operations have greatly simplified it.

**Theorem 3.10.** Let $A$ be an $n$-by-$m$ matrix and let $E$ be an $n$-by-$n$ elementary matrix. Then $\text{rank } A = \text{rank}(EA)$.

**Proof.** For any scalars $a_1, \ldots, a_m$, since elementary matrices are invertible we see that

$$a_1 C_1(A) + \cdots + a_m C_m(A) = 0 \iff A \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = 0 \iff EA \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = 0 \iff a_1 C_1(EA) + \cdots + a_m C_m(EA) = 0.$$ 

Moreover, as proved by Lemma 3.6 and Corollary 3.8, we have that for any matrix $B$

$$\text{rank } B = \text{size of the largest linearly independent subset of the columns of } B.$$ 

By applying this to $B = A$ and $B = EA$ and using the observation above concerning their columns we deduce that indeed the rank of $A$ is equal to the rank of $EA$. \quad \Box
3.3 Linear systems: Theoretical aspects

This section and the next form a bit of an interlude from our endeavours to understand linear maps better using computations involving their matrices. In this section and the next we will focus solely on linear systems.

The setup is the following: given an \( n \)-by-\( m \) matrix \( A \) over a field \( F \) and a vector \( b \in F^n \) we look for \( x \in F^m \) such that \( x \) solves \( Ax = b \), i.e.

\[
\begin{align*}
    a_{11}x_1 + \cdots + a_{1m}x_m &= b_1 \\
    \vdots \\
    a_{n1}x_1 + \cdots + a_{nm}x_m &= b_m
\end{align*}
\]

In this section we will discuss some theoretical aspects of linear systems, answering the two questions below.

- What does the set of solutions of \( Ax = b \) look like?
- When can we guarantee that \( Ax = b \) has a unique solution?

In the next section we will discuss some computational aspects of linear systems, namely discussing a recipe allowing us to systematically solve any linear system.

**Motivating example 3.11** (The set of solutions to a linear system). Consider

\[
A = \begin{pmatrix}
2 & -2 & -1 \\
-1 & 1 & 1 \\
1 & -1 & 1
\end{pmatrix}
\quad \text{and} \quad
b = \begin{pmatrix}
1 \\
0 \\
2
\end{pmatrix}.
\]

We seek to describe the set of solutions of \( Ax = b \). One can compute (we will discuss how to do this systematically in the next section) that any solution \( x \) of \( Ax = b \) takes the form

\[
x = \begin{pmatrix}
1 \\
0 \\
1
\end{pmatrix} + a \begin{pmatrix}
1 \\
1 \\
0
\end{pmatrix}
\]

for some \( a \in \mathbb{R} \), where \( x_* \) is called a particular solution and \( x_0 \) is called a homogeneous solution.

Indeed, we note that \( Ax_* = b \) while \( Ax_0 = 0 \), and hence that \( Ax = b \). The set of solutions can therefore be viewed as a translation of the kernel of \( A \) (to which \( x_0 \) belongs), translated by \( x_* \). This is depicted in Figure 13.

The fact that the solution to a linear system can be split into two parts, namely a particular solution and a homogeneous solution, is a general feature of linear systems, as recorded below.

**Theorem 3.12** (The solution of a linear system). Let \( A \) be an \( n \)-by-\( m \) matrix over a field \( F \) and let \( b \in F^n \). If \( x_* \) is a solution of \( Ax = b \) then the set of all solutions of \( Ax = b \) is given by

\[
x_* + \ker A,
\]
Figure 13: The set of solutions to $Ax = b$ for $A$ and $b$ as in Example 3.11.

i.e. $\{ x : Ax = b \} = \{ x_\ast + x_0 : x_0 \in \ker A \}$. We call $x_\ast$ a particular solution and $x_0 \in \ker A$ a homogeneous solution.

Proof. First we show that $\{ x : Ax = b \} \subseteq \{ x_\ast + x_0 : x_0 \in \ker A \}$. Let $x$ solve $Ax = b$. Then $x - x_\ast$ belongs to the kernel of $A$ since

$$A(x - x_\ast) = Ax - Ax_\ast = b - b = 0.$$  

So indeed $x = x_\ast + x_0$ for some $x_0 \in \ker A$.

Second we show that $\{ x_\ast + x_0 : x_0 \in \ker A \} \subseteq \{ x : Ax = b \}$. Let $x_0 \in \ker A$. Then

$$A(x_\ast + x_0) = Ax + Ax_0 = b + 0 = b$$

such that indeed $x = x_\ast + x_0$ solves $Ax = b$. \qedhere

Above we answered the question: what does the set of solutions to $Ax = b$ look like? We now answer the second question we sought to answer in this section, namely: when can we guarantee that $Ax = b$ has a unique solution.

**Theorem 3.13** (Invertibility and unique solvability). Let $A$ be a $n$-by-$m$ matrix over a field $\mathbb{F}$. $A$ is invertible if and only if, for every $b \in \mathbb{F}^n$, there is a unique solution of $Ax = b$.

Proof. Suppose that $A$ is invertible. Then (by applying $A^{-1}$ to go from left to right and applying $A$ to go from right to left)

$$Ax = b \iff x = A^{-1}b.$$  

So the unique solution is $x = A^{-1}b$.

Conversely, suppose that, for every $b \in \mathbb{F}^n$, there is a unique solution of $Ax = b$. Viewed as a linear map, $A : \mathbb{F}^m \to \mathbb{F}^n$ satisfies

1. $\text{im } A = \mathbb{F}^n$ since, for every $b \in \mathbb{F}^n$, there exists $x \in \mathbb{F}^m$ such that $Ax = b$ and
2. \( \ker A = \{0\} \) since \( Ax = 0 \) has only one solution, namely \( x = 0 \).

Since \( \text{im} \ A = \mathbb{F}^n \) and \( \ker A = \{0\} \) we deduce that \( A \) is invertible (this was proved in assignment 6).

3.4 Linear systems: Computational aspects

As promised, after discussing theoretical aspects of linear systems in the previous section we now turn our attention towards computational aspects. Specifically, we will answer the following question:

- Given a matrix \( A \) and a vector \( b \), how we actually find a solution \( x \) of \( Ax = b \).

The answer will come by way of elementary row operations which will turn the augmented matrix \( (A|b) \) into a much simpler-looking matrix. We give a name to the form that we wish that matrix to take.

**Definition 3.14** (Upper echelon form and reduced upper echelon form). A matrix is said to be in upper echelon form if

1. all zero rows are below all non-zero rows and
2. the first non-zero entry of each row is strictly to the right of the first non-zero entry of the row above.

A matrix is said to be in reduced upper echelon form if it is in upper echelon form and

3. the first non-zero entry of each row is 1 and
4. all entries in the same column as the first non-zero entry of a row are zero.

**Example 3.15** (Upper echelon forms and reduced upper echelon forms). Here are examples and non-examples of matrices in upper echelon form and reduced upper echelon form.

1. The matrix

\[
\begin{pmatrix}
2 & 1 & 3 \\
0 & 1 & 2 \\
0 & 0 & 0
\end{pmatrix}
\]

is in upper echelon form but not in reduced upper echelon form (since the first non-zero entry of the second row is not equal to 1, i.e. condition (3) fails in Definition 3.14).

2. The matrix

\[
\begin{pmatrix}
0 & 0 & 0 & 0 \\
1 & 2 & 1 & 1 \\
0 & 0 & 1 & -1
\end{pmatrix}
\]

is not in upper echelon form (since there is a zero row above some non-zero rows, i.e. condition (1) fails in Definition 3.14).
3. The matrix
\[
\begin{pmatrix}
1 & 2 & 0 & 2 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]
is in upper echelon form but not in reduced upper echelon form (since the entries in the fourth column above the first non-zero entry of the third row are non-zero, i.e. condition (4) fails in Definition 3.14).

4. The matrix
\[
\begin{pmatrix}
1 & 2 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\]
is in reduced upper echelon form.

We are now ready to state the recipe we will use to systematically solve linear systems.

**Recipe 3.16** (Systematically solving linear systems). To solve $Ax = b$ we put the augmented matrix $(A|b)$ in reduced upper echelon form.

**Example 3.17** (Solving linear systems). Here are some examples of linear systems being solved using Recipe 3.16.

1. Consider the matrix
   \[
   A = \begin{pmatrix}
   1 & -1 & 2 & 4 \\
   2 & -2 & 1 & 5 \\
   -1 & 1 & 3 & 1
   \end{pmatrix}
   \quad \text{and} \quad
   b = \begin{pmatrix}
   2 \\
   6 \\
   -3
   \end{pmatrix}
   \]

   We wish to solve $Ax = b$. We begin with the augmented matrix $(A|b)$ and perform elementary row operations on $A$ in order to put it in reduced upper echelon form.

   \[
   \begin{pmatrix}
   1 & -1 & 2 & 4 & 3 \\
   2 & -2 & 1 & 5 & 6 \\
   -1 & 1 & 3 & 1 & -3
   \end{pmatrix}
   \xrightarrow{R_2 \rightarrow R_2 - 2R_1}
   \begin{pmatrix}
   1 & -1 & 2 & 4 & 3 \\
   0 & 0 & -3 & -3 & 0 \\
   0 & 0 & 5 & 5 & 0
   \end{pmatrix}
   \xrightarrow{R_3 \rightarrow R_3 + R_1}
   \begin{pmatrix}
   1 & -1 & 2 & 4 & 3 \\
   0 & 0 & -3 & -3 & 0 \\
   0 & 0 & 5 & 5 & 0
   \end{pmatrix}
   \xrightarrow{R_2 \rightarrow \frac{1}{4}R_2}
   \begin{pmatrix}
   1 & -1 & 2 & 4 & 3 \\
   0 & 0 & 1 & 1 & 0 \\
   0 & 0 & 0 & 0 & 0
   \end{pmatrix}
   \xrightarrow{R_3 \rightarrow R_3 + \frac{1}{5}R_2}
   \begin{pmatrix}
   1 & -1 & 0 & 2 & 3 \\
   0 & 0 & 1 & 1 & 0 \\
   0 & 0 & 0 & 0 & 0
   \end{pmatrix}
   \xrightarrow{R_1 \rightarrow R_1 - 2R_2}
   \begin{pmatrix}
   1 & -1 & 0 & 2 & 3 \\
   0 & 0 & 1 & 1 & 0 \\
   0 & 0 & 0 & 0 & 0
   \end{pmatrix}
   \]

   Translating this back into the language of a linear system we see that $Ax = b$ can be simplified to
   \[
   x_1 - x_2 + 2x_4 = 3 \quad \text{and} \quad
   x_3 + x_4 = 0
   \]
The second equation has two unknowns \((x_3\text{ and } x_4)\), so one of them gets to be chosen for free. Let’s write this as \(x_4 = b\) for \(\text{any } b \in \mathbb{R}\). Similarly, once we plug \(x_4 = b\) into the first equation, there are two unknowns left \((x_1\text{ and } x_2)\), so once again one of these unknowns gets to be chosen for free – we write \(x_2 = a\) for \(\text{any } a \in \mathbb{R}\).

What did we just discover? By setting \(x_2 = a\) and \(x_4 = b\) we deduce that

\[
\begin{align*}
  x_1 &= 3 + a - 2b \\
  x_3 &= -b,
\end{align*}
\]

i.e.

\[
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4
\end{pmatrix} = \begin{pmatrix} 3 \end{pmatrix} + a \begin{pmatrix} 1 \\ -1 \end{pmatrix} + b \begin{pmatrix} -2 \\ 0 \end{pmatrix}.
\]

This means that the set of solutions of \(Ax = b\) is a two-dimensional \textit{plane} of solutions in \(\mathbb{R}^2\).

Let’s stop for a second and think about how we read off the solution from the reduced upper echelon form.

- Each column which does \textit{not} have a leading non-zero entry contributes to the \textit{kernel} of \(A\) in the form of the corresponding unknown being chosen to be equal to some arbitrary parameter (called \(a\) and \(b\) above) – this was the case for the second and fourth columns above.
- Each column with \textit{does} have a leading non-zero entry is entirely determined in terms of a constant term and the parameters – this was the case for the first and third columns above.

2. Consider the same matrix \(A\) as above but with \(b = \begin{pmatrix} 3 \\ 6 \\ -2 \end{pmatrix}\). Performing the \textit{same} elementary row operations as above we deduce that

\[
(A|b) \rightarrow \begin{pmatrix} 1 & -1 & 2 & 4 & 3 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.
\]

We are in serious trouble: the last row of the matrix above corresponds to the equation \(0 = 1!\) This is impossible! This tells us that the system \(Ax = b\) is \textit{inconsistent} and has no solutions.

3. Consider now the matrix

\[
A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.
\]
We skip the details and note that, after some elementary row operations, the augmented matrix \((A|b)\) becomes

\[
\begin{pmatrix}
1 & 2 & 0 & 1 \\
2 & 1 & 0 & 2 \\
-1 & 3 & 1 & 3
\end{pmatrix}
\rightarrow
\begin{pmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 4
\end{pmatrix}
\]

which tells us that \(Ax = b\) has a unique solution given by \(x = \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix}\).

4 Determinants

In the previous section we saw a few examples of how computations involving matrices could be used to deduce properties of the corresponding linear maps. In this section we introduce a new object which will also allow us to turn computations about matrices into insights about linear maps, namely the determinant. Determinants will be a way to assign, to any square matrix, a scalar. Since this definition is somewhat involved for generic \(n\)-by-\(n\) matrices we first consider the case of 2-by-2 matrices.

4.1 Determinants of 2-by-2 matrices

Definition 4.1 (Determinant of a 2-by-2 matrix). The determinant of a 2-by-2 matrix \(A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\), denoted \(\det A\), is defined as \(\det A = ad - bc\).

Example 4.2. The determinants of \(A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\) and \(B = \begin{pmatrix} 3 & 2 \\ 6 & 4 \end{pmatrix}\) are \(\det A = 4 - 6 = -2\) and \(\det B = 12 - 12 = 0\) but the determinant of \(A + B\) is not equal to the sum of these determinants since \(\det(A + B) = \det \begin{pmatrix} 4 & 4 \\ 9 & 8 \end{pmatrix} = 8 \cdot 4 - 9 \cdot 4 = -4 \neq \det A + \det B\).

The example above shows that the determinant, as a function from the space of 2-by-2 matrices to the underlying field, is not linear. That being said, there is a way to view the determinant as a linear map, and that is to vary only one column of the matrix at a time. This and another very important property of 2-by-2 determinants is recorded below. In order to phrase these properties properly we must recall first that, for any two sets \(A\) and \(B\), their product \(A \times B\) is defined as \(A \times B = \{(a, b) : a \in A\ and\ b \in B\}\).
Theorem 4.3 (Fundamental properties of 2-by-2 determinants). Let $F$ be a field. Viewed as a map $\det : F^2 \times F^2 \to F$, i.e. for $u, v \in F^2$ we write $\det(u|v)$ to mean $\det \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$, the determinant satisfies

1. bilinearity, i.e. $\det(\cdot|v)$ and $\det(u|\cdot)$ are linear for any fixed $u, v \in F^2$ and

2. anti-symmetry, i.e. $\det(u|v) = -\det(v|u)$ for any $u, v \in F^2$.

Proof. We first show that (2) holds, which follows from a direct computation:

$$\det(u|v) = \det \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix} = u_1 v_2 - u_2 v_1 = -(v_1 u_2 - v_2 u_1) = -\det \begin{pmatrix} v_1 & u_1 \\ v_2 & u_2 \end{pmatrix} = -\det(v|u).$$

Now, using (2) we note that, in order to prove (1) it suffices to prove that $\det(\cdot|v)$ is linear since then, by (2),

$$\det(u|\cdot) = -\det(\cdot|u)$$

is also linear. So let us show that $\det(\cdot|v)$ preserves vector addition and scalar multiplication.

• For any $u, w \in F^2$,

$$\det(u+w|v) = \det \begin{pmatrix} u_1 + w_1 & v_1 \\ u_2 + w_2 & v_2 \end{pmatrix} = (u_1 + w_1)v_2 - (u_2 + w_2)v_1 = (u_1 v_2 - u_2 v_1) + (w_1 v_2 - w_2 v_1) = \det(u|v) + \det(w|v).$$

• For any $s \in F$ and any $u \in F^2$,

$$\det(su|v) = \det \begin{pmatrix} su_1 & v_1 \\ su_2 & v_2 \end{pmatrix} = su_1 v_2 - su_2 v_1 = s(u_1 v_2 - u_2 v_1) = s\det(u|v)$$

This shows that $\det(\cdot|v)$ is linear for any fixed $v \in F^2$, which concludes the proof.

As mentioned at the start of this section, determinants will help turn computations involving matrices into insights concerning their corresponding linear map. The result below is an example of this where computing the determinant tells us whether or not a matrix (and hence its associated linear map) is invertible.

Theorem 4.4. A 2-by-2 matrix $A$ is invertible if and only if $\det A \neq 0$. 

39
Proof. If \( \det A \neq 0 \) then we may define
\[
B = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.
\]
It is then a straightforward computation (which we omit here) to show that \( AB = BA = I \), which proves that \( B \) is the inverse of \( A \) and thus that \( A \) is invertible.

Conversely, suppose that \( \det A = 0 \). Writing \( A = (u|v) \) this means that \( u_1v_2 = u_2v_1 \). We may now split into two cases.

- **Case 1**: \( u_1 = 0 \) or \( v_1 = 0 \). If \( u_1 = 0 \) then, since \( u_1v_2 = u_2v_1 \), we deduce that either \( u_2 = 0 \) or \( v_1 = 0 \), which means that one of the rows or one of the columns of \( A \) is equal to zero. Either way this tells us that rank \( A \leq 1 \), and hence that \( A \) is not invertible.

  Similarly: if \( v_1 = 0 \) then \( u_1 = 0 \) or \( v_2 = 0 \) and hence, as above, rank \( A \leq 1 \) and \( A \) fails to be invertible.

- **Case 2**: \( u_1 \) and \( v_1 \) are nonzero. Then we may deduce from \( u_1v_2 = u_2v_1 \) that
\[
\frac{u_2}{u_1} = \frac{v_2}{v_1} = s
\]
for some scalar \( s \). Therefore \( A \) takes the form
\[
A = (u|v) = \begin{pmatrix} u_1 & v_1 \\ su_1 & sv_1 \end{pmatrix}
\]
such that the columns of \( A \) are linearly dependent (they are scalar multiples of one another). This means that nullity \( A \geq 1 \), and hence \( A \) is not invertible.

In both cases we have deduced that \( A \) is not invertible, which concludes the proof.

Having established various properties of 2-by-2 determinants, an important question: what is the interpretation of the determinant? And how would anyone think of defining such a thing. The answer lies in the geometric interpretation of 2-by-2 determinants provided below.

**Motivating example 4.5** (A geometric interpretation of 2-by-2 determinants). Define, for \( u, v \in \mathbb{R}^2 \),
\[
\text{Area}(u, v) = \text{signed area of the parallelogram generated by } u \text{ and } v.
\]
Then the function \( \text{Area}(\cdot, \cdot) : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R} \) is bilinear (as we will prove below) and anti-symmetric (this is by definition of what we mean by a *signed* area: the sign of the area changes when we change the orientation by swapping the two vectors).

40
Therefore, for any \( u, v \in \mathbb{R}^2 \), if we denote by \( e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \) and \( e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \) the standard basis vectors of \( \mathbb{R}^2 \) then we see that

\[
\text{Area} (u, v) = \text{Area} (u_1 e_1 + u_2 e_2, v) = u_1 \text{Area} (e_1, v) + u_2 \text{Area} (e_2, v) \\
= u_1 v_1 \text{Area} (e_1, e_1) + u_1 v_2 \text{Area} (e_1, e_2) + u_2 v_1 \text{Area} (e_2, e_1) + u_2 v_2 \text{Area} (e_2, e_2)
\]

where, by anti-symmetry of \( \text{Area} (\cdot, \cdot) \),

\[
\text{Area} (e_1, e_1) = \text{Area} (e_2, e_2) = 0 \quad \text{and} \quad \text{Area} (e_2, e_1) = -\text{Area} (e_1, e_2)
\]

such that

\[
\text{Area} (u, v) = (u_1 v_2 - u_2 v_1) \underbrace{\text{Area} (e_1, e_2)}_{=1} = \det(u|v).
\]

In other words: since the signed area is bilinear and antisymmetric, just like 2-by-2 determinants, it follows that the determinant of \((u|v)\) is precisely the signed area of the parallelogram generated by \( u \) and \( v \)!

It remains to verify that \( \text{Area} (\cdot, \cdot) \) is bilinear. This is depicted in Figure 14.

### 4.2 Interlude: Permutations

Having discussed 2-by-2 determinants we seek to extend that discussion to the case of \( n \)-by-\( n \) determinants, i.e. determinants of square matrices of arbitrary size. However, in order to be able to do that in a clean way we will need to first introduce permutations and their sign.

**Motivating example 4.6** (Determinants and permutations). When computing the determinant of a 2-by-2 matrix, i.e.

\[
\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc
\]

we see that this determinant is the sum of products of entries of the matrix coming from *distinct* rows and columns (where each product is then provided a sign in an appropriate way).

In general, therefore, in order to list all of these possible products we must find all the way to lists all of the numbers \( 1, \ldots, n \) *once and exactly one*. In other words: we are looking for an invertible map from \( \{1, \ldots, n\} \) to itself. This motivates our definition of a permutation below.

**Definition 4.7** (Permutation). A *permutation* is an invertible map \( \sigma : \{1, \ldots, n\} \to \{1, \ldots, n\} \). The set of all permutations of \( \{1, \ldots, n\} \) is denoted by \( S_n \).

**Example 4.8** (Permutations). Here are examples of permutations. In particular we list all of the elements of \( S_2 \) and \( S_3 \).
Figure 14: A sketch arguing that the signed area is bilinear. The key is that the area generated by \( u + w \) and \( v \) (in \textcolor{orange}{orange}) can be rearranged into the union of the areas generated by \( u \) and \( w \) and by \( v \) and \( w \) (in \textcolor{pink}{pink} and \textcolor{green}{green}, respectively) by cutting and pasting a triangle (in \textcolor{blue}{blue}).
1. $S_2$ contains two elements: the identity and the permutation which swaps 1 and 2, i.e. the two permutations are given by

\[
\begin{align*}
1 & \rightarrow 1 \\
2 & \rightarrow 2 \\
\end{align*}
\quad \text{and} \quad 
\begin{align*}
1 & \rightarrow 1 \\
2 & \rightarrow 2 \\
\end{align*}
\]

2. $S_3$ contains six elements, all represented below.

\[
\begin{array}{ccc}
1 & \rightarrow 1 & 1 & \rightarrow 1 & 1 & \rightarrow 1 \\
2 & \rightarrow 2 & 2 & \rightarrow 2 & 2 & \rightarrow 2 \\
3 & \rightarrow 3 & 3 & \rightarrow 3 & 3 & \rightarrow 3 \\
\end{array}
\]

3. Here is one element of $S_5$:

\[
\begin{array}{ccc}
1 & \rightarrow 1 & 1 & \rightarrow 1 & 1 & \rightarrow 1 \\
2 & \rightarrow 2 & 2 & \rightarrow 2 & 2 & \rightarrow 2 \\
3 & \rightarrow 3 & 3 & \rightarrow 3 & 3 & \rightarrow 3 \\
\end{array}
\]

As we will soon see, some permutations are particularly important. They are introduced below.

**Definition 4.9 (Cycles and transpositions).** Let $i_1, \ldots, i_k$ be distinct elements of $\{1, \ldots, n\}$. The $k$-cycle denoted by $\sigma = (i_1 \ldots i_k)$ is the element $\sigma$ of $S_n$ defined by

\[
\begin{cases}
\sigma(i_j) = i_{j+1} & \text{if } j < k \\
\sigma(i_k) = i_1.
\end{cases}
\]

2-cycles are called *transpositions*.

**Example 4.10 (Cycles and transpositions).** Here are examples of cycles, transpositions, and compositions of the two.

1. The 4-cycle $\sigma = (1423) \in S_5$ corresponds to

\[
\begin{array}{ccc}
1 & \rightarrow 1 & 1 & \rightarrow 1 & 1 & \rightarrow 1 \\
2 & \rightarrow 2 & 2 & \rightarrow 2 & 2 & \rightarrow 2 \\
3 & \rightarrow 3 & 3 & \rightarrow 3 & 3 & \rightarrow 3 \\
\end{array}
\]
2. The permutation $\sigma = (132)(45)$ is the composition of a transposition and a 3-cycle. It is actually a permutation we have seen before, namely in Example 4.8, since it corresponds to

\[
\begin{array}{c}
1 \\
2 \\
3 \\
4 \\
5 \\
\end{array} \quad \xrightarrow{\sigma} \quad \begin{array}{c}
1 \\
2 \\
3 \\
4 \\
5 \\
\end{array}
\]

The reason we care about cycles and transpositions is because all permutations can be decomposed into cycles, and all cycles can be decomposed into transpositions (such that all permutations can be decomposed into transpositions). We state these facts below without proof.

**Fact 4.11** (Decomposition of permutations). *The following hold.*

1. All permutations are compositions of cycles.
2. All cycles are compositions of transpositions.
3. Any permutation can be written as a composition of an odd or even number of transpositions (but not both).

In particular, since all permutations can be written as a composition of an odd or even number of transpositions, this allows us to define the *sign* of a permutation. This is essential to allow us to define determinants: remember that we want determinants to be sum of products taken over entries coming from distinct rows and columns, *with an appropriate sign* associated to each product. This “appropriate sign” is precisely the sign of the corresponding permutation.

**Definition 4.12** (Sign of a permutation). A permutation $\sigma$ is called *even* (or respectively *odd*) if it can be written as a composition of an even (or respectively odd) number of transpositions. We define

\[
\text{sign } \sigma = \begin{cases} 
+1 & \text{if } \sigma \text{ is even and} \\
-1 & \text{if } \sigma \text{ is odd.}
\end{cases}
\]

**Example 4.13.** Recall that $S_2 = \{\text{id}, \sigma\}$ where

\[
\begin{cases} 
\sigma(1) = 2 \text{ and} \\
\sigma(2) = 1.
\end{cases}
\]

Then

\[
\begin{cases} 
\text{sign id} = 1 \text{ and} \\
\text{sign } \sigma = -1.
\end{cases}
\]
The key observation that will conclude this section and allow us to define determinants in the next section is the following. For a 2-by-2 matrix $A$,

$$\sum_{\sigma \in S_2} \text{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} = a_{11} a_{22} - a_{12} a_{21} = \det A.$$ 

In other words: we can use permutations, and their sign, to define the determinant!

### 4.3 Determinants of $n$-by-$n$ matrices

The entire purpose of the previous section, in which we introduced permutations, was to allow us to make sense of determinants of $n$-by-$n$ matrices. We can now do that by extending the formula obtained at the end of Example 4.13 above to the case of $n$-by-$n$ matrices.

**Definition 4.14** ($n$-by-$n$ determinant). The determinant of an $n$-by-$n$ matrix $A$ is defined to be

$$\det A = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$ 

**Example 4.15.** We consider the cases of 2-by-2 and 3-by-3 determinants explicitly.

1. When $n = 2$, $S_2 = \{\text{id}, (12)\}$, and hence

$$\det A = \text{sign(id)} a_{11} a_{22} + \text{sign(12)} a_{12} a_{21} = a_{11} a_{22} - a_{12} a_{21},$$

as before!

2. When $n = 3$, $S_3 = \{\text{id}, (12), (13), (23), (123), (132)\}$, and hence

$$\det A = a_{11} a_{22} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - a_{12} a_{21} a_{33} - a_{13} a_{22} a_{31} - a_{11} a_{23} a_{32}.$$ 

For example, if

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

then

$$\det A = 1 \cdot 1 \cdot 1 + 3 \cdot 4 \cdot 1 + 2 \cdot 2 \cdot 1 - 1 \cdot 1 \cdot 2 \cdot 4 \cdot 1 - 2 \cdot 3 \cdot 1 = 1 + 12 + 4 - 1 - 8 - 6 = 2.$$ 

We are now ready to prove elementary properties of determinants such as its linearity in each row and its anti-symmetry under row swaps, from which we can deduce the effect that elementary row operations have on determinants.
Theorem 4.16 (Linearity in the rows and the effect of elementary row operations on determinants). Let $A$ be an $n$-by-$n$ matrix.

1. $\det A$ is linear in the rows of $A$. In particular if $B$ is obtained by multiplying a row of $A$ by a scalar $s$ then $\det B = s \det A$.

2. If $B$ is obtained by swapping two rows of $A$ then $\det B = -\det A$.

3. If $A$ has two equal rows then $\det A = 0$.

4. If $B$ is obtained by adding a scalar multiple of one row of $A$ to another row of $A$ then $\det B = \det A$.

Proof. We will prove (1) during a problem session. Note that (4) follows immediately from (1) and (3) since:

$$
\det \begin{pmatrix}
R_1 \\
R_2 + sR_1 \\
R_3 \\
\vdots
\end{pmatrix} = \det \begin{pmatrix}
R_1 \\
R_2 \\
R_3 \\
\vdots
\end{pmatrix} + s \det \begin{pmatrix}
R_1 \\
R_1 \\
R_3 \\
\vdots
\end{pmatrix} = 0.
$$

To prove (2) suppose for simplicity that we swap the first two rows. Then

$$
\det B = \sum_{\sigma \in S_n} \text{sign}(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)}
$$

$$
= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\sigma(2)} a_{2\sigma(1)} a_{3\sigma(3)} \cdots a_{n\sigma(n)}
$$

$$
= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\tau(1)} a_{2\tau(2)} a_{3\tau(3)} \cdots a_{n\tau(n)}
$$

$$
= \sum_{\tau \in S_n} -\text{sign}(\tau) a_{1\tau(1)} a_{2\tau(2)} a_{3\tau(3)} \cdots a_{n\tau(n)} = -\det A
$$

for $\tau = \sigma \circ (12)$, where crucially we have used the fact that as $\sigma$ runs through all permutations of $S_n$, then so does $\tau$ (but in a different order), such that indeed $\sum_{\sigma \in S_n} = \sum_{\tau \in S_n}$.

Finally to prove (3) we once again suppose for simplicity that the first two rows are the same. Then each term

$$
\text{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}
$$

cancels out with the term

$$
\text{sign}(\tau) a_{1\tau(1)} a_{2\tau(2)} \cdots a_{n\tau(n)}
$$

for $\tau = \sigma \circ (12)$ as above since

$$
\begin{cases}
  a_{1\sigma(1)} a_{2\sigma(2)} \cdots = a_{1\sigma(2)} a_{2\sigma(1)} \cdots = a_{1\tau(1)} a_{2\tau(2)} \cdots \\
  \text{sign} \sigma = -\text{sign} \tau.
\end{cases}
$$

46
So indeed det A since terms in the sum defining det A come in pairs that cancel each other out.

4.4 Properties of determinants

Having introduced \( n \times n \) determinants in the preceding section and recorded some of their elementary properties we now dig a bit deeper to identify some other important properties of determinants.

First we introduce the transpose, since it will ultimately allow us to treat the words “rows” and "columns" fairly interchangeably when dealing with determinants.

Definition 4.17 (Transpose of a matrix). Let \( A \) be an \( n \times m \) matrix. Its transpose, denoted \( A^T \), is the \( m \times n \) matrix whose entries are \( A^T_{ij} = A_{ji} \) (i.e. the columns and rows are swapped).

Example 4.18.

\[
\begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}
^{T}
= 
\begin{pmatrix}
1 & 4 & 7 \\
2 & 5 & 8 \\
3 & 6 & 9
\end{pmatrix}
\]

The first property of determinants we record in this section is that determinants does not care much for transposes.

Theorem 4.19. For any \( n \times n \) matrix \( A \), \( \det A = \det(A^T) \).

Proof. This follows from a direct computation:

\[
\det(A^T) = \sum_{\sigma \in S_n} \text{sign}(\sigma) A^T_{1\sigma(1)} \cdots A^T_{n\sigma(n)} \\
= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{\sigma(1)} \cdots a_{\sigma(n)} \\
= \sum_{\sigma \in S_n} \text{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)},
\]

where we have used the fact that \( \text{sign}(\sigma) = \text{sign}(\sigma^{-1}) \) since, for transpositions \( \tau_1, \ldots, \tau_k \),

\[
(\tau_k \cdots \tau_1)^{-1} = \tau_k \cdots \tau_1.
\]

Crucially: as \( \sigma \) runs through \( S_n \), so does \( \sigma^{-1} \) (albeit in a different order), i.e. \( \sum_{\sigma \in S_n} = \sum_{\sigma^{-1} \in S_n} \). We therefore conclude that

\[
\det(A^T) = \sum_{\sigma \in S_n} \text{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)} \\
= \sum_{\sigma^{-1} \in S_n} \text{sign}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)} = \det A.
\]
The second property of determinants we consider is the fact that, while the determinant does not preserve addition of matrices, in the sense that in general \( \det(A + B) \neq \det A + \det B \), the determinant does behave very nicely with respect to multiplication.

**Theorem 4.20.** For any \( n \)-by-\( n \) matrices \( A \) and \( B \), \( \det(AB) = \det(A) \det(B) \).

**Proof.** We split into two cases.

- **Case 1:** rank \( A < n \). Then the reduced upper echelon form of \( A \) must have a zero row, and hence \( \det A = 0 \). In the same way: rank \( AB \leq \text{rank} \ A < n \) (since \( \text{im} \ AB \subseteq \text{im} \ A \)) and hence \( \det(AB) = 0 \).

- **Case 2:** rank \( A = n \). The Rank-Nullity Theorem tells us that nullity \( A = 0 \) such that \( A \) is invertible, and hence a product of elementary matrices. Crucially: \( \det(EC) = \det(E) \det(C) \) for any elementary matrix \( E \) and any \( n \)-by-\( n \) matrix \( C \). Therefore, writing \( A = E_1 \ldots E_n \),

\[
\det(AB) = \det(E_1 \ldots E_n B) = \det(E_1) \ldots \det(E_n) \det(B) \\
= \det(E_1 \ldots E_n) \det(B) = \det(A) \det(B).
\]

The third (and last) property of determinants we discuss in this section is their relation to invertibility.

**Theorem 4.21.** An \( n \)-by-\( n \) matrix \( A \) is invertible if and only if \( \det A \neq 0 \).

**Proof.** Suppose \( A \) is invertible. Then

\[
1 = \det I = \det(AA^{-1}) = \det A \det(A^{-1})
\]

and hence \( \det A \neq 0 \). Conversely, suppose that \( A \) is not invertible. Then either rank \( A < n \) or nullity \( A > 0 \) and, by the Rank-Nullity Theorem, rank \( A < n \) in both cases. Then the reduced upper echelon form of \( A \) must have a zero row, and hence \( \det A = 0 \).

---

5 Diagonalisation

5.1 Eigenvalues and eigenvectors

**Motivating example 5.1.** As we saw in Example 2.26 the basis \( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \) is in some sense the right basis for \( A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \). Why is that?

Well, it is because

\[
A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix}
\]

48
and
\[ A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix} = -2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \]
such that, along these two particular directions, matrix multiplication by \( A \) acts like scalar multiplication (by 4 or \(-2\)). This behaviour is particularly useful and so we give it a name in the definition below.

**Definition 5.2** (Eigenvalues and eigenvectors). Let \( T : V \to V \) be a linear map. Let \( v \) be a nonzero vector in \( V \) and let \( \lambda \) be a scalar such that \( T(v) = \lambda v \). We call \( v \) an eigenvector of \( T \) and call \( \lambda \) and eigenvalue of \( T \).

**Example 5.3.** Here are examples of eigenvalues and eigenvectors.

1. The matrix \( A \) has two eigenvalues: \( \lambda_1 = 4 \) and \( \lambda_2 = -2 \). Corresponding eigenvectors are \( v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \), or \( \tilde{v}_1 = \begin{pmatrix} 6 \\ 6 \end{pmatrix} \) (nonzero scalar multiples of an eigenvector are eigenvectors corresponding to the same eigenvalue), and \( v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \).

2. Consider the vector space \( V \) of infinitely differentiable functions from \( \mathbb{R} \) to \( \mathbb{R} \) and consider the linear map \( T : V \to V \) defined by \( T(f) = f' \). Every \( \lambda \in \mathbb{R} \) is an eigenvalue of \( T \), with eigenvector \( f(x) = e^{\lambda x} \), since \( T(f) = f' = \lambda f \).

**Theorem 5.4.** Let \( A \) be an \( n \times n \) matrix and let \( \lambda \) be a scalar. \( \lambda \) is an eigenvalue of \( A \) if and only if \( \det(A - \lambda I) = 0 \).

**Proof.** The following chain of equivalences holds, where for the third one we use the Rank-Nullity Theorem:

\[ \lambda \text{ is an eigenvector of } A \iff \text{there is a nonzero } v \in \mathbb{F}^n : (A - \lambda I)v = 0 \]
\[ \iff \text{nullity}(A - \lambda I) > 0 \]
\[ \iff A - \lambda I \text{ is not invertible} \]
\[ \iff \det(A - \lambda I) = 0. \]

**Example 5.5.** We use the theorem above to find all eigenvalues of some matrices.

1. Consider \( A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \) again. Then
\[
\det(A - \lambda I) = \det \begin{pmatrix} 1 - \lambda & 3 \\ 3 & 1 - \lambda \end{pmatrix}
= (1 - \lambda)^2 - 9 = \lambda^2 - 2\lambda - 8 = (\lambda - 4)(\lambda + 2),
\]
which tells us that the only eigenvalues of \( A \) are \(-2\) and \(4\).
2. Consider \( A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \). Then

\[
\det(A - \lambda I) = \det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1.
\]

- Viewed as a matrix over the field of real numbers, this matrix does not have any eigenvalues.
- Viewed as a matrix over the field of complex numbers, this matrix has two eigenvalues, namely \( \pm i \).

As can be seen readily from the definition of a determinant, the expression \( \det(A - \lambda I) \) is a polynomial. This polynomial is central to the study of eigenvalues and eigenvectors and so it too warrants a name of its own.

**Definition 5.6** (Characteristic polynomial). Let \( A \) be a square matrix. The polynomial \( p_A(\lambda) = \det(A - \lambda I) \) is called the characteristic polynomial of \( A \).

**Corollary 5.7.** Let \( A \) be an \( n \times n \) matrix. \( A \) has at most \( n \) distinct eigenvalues.

**Proof.** The only term of degree \( n \) in \( \det(A - \lambda I) \) is \( (-1)^n \lambda^n \) (this follows from the definition of the determinant), so the characteristic polynomial has degree \( n \), and hence has at most \( n \) distinct roots. \( \square \)

### 5.2 Diagonalisability

**Definition 5.8** (Diagonal, diagonalisable). A square matrix \( A \) is called

1. diagonal if \( a_{ij} = 0 \) when \( i \neq j \) and
2. diagonalisable if it is similar to a diagonal matrix.

**Theorem 5.9** (Characterisation of diagonalisability). An \( n \times n \) matrix \( A \) over a field \( F \) is diagonalisable if and only if \( F^n \) has a basis of eigenvectors of \( A \).

**Proof.** Write \( T_A : F^n \to F^n \) for the linear map \( T_A(v) = Av \). \( A \) is diagonalisable if and only if there is a basis \( v_1, \ldots, v_n \) of \( F^n \) such that the matrix of \( T_A \) in that basis is a diagonal matrix

\[
D = \begin{pmatrix} d_{11} & & \\ & \ddots & \\ & & d_{nn} \end{pmatrix}.
\]

In other words: \( Av_i = T_A(v_i) = d_{ii}v_i \), i.e. each of the \( v_i \)'s is an eigenvector of \( A \) (with eigenvalue \( d_{ii} \)). \( \square \)
\[
\begin{array}{c}
\text{Coord. from } e_1, e_2 \xrightarrow{A} \text{Coord. from } e_1, e_2 \\
\downarrow Q^{-1} \\
\text{Coord. from } v_1, v_2 \xrightarrow{D} \text{Coord. from } v_1, v_2
\end{array}
\]

Figure 15: How to use the eigenvectors of \(A\) to change basis and produce the diagonal matrix \(D\) (which is similar to \(A\)). Recall that \(e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\) and \(e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\).

**Example 5.10** (Diagonalisation). The matrix \(A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}\) has eigenvectors 
\(v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\) and \(v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}\) (with eigenvalues \(\lambda_1 = 4\) and \(\lambda_2 = -2\)), which form a basis of \(\mathbb{R}^2\). So \(A\) is diagonalisable. Can we diagonalise it explicitly?

Yes! As shown in Figure 15, \(A = QDQ^{-1}\) for
\[
Q = (v_1 | v_2) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}
\]

As a sanity check we may compute that \(Q^{-1} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}\) and hence
\[
QDQ^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = A.
\]

Now the question is: when is a matrix guaranteed to be diagonalisable? We will dig into this more later in the course, for specific classes of matrices. For now we have the following result.

**Theorem 5.11.** Let \(\lambda_1, \ldots, \lambda_r\) be distinct eigenvalues of a linear map \(T\) with corresponding eigenvectors \(v_1, \ldots, v_r\). These eigenvectors are linearly independent.

**Proof.** This is true when \(r = 1\) (since eigenvectors are nonzero). Now suppose that this is true for some \(r \geq 1\). We now show that it must then be true for \((r + 1)\)-many eigenvectors associated to distinct eigenvalues. So suppose
\[
a_1v_1 + \cdots + a_r v_r + a_{r+1}v_{r+1} = 0.
\]
Applying \(T\) tells us that
\[
a_1\lambda_1v_1 + \cdots + a_r \lambda_r v_r + a_{r+1}\lambda_{r+1}v_{r+1} = 0
\]
while multiplying by $\lambda_{r+1}$ tells us that
\[ a_1 \lambda_{r+1} v_1 + \cdots + a_r \lambda_{r+1} v_r + a_{r+1} \lambda_{r+1} v_{r+1} = 0. \]

Subtracting the second equation from the first yields
\[ a_1 (\lambda_1 - \lambda_{r+1}) v_1 + \cdots + a_r (\lambda_r - \lambda_{r+1}) v_r = 0. \]

Since $v_1, \ldots, v_r$ are, by assumption, linearly independent, we deduce that
\[ a_i (\lambda_i - \lambda_{r+1}) = 0 \text{ for every } i \leq r. \]

Since the eigenvalues are distinct we know that $\lambda_i - \lambda_{r+1} \neq 0$, and hence $a_i = 0$ for every $i \leq r$. So finally: $a_{r+1} v_{r+1} = 0$, but $v_{r+1} \neq 0$, and hence $a_{r+1} = 0$ as well, such that indeed the $(r+1)$-many eigenvectors are linearly independent. \( \square \)

The power of this theorem is that it gives us a criterion for diagonalisability.

**Corollary 5.12.** An $n$-by-$n$ matrix with $n$ distinct eigenvalues is diagonalisable.

**Proof.** By Theorem 5.11 such a matrix has $n$ linearly independent eigenvectors, which thus form a basis of $\mathbb{F}^n$. \( \square \)

### 5.3 Bonus: Applications of eigenvalues and eigenvectors

Linear algebra is at the heart of many disciplines leveraging mathematical tools, ranging from economics to physics and biology. In particular it is often the case that eigenvalues and eigenvectors play a central role. Here we verify this paradigm by studying two examples:

- systems of differential equations and
- random processes.

**Example 5.13 (Systems of differential equations).** In a differential equation the unknown is a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which must satisfy
\[
\begin{cases}
 f'(t) = \ldots, \text{ i.e. the differential equation itself, and} \\
 f(0) = \ldots, \text{ which is referred to as an initial condition.}
\end{cases}
\]

In other words, for some unknown function $f$, we are specifying (1) $f$’s starting value (this is the initial condition) as well as (2) how $f$ varies over time. It thus stands to reason that, as long as the specified dependence of $f'$ on $f$ and $t$ is not too horrible, we will be able to reconstruct $f$ from these two pieces of information. For example
\[
\begin{cases}
 f'(t) = 5f(t), \\
 f(0) = 3
\end{cases}
\]

has solution $f(t) = 3e^{5t}$. Such an equation can be thought of as modelling the early stages of population development, where the number of new births is proportional to the population itself, thus leading to exponential growth.
In practice we often wish to model situations where more than one unknown function is present. For example: what if we are modelling the population of two species where at least one species feeds off on the other (such as modelling the population of foxes and rabbits, for example)? It turns out that in that context eigenvalues and eigenvectors will play a crucial role in helping us determine and interpret the solution.

To make these ideas concrete we will consider the following system:

\[
\begin{align*}
    f'(t) &= -\frac{7}{5} f(t) + \frac{6}{5} g(t) \quad \text{and} \\
    g'(t) &= \frac{6}{5} f(t) + \frac{2}{5} g(t)
\end{align*}
\] (1)

subject to the initial conditions

\[
\begin{align*}
    f(0) &= 1 \quad \text{and} \\
    g(0) &= 7.
\end{align*}
\]

There are now two unknown functions to determine, namely \( f \) and \( g \).

Before we start, let us consider the simplified scenario where the mutual dependence of \( f \) on \( g \) and \( g \) on \( f \) is removed, such that the model reads

\[
\begin{align*}
    f'(t) &= -\frac{7}{5} f(t) \\
    g'(t) &= \frac{6}{5} g(t)
\end{align*}
\]

In that case we see that the population modelled by \( f \) with die out (or, more precisely, decrease exponentially fast) while the population modelled by \( g \) will thrive, growing exponentially, since the solution in this simplified model is

\[
\begin{align*}
    f(t) &= e^{-7t/5} \\
    g(t) &= 7e^{2t/5}.
\end{align*}
\]

So let us now come back to the full model (1) to see how the interdependence between the two populations affects their dynamics. Before computing anything, note that we can expect the following:

- the population \( f \) will do better once the interdependence is taken into account thanks to the term \( \frac{6}{5} g(t) \) present in the equation for \( f'(t) \) which tells us that the population \( f \) benefits from the population \( g \) being higher, and

- the population \( g \) might also do better once the interdependence is taken into account since there is a factor of \( \frac{6}{5} \) in front of \( f \) in the equation for \( g' \) while there is only a factor of \( \frac{2}{5} \) in front of \( g \) in that same equation.

In other words: it seems that the system (1) might model a **symbiotic** relationship between the two populations, where each benefits from the other. Let’s see if this indeed comes out when we solve the system.
The starting point in solving this system is realizing that we may write it equivalently as
\[
\begin{pmatrix} f'(t) \\ g'(t) \end{pmatrix} = \begin{pmatrix} \frac{-7}{5} & \frac{6}{5} \\ \frac{6}{5} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} f(t) \\ g(t) \end{pmatrix}.
\]

We may now compute the eigenvalues and eigenvectors of \( A \), which will be of tremendous help very soon. The characteristic polynomial of \( A \) is
\[
\det(A - \lambda I) = \lambda^2 + \lambda - 2 = (\lambda + 2)(\lambda - 1)
\]
such that the eigenvalues of \( A \) are \( \lambda_1 = -2 \) and \( \lambda_2 = 1 \). We may then find associated eigenvectors of \( A \) to be
\[
v_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix},
\]
respectively.

The eigenvalues of \( A \) now help us define “good unknowns”, i.e. new unknown functions which will satisfy much nicer differential equations than those satisfied by \( f \) and \( g \). So, for \( Q = (v_1 | v_2) \) let us define
\[
\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = Q^{-1} \begin{pmatrix} f(t) \\ g(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{5}(f + 2g) \\ \frac{1}{5}(2f - g) \end{pmatrix}.
\]

Where did these new unknown come from? Well we can rewrite the system of differential equations, for \( w = (f, g) \), as
\[
w' = Aw.
\]

Since \( A \) is diagonalisable, we may write \( A = QDQ^{-1} \) of a diagonal matrix \( D \) such that the system immediately above becomes
\[
w' = QDQ^{-1}w \iff (Q^{-1}w)' = D (Q^{-1}w).
\]

Crucially: since \( D \) is diagonal, this will lead to \( Q^{-1}w \) solving a much simpler system of differential equations! This is precisely why we define \( u \) and \( v \) as \( (u, v) = Q^{-1}(f, g) \).

Now we seek to verify that indeed \( u \) and \( v \) satisfy simpler equations than \( f \) and \( g \) did. We compute that
\[
u' = \frac{1}{5}f' + \frac{2}{5}g' = \frac{1}{5} \left( -\frac{7}{5}f + \frac{6}{5}g \right) + \frac{2}{5} \left( \frac{6}{5}f + \frac{2}{5}g \right) = \frac{1}{5}f + \frac{2}{5}g = u
\]
while
\[
v' = \frac{2}{5}f' - \frac{1}{5}g' = \frac{2}{5} \left( -\frac{7}{5}f + \frac{6}{5}g \right) - \frac{1}{5} \left( \frac{6}{5}f + \frac{2}{5}g \right) = \frac{-4}{5}f + \frac{2}{5}g = -2v,
\]

54
i.e. $u$ and $v$ solve the much nicer-looking system
\[ \begin{cases} 
  u'(t) = u(t) \text{ and} \\
  v'(t) = -2v(t). 
\end{cases} \]

Their initial conditions are easy enough to determine:
\[ \begin{cases} 
  u(0) = \frac{1}{5}f(0) + \frac{2}{5}g(0) = 3 \text{ and} \\
  v(0) = \frac{2}{5}f(0) - \frac{1}{5}g(0) = -1. 
\end{cases} \]

We may therefore deduce that $u(t)$ and $v(t)$ are given by
\[ \begin{cases} 
  u(t) = 3e^t \text{ and} \\
  v(t) = -e^{-2t}. 
\end{cases} \]

So finally, since the unknowns we care about are really $f$ and $g$, not $u$ and $v$, we recover the former from the latter via
\[ \begin{pmatrix} f \\ g \end{pmatrix} = Q \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u + 2v \\ 2u - v \end{pmatrix} \]
such that
\[ \begin{cases} 
  f(t) = u(t) + 2v(t) = 3e^t - 2e^{-2t} \text{ and} \\
  g(t) = 2u(t) - v(t) = 6e^t + e^{-2t}. 
\end{cases} \]

Remember how we said above that the system of differential equations (1) modelled two populations in *symbiosis*, both benefiting from each other? Well we can verify this now! Remember that, without the interdependence between species, $f(t)$ decayed exponentially fast whereas $g(t)$ grew at a rate of $e^{2t/5}$. But now they both grow at a rate of $e^t$! This is an improvement for both species, which shows us that this is indeed a relationship that benefits to both species.

We now turn our attention to another example

**Example 5.14** (Random processes). Consider a random process which is described by two states such that, at each time step (say every minute, every hour, or every million years, depending on the application), an agent in either of the states will either stay in its current state or change state with some prescribed probability. We will discuss a random process where these transition probabilities are given according to the following diagram:
The particularly neat aspect of such models is that their behaviour can be fully encapsulated by a so-called *transition matrix*. For our example the transition matrix is given by

\[
M = \begin{pmatrix}
\frac{1}{2} & 3/4 \\
\frac{1}{2} & 1/4 
\end{pmatrix}.
\]

Why is this the appropriate transition matrix? Well, let \((p_1, p_2)\) be the distribution of agents among the two states at some time. For example of \(p_1 = 0.2\) and \(p_2 = 0.8\) this means that 20% of agents are in state 1 and 80% of agents are in state 2. What is the distribution of agents at the next time step? Remember that, to enter state 1 at the next time step, agents must be either be

- from the half of agents previously in state 1 which are *not* switching state, or
- from the three-quarters of agents previously in state 2 which *are* switching state.

In other words, the proportion of agents in state 1 after one time step is

\[
\frac{1}{2}p_1 + \frac{3}{4}p_2.
\]

With exactly the same reasoning we may deduce that the proportion of agents in state 2 after one time step is

\[
\frac{1}{2}p_1 + \frac{1}{4}p_2.
\]

This can all be written more succinctly by noting that

\[
\begin{pmatrix}
\frac{1}{2}p_1 + \frac{3}{4}p_2 \\
\frac{1}{2}p_1 + \frac{1}{4}p_2
\end{pmatrix} = M \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}
\]

such that the matrix tells us how to determine the distribution of agents among states given their distribution at the previous step!

A natural question that then arises is the following: after many time steps, what does the distribution of agents look like? Let’s answer that question in the case where the initial distribution of agents is

\[
d_0 \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}.
\]

To compute \(M^nd_0\), which is the distribution of agents after \(n\) time steps, it will be particularly convenient to write the vector \((p_1, p_2)\) as a linear combination of the eigenvectors of \(M\).

We may compute that the eigenvalues of \(M\) are \(\lambda_1 = -\frac{1}{4}\) and \(\lambda_2 = 1\), with corresponding eigenvectors

\[
v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } v_2 = \begin{pmatrix} 3/5 \\ 2/5 \end{pmatrix}.
\]
Then we may write the initial distribution of agents $d_0$ as a linear combination of $v_1$ and $v_2$ as

$$d_0 = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 3/5 \\ 2/5 \end{pmatrix} - \frac{4}{15} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$  

Computing $M^n d_0$ is now very easy:

$$M^n d_0 = 1^n \begin{pmatrix} 3/5 \\ 2/5 \end{pmatrix} - \frac{4}{15} \left( \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)^n \to 0,$$

and we easily see that the second term approaches zero as $n$ approaches infinity. In other words, as time goes on,

$$M^n d_0 \to \begin{pmatrix} 3/5 \\ 2/5 \end{pmatrix},$$

which is precisely the eigenvector of $M$ corresponding to the eigenvalue $\lambda_2 = 1$.

Qualitatively this result is not too surprising: agents are more likely to go from state 2 to state 1 than vice-versa, and so it is expected that, as time goes on, we find more agents in state 1.

## 6 Inner product spaces

We now begin the last chapter of the course and introduce the notion of an inner product. We will define that term shortly, but first let us mention that things that this new notion will allow us to do.

- We will be able to measure the length of vectors, whether they are geometric vectors in $\mathbb{R}^3$ or $\mathbb{R}^3$, or whether the vectors are really functions or polynomials.
- We will be able to determine when two vectors are perpendicular to each other, once again whether these vectors are in $\mathbb{R}^2$, $\mathbb{R}^3$, or in a vector space of functions.
- We will obtain alternative (and simpler!) criteria allowing us to determine whether or not a given matrix is diagonalisable.

### 6.1 Inner products and norms

**Definition 6.1 (Inner product).** Let $V$ be a vector space over a field $\mathbb{F}$ where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. An inner product on $V$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ satisfying the following properties.

1. Linearity in the first component: $\langle sx + y, z \rangle = s \langle x, z \rangle + \langle y, z \rangle$ for every $x, y, z \in V$ and every $s \in \mathbb{F}$. 


2. If $F = \mathbb{R}$, linearity in the second component: $\langle x, sy + z \rangle = s \langle x, y \rangle + \langle x, z \rangle$ for every $x, y, z \in V$ and $s \in \mathbb{R}$.

If $F = \mathbb{C}$, conjugate linearity in the second component: $\langle x, sy + z \rangle = \overline{s} \langle x, y \rangle + \langle x, z \rangle$ for every $x, y, z \in V$ and $s \in \mathbb{C}$.

3. If $F = \mathbb{R}$, symmetry: $\langle x, y \rangle = \langle y, x \rangle$ for every $x, y \in V$.

If $F = \mathbb{C}$, Hermitian symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for every $x, y \in V$.

4. Positive-definiteness: $\langle x, x \rangle \geq 0$ for every $x \in V$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

**Example 6.2** (Inner products). Here are some examples of inner products, the first two examples being familiar ones (even though we may not have called them inner products when we encountered them before).

1. The dot product: for $x, y \in \mathbb{R}^n$ we may consider the inner product

$$\langle x, y \rangle = x \cdot y = x_1 y_1 + \cdots + x_n y_n.$$ 

2. We can define an inner product on $\mathbb{C}$ as follows: for any two complex numbers $x, y \in \mathbb{C}$ we define

$$\langle x, y \rangle = xy.$$ 

Note that this is the example to remember when trying to understand why we ask conjugate linearity or Hermitian symmetry of inner products for vector spaces over the field of complex numbers.

3. For $f, g \in C[0, 1]$, the vector space of continuous real-valued functions on $[0, 1]$, we may consider the inner product

$$\langle f, g \rangle = \int_0^1 fg = \int_0^1 f(x)g(x)dx.$$ 

This is known as the $L^2$ inner product.

It is important to note that the same vector space can be equipped with different inner products. For example we could consider a weighted dot product

$$\langle x, y \rangle = 5x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$ 

for any two $x, y \in \mathbb{R}^n$. This is still an inner product, but it is different from the standard dot product. This is why we must always specify which inner product we are working with. This motivates the following definition.

**Definition 6.3** (Inner product space). An inner product space is a pair $(V, \langle \cdot, \cdot \rangle)$ where $\langle \cdot, \cdot \rangle$ is an inner product on $V$.

As promised, inner products allow us to define the length of a vector. Although we will use the word norm instead of length.
Definition 6.4 (Norm). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space. For any \(x \in V\) the norm of \(x\), denoted \(||x||\), is defined to be \(||x|| = \sqrt{\langle x, x \rangle}||

Example 6.5 (Norms). We revisit Example 6.2 and record some of the norms associated with the inner products introduced in that example.

1. In \(\mathbb{R}^n\) equipped with the dot product:
   \[
   ||x|| = \sqrt{x \cdot x} = \sqrt{x_1^2 + \cdots + x_n^2}.
   \]

2. In \((C[0,1], L^2)\) the norm of a continuous function \(f : [0,1] \rightarrow \mathbb{R}\) is given by
   \[
   ||f|| = \left( \int_0^1 f^2 \right)^{1/2}.
   \]
   Note that it is because of the \(f^2\) that appears in the norm of \(f\) that the underlying inner product is called the \(L^2\) inner product.

To wrap up our first foray into the land of inner products we record two simple by fundamental inequalities.

Theorem 6.6. Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space. For every \(x, y \in V\),

1. \(\langle x, y \rangle \leq ||x|| \cdot ||y||\) (this is known as the Cauchy-Schwarz inequality) and
2. \(||x + y|| \leq ||x|| + ||y||\) (this is known as the triangle inequality, which is depicted in Figure 16).

![Figure 16: A pictorial representation of the triangle inequality.](image)

Proof. 1. The result is immediate if \(y = 0\) so suppose \(y \neq 0\). For any \(s \in \mathbb{F}\),
   \[
   0 \leq \langle x - sy, x - sy \rangle = \langle x, x \rangle - s\langle x, y \rangle - \overline{s}\langle y, x \rangle + s\overline{s}\langle y, y \rangle.
   \]
   Choosing
   \[
   s = \frac{\langle x, y \rangle}{\langle y, y \rangle}
   \]
   we see that the last three terms in the sum above are identical (up to sign), such that
   \[
   0 \leq \langle x, x \rangle - \frac{|\langle x, y \rangle|^2}{\langle y, y \rangle} = ||x||^2 - \frac{|\langle x, y \rangle|^2}{||y||^2},
   \]
   as desired.
2. This result follows from simply expanding out $\|x + y\|^2$ and completing the square:

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle$$

where

$$\langle x, y \rangle + \langle y, x \rangle = \langle x, y \rangle + \langle x, y \rangle = 2 \text{Re} \langle x, y \rangle$$

such that, by the Cauchy-Schwarz inequality proved above,

$$\|x + y\|^2 = \|x\|^2 + 2 \text{Re} \langle x, y \rangle + \|y\|^2$$

$$\leq \|x\|^2 + 2 \|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

as desired.

\[\square\]

### 6.2 Orthogonality and the Gram-Schmidt process

When motivating the concept of an inner product we mentioned that it could be used to generalise the notion of “perpendicular”. Indeed, we know that two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are perpendicular precisely when their dot product is equal to zero, and since inner products are generalisations of dot products, we may now, given any inner product, determine whether or not two vectors are perpendicular. Although, in this broader context the term “perpendicular” is not used and the term “orthogonal” is used instead.

**Definition 6.7** (Orthogonality, unit vector, and orthonormality). Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space.

1. Two vectors $x, y \in V$ are said to be **orthogonal** if $\langle x, y \rangle = 0$.
2. A vector $x \in V$ is said to be a **unit vector** if $\|x\| = 1$.
3. A set $S \subseteq V$ is said to be **orthonormal** if the vectors in $S$ are mutually orthogonal unit vectors, i.e. $\langle x, y \rangle = 0$ and $\|x\| = 1$ for every distinct $x, y \in S$.

Given any vector $v$ we may always find a unit vector with the same direction (i.e. the same span) by considering $\frac{v}{\|v\|}$. This process is called **normalisation**, which explains why the term “orthonormal” is used.

**Example 6.8** (Orthogonality and orthonormality). Here are examples of orthogonal and orthonormal vectors in various inner product spaces.

1. The vectors $\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ and $\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ form an orthonormal set in $\mathbb{R}^2$ equipped with the dot product.
2. The functions \( f(x) = 1 \) and \( g(x) = 2x - 1 \) are orthogonal in the inner product space \((C[0,1], L^2)\) since

\[
\langle f, g \rangle = \int_0^1 f(x)g(x)dx = \int_0^1 (2x - 1)dx = (x^2 - x) \big|_{x=0}^{x=1} = 0.
\]

Orthogonality is a useful notion of a myriad of reasons. Once of those is that bases of orthogonal vectors are particularly convenient to work with. In order to convince ourselves of that fact, let us first consider the case of a spanning set of mutually orthogonal vectors.

**Theorem 6.9.** Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space and let \(v_1, \ldots, v_n \in V\) be mutually orthogonal nonzero vectors. If \(v \in \text{span} \{(v_1, \ldots, v_n)\}\) then

\[
v = \frac{\langle v, v_1 \rangle}{\|v_1\|^2} v_1 + \cdots + \frac{\langle v, v_n \rangle}{\|v_n\|^2} v_n.
\]

Why is this result particularly useful? In general, given

\[v \in \text{span} \{(v_1, \ldots, v_n)\},\]

finding the appropriate scalars \(a_i\) that let us write \(v = a_1v_1 + \cdots + a_nv_n\) require us to solve a linear system, namely the system

\[
\begin{pmatrix}
v_1 & \cdots & v_n
\end{pmatrix}
\begin{pmatrix}
a_1 \\
\vdots \\
a_n
\end{pmatrix}
= v.
\]

What **Theorem 6.9** tells us is that, if the vectors \(v_1, \ldots, v_n\) are mutually orthogonal, then we do not need to solve anything, we only need to compute \(a_i = \frac{\langle v, v_i \rangle}{\|v_i\|^2}\). (And it turns out that computing is much easier than solving ... this is why differentiation is a breeze compared with integration!)

**Proof of Theorem 6.9.** If \(v \in \text{span} \{(v_1, \ldots, v_n)\}\) then

\[v = a_1v_1 + \cdots + a_nv_n\]

for some scalars \(a_i\). Then, by mutual orthogonality of the \(v_i\)'s and linearity in the first component of the inner product, we see that, for any \(i\),

\[
\langle v, v_i \rangle = a_1\langle v_1, v_i \rangle + \cdots + a_n\langle v_n, v_i \rangle = a_i\langle v_i, v_i \rangle = a_i \|v_i\|^2
\]

such that indeed

\[a_i = \frac{\langle v, v_i \rangle}{\|v_i\|^2},\]

where note that \(\|v_i\| \neq 0\) since \(v_i \neq 0\).

We can also deduce the following result using the same idea that was used in the proof of **Theorem 6.9**
Corollary 6.10. A mutually orthogonal set of nonzero vectors is linearly independent.

Proof. Let \( v_1, \ldots, v_n \) be mutually orthogonal and nonzero and suppose that
\[
a_1 v_1 + \cdots + a_n v_n = 0.
\]
Then
\[
a_i = \frac{\langle 0, v_i \rangle}{||v_i||^2} = 0
\]
such that indeed the vectors \( v_1, \ldots, v_n \) are linearly independent.

In particular, Corollary 6.10 tells us that any set of mutually orthogonal vectors forms a basis of its span. In light of Theorem 6.9, such bases are particularly convenient to work with, so the question is now: can we always find such bases of mutually orthogonal vectors? The answer is yes, and moreover this can be done in a very constructive manner: not only do such bases always exist, but they can easily be computed from another basis, using the so-called Gram-Schmidt process described below.

Theorem 6.11 (Gram-Schmidt). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space and let \( \{v_1, \ldots, v_n\} \) be a linearly independent subset of \( V \). Define
\[
w_1 = v_1 \text{ and } w_k = v_k - \sum_{i=1}^{k-1} \frac{\langle v_k, w_i \rangle}{||w_i||^2} w_i \text{ for } k = 2, \ldots, n.
\]
Then \( w_1, \ldots, w_n \) are mutually orthogonal and
\[
\text{span} (\{w_1, \ldots, w_n\}) = \text{span} (\{v_1, \ldots, v_n\}).
\]

A key consequence of Theorem 6.11 is the following.

Corollary 6.12. Every finite-dimensional inner product space has an orthonormal basis.

Proof of Corollary 6.12. It suffices to apply the Gram-Schmidt process to a basis and consider the normalised vectors \( \tilde{w}_i = \frac{w_i}{||w_i||} \).

We will not discuss the proof of Theorem 6.11 in class (although the proof is below, and purely optional, for those who are interested). Instead we will try and convince ourselves that the Gram-Schmidt process works by looking at an example.

Example 6.13 (Gram-Schmidt). We apply the Gram-Schmidt process to
\[
v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and } v_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.
\]
At the first step we consider
\[ w_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}. \]

At the second step consider
\[ w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = v_2 - \frac{2}{2} w_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \]

Finally, at the third step we compute that
\[ w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2 \
= v_3 - \frac{1}{2} w_1 - \frac{1}{1} w_2 \
= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix} - \frac{1}{1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1/2 \\ 0 \end{pmatrix}. \]

**Example 6.14** (Another example of Gram-Schmidt, one we can actually draw out). Consider the vectors
\[ v_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1/2 \\ 0 \end{pmatrix}, \quad \text{and} \quad v_3 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}. \]

Applying the Gram-Schmidt orthogonalisation process to these vectors produces the following vectors: (we skip the computations here to emphasize the geometric picture instead – it is a good exercise to verify these computations)
\[ w_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 \\ -1/2 \\ 0 \end{pmatrix}, \quad \text{and} \quad w_3 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}. \]

This is represented pictorially in Figures 17 and 18 where the former depicts how \( w_2 \) is obtained by decomposing \( v_2 \) along \( w_1 \) and the latter depicts how \( w_3 \) is obtained by decomposing \( v_3 \) along the plane spanned by \( w_1 \) and \( w_2 \).

What follows was not covered in lecture and is only provided if you are curious as to how one would prove Theorem 6.11.

**Proof of Theorem 6.11.** First we verify the mutual orthogonality. Clearly \( w_1 \) and \( w_2 \) are orthogonal since
\[ \langle w_2, w_1 \rangle = \langle v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1, w_1 \rangle = \langle v_2, w_1 \rangle - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} \langle w_1, w_1 \rangle = \langle v_2, w_1 \rangle - \langle v_2, w_1 \rangle = 0. \]
Figure 17: How the vector $w_2$ in Example 6.14 is determined by decomposing $v_2$ along $w_1$.

Figure 18: How the vector $w_3$ in Example 6.14 is determined by decomposing $v_3$ along the plane spanned by $w_1$ and $w_2$. 
Now we will show that if the first $k$ vectors $w_1, \ldots, w_k$ are mutually orthogonal then $w_{k+1}$ is orthogonal with $w_1, \ldots, w_k$. We compute, since $\langle w_i, w_j \rangle = 0$ for any distinct $i, j = 1, \ldots, k$, that for any $j = 1, \ldots, k$,

$$
\langle w_{k+1}, w_j \rangle = \langle w_{k+1}, w_j \rangle - \sum_{i=1}^{k} \frac{\langle w_{k+1}, w_i \rangle}{\|w_i\|^2} \langle w_i, w_j \rangle
= \langle w_{k+1}, w_j \rangle - \frac{\langle w_{k+1}, w_j \rangle}{\|w_j\|^2} \langle w_j, w_j \rangle
= \langle w_{k+1}, w_j \rangle - \langle w_{k+1}, w_j \rangle = 0.
$$

So indeed $w_1, \ldots, w_n$ are mutually orthogonal. In particular, this means that they must be linearly independent. On the other hand, by definition the $w_i$’s belong to $\text{span} \{v_1, \ldots, v_n\} = V_0$. Since the $v_i$’s are linearly independent we know that the dimension of $V_0$ is equal to $n$, and so the new vectors $w_1, \ldots, w_n$ are $n$ linearly independent vectors in a subspace $V_0$ of dimension $n$, which tells us that they must form a basis of $V_0$ and in particular must span $V_0$. So indeed

$$\text{span} \{w_1, \ldots, w_n\} = V_0 = \text{span} \{v_1, \ldots, v_n\}.$$

This is the end of the portion of the notes covering extra material not discussed in lectures.

### 6.3 The adjoint of a linear map

We motivated the notion of an inner product by pointing out that it would help us do two things:

1. generalise geometric notions, such as lengths and perpendicularity, to vector spaces (or rather to inner product spaces) and
2. provide an alternative, and simpler, criterion for diagonalisability.

It will turn out that this criterion can be phrased, in terms of matrices, using the transpose. For example so-called symmetric matrices which satisfy $A = A^T$ will be guaranteed to be diagonalisable.

However, remember that matrices are only representations of linear maps once we have chosen a basis (or bases). We would therefore like to have a notion comparable to the notion of the transpose, but applicable to linear maps instead. This is known as the adjoint.

**Definition 6.15 (Adjoint).** Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and let $T : V \to V$ be a linear map. We call $T^* : V \to V$ an adjoint of $T$ if, for every $x, y \in V$, $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$. 

65
Note that, although the definition above leave this point ambiguous, adjoints are actually unique, and always guaranteed to exist. We will prove that below in Theorem 6.19.

**Example 6.16.** Here are some example of linear maps and their respective adjoints.

1. Let $A$ be a real $n$-by-$n$ matrix. With respect to the dot product, the adjoint of the linear map on $\mathbb{R}^n$ corresponding to multiplication by $A$ is multiplication by its transpose $A^T$. Indeed:

\[
(Ax) \cdot y = \sum_i (Ax)_iy_i = \sum_{i,j} A_{ij}x_j y_i = \sum_{i,j} x_j A^T_{ij}y_i = \sum_j x_j (A^T y)_j = x \cdot (A^T y).
\]

2. Let $V$ denote the vector space of continuous real-valued functions on $[0, 1]$ which are infinitely differentiable in $(0, 1)$ and for which all derivatives vanish at 0 and 1, i.e. $f : [0, 1] \rightarrow \mathbb{R}$ satisfies $f^{(k)}(0) = f^{(k)}(1) = 0$ for all $k = 1, 2, \ldots$. Consider $T : V \rightarrow V$ defined by $T(f) = f'$. Using the $L^2$ inner product (and integrating by parts):

\[
\langle T(f), g \rangle = \int_0^1 f'g = fg\bigg|_0^1 - \int_0^1 fg' = -\int_0^1 fg' = \langle f, -Tg \rangle,
\]

i.e. $T^* = -T$.

We now build the necessary tools to prove that the adjoint of a linear map is always guaranteed to exist, and is uniquely determined by the linear map in question. First we prove a representation lemma.

**Lemma 6.17.** Let $(V, \langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product space over a field $\mathbb{F}$ and let $\phi : V \rightarrow \mathbb{F}$ be a linear map. There exists a unique $v \in V$ such that $\phi(x) = \langle x, v \rangle$ for every $x \in V$.

This result tells us that linear functions of a particular kind, namely those which map the inner product space to its underlying field, can all be represented as the inner product with a fixed vector (the fixed vector being uniquely determined by the linear function in question).

---

2Such linear functions are often called functionals. This is because historically they first arose when the inner product space in question contained functions. This linear function from the inner product space to its underlying field is then a function whose input are themselves functions, a sufficiently interesting phenomenon to warrant its own name, namely that of a functional.
Example 6.18. Consider $\phi : \mathbb{R}^2 \to \mathbb{R}$ given by $\phi(x, y) = x - 3y$. Using the dot product we may then write
\[ \phi(x) = x - 3y = \langle (1, -3), (x, y) \rangle = \langle v, (x, y) \rangle \]
for $v = (1, -3)$.

Proof of Lemma 6.17. Let $v_1, \ldots, v_n$ be an orthonormal basis of $V$ and define
\[ v = \overline{\phi(v_1)}v_1 + \cdots + \overline{\phi(v_n)}v_n. \]
The map $x \mapsto \langle x, v \rangle$ is linear so we only have to show that $\phi(v_i) = \langle v_i, v \rangle$ for every basis vector $v_i$ (since two linear maps which agree on the basis vectors must agree everywhere). Well indeed, since the $v_i$’s are mutually orthogonal:
\begin{align*}
\langle v_i, v \rangle &= \langle v_i, \overline{\phi(v_1)}v_1 + \cdots + \overline{\phi(v_n)}v_n \rangle \\
&= \overline{\phi(v_1)}\langle v_i, v_1 \rangle + \cdots + \overline{\phi(v_n)}\langle v_i, v_n \rangle = \overline{\phi(v_i)} \langle v_i, v_i \rangle \\
&= 1.
\end{align*}
The uniqueness of $v$ is proved in assignment 10. \qed

We are now ready to prove the existence and uniqueness of adjoints.

Theorem 6.19 (Existence and uniqueness of adjoints). Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Every linear map $T : V \to V$ has a unique adjoint.

Proof. Uniqueness is immediate: if $S : V \to V$ satisfies $\langle T(x), y \rangle = \langle x, S(y) \rangle$ for every $x, y \in V$ then $\langle x, T^*(y) \rangle = \langle x, S(y) \rangle$ for every $x, y \in V$, so $T^*(y) = S(y)$ for every $y \in V$, which is a long-winded way of saying that $T^* = S$. Now we prove the existence of an adjoint. Fix $y \in V$ and define
\[ \phi(x) = \langle T(x), y \rangle. \]
Clearly $\phi : V \to \mathbb{F}$ is linear, so by Lemma 6.17 above there exists a unique $v \in V$ such that
\[ \phi(x) = \langle x, v \rangle \]
for every $x \in V$.
Since $\phi(x) = \langle T(x), y \rangle$, we want $T^*$ to satisfy the equation
\[ \langle x, T^*(y) \rangle = \langle T(x), y \rangle = \phi(x) = \langle x, v \rangle \]
for every $x \in V$ and so we may define $T^*(y) = v$. \qed

6.4 A prelude to diagonalisability via orthonormal bases: Schur’s Lemma

Our goal as we begin to wrap-up Section 6 is to obtain a characterisation of linear maps that can be represented as a diagonal matrix with respect to some basis, and not any basis, since we will ask that basis to be orthonormal. Although the statement of the characterisation will ultimately be very simple, the legwork required to be able to prove that characterisation is somewhat technical. This section is devoted entirely to those technicalities.

First we introduce some necessary vocabulary.
Definition 6.20 (Splitting polynomial; upper triangular matrix). 1. We say that a polynomial $p$ splits over a field $F$ if

$$p(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

for some possibly repeated $\lambda_i \in F$.

2. A square matrix $A$ is called upper-triangular if $a_{ij} = 0$ whenever $i > j$.

Example 6.21. 1. The polynomial $x^2 - 2x + 1 = (x - 1)^2$ splits over $\mathbb{R}$ but the polynomial $x^2 + 1$ does not, although both polynomials split over $\mathbb{C}$.

2. The matrix

$$
\begin{pmatrix}
0 & 1 & 3 \\
0 & 2 & 4 \\
0 & 0 & 5
\end{pmatrix}
$$

is upper-triangular.

The first technical result we need can be thought of as a weaker version of Corollary 5.12 which told us that if a matrix had $n$ distinct eigenvalues, i.e. if its characteristic polynomial split into $n$ distinct linear factors of the form $x - \lambda_i$, then it was diagonalisable. Here we assume less since we only ask for the characteristic polynomial to split, allowing it to have repeated roots, and also obtain less since we only guaranteed that the matrix is similar to an upper-triangular matrix, and not necessarily to a diagonal matrix.

Lemma 6.22. A matrix whose characteristic polynomial splits is similar to an upper-triangular matrix.

Proof. This is true for 1-by-1 matrices so we will show that if it is true for $(n-1)$-by-$(n-1)$ matrices then it is also true for $n$-by-$n$ matrices.

Let $A$ be an $n$-by-$n$ matrix whose characteristic polynomial splits. Then $A$ must have (at least) one eigenvalue $\lambda$, so let us denote by $v_1$ an associated eigenvector. Extending to a basis $v_1, \ldots, v_n$ of $V$, the matrix $P = (v_1 \mid \ldots \mid v_n)$ is invertible and so $P^{-1}AP$ has the same characteristic polynomial as $A$, which splits. In particular:

$$AP = A(v_1 \mid \ldots \mid v_n) = (Av_1 \mid \ldots \mid Av_n) = (\lambda v_1 \mid Av_2 \mid \ldots \mid Av_n).$$

Also, since $Pe_1 = (v_1 \mid \ldots \mid v_n)e_1 = v_1$, we have that $P^{-1}v_1 = e_1$ and so

$$P^{-1}AP = P^{-1}(\lambda v_1 \mid Av_2 \mid \ldots \mid Av_n) = (\lambda P^{-1}v_1 \mid \ldots) = (\lambda e_1 \mid \ldots) = \begin{pmatrix} \lambda & w^T \\ 0 & B \end{pmatrix}$$

for some matrix $B$ and some vector $w$. Crucially:

$$\det(P^{-1}AP - xI) \overset{\text{splits}}{=} \det \begin{pmatrix} \lambda - x & w^T \\ 0 & B - xI \end{pmatrix} = (\lambda - x) \det(B - xI)$$

68
and so the characteristic polynomial of $B$ splits. By assumption this means that $B = QUQ^{-1}$ for some upper-triangular matrix $Q$. So finally we may define

$$R = \begin{pmatrix} 1 & 0^T \\ 0 & Q \end{pmatrix}$$

such that

$$(PR)^{-1}A(PR) = R^{-1}(P^{-1}AP)R = \begin{pmatrix} 1 & 0^T \\ 0 & Q^{-1} \end{pmatrix} \begin{pmatrix} \lambda & w^T \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0^T \\ 0 & Q \end{pmatrix}$$

$$= \begin{pmatrix} \lambda & w^T \\ 0 & Q^{-1}B \end{pmatrix} \begin{pmatrix} 1 & 0^T \\ 0 & Q \end{pmatrix}$$

where $Q^{-1}BQ = U$, which is upper-triangular, and so $A$ is indeed similar to an upper-triangular matrix. \qed

In other words, the result above tells us that if the characteristic polynomial of a linear map splits then there exists a basis with respect to which this linear map can be represented as an upper-triangular matrix. We now seek to improve this by showing that, under the same conditions, we can actually find an orthonormal basis. Before we can prove that we need a better understanding of the geometry underlying upper-triangular matrices.

**Lemma 6.23.** Let $T : V \rightarrow V$ and let $v_1, \ldots, v_n$ be a basis of $V$. The following are equivalent.

(a) The matrix of $T$ with respect to $v_1, \ldots, v_n$ is upper-triangular

(b) $Tv_j \in \text{span}(v_1, \ldots, v_j)$ for every $j = 1, \ldots, n$.

(c) $S_j = \text{span}(v_1, \ldots, v_j)$ is invariant under $T$ for every $j = 1, \ldots, n$, i.e. $T(w) \in S_j$ for every $w \in S_j$.

**Proof.** Clearly (a) and (b) are equivalent, see for example the 3-by-3 case

$$\begin{pmatrix} X & X & X \\ 0 & X & X \\ 0 & 0 & X \end{pmatrix}.$$

The second column of that matrix corresponds to the coordinates of $T(v_2)$, and since only the first two coordinates are nonzero this tells us that indeed $T(v_2)$ is a linear combination of $v_1$ and $v_2$, i.e. $T(v_2) \in \text{span}(v_1, v_2)$. We also see that (c) implies (b) and so it suffices to prove that (b) implies (c).

Suppose (b) holds and let $w \in S_j$. Then

$$w = a_1v_1 + \cdots + a_jv_j$$

and hence

$$T(w) = a_1T(v_1) + \cdots + a_jT(v_j) \in \text{span}(v_1, \ldots, v_j).$$

\qed
We are now ready prove Schur’s Lemma, the eponymous (and thus key) result of this section.

**Theorem 6.24** (Schur’s Lemma). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space and let \(T : V \rightarrow V\) be a linear map whose characteristic polynomial splits. There is an orthonormal basis of \(V\) with respect to which \(T\) is upper-triangular.

**Proof.** Since the characteristic polynomial of \(T\) splits there is a basis \(v_1, \ldots, v_n\) with respect to which the matrix of \(T\) is upper-triangular, and so

\[
S_j = \text{span}(v_1, \ldots, v_j)
\]

is invariant under \(T\). Applying Gram-Schmidt to \(v_1, \ldots, v_n\) we obtain mutually orthogonal vectors \(w_1, \ldots, w_n\) for which

\[
\text{span}(w_1, \ldots, w_j) = \text{span}(v_1, \ldots, v_j) = S_j.
\]

Therefore, since \(S_j\) is invariant under \(T\) for every \(j\), the matrix of \(T\) with respect to \(w_1, \ldots, w_n\) is also upper-triangular. To conclude it suffices to normalise and use the basis \(\tilde{w}_i = \frac{w_i}{||w_i||}\).

**6.5 Self-adjoint and normal operators**

Now that we have proved Schur’s Lemma we are in a position to wrap-up the course in a very neat way by combining two important ideas: diagonalisability and orthonormal bases. On the one hand diagonalisable linear maps are particularly nice because computing with them boils down to multiplying by scalars along appropriate directions while on the other hand orthonormal bases are particularly nice because they make a lot of computations much easier to carry out (compared with non-orthonormal bases). In this section we will characterise the linear maps (on finite-dimensional inner product spaces) which may be diagonalised using an orthonormal basis. First we introduce some new vocabulary.

**Definition 6.25** (Self-adjoint). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space and let \(T : V \rightarrow V\) be a linear map. We say that \(T\) is self-adjoint if \(T = T^*\).

**Lemma 6.26.** The eigenvalues of self-adjoint linear maps are real.

**Proof.** Let \(\lambda\) be an eigenvalue of a self-adjoint linear map \(T\) with associated eigenvector \(v\). Then

\[
\begin{align*}
\langle T(v), v \rangle &= \langle \lambda v, v \rangle = \lambda ||v||^2 \\
\langle T(v), v \rangle &= \langle v, T^*(v) \rangle = \langle v, T(v) \rangle = \langle v, \lambda v \rangle = \overline{\lambda} ||v||^2
\end{align*}
\]

such that \(\lambda = \overline{\lambda}\) and so indeed \(\lambda\) must be real. \(\square\)

While interesting in its own right, the importance of Lemma 6.26 for the purpose of this section comes from the fact that it ensures that the characteristic polynomial of real self-adjoint operators splits. Combined with Schur’s Lemma, this will thus guaranteed the orthonormal diagonalisability of self-adjoint linear maps.
Theorem 6.27 (Characterisation of linear maps diagonalisable via an orthonormal basis over the reals). Let \((V, \langle \cdot, \cdot \rangle)\) be a finite-dimensional inner product space over \(\mathbb{R}\) and let \(T : V \to V\) be a linear map. \(T\) is self-adjoint if and only if there exists an orthonormal basis of \(V\) of eigenvectors of \(T\).

Proof. Suppose that there exists an orthonormal basis of \(V\) of eigenvectors of \(T\). In that basis the matrix \(A\) of \(T\) is diagonal and the matrix of \(T^*\) is \(A^T\). \(A\) is diagonal so \(A = A^T\) and indeed \(T = T^*\).

Conversely, suppose that \(T\) is self-adjoint. Its eigenvalues must be real and so its characteristic polynomial splits over \(\mathbb{R}\). By Schur’s Lemma there exists an orthonormal basis \(v_1, \ldots, v_n\) of \(V\) with respect to which the matrix \(A\) of \(T\) is upper-triangular. But then the matrix of \(T^*\) is \(A^T = B\) which is lower-triangular, i.e. \(b_{ij} = 0\) if \(i < j\). Since \(T\) is self-adjoint: \(A = A^T\), where \(A\) is lower-triangular and \(A^T\) is upper-triangular, and so \(A\) must be diagonal. This proves that the basis vectors \(v_1, \ldots, v_n\) are eigenvectors of \(T\). \(\square\)

We have characterised the linear maps that can be diagonalised over the reals using orthonormal bases, so now we turn our attention to the complex case. Once again we begin by introducing appropriate vocabulary.

Definition 6.28 (Normal). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space and let \(T : V \to V\) be a linear map. We say that \(T\) is normal if \(T^* T = T T^*\).

Lemma 6.29. Let \(T : V \to V\) be normal. Then \(|\|Tv\|\| = |\|T^*v\|\|\) for all \(v \in V\).

Proof. This follows from a direct computation:

\[ |\|T(v)\|\|^2 = \langle T(v), T(v) \rangle = \langle T^*T(v), v \rangle = \langle TT^*(v), v \rangle = |\|T^*(v)\|\|^2. \]

We are now ready to prove the analog of Theorem 6.27 but for diagonalisation over the complex numbers.

Theorem 6.30 (Characterisation of linear maps diagonalisable via an orthonormal basis over the complex numbers). Let \((V, \langle \cdot, \cdot \rangle)\) be an inner product space over \(\mathbb{C}\) and let \(T : V \to V\) be a linear map. \(T\) is normal if and only if there exists an orthonormal basis of \(V\) of eigenvectors of \(T\).

Proof. If there is an orthonormal basis of \(V\) of eigenvectors of \(T\) then the matrices \(A\) and \(A^T\) of \(T\) and \(T^*\) are diagonal and \(AA^T = A^T A\), so \(TT^* = T^* T\), since diagonal matrices commute.

Conversely, if \(T\) is normal then by Schur’s Lemma there is an orthonormal basis \(v_1, \ldots, v_n\) of \(V\) such that its matrix \(A\) is upper-triangular, i.e.

\[
A = \begin{pmatrix}
  a_{11} & \cdots & a_{1n} \\
  \vdots & \ddots & \vdots \\
  0 & \cdots & a_{nn}
\end{pmatrix}
\quad \text{and} \quad
A^T = \begin{pmatrix}
  \bar{a}_{11} & \cdots & 0 \\
  \vdots & \ddots & \vdots \\
  \bar{a}_{1n} & \cdots & \bar{a}_{nn}
\end{pmatrix}
\]

Since \(|\|T(v_1)\|| = |\|T^*(v_1)\||\), where

\[ T(v_1) = a_{11}v_1 \quad \text{and} \quad T^*(v_1) = \bar{a}_{11}v_1 + \cdots + \bar{a}_{1n}v_n, \]

71
we have that
\[
\begin{align*}
|T(v_1)|^2 &= |a_{11}|^2 \\
|T^*(v_1)|^2 &= |a_{11}|^2 + \cdots + |a_{nn}|^2
\end{align*}
\]
and so we deduce that the first row of \( A \) must be equal to
\[
(a_{11} \ 0 \ \cdots \ 0).
\]
Repeating this computation \( n \) times where \( |T(v_i)| = |T^*(v_i)| \) for every \( i \) we conclude that \( A \) is diagonal, so that indeed the basis vectors \( v_1, \ldots, v_n \) are eigenvectors of \( A \).

6.6 Orthogonal and unitary matrices

In the realm of “vanilla” vector spaces, before inner products where introduced, we saw that we could characterise a linear map \( T \) being diagonalisable in two ways:

- there is a basis in which the matrix of that linear map \( T \) is diagonal, or
- in any basis, the matrix \( A \) of that linear map is similar to a diagonal matrix, i.e. \( A = QDQ^{-1} \) for some diagonal matrix \( D \).

In the realm of inner product spaces, as we saw in the previous section, we are looking for something a bit more restrictive, namely looking for linear maps which can be diagonalised using orthonormal bases. So what is the appropriate notion of “similarity” in that case? In order to answer that question we need some new vocabulary.

**Definition 6.31** (Orthogonal and unitary matrices).

1. A real matrix \( A \) is called **orthogonal** if \( A^T = A^{-1} \).
2. A complex matrix \( A \) is called **unitary** if \( \overline{A^T} = A^{-1} \).

**Example 6.32.** Here are some examples of orthogonal and unitary matrices

1. The matrix \( Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \) is orthogonal since
\[
\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = I.
\]
2. The matrix \( U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & -i \end{pmatrix} \) is unitary since
\[
\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & -i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = I.
\]
To answer the question “What does similarity look like if we require the bases to be orthonormal?” we need a better understanding of orthogonal and unitary matrices, which the following result provides.

**Lemma 6.33.**

1. A real matrix $A$ is **orthogonal** if and only if its columns form an orthonormal basis of $\mathbb{R}^n$ equipped with the dot product.
2. A complex matrix $A$ is **unitary** if and only if its columns form an orthonormal basis of $\mathbb{C}^n$ equipped with the complex dot product.

**Proof.** We only prove (2) since (1) follows in the same way, only neglecting all complex conjugates. Let $A$ be a complex $n$-by-$n$ matrix and write

$$A = (v_1 | \ldots | v_n).$$

Then

$$A^T A = \begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix} (v_1 | \ldots | v_n) = \begin{pmatrix} v_1^T v_1 & \ldots & v_1^T v_n \\ \vdots & \ddots & \vdots \\ v_n^T v_1 & \ldots & v_n^T v_n \end{pmatrix}.$$

So indeed $A^T A = I$ if and only if $v_i^T v_j = v_j \cdot v_i = \begin{cases} 1 & \text{if } i = j \\
0 & \text{if } i \neq j \end{cases}$, i.e. $A$ is unitary if and only if the $v_i$’s form an orthonormal basis. \qed

We are almost ready to phrase the characterisation of “similarity” pertaining to orthonormal bases. First we need to translate the words “self-adjoint” and “normal” from the world of linear maps into the world of matrices. Only some new vocabulary is needed for that.

**Definition 6.34** (Symmetric and normal matrices).

1. A real matrix $A$ is called **symmetric** if $A^T = A$.
2. A complex matrix $A$ is called **normal** if $A^T A = A A^T$.

Using all the vocabulary introduced in this section we can now rephrase the main results of the previous section in terms of matrices.

**Corollary 6.35** (Orthogonal and unitary similarity).

1. A real matrix $A$ is **symmetric** if and only if it is orthogonally similar to a diagonal matrix, i.e. there exists an orthogonal matrix $Q$ and a diagonal matrix $D$ such that $A = Q D Q^T$. 

73
2. A complex matrix $A$ is normal if and only if it is unitarily similar to a diagonal matrix, i.e. there exists a unitary matrix $U$ and a diagonal matrix $D$ such that

$$A = UDU^T.$$  

Proof. We write $T$ for the linear map corresponding to multiplication by $A$ (acting on $\mathbb{R}^n$ in the first case, acting on $\mathbb{C}^n$ in the second case). Then, in the real case,

$$A \text{ is symmetric} \iff T \text{ is self-adjoint} \iff \text{there is an orthonormal basis of eigenvectors of } T \iff A \text{ is orthogonally similar to a diagonal matrix},$$

and, in the complex case

$$A \text{ is normal} \iff T \text{ is normal} \iff \text{there is an orthonormal basis of eigenvectors of } T \iff A \text{ is unitarily similar to a diagonal matrix}.$$  

Example 6.36. Here are examples of matrices that are orthogonally or unitarily diagonalisable, i.e. orthogonally or unitarily similar to a diagonal matrix.

1. The matrix $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ is orthogonally similar to a diagonal matrix since

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$  

2. The matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is unitarily similar to a diagonal matrix since

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$  

3. The matrix $A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$ is unitarily similar to a diagonal matrix since

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \begin{pmatrix} 2 - 3i & 0 \\ 0 & 2 + 3i \end{pmatrix}.$$  

74
Summary

Now that we come to the end of our (introductory) journey through linear algebra it is a good time to look back and take stock of what we’ve discussed.

The fundamental observation was that, once we fix a basis \( v_1, \ldots, v_n \) of a vector space \( V \), we have access to coordinates since for every \( v \in V \) there exists a unique \( x \in \mathbb{F}^n \) such that

\[
v = x_1 v_1 + \cdots + x_n v_n,
\]

where we refer to \( x \) as the coordinates of \( v \). (Note that the “existence” part comes from the fact that bases span the vector space while the “uniqueness” part comes from the fact that bases are linearly independent.)

This single correspondence, between a vector \( v \in V \) and its coordinates \( x \in \mathbb{F}^n \), then gives rise to a slew of correspondences between “vector–land” and “matrix–land”:

\[
\begin{align*}
    v \in V & \leftrightarrow x \in \mathbb{F}^n \\
    \text{linear map } T & \leftrightarrow \text{matrix } A \\
    \ker T & \leftrightarrow \{ x \in \mathbb{F}^n : Ax = 0 \} \\
    \im T & \leftrightarrow \{ b \in \mathbb{F}^n : Ax = b \text{ has a solution } \} \\
    S \circ T & \leftrightarrow BA (\det BA = \det B \det A) \\
    T^{-1} & \leftrightarrow A^{-1}.
\end{align*}
\]

Since invertibility is such an important notion, it is worth stopping here and talking about that some more, noting that altogether we have obtained several equivalent characterisations of invertibility:

\[
A \text{ is invertible } \iff \det A = 0 \\
\iff \im A = \mathbb{F}^n \text{ and } \ker A = \{0\} \\
\iff \text{rank } A = n \text{ and nullity } A = 0 \\
\iff \text{rank } A = n \text{ or nullity } A = 0.
\]

Finally we considered eigenvalues and eigenvectors, where once more the correspondence holds:

\[
T(v) = \lambda v \leftrightarrow Ax = \lambda x,
\]

and moreover we know that the eigenvalues are roots of the characteristic polynomial

\[
p_A(\lambda) = \det(A - \lambda I).
\]

We then said “let there be light angles and lengths” and introduced inner products. By Gram-Schmidt we quickly discovered that in any inner product space had an orthonormal basis. Using orthonormal bases allowed us to continue adding more items to our already long list of concepts that have manifestations...
in both “vector–land” and “matrix–land”:

\[ v, w \in V \quad \longleftrightarrow \quad x, y \in \mathbb{F}^n \ (\mathbb{R}^n \text{ or } \mathbb{C}^n) \]
\[ \langle v, w \rangle = x \cdot \overline{y} \]
\[ T \quad \longleftrightarrow \quad A \]
\[ T^* \quad \longleftrightarrow \quad A^T. \]

The culmination of our journey was the characterisation of the linear maps which could be diagonalised by way of orthonormal bases:

\[ T \text{ has an orthonormal basis of eigenvectors} \iff \begin{cases} T \text{ is self-adjoint (over } \mathbb{R}) \\ T \text{ is normal (over } \mathbb{C}) \end{cases} \iff \begin{cases} A = QDQ^T \ (\text{over } \mathbb{R}) \\ A = UDU^T \ (\text{over } \mathbb{C}) \end{cases} \]