# Linear Algebra I
## Prereading

Luke Phillips

October 2024

# Contents

# 1 The vector space $\mathbb{R}^n$

**Definition 1.0.1** *(n-dimensional real space).* The space $\mathbb{R}^n$ (n-dimensional real space) is the collection of all (column) vectors of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Vectors are typically denoted by bold letters such as $\mathbf{x}$ and sometimes using underlined letters $\underline{x}$ to differentiate them from ordinary real numbers.

We write $\mathbf{0}$ for the zero-vector

$$\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

this is the point of origin of $\mathbb{R}^n$.

There are two important operations which can be performed on vectors.

**Addition**: for any vectors $\mathbf{v}$ and $\mathbf{w}$, we define

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{pmatrix}$$

this is component-wise addition.

**Scalar multiplication**: for any vector $\mathbf{v} \in \mathbb{R}^n$ and real number $\lambda \in \mathbb{R}$ we define

$$\lambda \mathbf{v} = \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix}$$

this is component-wise multiplication by $\lambda$.

**Proposition 1.0.2.** *These operations satisfy the axioms of a real vector space*

## Axioms for Addition

(i) existence of additive identity: there is a vector $\mathbf{0}$ in $\mathbb{R}^n$ so that $\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}$ for all $\mathbf{v}$ in $\mathbb{R}^n$.

(ii) commutativity: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ for all $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$

(iii) existence of additive inverses: for each $\mathbf{v}$ in $\mathbb{R}^n$ there is an element $-\mathbf{v}$ in $\mathbb{R}^n$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0} = (-\mathbf{v}) + \mathbf{v}$

(iv) associativity: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$

these axioms can be summarised by saying that $\mathbb{R}^n$ is an abelian group under addition.

## Axioms for Scalar Multiplication

(i) multiplication by 0: $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v}$ in $\mathbb{R}^n$

(ii) multiplication by 1: $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$ in $\mathbb{R}^n$

(iii) associativity: $(\lambda\mu)\mathbf{v} = \lambda(\mu\mathbf{v})\ \forall \lambda, \mu \in \mathbb{R}$ and $\forall \mathbf{v} \in \mathbb{R}^n$

(iv) distributivity: $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$ and $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}\ \forall \lambda, \mu \in \mathbb{R}$ and $\forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$

*Proof.* The above axioms ate trivially satisfied by $\mathbb{R}$ (the case $n = 1$), many of the proofs follow from this observation. Once we set

$$\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad -\mathbf{v} = -\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} -v_1 \\ \vdots \\ -v_n \end{pmatrix} = (-1)\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

then the proofs are very straight forward $\qquad\qquad\qquad\qquad\qquad \square$

**Definition 1.0.3** *(Standard basis vectors).* We set

$$\mathbf{e_1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathbf{e_2} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \ \mathbf{e_n} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

and call these vectors the standard basis vectors of $\mathbb{R}^n$.

Sometimes the entries of a vector are referred to as the Cartesian coordinates of a point in space.

Addition can be thought of as an operation that inputs a pair of vectors, each from $\mathbb{R}^n$, and outputs a single vector in $\mathbb{R}^n$. In other words, addition is a function $\mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$. Similarly, scalar multiplication is a function $\mathbb{R} \times \mathbb{R}^n \mapsto \mathbb{R}^n$. The domains of each of these functions are called Cartesian products.

## 1.1 The scalar product in $\mathbb{R}^n$

**Definition 1.1.1** *(Scalar product).* The scalar (or dot) product is a function

$$\mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$$

defined as follows. For $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ set

$$\mathbf{v} \cdot \mathbf{w} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \cdots + v_n w_n.$$

This is a real number, a scalar.

The transpose $\mathbf{v}^t$ of $\mathbf{v}$ is defined to be the row vector with the same entries as the column vector $\mathbf{v}$. So if

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \qquad \text{then} \qquad \mathbf{v}^t = (v_1, \ldots, v_n).$$

The dot product of $\mathbf{v}$ and $\mathbf{w}$ is then obtained by matrix multiplying $\mathbf{v}^t$ and $\mathbf{w}$.

**Proposition 1.1.2** (Properties of the scalar product)**.**

(i) *Symmetry*

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \qquad \textit{for all } \mathbf{u}, \mathbf{v} \textit{ in } \mathbb{R}^n$$

(ii) *Linearity in first factor*

(a) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \qquad \textit{for all } \mathbf{u}, \ \mathbf{v}, \ \mathbf{w} \textit{ in } \mathbb{R}^n$

(b) $(\lambda\mathbf{u}) \cdot \mathbf{v} = \lambda(\mathbf{u} \cdot \mathbf{v}) \qquad \textit{for all } \lambda \in \mathbb{R} \textit{ and all } \mathbf{u}, \ \mathbf{v} \textit{ in } \mathbb{R}^n$

(iii) *Linearity in the second factor*

(a) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{w} + \mathbf{u} \cdot \mathbf{v} \qquad \textit{for all } \mathbf{u}, \ \mathbf{v}, \ \mathbf{w} \textit{ in } \mathbb{R}^n$

(b) $\mathbf{u} \cdot (\lambda\mathbf{v}) = \lambda(\mathbf{u} \cdot \mathbf{v}) \qquad \textit{for all } \lambda \in \mathbb{R} \textit{ and all } \mathbf{u}, \ \mathbf{v} \textit{ in } \mathbb{R}^n$

(iv) *Positivity*

$\mathbf{v} \cdot \mathbf{v} \geq 0$ *for all $v$, and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$.*

*Proof.*

(i) First we can determine some expressions for $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{v} \cdot \mathbf{u}$

$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$ and

$\mathbf{v} \cdot \mathbf{u} = v_1 u_1 + v_2 u_2 + \ldots + v_n u_n$.

Then by the commutativity of multiplication we have

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= u_1 v_1 + u_2 v_2 + \ldots + u_n v_n \\ &= v_1 u_1 + v_2 u_2 + \ldots + v_n u_n \\ &= \mathbf{v} \cdot \mathbf{u} \end{aligned}$$

which proves our goal as required

$\square$

**Definition 1.1.3** *(Magnitude)*. Suppose

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

then $\mathbf{v} \cdot \mathbf{v} = v_1^2 + \ldots + v_n^2$ which is a non-negative real number. We define the magnitude (or length) of $\mathbf{v}$, denoted $|\mathbf{v}|$, by

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

So long as the $n$ coordinate directions are mutually orthogonal, the magnitude

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{x_1^2 + \cdots + v_n^2}$$

is the euclidean distance from the origin to the point with coordinates given by $\mathbf{v}$. We say that a vector is non-trivial if it has positive length. It is easy to see that the only trivial vector in $\mathbb{R}^n$ is the zero vector $\mathbf{0}$.

We can describe the position of a point by what direction it lies and what distance away from us it is, using polar coordinates.

Using the length of a vector in $\mathbb{R}^2$ we define polar co-ordinates on $\mathbb{R}^2$ as follows. Let $\mathbf{v} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ be a non-trivial vector with length $|\mathbf{v}| = r > 0$. Let $\theta \in [0, 2\pi)$ be the angle at origin $\mathbf{0}$ measured in an anticlockwise direction from the $x$-axis to $\mathbf{v}$. By basic trigonometry

$$\cos(\theta) = \frac{x}{r}, \qquad \sin(\theta) = \frac{y}{r}.$$

We define the polar coordinates of $\mathbf{v} \neq \mathbf{0}$ to be the unique $(r, \theta)$ in $\mathbb{R}_+ \times [0, 2\pi)$ so that

$$\mathbf{v} = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix} = r\left(\cos(\theta)\sin(\theta)\right)$$

**Lemma 1.1.4.** *Suppose that $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^2$ of length $r > 0$ and $s > 0$ respectively. Let $\theta$ be the angle between them. Then*

$$\mathbf{v} \cdot \mathbf{w} = rs\cos(\theta).$$

*In particular, we have*

$$|\mathbf{v} \cdot \mathbf{w}| \leq |\mathbf{v}||\mathbf{w}|$$

*with equality if and only if $\mathbf{v}$ and $\mathbf{w}$ are collinear, i.e. are scalar multiples of each other.*

*Proof.* Suppose that $\mathbf{v}$ has polar coordinates $(r, \gamma)$ and $\mathbf{w}$ has polar coordinates $(s, \rho)$. The angle between $\mathbf{v}$ and $\mathbf{w}$ is $\theta = |\gamma - \rho|$. Then

$$\mathbf{v} = \begin{pmatrix} r\cos\gamma \\ r\sin\gamma \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} s\cos\rho \\ s\sin\rho \end{pmatrix}.$$

Applying the dot product using these definitions of $\mathbf{v}$ and $\mathbf{w}$ we obtain the following, applying trigonometric addition formula and the fact that cosine is even we get,

$$\mathbf{v} \cdot \mathbf{w} = rs\cos\gamma\cos\rho + rs\sin\gamma\sin\rho = rs(\cos(\gamma - \rho)) = rs\cos(|\gamma - \rho|) = rs\cos\theta.$$

$\square$

Note that using this lemma we can obtain an expression for the angle between two vectors

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}||\mathbf{w}|}$$

**Theorem 1.1.5** (*Cauchy-Schwarz inequality*). *Suppose that $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^n$. Then*

$$|\mathbf{v} \cdot \mathbf{w}| \leq |\mathbf{v}||\mathbf{w}|$$

*with equality if and only if $\mathbf{v}$ and $\mathbf{w}$ are collinear*

*Proof.* $\square$

**Definition 1.1.6** *(Angle between two vectors).* Suppose that $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^n$ of length $r > 0$ and $s > 0$ respectively. Then we define the angle $\theta \in [0, \pi]$ between them by

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{rs}.$$

We say that the vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal (or perpendicular) if the angle between them is $\frac{\pi}{2}$.

**Corollary 1.1.7.** *The non-trivial vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ are orthogonal if and only if we have $\mathbf{v} \cdot \mathbf{w} = 0$.*

*Proof.* By Definition 1.1.6 we have that[1]

$$\mathbf{v} \cdot \mathbf{w} = 0 \iff \cos \theta = 0 \iff \theta = \frac{\pi}{2}.$$

$\square$

## 1.2 The vector product in $\mathbb{R}^3$

**Definition 1.2.1** *(Vector product in $\mathbb{R}^3$).* The vector (or cross) product is a function

$$\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$$

defined as follows. If $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, their cross product is given by

$$\mathbf{v} \times \mathbf{w} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - w_1 v_2 \end{pmatrix}$$

**Example 1.2.2.**

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 - 6 \\ 9 - 1 \\ 2 - 6 \end{pmatrix} = \begin{pmatrix} -4 \\ 8 \\ -4 \end{pmatrix}$$

**Proposition 1.2.3** (Properties of the cross product)**.** *(i) Anti-Symmetry:*

$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u} \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^3$$

*(ii) Linearity in first factor*

    *(1)* $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w} \quad \text{for all } \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$

    *(2)* $(\lambda \mathbf{u}) \times \mathbf{v} = \lambda(\mathbf{u} \times \mathbf{v}) \quad \text{for all } \lambda \in \mathbb{R} \text{ and all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^3$

*(iii) Linearity in the second factor*

    *(1)* $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{w} + \mathbf{u} \times \mathbf{v} \quad \text{for all } \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$

    *(2)* $\mathbf{u} \times (\lambda \mathbf{v}) = \lambda(\mathbf{u} \times \mathbf{v}) \quad \text{for all } \lambda \in \mathbb{R} \text{ and all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^3$

---

[1]Since $\cos \theta$ can only be 0 when $\theta = \frac{\pi}{2}$ then they have to be orthogonal and the converse is true too.

*(iv) Orthogonality to the input*

$$\mathbf{v} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{w} \cdot (\mathbf{v} \times \mathbf{w}) = 0 \quad \text{for all vectors } \mathbf{v} \text{ and } \mathbf{w} \text{ in } \mathbb{R}^3$$

**Lemma 1.2.4.** *Suppose that* $\mathbf{v}$ *and* $\mathbf{w}$ *are vectors of length* $r > 0$ *and* $s > 0$ *respectively. Let* $\theta \in [0, \pi]$ *be the angle between them. Then*

$$|\mathbf{v} \times \mathbf{w}| = rs \sin(\theta).$$

*In particular, we have*

$$\mathbf{v} \times \mathbf{w} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{v} \text{ and } \mathbf{w} \text{ are collinear.}$$

*Proof.* We have

$$
\left| \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - w_1 v_2 \end{pmatrix} \right|^2 = (v_2 w_3 - v_3 w_2)^2 + (v_3 w_1 - v_1 w_3)^2 + (v_1 w_2 - w_1 v_2)^2
$$

$$
= (v_1^2 + v_2^2 + v_3^2)(w_1^2 + w_2^2 + w_3^2) - (v_1 w_1 + v_2 w_2 + v_3 w_3)^2
$$

$$
= |\mathbf{v}|^2 |\mathbf{w}|^2 - (\mathbf{v} \cdot \mathbf{w})^2
$$

$$
= |\mathbf{v}|^2 |\mathbf{w}|^2 - (|\mathbf{v}||\mathbf{w}| \cos(\theta))^2
$$

$$
= r^2 s^2 (1 - \cos(\theta))^2
$$

$$
= (rs \sin(\theta))^2.
$$

Since $\sin \theta \geq 0$ for $\theta \in [0, \pi]$, we are done. $\qquad\square$

## 1.3 Planes and Lines in $\mathbb{R}^3$

### 1.3.1 Planes

**Parametric Form**

A plane $\Pi$ in $\mathbb{R}^3$ can be described in parametric form as the collection of all points corresponding to vectors of the form

$$\mathbf{a} + \lambda \mathbf{d_1} + \mu \mathbf{d_2}$$

where $\lambda$ and $\mu$ each run through $\mathbb{R}$. $\mathbf{a}$ is any point on the plane and $\mathbf{d_1}$ and $\mathbf{d_2}$ are any two direction vectors for $\Pi$, which lie on the plane and are not collinear.

For example, if $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ are all points on $\Pi$ which are not collinear then we could take $\mathbf{d_1} = \mathbf{b} - \mathbf{a}$ and $\mathbf{d_2} = \mathbf{c} - \mathbf{a}$.

We write

$$\Pi = \{\mathbf{a} + \lambda \mathbf{d_1} + \mu \mathbf{d_2} : \lambda, \mu \in \mathbb{R}\}.$$

**Using the normal vector**

We can describe a plane via its normal vector.

$$\mathbf{d_1} \cdot (\mathbf{d_1} \times \mathbf{d_2}) = 0 = \mathbf{d_2} \cdot (\mathbf{d_1} \times \mathbf{d_2})$$

that is, $\mathbf{d_1} \times \mathbf{d_2}$ is orthogonal to both $\mathbf{d_1}$ and $\mathbf{d_2}$.

Let us write $\mathbf{n}$ for $\mathbf{d}_1 \times \mathbf{d}_2$. Additionally we can write the plane in terms of the dot product

$$\mathbf{n} \cdot (\mathbf{x} - \mathbf{a}) = 0.$$

We can rearrange this to have the normal form of the plane $\Pi$: the plane $\Pi$ is the set of vectors satisfying

$$\mathbf{n} \cdot \mathbf{x} = \ell$$

where $\ell = \mathbf{n} \cdot \mathbf{a} \in \mathbb{R}$ since $\mathbf{n} \cdot \mathbf{a}$ doesn't depend on $\mathbf{x}$ it is just a real number.

**Cartesian form**

A variation of the normal form, really just the same description but now written in Cartesian coordinates immediately follows. Set $\mathbf{n} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

Then the normal form above becomes the Cartesian (or the Cartesian form) of the plane:

$$ax + by + cz = \ell$$

Thus we have

$$\Pi = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{x} = \ell\} = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} : ax + by + cz = \ell \right\}$$

### 1.3.2 Lines

**Parametric form**

A line $L$ in $\mathbb{R}^3$ can be described in parametric form as the collection of all points corresponding to vectors of the form

$$\mathbf{a} + t\mathbf{d}$$

where $t$ runs through $\mathbb{R}$.

$$L = \{\mathbf{a} + t\mathbf{d} : t \in \mathbb{R}\}.$$

If $\mathbf{b} \neq \mathbf{a}$ is another point on $L$, then we could take $\mathbf{d} = \mathbf{b} - \mathbf{a}$.

**Cartesian form**

Suppose the line is given in parametric form by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + t \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}.$$

If all $d_i \neq 0$ then we may solve for $t$ in each component and obtain

$$\frac{x - a_1}{d_1} = \frac{y - a_2}{d_2} = \frac{z - a_3}{d_3}.$$

If one of the $d_i = 0$, such as $d_3 = 0$ then we would get two equations

$$\frac{x - a_1}{d_1} = \frac{y - a_2}{d_2} \qquad \text{and} \qquad z = a_3$$

**Solution set for a pair of linear equations** A pair of planes in $\mathbb{R}^3$ is either parallel, in which case they have the same normal vector, or else they intersect in a line.

$$ax + by + cz = \ell$$
$$dx + ey + fz = m$$

the either the two normal vectors

$$\mathbf{n}_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad \mathbf{n}_2 = \begin{pmatrix} d \\ e \\ f \end{pmatrix}$$

are multiples of each other or else the solution is a line.

**Definition 1.3.1** *(Scalar triple product).* The scalar triple product is a function

$$\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}.$$

Given three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in $\mathbb{R}^3$ we define their scalar triple product $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ by

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

**Lemma 1.3.2** (Invariance under cyclic permutation)**.** *For any three vectors* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *in* $\mathbb{R}^3$ *we have*

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] = [\mathbf{b}, \mathbf{c}, \mathbf{a}] = [\mathbf{c}, \mathbf{a}, \mathbf{b}] = -[\mathbf{a}, \mathbf{b}, \mathbf{c}] = -[\mathbf{b}, \mathbf{c}, \mathbf{a}] = -[\mathbf{c}, \mathbf{a}, \mathbf{b}].$$

# 2  Matrix algebra

## 2.1  What is a matrix?

**Definition 2.1.1** *(Matrix).* A matrix is a rectangular array of numbers (or symbols, or functions) arranged into rows and columns.

**Notation and Terminology:**

- A matrix which has $m$ rows and $n$ columns is referred to as an $m \times n$ matrix. The individual entries in the matrix are sometimes referred to as its elements. The entry in the $i$th row and $j$th column of a matrix is referred to as the $(i, j)$th element of a matrix $M$, usually denoted $M_{ij}$.

- Matrices which consist of a single column are (column) vectors. Matrices which consist of a single row are row vectors.

**Definition 2.1.2** *(Transpose).* Given an $m \times n$ matrix $W$ we define $W^t$, the transpose of $W$, to be the $n \times m$ matrix whose $(i, j)$th entry is the $(j, i)$th entry of $W$; that is the matrix for which

$$(W^t)_{ij} = W_{ji}.$$

A matrix which is equal to its transpose is said to be symmetric.

One way of describing or otherwise specifying the elements of a matrix is as follows. If $f(i, j)$ is some expression depending on $i$ and $j$ then $(f(i, j))_{m \times n}$ stands for the $m \times n$ matrix whose $(i, j)$th entry is $f(i, j)$. We often write an $m \times n$ matrix as $X = (x_{ij})_{m \times n}$, or $X = (x_{ij})$ if the size is obvious.

We write $O_{m \times n}$ for $(0)_{m \times n}$ that is the $m \times n$ matrix all of whose entries are zero, this is called the $m \times n$ zero matrix.

For $r$ and $s$ integers, we define $\delta_{r\,s}$ to be 1 if $r = s$ and to be 0 if $r \neq s$. (This is called the Kronecker delta.) The matrix $(\delta_{ij})_{m \times n}$ is denoted $I_n$ and is referred to as the $n \times n$ identity matrix.

In general, we write $M_{m \times n}(\mathbb{R})$ or $M_{m,n}(\mathbb{R})$ for the set of all real $m \times n$ matrices. We also write $M_n(\mathbb{R})$ for the set of square matrices of size $n$, so $M_n(\mathbb{R}) = M_{n,n}(\mathbb{R})$. $M_{m,1}(\mathbb{R})$ is the same set as $\mathbb{R}^m$.

## 2.2 Matrix operations I: matrix addition and scalar multiplication

(i) Matrix addition. This is a function

$$M_{m \times n}(\mathbb{R}) \times M_{m \times n}(\mathbb{R}) \to M_{m \times n}(\mathbb{R}).$$

For $X = (x_{ij})_{m \times n}$ and $Y = (y_{ij})_{m \times n}$ both $m \times n$ matrices, we define

$$X + Y = (x_{ij})_{m \times n} + (y_{ij})_{m \times n} = (x_{ij} + y_{ij})_{m \times n},$$

that is, matrix addition is defined by component wise addition. Matrix addition isn't defined when the matrices are of different sizes.

(ii) Scalar multiplication. This is a function

$$\mathbb{R} \times M_{m \times n}(\mathbb{R}) \to M_{m \times n}(\mathbb{R}).$$

For $X = (x_{ij})_{m \times n}$ in $M_{m \times n}(\mathbb{R})$ and $\lambda \in \mathbb{R}$ we set

$$\lambda X = (\lambda x_{ij})_{m \times n}$$

that is, we multiply every entry of the matrix by $\lambda$.

We write $-X$ for $(-1)X$. If $X^t = -X$ we say that $X$ is skew-symmetric (or anti-symmetric).

**Proposition 2.2.1** (Properties of matrix addition and scalar multiplication)**.** *Matrix addition and scalar multiplication for $M_{m,n}(\mathbb{R})$ satisfy exactly the same properties as vector addition and scalar multiplication for $\mathbb{R}^n$. That is, the operations satisfy the axioms of a (real) vector space.*

### Axioms for addition

(i) *existence of additive identity: there is a matrix $O_{m \times n}$ in $M_{m,n}(\mathbb{R})$ so that for all $X$ in $M_{m,n}(\mathbb{R})$ we have $X + O_{m \times n} = X = O_{m \times n} + X$*

(ii) *commutativity: $X + Y = Y + x$ for all $X$ and $Y$ in $M_{m,n}(\mathbb{R})$*

(iii) *existence of additive inverses: for each $X$ in $M_{m,n}(\mathbb{R})$ there is an element $-X$ in $M_{m,n}(\mathbb{R})$ such that $X + (-X) = O_{m \times n} = (-X) + X$*

*(iv) associativity:* $(X + Y) + Z = X + (Y + Z)$ *for all* $X, Y, Z \in M_{m,n}(\mathbb{R})$

*(As before this makes $M_{m,n}(\mathbb{R})$ an abelian group under addition)*

### Axioms for Scalar Multiplication

(i) $0X = O_{m \times n}$ *for all $X$ in $M_{m,n}(\mathbb{R})$*

(ii) $1X = X$ *for all $X$ in $M_{m,n}(\mathbb{R})$*

(iii) *associativity:* $(\lambda \mu)X = \lambda(\mu X)$ *for all $\lambda, \mu \in \mathbb{R}$ and for all $X \in M_{m,n}(\mathbb{R})$*

(iv) *distributivity:* $(\lambda + \mu)X = \lambda X + \mu X$ *and* $\lambda(X + Y) = \lambda X + \lambda Y$ *for all real $\lambda, \mu$ and for all $X$ and $Y$ in $M_{m,n}(\mathbb{R})$.*

*Proof.* Do $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

## 2.3  Matrix Operations II: Matrix multiplication

Matrix multiplication: This is a function

$$M_{m \times n}(\mathbb{R}) \times M_{n \times p}(\mathbb{R}) \to M_{m \times p}(\mathbb{R}).$$

If $X = (x_{ij})_{m \times n}$ is $m \times n$ and $Y = (y_{ij})_{n \times p}$ is $n \times p$ (if the number of columns of $X$ equals the number of rows of $Y$) then we may form the product matrix $XY$. This is defined to be the $m \times p$ matrix whose $(i, j)$th element $(XY)_{ij}$ is

$$(XY)_{ij} = x_{i\,1}y_{1\,j} + x_{i\,2}y_{2\,j} + \cdots + x_{i\,n}x_{n\,j} = \sum_{k=1}^{n} x_{ik}y_{kj}.$$

If $\mathbf{x^t}_i$ denotes the $i$-th row of $X$ and $\mathbf{y}_j$ denotes the $j$-th column of $Y$ then $(XY)_{ij} = \mathbf{x}_i \cdot \mathbf{y}_j$ (scalar product of column vectors).

**Proposition 2.3.1** (Multiplication rules)**.** *Let $X$ and $X'$ be $m \times n$, $Y$ and $Y'$ be $n \times p$ and $Z$ be $p \times q$ matrices. Let $\lambda$ and $\lambda'$ be scalars. Then*

(i) $O_{p \times m}X = O_{p \times n}$ *and* $XO_{n \times q} = O_{m \times q}$.

(ii) $I_m X = X = XI_n$.

(iii) $\lambda(XY) = (\lambda X)Y = X(\lambda Y)$.

(iv) *associativity:* $(XY)Z = X(YZ)$

(v) *distributive rules:*

(a) $(X + X')Y = XY + X'Y$ *and* $X(Y + Y') = XY + XY'$.

(b) $\lambda(X + X') = \lambda X + \lambda X'$ *and* $(\lambda + \lambda')X = \lambda X + \lambda'X$.

*Proof.* For (iv), the $k$th entry in the $i$th row of $XY$ is given by $\sum_{r=1}^{n} x_{ir}y_{rk}$, with $k = 1, \ldots, p$. Hence the $(i, j)$th entry of $(XY)Z$ is given by

$$((XY)Z)_{ij} = (i\text{th row of } XY) \cdot (j\text{th column of } Z) = \sum_{k=1}^{p} \left( \sum_{r=1}^{n} x_{ir}y_{rk} \right) z_{kj}.$$

On the other hand, the $r$th entry in the $j$th column of $YZ$ is given by $\sum_{k=1}^{p} y_{rk} z_{kj}$, with $r = 1, \ldots, n$. Hence the $(i, j)$th entry of $X(YZ)$ is given by

$$(X(YZ))_{ij} = (i\text{th row of } X) \cdot (j\text{th column of } YZ) = \sum_{r=1}^{n} x_{ir} \left( \sum_{k=1}^{p} y_{rk} z_{kj} \right).$$

Switching the order of summation using associativity of the usual multiplication given $(XY)Z = X(YZ)$.

**Prove rest!** □

**Warning**: We do not necessarily have $XY = YX$ in general.

**Warning**: It is possible to have $XY = O_{m \times p}$ even though both $X$ and $Y$ are non-zero.

**A useful trick for sometimes multiplying big matrices more easily**

Let $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in M_{m,n}(\mathbb{R})$ and $\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} \in M_{n,p}(\mathbb{R})$ be two matrices given by blocks $A \in M_{m_1,n_1}(\mathbb{R})$, $B \in M_{m_1,n_2}(\mathbb{R})$, $C \in M_{m_2,n_1}(\mathbb{R})$, $D \in M_{m_2,n_2}(\mathbb{R})$, $A' \in M_{n_1,p_1}(\mathbb{R})$, $B' \in M_{n_1,p_2}(\mathbb{R})$, $C' \in M_{n_2,p_1}(\mathbb{R})$, $D' \in M_{n_2,p_2}(\mathbb{R})$ (so-called block matrices) with $m_1 + m_2 = m$ and $n_1 + n_2 = n$ and $p_1 + p_2 = p$. It can easily shown that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} AA' + BC' & AB' + BD' \\ CA' + DC' & CB'DD' \end{pmatrix} \in M_{m,p}(\mathbb{R})$$

that is, if matrix sizes match correctly the formula for the multiplication of block matrices is the same as one of "usual two by two matrices"

For example, if $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ (so that $A^2 = O_{2 \times 2}$) then

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} I_2 & I_2 \\ A & A \end{pmatrix}^2 = \begin{pmatrix} I & I \\ A & A \end{pmatrix} \begin{pmatrix} I & I \\ A & A \end{pmatrix} = \begin{pmatrix} I + A & I + A \\ A + A^2 & A + A^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

## 2.4    Inverse matrix

**Definition 2.4.1.** Let $A$ be a square $(n \times n)$ matrix. We say $A$ is invertible (or non-singular) if there exists an $n \times n$ matrix $B$ such that

$$AB = BA = I_n.$$

If such a matrix $B$ exists we call $B$ the inverse of $A$ and we write $A^{-1}$ for this inverse. (there can only exist at most one inverse). If no such inverse $B$ exists for a matrix $A$, we say $A$ is non-invertible (or singular).

**Proposition 2.4.2** (Properties of the inverse and transpose). *(i) A matrix has at most one inverse.*

(ii) *Assume $A$ and $B$ are invertible $n \times n$ matrices. Then the product $AB$ is invertible and*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

(iii) *We have $(AB)^t = B^t A^t$ and $(\lambda A)^t = \lambda A^t$ for any matrix $A$ and any real constant $\lambda$.*

(iv) *Assume $A$ is invertible. then the transpose $A^t$ is invertible and*

$$(A^t)^{-1} = (A^{-1})^t.$$

(v) *Let $A \in M_{n,n}(\mathbb{R})$ and assume $BA = I_n$ for some $B \in M_{n,n}(\mathbb{R})$. Then $AB = I_n$ as well as vice versa. So the "left inverse" is automatically the "right inverse".*

*Proof.*    (i) If $B'$ were a second inverse for $A$ then we would have

$$B' = B' I_n = B'(AB) = (B'A)B = I_n B = B.$$

$\square$

**Remark 2.4.3.** *We also have*

$$(X + Y)^t = X^t + Y^t,$$

*but it is not the case in general that*

$$(X + Y)^{-1} = X^{-1} + Y^{-1}.$$

# 3   Gauss-Jordan elimination

## 3.1   Solving equations and Gauss-Jordan elimination

We will turn to solving the general problem of solving $m$ linear equations in $n$ variables. To handle this, we need a systematic approach.

We begin by setting up our notation. We shall denote our variables by $x_1, x_2, \ldots, x_n$, and consider the solution of the system of $m$ linear equations

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 & \quad + \cdots + a_{1n}x_n & = b_1; \\
a_{21}x_1 + a_{22}x_2 & \quad + \cdots + a_{2n}x_n & = b_2; \\
\vdots & \quad \ddots \vdots & \vdots; \\
a_{m1}x_1 + a_{m2}x_2 & \quad + \cdots + a_{mn}x_n & = b_m.
\end{aligned}
$$

For short we can write

$$
\sum_{j=1}^{n} a_{ij}x_j = b_i \quad \text{for } i = 1, \ldots, m.
$$

In vector notation we can write this using the dot product

$$
\mathbf{a}_i \cdot \mathbf{x} = \begin{pmatrix} a_{i1} \\ \vdots \\ a_{in} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = b_i \quad \text{for } i = 1, \ldots, m.
$$

Each equation gives a so-called hyperplane with normal vector $\mathbf{a}_i$.

To solve the equations we shall go through the following process.

- First formalise the system of equations as a single equation of matrices, this is for notational convenience and systematises our process.

- Perform series of matrix operations (elementary row operations (ERO's)) that in effect change the original set of equations into a different set of linear equations, but ones that have the same solution set and are much easier to solve. This process is called Gauss-Jordan elimination, and the sort of matrix that represents the equations that are easier to solve is said to be in row reduced echelon form (RREF).

The first step we have essentially already done. Using the notation from before, the system of equations can be written as the matrix equation

$$
A\mathbf{x} = \mathbf{b}
$$

where $A$ is the matrix

$$
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & & \vdots \\
a_{m1} & a_{m2} \cdots a_{mn} &
\end{pmatrix}
$$

and $\mathbf{b}$ is the column vector

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

The augmented matrix associated with a linear system of equations is a rectangular array of numbers, the coefficients of each equation are placed in order in each row, followed by a vertical bar where the equal signs are with the $a_{ij}$ to the left, and then the $b_i$ recorded to the right. In other words

$$\left( \begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right).$$

For example, the system

$$3x_1 - 4x_2 + 3x_3 = 1$$
$$2x_1 - x_2 + 2x_3 = 2$$

has augmented matrix $\left( \begin{array}{ccc|c} 3 & -4 & 3 & 1 \\ 2 & -1 & 2 & 2 \end{array} \right).$

If as before we write the matrix

$$\left( \begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right),$$

which we shall sometimes call the coefficient matrix of the linear system, and as above write

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

then the augmented matrix cab we written as $(A|\mathbf{b})$.

We call a linear system homogenous if $\mathbf{b} = \mathbf{0}$, and inhomogeneous if $\mathbf{b} \neq \mathbf{0}$. Note that a homogeneous system always has at least one solution: $\mathbf{x} = \mathbf{0}$.

**Lemma 3.1.1.**

(i) *Let $A \in M_{n,n}(\mathbb{R})$ be a square matrix and let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector such that $A\mathbf{v} = \mathbf{0}$. Then $A$ is singular.*

(ii) *Let $A \in M_{n,n}(\mathbb{R})$ be an invertible square matrix. Then the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ has a single unique solution. In particular, the only solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.*

*Proof.* For (i), see example 2.4.2(iii) for an easily generalisable argument. For (ii), if $A$ is invertible then

$$\mathbf{x} = I_n\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b}.$$

$\square$

## 3.2 Echelon form and its application to linear systems

**Definition 3.2.1.** A matrix of numbers is said to be in row reduced echelon form (RREF) if

(i) The first (leftmost) non-zero entry in any non-zero row is a 1 (called its leading 1).

(ii) If a row has its leading 1 in the $j$th column then

   (a) all the other entries in the $j$th column are 0; and

   (b) the leading 1s of subsequent rows are in columns to the right of the $j$th column.

(iii) Any row(s) of zeros come after all the rows with non-zero entries.

For example, the matrices

$$\begin{pmatrix} 1 & 4 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

are all in RREF.

In general, an RREF matrix looks as follows:

$$\begin{pmatrix} 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \end{pmatrix}$$

**Example 3.2.2.** Consider these three RREF's for a non-zero $2 \times 2$ matrix.

$$(a): \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix}, \qquad (b): \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (c): \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Here $\alpha$ is some scalar. Viewing these as the coefficient matrix for a linear system we have the following augmented matrices:

$$(a): \left( \begin{array}{cc|c} 1 & \alpha & b_1 \\ 0 & 0 & b_2 \end{array} \right), \qquad (b): \left( \begin{array}{cc|c} 1 & 0 & b_1 \\ 0 & 1 & b_2 \end{array} \right), \qquad (c): \left( \begin{array}{cc|c} 0 & 1 & b_1 \\ 0 & 0 & b_2 \end{array} \right).$$

For $(a)$ and $(c)$ the correspond system has clearly only a solution if $b_2 = 0$ (because the second equation now reads $0x_1 + 0x_2 = 1$). In these cases we obtain for the solution sets

$$(a): x_1 = b_1 - \alpha x_2, \quad (b): x_1 = b_1, \, x_2 = b_2, \quad (c): x_2 = b_1$$

note for $(b)$ we have only have a unique solution, while for $(a)$ and $(c)$ the solution set is a line. More precisely, it is a non-horizontal line for $(a)$, and a horizontal one for $(c)$.

## 3.3 The Elementary Row Operations and the big theorem

The ERO's are simple operations you can do to a matrix to turn it into another matrix. There are three fundamental operations, which are as follows,

(a) $P_{rs}$: switch row $r$ with row $s$;

(b) $M_r(\boldsymbol{\lambda})$: multiple (all the entries in) row $r$ by the number $\lambda \neq 0$;

(c) $A_{rs}$: add (each entry in) row $r$ to (the corresponding entry in) row $s$.

(c') $A_{rs}(\boldsymbol{\lambda})$: add $\lambda$ times (each entry in) row $r$ to (the corresponding entry in) row $s$.

Note that $A_{rs}(\lambda) = M_r\left(\frac{1}{\lambda}\right) \circ A_{rs} \circ M_r(\lambda)$, read from right to left (in the same way as how $f \circ g$ means do $g$ then $f$).

**Example 3.3.1.**

$$
\begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & -1 \\ -1 & 1 & 2 \end{pmatrix}
\xrightarrow[A_{13}(1)]{A_{12}(-2)}
\begin{pmatrix} 1 & 3 & 2 \\ 0 & -5 & -5 \\ 0 & 4 & 4 \end{pmatrix}
\xrightarrow[M_3(1/4)]{M_2(-1/5)}
\begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}
$$

$$
\xrightarrow{A_{23}(-1)}
\begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}
\xrightarrow{A_{21}(-3)}
\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}
$$

### 3.3.1 The basic routine

(a) Find the first non-zero column (from the left).

(b) Go down the first non-zero column and find the first non-zero entry.

(c) If this is in the $r$-th row with $r \neq 1$ apply the elementary row operation $P_{1r}$ to get a non-zero entry, $\lambda$ say, at the top of the first non-zero column.

(d) Apply the elementary row operation $M_1(1/\lambda)$ to make the first non-zero entry of the first non-zero column a 1.

(e) Using elementary row operation of the form $A_{1i}(\mu)$ clear all other non-zero entries in the first non-zero column.

The resulting matrix is of the form

$$
\begin{pmatrix}
1 & * & \cdots & * \\
0 & * & \cdots & * \\
\vdots & \vdots & & \vdots \\
0 & * & \cdots & *
\end{pmatrix}
$$

or, if the first non-zero column is not the first one,

$$
\begin{pmatrix}
0 & \cdots & 0 & 1 & * & \cdots & * \\
0 & \cdots & 0 & 0 & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & * & \cdots & *
\end{pmatrix}
$$

### 3.3.2  The algorithm

Let $A$ be given $m \times n$ matrix. The process of Gauss-Jordan elimination applied to $A$ is described by repeatedly applying the Basic Routine in stages. (For simplicity of description we will assume that the first column of $A$ is non-zero).

**Stage 1**

Apply the basic routine to the given $m \times n$ matrix $A$. This gives a new $m \times n$ matrix $A_1$ of the form:

$$A_1 = \begin{pmatrix} 1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix}$$

**Stage 2**

If all of the remaining rows (below row 1) are zero, it is easy to see that the matrix is now in RREF. Otherwise:

(i). Apply the basic routine to the submatrix of $A_1$ obtained by ignoring the first row. This gives a new $m \times n$ matrix $\tilde{A}_2$ of the form:

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix}$$

(ii). Apply an elementary row operation of the form $A_{21}(\mu)$, if necessary, to $\tilde{A}_2$ to make the entry above the first 1 in the second row a zero. This gives a new $m \times n$ matrix $A_2$ of the form:

$$A_2 = \begin{pmatrix} 1 & * & \cdots & * & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix}.$$

Now continue inductively. Assume stages 1 to $k-1$ have been completed (for $k-1 < m$), resulting in a matrix $A_{k-1}$.

**Stage k (with $k \leq m$)**

If rows $k, k+1, \ldots, m$ of $A_{k-1}$ are all zero then this matrix is in RREF and the algorithm terminates. Similarly, if the leading 1 in the $(k-1)$th row of $A_{k-1}$ lies in the $n$th column then it is easy to check that $A_{k-1}$ is in RREF. Otherwise:

(i). Apply the basic routine to the submatrix of $A_{k-1}$ obtained by ignoring the

first $k - 1$ rows. This gives a new $m \times n$ matrix $\tilde{A}_k$ of the form:

$$\tilde{A}_k = \begin{pmatrix} 1 & 0 & \cdots & * & 0 & * & \cdots & * & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & * & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * & \cdots & * \end{pmatrix}.$$

(ii). Apply elementary row operations of the form $A_{kr}(\mu)$ (for $r = 1, 2, \ldots, k-1$), if necessary, to $\tilde{A}_k$; this ensures all remaining non-zero entries in the column containing the first 1 in the $k$-th row equal to zero and gives a new $m \times n$ matrix $A_k$ of the form:

$$\tilde{A}_k = \begin{pmatrix} 1 & 0 & \cdots & * & 0 & * & \cdots & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 1 & * & \cdots & 0 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & * & \cdots & * \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & * & \cdots & * \end{pmatrix}.$$

By induction, this process must terminate either before Stage $m$ or after Stage $m$. In the former case, the reason for early termination must have been that the matrix had reached RREF. In the latter case we must have reached a matrix $A_m$ with a leading 1 on every row. It is then easy to check that $A_m$ is in RREF.

## 3.4 Using G-J to solve a system of linear equations

Recall, we are considering the solution set of a general system of $m$ linear equations in $n$ variables.

Two such systems are said to be equivalent if they have the same solution set. We say an augmented matrix is in row reduced echelon form if the associated coefficient matrix is in row reduced echelon form. Applying the elementary row operations to the augmented matrix $(A|\mathbf{b})$ allows us to pass to equivalent systems and eventually to a system for which it is easy to spot the solutions. This is justified by

**Theorem 3.4.1.** *Applying ERO's to the augmented matrix of a linear system simply changes the system to an equivalent one; that is, the solution set remains unchanged.*

If a row is of the form $(0 \cdots \mid k)$ for $k \neq 0$ this would mean that there are no solutions, we call a system like this inconsistent.

A quick way to solve a linear system, there is no need to run G-J fully. One can do the following:

1) : Perform ERO's to bring coefficient matrix $A$ into upper triangular form, that is, all the entries below the diagonal vanish. (Also not necessary to scale the "leading ones" to be actually equal to 1.)

2) : Then go back to the linear system and use back substitution from bottom to top to solve the system.

**Example 3.4.2.**
$$x_1 + 2x_2 + 3x_3 = 1$$
$$2x_1 + 3x_2 + 2x_3 = 1$$
$$3x_1 + 2x_2 + \ x_3 = 1.$$

We put this into matrix form then EROs to get it into upper triangular form

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 2 & 3 & 2 & | & 1 \\ 3 & 2 & 1 & | & 1 \end{pmatrix} \xrightarrow[A_{13}(-3)]{A_{12}(-2)} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -1 & -4 & | & -1 \\ 0 & -4 & -8 & | & -2 \end{pmatrix}$$

$$\xrightarrow{A_{23}(-4)} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -1 & -4 & | & -1 \\ 0 & 0 & 8 & | & 2 \end{pmatrix}.$$

Putting this back into our equations gives us:

3rd equation: $x_3 = \frac{1}{4}$;

2nd equation: $x_2 = -4x_3 + 1 = -4\frac{1}{4} + 1 = 0$;

1st equation: $x_1 = 1 - 2x_2 - 3x_3 = 1 - \frac{3}{4} = \frac{1}{4}$, i.e.,

$$\begin{pmatrix} \frac{1}{4} \\ 0 \\ \frac{1}{4} \end{pmatrix}.$$

## 3.5 Inverse matrices and linear systems

**Lemma 3.5.1.** *Consider an arbitrary $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$. Let $B \in M_{m \times m}(\mathbb{R})$ be an invertible matrix. Then the new system $(BA)\mathbf{x} = B\mathbf{b}$ has the same solution set as the original system.*

*Proof.* We need to show that the two systems have the same solution set. First assume that $\mathbf{v}$ is a solution to the system $A\mathbf{x} = \mathbf{b}$, assume $A\mathbf{v} = \mathbf{b}$. Then, $B\mathbf{b} = B(A\mathbf{v}) = (BA)\mathbf{v}$ as required. Conversely, assume $(BA)\mathbf{v} = B\mathbf{b}$. Then

$$\mathbf{b} = I_m\mathbf{b} = (B^{-1}B)\mathbf{b} = B^{-1}(B\mathbf{b}) = B^{-1}(BA\mathbf{v}) = (B^{-1}B)A\mathbf{v} = A\mathbf{v},$$

as required. □

**Theorem 3.4.1** *(continuing from p. 21).* *Proof.* Suppose we want to perform elementary row operations on a matrix with $m$ rows. Then we can do this using the following $m \times m$ matrices.

(a) The row operation $P_{rs}$ can be performed by left multiplication by the $m \times m$ matrix, which we also call $P_{rs}$, obtained from the identity by swapping the $r$th and $s$th columns. That is $(p_{ij})_{m \times m}$ with $p_{ii} = 1$ when $i \neq r, s, p_{rr} = p_{ss} = 0$; $p_{ij} = 0$ when $i \neq j$ and $\{i, j\} \neq \{r, s\}, p_{rs} = p_{sr} = 1$. That is,

$$P_{rs} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

One can check that we have $P_{rs}^{-1} = P_{rs}$ (swapping twice leaves you back where you started).

(b) The row operation $M_r(\lambda)$ can be performed by left multiplication by the $m \times m$ matrix, which we also call $M_r(\lambda)$, obtained from the identity by replacing the $r$th diagonal entry with $\lambda$. That is $(m_{ij})_{m \times m}$ with $m_{ii} = 1$ when $i \neq r$ and $m_{rr} = \lambda$; $m_{ij} = 0$ when $i \neq j$. That is,

$$M_r(\lambda) = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

We have $M_r(\lambda)^{-1} = M_r\left(\frac{1}{\lambda}\right)$.

(c) The row operation $A_{rs}(\lambda)$ has be performed by left multiplication by the $m \times m$ matrix, which we also call $A_{rs}(\lambda)$, obtained from the identity matrix by replacing the entry in the $r$th column and $s$th row with $\lambda$. That is $(a_{ij})_{m \times m}$ where $a_{ii} = 1$; $a_{ij} = 0$ when $i \neq j$ and $(i, j) \neq (s, r)$; $a_{sr} = \lambda$. That is,

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

We have $A_{rs}(\lambda)^{-1} = A_{rs}(-\lambda)$.

We call any of these matrices an elementary matrix. Then if $E$ is an elementary $m \times m$ matrix the corresponding row operation is given by $EA\mathbf{x} = E\mathbf{b}$. $\quad\square$

**Remark 3.5.2.** *Note that every elementary matrix is not only invertible, but the inverse of any elementary matrix is also an elementary matrix.*

## 3.6 The fundamental theorem of invertible matrices

**Lemma 3.6.1.** *Let $A \in M_{n \times n}(\mathbb{R})$ be a square matrix. Then the following are (pairwise) equivalent, that is they are either all true or all false.*

(i) *In each column of RREF of A there exists a leading 1.*

(ii) *The RREF of A is the identity matrix* $I_n = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$.

(iii) *The only solution to the system $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.*

*Proof.* The equivalence of (i) and (ii) is clear: if the RREF is the identity then each column has a leading 1. Conversely, if each column of the RREF has a leading 1, then the first leading 1 is in the $1,1$-entry, the second must be in the $2,2$-entry and so on. Then all the off-diagonal entries must be zero, that is, the RREF is indeed $I_n$.

Next we show that (ii) implies (iii). So let's assume (ii). But by Lemma 3.5.1 we can replace the equation $A\mathbf{x} = \mathbf{0}$ with the RREF of $A$, that is, we can consider $I_n\mathbf{x} = \mathbf{0}$, which means $\mathbf{x} = \mathbf{0}$. So (iii) holds.

What remains is to show that (iii) implies (i). So let's assume (iii), that is, the only solution to the system $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. Let's assume for a contradiction that the number of leading 1's in the RREF of $A$ is equal to $r < n$. This means that there are $n - r > 0$ free variables (one for each column without a leading 1). To be explicit, after permutation of columns we can assume the RREF of $A$ looks as follows

$$\begin{pmatrix} I_r & A'_{r,n-r} \\ \mathbf{0}_{n-r,r} & \mathbf{0}_{n-r,n-r} \end{pmatrix}$$

for some $r \times (n - r)$ matrix $A'$. But then any choice of real value for each of the free variables will yield a solution t the system $A\mathbf{x} = \mathbf{0}$. In particular there exists a non-trivial solution to the system. This is a contradiction, therefore there must have been $r = n$ leading 1's after all. $\qquad\square$

**Theorem 3.6.2** (*Fundamental Theorem of Invertible Matrices*)**.** *A square matrix $A \in M_{n \times n}(\mathbb{R})$ is invertible if and only if any one of the conditions (i), (ii), (iii) of Lemma 3.6.1 holds.*

*Proof.* Suppose $A$ is invertible. Then by Lemma 3.1.1(ii) the only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$, which is Lemma 3.6.1(iii).

Now assume that any one of (i), (ii), and (iii) hold. By (ii), there exist elementary matrices $E_1, E_2, \ldots, E_\ell$ such that

$$E_\ell E_{\ell-1} \cdots E_1 A = I_n.$$

Thus $(E_\ell E_{\ell-1}\cdots E_1)$ is a left sided inverse of $A$. We must show it is a right sided inverse too. To do this, note that

$$\begin{aligned}
AE_\ell E_{\ell-1}\cdots E_1 &= I_n A E_\ell E_{\ell-1}\cdots E_1 \\
&= (E_1^{-1} E_2^{-1}\cdots E_\ell^{-1} E_\ell E_{\ell-1}\cdots E_1)AE_\ell E_{\ell-1}\cdots E_1 \\
&= E_1^{-1} E_2^{-1}\cdots E_\ell^{-1}(E_\ell E_{\ell-1}\cdots E_1 A)E_\ell E_{\ell-1}\cdots E_1 \\
&= E_1^{-1} E_2^{-1}\cdots E_\ell^{-1} E_\ell E_{\ell-1}\cdots E_1 = I_n.
\end{aligned}$$

So $A$ is invertible with $A^{-1} = E_\ell E_{\ell-1}\cdots E_1$. $\qquad\square$

This proof immediately gives us

**Corollary 3.6.3.** *Any invertible square matrix is the product of elementary matrices (and vice versa).*

*Proof.* Simply notice that $E_1^{-1} E_2^{-1}\cdots E_i^{-1}$ is an inverse of $A^{-1} = E_1 E_2\cdots E_i$. But, so is the matrix $A$. So, by uniqueness of the inverses, we must have $A = E_1^{-1} E_2^{-1}\cdots E_\ell^{-1}$. $\qquad\square$

**Remark 3.6.4.** *Note that the product above is not necessarily unique.*

**Remark 3.6.5.** *We can also now show that the "left inverse" of a square matrix $A$ is also the "right inverse" of $A$, that is if $BA = I_n$, then also $AB = I_n$*

**Proposition 2.4.2** (continuing from p. 14)**.** *Proof of (v).* Suppose that $BA = I_n$ and consider $A\mathbf{x} = \mathbf{0}$. Left multiplication by $B$ shows that the only solution is $\mathbf{x} = \mathbf{0}$. Hence $A$ is invertible by Theorem 3.6.2. In particular, $A$ has an inverse $A^{-1}$ and we have

$$AB = ABAA^{-1} = A(BA)A^{-1} = AA^{-1} = I_n.$$

By the uniqueness of inverses, this shows that $A^{-1} = B$. $\qquad\square$

## 3.7 An algorithm to compute the inverse of a matrix

We run G-J for the matrix $A$ and keep track of the elementary operations/matrices we used. Concretely, we run G-J on the augmented or double matrix

$$(A\mid I_n) = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\
a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 1
\end{pmatrix}.$$

Now perform G-J as always and get (if $A$ is invertible)

$$(I_n\mid E_\ell\cdots E_1),$$

and $A^{-1} = E_\ell\cdots E_1$. If $A$ is singular, you will not obtain the identity matrix on the left hand side.

# 4   Determinants

The determinant of a square matrix is a scalar. The determinant of a matrix $A$ with real entries is a real number. So we can think of this as a function

$$\det : M_{n \times n}(\mathbb{R}) \to \mathbb{R}.$$

## 4.1   Small values of $n$

We look at particular cases of the function

$$\det : M_{n \times n}(\mathbb{R}) \to \mathbb{R}.$$

$n = 1$: The equation $ax = b$ has a unique solution if and only if $a \neq 0$. We set

$$\det(a) = a.$$

$n = 2$: We know $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. So, set

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

Then the matrix equation (system of two linear equations)

has a unique solution if and only if $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$

## 4.2   General definition of determinant

We define the determinant of an $n \times n$ matrix inductively in terms of the determinants of certain $(n-1) \times (n-1)$ matrices. Let $A = (a_{ij})_{n \times n}$

- For integers $r$ and $s$ between 1 and $n$ write $A_{r,s}$ for the $(n-1) \times (n-1)$ matrix obtained from $A$ be deleting its $r$th row and its $s$th column. Then $A_{r,s}$ is called the $(r,s)$th minor [2] of $A$.

- The value of $\det(A_{r,s})$ i.e. the determinant of the $(r,s)$th minor of $A$, is called the $(r,s)$th unsigned cofactor of $A$ and the value $(-1)^{r+s} \det(A_{r,s})$ is called the $(r,s)$th (signed) cofactor of $A$ [3] .

For any matrix $A \in M_{n \times n}(\mathbb{R})$ the determinant of $A$ is the defined inductively via expansion by the first row, in terms of the signed cofactors:

$$\det A = \sum_{k=1}^{n} (-1)^{1+k} a_{1k} \det(A_1, k).$$

An important observation is that we could have defined it via expansion by any given row.

---

[2] The minor is called the 'unsigned cofactor'. We have defined the minor as simply a submatrix.

[3] Some texts write 'adjunct' instead of 'cofactor'.

**Lemma 4.2.1.** *If $A$ is an $n \times n$ matrix then for any $1 \leq i \leq n$ we have*

$$\det(A) = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det(A_i, k).$$

*That is, the determinant formula remains valid via expansion along the $i$th row (note however the signs).*

*Proof.* The proof is a very long (and involved) inductive one so... $\qquad\square$

## 4.3  Fundamental properties

In this section, for notational convenience, we write in terms of the row vector $\mathbf{a}_r$ of an $n \times n$ matrix $A = (a_{ij})$ that is $\mathbf{a}_r = (a_{r1} \; a_{r2} \; \cdots \; a_{rn})$

(i) $\det I_n = 1$.

(ii) If one multiplies (all the elements in) the $r$th row by some scalar $\lambda$ (that is, if we apply the ERO $M_r(\lambda)$), then the determinant changes by precisely this factor:

$$\det(M_r(\lambda)A) = \det \begin{pmatrix} \mathbf{a} \\ \vdots \\ \lambda\mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \lambda \det A.$$

(iii) For any row vector $\mathbf{b}_r$ we have

$$\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r + \mathbf{b}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} + \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{b}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

Note that (ii) and (iii) can be summarised as linearity in each row.

(iv) Anti-symmetry in rows. The determinant switches signs if one swaps a pair of rows (i.e. if one applies the ERO $P_{rs}$); that is,

$$\det(A) = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = -\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_s \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = -\det(P_{rs}A).$$

(v) $\det A = 0$ if $A$ has two rows that are equal (or even two rows that are collinear - follows simply by property (ii)).

(vi) $\det A = 0$ if $A$ has a row of zeros.

(vii) The determinant does not change under the ERO $A_{rs}(\lambda)$; that is,

$$
\det(A_{rs}(\lambda)A) = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s + \lambda \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det A.
$$

*Proof.* (i) We perform induction on $n$. The case $n = 1$ is obvious: $\det(1) = 1$. So now assume the claim holds for some $n \geq 1$, that $\det I_n = 1$. We need to establish the claim for $n + 1$. But now by the definition of the determinant (say, by expanding by the first row) we immediately see $\det(I_{n+1}) = 1 \cdot \det(I_n) = 1 \cdot 1 = 1$, as desired.

(ii) This is straightforward, by expansion along the $r$th row (by Lemma 4.2.1):

$$
\det(M_r(\lambda)A) = \sum_{k=1}^{n}(-1)^{r+k}(\lambda a_{rk})\det(A_r, k) = \lambda \sum_{k=1}^{n}(-1)^{r+k}a_{rk}\det(A_{r,k}) = \lambda \det A.
$$

(iii) is similar to (ii), by expansion along the $r$th row.

(iv)

(v) This follows directly from (iv): Assume row $r$ is repeated in row $s$ then

$$
\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r + \mathbf{b}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = -\det \left( P_{rs} \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r + \mathbf{b}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \right) = -\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r + \mathbf{b}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix},
$$

and so the value of the determinant must be zero.

(vi) This follows directly from (iii) by writing $\mathbf{0} = \mathbf{a}_r - \mathbf{a}_r$ (or from (ii) by writing $\mathbf{0} = 0 \cdot \mathbf{a}_r$ for some vector $\mathbf{a}_r$). Indeed, if the row of zeros appears in row $r$ then we have

$$
\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{0} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{a}_r - \mathbf{a}_r \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{a}_r \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} - \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{a}_r \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = 0.
$$

(vii) This follows from (ii), (iii), and (iv):

$$\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s + \lambda \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s \\ \vdots \\ \mathbf{a}_n \end{pmatrix} + \lambda \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_s \\ \vdots \\ \mathbf{a}_n \end{pmatrix}.$$

$\square$

## 4.4  Uniqueness of the determinant

We will now prove that properties (i)-(vii) completely characterise the determinant. We could define the function

$$\det : M_{n \times n}(\mathbb{R}) \to \mathbb{R}$$

as the unique function that satisfies these four properties.

**Theorem 4.4.1.** *Let* $f : M_n(\mathbb{R}) \to \mathbb{R}$ *be a function that satisfies properties (i)-(iv). Then*

$$f(A) = \det(A).$$

*In other words, the determinant is the only function with the properties (i)-(vii).*

*Proof.* Apply Gauss-Jordan to the matrix $A$ to get the row reduced echelon form $A'$. As det satisfies properties (ii), (iv) and (vii) and we are assuming $f$ does too, during this process det and $f$ change value in exactly the same way. So

$$\det A' = \lambda_1 \cdots \lambda_k \det A$$
$$f(A') = \lambda_1 \cdots \lambda_k f(A)$$

for some non-zero scalars $\lambda_i$. But $A'$ is either equal to $I_n$ in which case $f(A') = \det A' = 1$ (by property (i)), or it contains a zero row in which case $f(A') = \det A' = 0$ by property (vi). So in either case $f(A') = \det A'$ and we obtain

$$\lambda_1 \cdots \lambda_k f(A) = \lambda_1 \cdots \lambda_k \det A,$$

which gives $f(A) = \det A$ since $\lambda_i \neq 0$ for all $i$.  $\square$

## 4.5  Determinant calculations

Properties (i)-(vii) tell us how the determinant changes under the G-J algorithm, and hence gives us a way to compute the determinant. There are two basic cases.

(I) The RREF $A'$ of $A$ contains a zero row. Then $\det(A) = \det(A') = 0$.

(II) The RREF $A'$ of $A$ is the identity matrix. Assume that we applied the operation $P_{rs}$ of switching rows $\ell$ times and applied the scaling operations $M_{r_1}(\lambda_1), M_{r_2}(\lambda_2), \ldots, M_{r_k}(\lambda_k)$ at some point during the algorithm. Then

$$1 = \det I_n = \det A' = (-1)^\ell \lambda_1 \cdots \lambda_k \det A.$$

So

$$\det(A) = (-1)^\ell \lambda_1^{-1} \cdots \lambda_k^{-1}.$$

**Proposition 4.5.1.** *Let $A$ be an upper (or lower) triangular $n \times n$ matrix with diagonal entries $a_{11}, a_{22}, \ldots, a_{nn}$. Then*

$$\det(A) = \prod_{k=1}^{n} a_{kk} = a_{11} \cdot a_{22} \cdots a_{nn}.$$

*Proof.* $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

## 4.6   Several Big Theorems

**Theorem 4.6.1.** *The $n \times n$ matrix $A$ has an inverse if and only if $\det(A) \neq 0$.*

*Proof.* By definition we already know this for the cases $n = 1$, $n = 2$ and $n = 3$. For $n > 3$ we apply G-J to the matrix $A$ to get to the RREF $A'$. As in the proof of Theorem 4.4.1, during this process det changes in a clear way using properties (ii), (iv), and (vii). We get

$$\det A' = \lambda_1 \cdots \lambda_k \det A$$

for some non-zero scalars $\lambda_i /$ But then

$$\det(A) \neq 0 \iff \det(A') \neq 0 \iff A' = I_n \iff A \text{ is non-singular.}$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Theorem 4.6.2.** *Let $A$ and $B$ be $n \times n$ matrices. Then*

$$\det(AB) = \det(A)\det(B).$$

*Proof.* We apply G-J to the matrix $A$ to get to the RREF $A'$. As before there are some non-zero scalars $\lambda_i$ such that

$$\det A' = \lambda_1 \cdots \lambda_k \det A \text{ so that } \det A = \lambda_1^{-1} \cdots \lambda_k^{-1} \det A'.$$

Applying the same ERO's to the matrix $AB$ we obtain the matrix $A'B$ (ERO's are given by left multiplication of elementary matrices) and hence also

$$\det(A'B) = \lambda_1 \cdots \lambda_k \det(AB).$$

We now have two cases:

(i) $A' = I_n$. Then

$$\det(AB) = \lambda_1^{-1} \cdots \lambda_k^{-1} \det(B) = \det(A)\det(B),$$

as desired.

(ii) $A'$ contains a zero row; so $\det A = \det A' = 0$. But then the product $A'B$ contains also a zero row which gives $\det(A'B) = 0$. Since $\det(A'B)$ and $\det(AB)$ only differ by non-zero scalars, we get $\det(AB) = 0$. But we also have $\det(A)\det(B) = 0$ since $\det A = 0$.

$\square$

**Theorem 4.6.3.** *We have*

$$\det(A^t) = \det(A).$$

*Proof.* Firstly, Proposition 2.4.2(iv) implies that either both $A$ and $A^t$ are invertible, or neither are. In the case $\det A = 0$ we have by Theorem 4.6.1 that $A$ is singular, and so $A^t$ is also singular. Applying Theorem 4.6.1 a second time shows that $\det(A^t) = 0$. If $\det A \neq 0$ then $A$ is invertible. But then Corollary 3.6.3 tells us we may write it as a product of EROs, say $A = E_1 E_2 \cdots E_\ell$. By combining determinant properties (ii), (iv) and (vii) it is easy to check that $\det(P_{rs}) = -1$, $\det(M_r(\lambda)) = \lambda$ and $\det(A_{rs}(\lambda)) = 1$ for every $r$ and $s$. Combining this with the observation that

$$(P_{rs})^t = P_{rs}, \quad (M_r(\lambda))^t = M_r(\lambda) \quad (A_{rs}(\lambda))^t = A_{sr}(\lambda),$$

it follows that $\det(E^t) = \det(E)$ for every elementary matrix $E$. Thus, by Proposition 2.4.2(iii) and Theorem 4.6.2 we have

$$\det(A^t) = \det((E_1 \cdots E_\ell)^t) \qquad = \det(E_\ell^t \cdots E_1^t) = \det(E_\ell^t) \cdots \det(E_1^t)$$
$$= \det(E_\ell) \cdots \det(E_1) \quad = \det(E_1) \cdots \det(E_\ell) = \det(E_1 \cdots E_\ell) \qquad = \det A.$$

$\square$

**Corollary 4.6.4.** *If $A$ is not singular, then*

$$\det(A^{-1}) = \det((A^t)^{-1}) = \frac{1}{\det(A)}.$$

*Proof.* Simply notice that $1 = \det(I_n) = \det(A^{-1}A) = \det(A^{-1})\det(A)$, and similarly for the transpose (using Proposition 2.4.2(iv)). $\square$

## 4.7 Determinants and inverses; Cramer's Rule

**Definition 4.7.1.** Given a square matrix $A$, we define the adjoint matrix $\text{Adj}(A)$ by

$$(\text{Adj}(A))_{ij} = (-1)^{i+j} \det(A_{j,i}).$$

---
| The adjoint is the transpose of the matrix of signed cofactors. |
---

**Proposition 4.7.2.** *Let $A \in M_n(\mathbb{R})$ and $\text{Adj}(A)$ be its adjoint. Then*

$$A \cdot \text{Adj}(A) = \det(A) \cdot I_n = \text{Adj}(A) \cdot A.$$

*Proof.* We first prove $A \cdot \text{Adj}(A) = \det(A) \cdot I_n$. The $(r, s)$-entry of $A \cdot \text{Adj}(A)$ is given by

$$\sum_{k=1}^n a_{rk}(\text{Adj}(A))_{ks} = \sum_{k=1}^n a_{rk}(-1)^{k+s} \det(A_{s,k}).$$

For $r = s$ this is the expansion of $\det A$ by the $s$-th row. While for $r \neq s$, this is the expansion of the determinant of $A$ by the $s$-th row with row $s$ replaced by the $r$th row $\mathbf{a}_r$. So we are looking at

$$\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_r \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = 0.$$

Hence in summary

$$(r, s) - \text{entry of } A \operatorname{Adj}(A) = \begin{cases} \det A & \text{if } r = s \\ 0 & \text{if } r \neq s \end{cases};$$

i.e., $A \cdot \operatorname{Adj}(A) = (\det A) I_n$. This is enough to conclude that $A^{-1} = \frac{1}{\det(A)} \operatorname{Adj}(A)$ and the second equation, that $\operatorname{Adj}(A) \cdot A = (\det A) I_n$ follows from the uniqueness of inverses. $\qquad \square$

**Theorem 4.7.3** *(Inverse by adjoint matrix).* *Let $A \in M_n(\mathbb{R})$ be non-singular. Then the inverse $A^{-1}$ is given by*

$$A^{-1} = \frac{1}{\det A} \operatorname{Adj}(A).$$

**Theorem 4.7.4** *(Cramer's rule).* *Let $A\mathbf{x} = \mathbf{b}$ be a linear system with a non-singular matrix $A \in M_n(\mathbb{R})$. The unique solution $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is given by*

$$x_i = \frac{\det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n)}{\det A}.$$

*Here $\mathbf{a}_\ell$ denotes (for a change) the $\ell$-th column vector of $A$.*

**Remark 4.7.5.** *For practical computations Theorem 4.7.3 and Theorem 4.7.4 are rather useless. But for theoretical considerations this is very useful. For example, since the determinant is a polynomial in the entries of the matrix $A$, the entries $A^{-1}$ (and also the solution to $A\mathbf{x} = \mathbf{b}$) depend continuously on the entries of $A$.*

## 4.8   Determinants, area and volume

### Area of parallelograms

Consider the pair of vectors

$$\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}, \ \mathbf{w} = \begin{pmatrix} c \\ d \end{pmatrix}.$$

Now form the parallelogram $P$ with vertices $0, \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w}$. Elementary trigonometry tells us

$$\text{area}(P) = |\mathbf{v}| \cdot |\mathbf{w}| \cdot \sin \theta, \tag{1}$$

where $\theta \in [0, \pi]$ is the angle between $\mathbf{v}$ and $\mathbf{w}$. Recall for two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, we have

$$|\mathbf{x} \times \mathbf{y}| = |\mathbf{x}||\mathbf{y}| \sin \theta. \tag{2}$$

We use this to rewrite (1) in the following way. We define two auxiliary vectors $\tilde{\mathbf{v}} = \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$ and $\tilde{\mathbf{w}} = \begin{pmatrix} c \\ d \\ 0 \end{pmatrix}$ in $\mathbb{R}^3$. We then have

$$\tilde{\mathbf{v}} \times \tilde{\mathbf{w}} = \begin{pmatrix} 0 \\ 0 \\ ad - bc \end{pmatrix}$$

So (2) becomes

$$|ad - bc| = |\tilde{\mathbf{v}}||\tilde{\mathbf{w}}| \sin \theta = |\mathbf{v}||\mathbf{w}| \sin \theta.$$

Combining this with (1) we have shown

**Theorem 4.8.1.** *For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ two vectors in the plane, form the parallelogram $P$ with vertices $\mathbf{0}, \mathbf{v}, \mathbf{w}$ and $\mathbf{v} + \mathbf{w}$. Then*

$$\text{area}(P) = |\det(\mathbf{v}, \mathbf{w})|.$$

### Area of triangles

The area of a triangle with vertices $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ is half the area of the parallelogram with vertices $\mathbf{a}, \mathbf{a} + (\mathbf{b} - \mathbf{a}), \mathbf{a} + (\mathbf{c} - \mathbf{a}) = \mathbf{c}$ and $\mathbf{a} + (\mathbf{b} - \mathbf{a}) + (\mathbf{c} - \mathbf{a}) = \mathbf{b} + \mathbf{c} - \mathbf{a}$. In other words, the area of this triangle is given by:

**Proposition 4.8.2.** *The triangle with vertices $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ has area*

$$\frac{1}{2} |\det(\mathbf{b} - \mathbf{a}, \mathbf{c} - \mathbf{a})|.$$

### Volume of parallelepipeds

A parallelepiped $P$ is a three dimensional polyhedron with six quadrilateral faces where opposite faces are parallel. A parallelepiped has twelve edges. These come in three families of four where all the vectors in each family are parallel; call a vector from the first family $\mathbf{u}$, one from the second $\mathbf{v}$ and one from the third $\mathbf{w}$ (we then say that $P$ is spanned by $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$). The parallelepiped $P$ in $\mathbb{R}^3$ spanned by the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ hence has vertices $\mathbf{0}, \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{w}, \mathbf{v} + \mathbf{w}, \mathbf{u} + \mathbf{v} + \mathbf{w}$. We have

$$\text{vol}(P) = \text{area}(\text{base}) \times \text{height}.$$

As base we take the parallelogram spanned by $\mathbf{v}$ and $\mathbf{w}$. Hence, as before

$$\text{area}(\text{base}) = |\mathbf{v} \times \mathbf{w}|.$$

For the height $h$ we have

$$h = |\mathbf{u}| \cdot |\sin(\text{angle}(\mathbf{u}, \text{base}))| = |\mathbf{u}| \cdot |\cos(\text{angle}(\mathbf{u}, \text{normal vector to base}))|.$$

A normal vector to the base is of course $\mathbf{v} \times \mathbf{w}$. Hence by definition of the angle between two vectors we see

$$|\cos(\text{angle}(\mathbf{u}, \mathbf{v} \times \mathbf{w}))| = \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})}{|\mathbf{u}| \cdot |\mathbf{v} \times \mathbf{w}|} = \frac{|\det(\mathbf{u}, \mathbf{v}, \mathbf{w})|}{|\mathbf{u}| \cdot |\mathbf{v} \times \mathbf{w}|}.$$

Combining all of this we obtain

**Proposition 4.8.3.** *The volume of the parallelepiped $P$ in $\mathbb{R}^3$ spanned by the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ is given by*

$$\text{vol}(P) = |\det(\mathbf{u}, \mathbf{v}, \mathbf{w})|.$$

**Other polyhedra**

Similar reasoning shows that the volume of a tetrahedron $T$ is one sixth of the determinant of the matrix whose columns are the vectors along edges of $T$ based at the same vertex.

**General $n$**

To define this phenomenon in higher dimensions we first need to have an idea what we mean by an $n$-dimension volume. The $n$-dimension box spanned by $a_1 \mathbf{e}_1, a_2 \mathbf{e}_2, \ldots, a_n \mathbf{e}_n$ we define to have volume $|a_1| \cdots |a_n|$. Note that $\left| \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \right| = |a_1| \cdots |a_n|$. With this definition one can show that the volume of the $n$-dimensional parallelepiped $P$ spanned by vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is given by

$$\text{vol}(P) = |\det(\mathbf{v}_1, \ldots, \mathbf{v}_n)|.$$

# 5 Spanning sets and linear independence in $\mathbb{R}^n$

## 5.1 Subspaces of $\mathbb{R}^n$

**Definition 5.1.1** *(Vector subspace of $\mathbb{R}^n$).* A (non-empty) subset $U$ of $\mathbb{R}^n$ is called a vector subspace (or linear subspace) of $\mathbb{R}^n$ if it satisfies the following three conditions:

(i) closure under addition: if $\mathbf{u}_1$ and $\mathbf{u}_2$ are in $U$, then $\mathbf{u}_1 + \mathbf{u}_2$ is also in $U$;

(ii) closure under scalar multiplication: if $\mathbf{u}$ is in $U$ and $\lambda$ is in $\mathbb{R}$, then $\lambda\mathbf{u}$ is in $U$;

(iii) existence of an origin: the vector $\mathbf{0}$ is in $U$.

**Simply**, a vector subspace is just a non-empty subset of $\mathbb{R}^n$ that satisfies some conditions.

**Example 5.1.2.** (i) $U = \{\mathbf{0}\}$ is clearly a vector subspace; it is called the trivial vector subspace.

(ii) Equally clearly $\mathbb{R}^n$ is a vector subspace of itself.

(iii) Let $L_0$ be a line in $\mathbb{R}^n$ through the origin $\mathbf{0}$ with direction vector $\mathbf{d}$. Then any point on $L_0$ has the form $\mathbf{x} = \mathbf{0} + \lambda\mathbf{d} = \lambda\mathbf{d}$. Clearly $L$ satisfies the three conditions above and so is a vector subspace of $\mathbb{R}^n$. If the line does not pass through $\mathbf{0}$ then it is not a vector subspace (it is called an affine line, which is a special case of an affine subspace).

(iv) Let $\Pi_0$ be the plane in $\mathbb{R}^n$ through the origin $\mathbf{0}$ with direction vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ (not collinear). Then we have

$$\Pi_0 = \{\lambda_1\mathbf{d}_1 + \lambda_2\mathbf{d}_2 \,:\, \lambda_1, \lambda_2 \in \mathbb{R}\},$$

it is clear that $\Pi_0$ satisfies the three conditions above and so is a vector subspace. If the plane does not pass through $\mathbf{0}$, then it is not a vector subspace. (It is called an affine plane, which is a special case of an affine subspace).

**Proposition 5.1.3.** *The solution set of an arbitrary $m \times n$ homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is a vector subspace of $\mathbb{R}^n$.*

*Proof.* Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be two solutions to the linear system and $\lambda \in \mathbb{R}$. Then

$$A(\mathbf{u}_1 + \mathbf{u}_2) = A\mathbf{u}_1 + A\mathbf{u}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0} \quad \text{and} \quad A(\lambda\mathbf{u}_1) = \lambda(A\mathbf{u}_1) = \lambda\mathbf{0} = \mathbf{0},$$

as claimed. Of course we also have $A\mathbf{0} = \mathbf{0}$. $\qquad\square$

Note that the solution set of an inhomogeneous system $A\mathbf{x} = \mathbf{b}$ (where $\mathbf{b} \neq \mathbf{0}$) is not a vector subspace since it does not contain the zero vector. Instead it forms an affine subspace of $\mathbb{R}^n$; that is, a subset of $\mathbb{R}^n$ of the form

$$\mathbf{c} + U = \{\mathbf{v} \in \mathbb{R}^n \,:\, \mathbf{v} = \mathbf{c} + \mathbf{u} \text{ with } \mathbf{u} \in U\},$$

where $U$ is a vector subspace.

The above is quite important as it has some relevance to some other definitions we will be making, specifically this includes linear independence and span.

## 5.2   Spanning sets

Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be vectors in $\mathbb{R}^n$. A linear combination of these vectors is a vector $\mathbf{u}$ of the form

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k$$

for some (real) scalars $\lambda_1, \ldots, \lambda_k$.

**Definition 5.2.1** *(Linear span).* Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ vectors in $\mathbb{R}^n$. Their (linear) span is the subset $U$ of $\mathbb{R}^n$ consisting of all linear combinations of these vectors. So

$$U = \{\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k \quad \lambda_1, \ldots, \lambda_k \in \mathbb{R}\}.$$

Conversely, we say that such a $U$ is spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and write $U = \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$. Equivalently, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is called a spanning set of $U$.

**Alternatively**, another simplification of this is to consider that the span is just a subset of $\mathbb{R}^n$ containing all of the combinations of the vectors. So for a set of vectors to span a subspace $U$, everything in $U$ must be able to be written as a combination of the vectors which span $U$.

**Proposition 5.2.2.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ be $k$ vectors, and let $A = (\mathbf{u}_1, \ldots, \mathbf{u}_k) \in M_{n,k}(\mathbb{R})$ be the associated matrix of column vectors. Then we have*

$$\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k = A\boldsymbol{\lambda},$$

*where $\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix} \in \mathbb{R}^k$ is the column vector of the $\lambda_j$. Moreover, we have*

$$\text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k) = \{\mathbf{u} \in \mathbb{R}^n : \text{The system } A\boldsymbol{\lambda} = \mathbf{u} \text{ has a solution}\} = \{A\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}^k\}.$$

The above sort of tells us that we can rewrite the span of a collection of vectors in terms of matrices.

**Lemma 5.2.3.** *The linear span $U$ of any collection $\mathbf{u}_1, \ldots, \mathbf{u}_k$ of vectors is a vector subspace of $\mathbb{R}^n$.*

*Proof.* We check the conditions of a vector space:

(i) If $\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k$ and $\mathbf{v} = \mu_1 \mathbf{u}_1 + \cdots + \mu_k \mathbf{u}_k$ then

$$\mathbf{u} + \mathbf{v} = (\lambda_1 + \mu_1)\mathbf{u}_1 + \cdots + (\lambda_k + \mu_k)\mathbf{u}_k \in U.$$

(ii) If $\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k$ and $\lambda \in \mathbb{R}$, then

$$\lambda\mathbf{u} = (\lambda\lambda_1)\mathbf{u}_1 + \cdots + (\lambda\lambda_k)\mathbf{u}_k \in U.$$

(iii) Clearly $\mathbf{0} = 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_k$ is in $U$.

$\square$

**Lemma 5.2.4.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ and assume that one of the vectors, say $\mathbf{u}_1$, can be expressed as a linear combination of the others. Then*

$$\mathrm{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = \mathrm{span}(\mathbf{u}_2, \ldots, \mathbf{u}_k).$$

*Proof.* For $\supseteq$: This is obvious since

$$\lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k = 0\mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k,$$

and so any element in the span of $\mathbf{u}_2, \ldots, \mathbf{u}_k$ is also in the span of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$. $\subseteq$: As $\mathbf{u}_1$ can be written in the span of the others, let us write $\mathbf{u}_1 = \mu_2 \mathbf{u}_2 + \cdots + \mu_k \mathbf{u}_k$ for some $\mu_j$. Then

$$\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k = (\lambda_2 + \lambda_1 \mu_2)\mathbf{u}_2 \cdots + (\lambda_k + \lambda_1 \mu_k)\mathbf{u}_k.$$

So an element in the span of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ is also in the span of $\mathbf{u}_2, \ldots, \mathbf{u}_k$. $\square$

**Idea:** This lemma tells us that the span of a set of vectors is equal to the span of a smaller set of vectors which doesn't include the vectors which can be made up of the other vectors. Effectively, I believe this is saying that we can keep reducing the spanning set until we get a smallest spanning set where each of the vectors cannot be written as a combination of the others.

## 5.3 Linear independence

**Definition 5.3.1** *(Linear independence).* The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are said to be linearly independent if the only solution to the equation

$$\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k = \mathbf{0} \quad \text{if} \quad \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0.$$

> This definition is saying that you cannot make one vector from the other. In other words, $\mathbf{v}$ cannot be made from $\lambda \mathbf{u}$ for some $\lambda \in \mathbb{R}$. Here is a more informative example
>
> $$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \text{ is not linearly independent to } \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

**Corollary 5.3.2.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ vectors in $\mathbb{R}^n$ and let $A = (\mathbf{u}_1, \ldots, \mathbf{u}_k) \in M_{n,k}(\mathbb{R})$ be the associated matrix of column vectors. Then*

$\mathbf{u}_1, \ldots, \mathbf{u}_k$ *are linear independent* $\iff$ *The system $A\boldsymbol{\lambda} = \mathbf{0}$ has no non-trivial[4] solution.*

*Proof.* We can see this from Proposition 5.2.2.

*As this states that $\boldsymbol{\lambda} = \mathbf{0}$ which is the same as $\lambda_1 = \lambda_2 = \cdots = \lambda_k$.* $\square$

---

[4]Has a solution where $\boldsymbol{\lambda} \neq 0$.

> The above corollary is very important, it states that if we cannot solve the matrix equation $A\boldsymbol{\lambda} = \mathbf{0}$ for non-trivial solutions then $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly independent.
>
> *I.e. we can use G-J to solve the matrix equation $A\boldsymbol{\lambda} = \mathbf{0}$ and if we do not get A (after G-J) in a form which shows $\lambda_1 = \cdots = \lambda_k = 0$ then we can say they are not linearly independent.*

If the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are not linearly independent we say that they are linearly dependent.

**Lemma 5.3.3.** *A pair of (non-zero) vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ are linearly dependent if and only if there is $\lambda \in \mathbb{R}$ so that $\mathbf{v} = \lambda \mathbf{u}$. In particular, any two (non-zero) non-collinear vectors are always linear independent.*

*Proof.* $\Leftarrow$: If $\mathbf{v} = \lambda \mathbf{u}$ then we can rewrite this as $1\mathbf{v} + (-\lambda)\mathbf{u} = \mathbf{0}$. So $\mathbf{u}$, and $\mathbf{v}$ are linearly dependent. $\Rightarrow$: Conversely if they are linearly dependent there are $\lambda_1, \lambda_2$ (not both zero, say $\lambda_2 \neq 0$) so that $\lambda_1 \mathbf{u} + \lambda_2 \mathbf{v} = \mathbf{0}$. Then $\mathbf{v} = \frac{-\lambda_1}{\lambda_2}\mathbf{u}$. $\square$

The above is simply saying if we can make one vector from another then they are not linearly independent, specifically if they are non-collinear then they are always linearly independent.

The following proposition says that for a set of linearly independent vectors, if we can choose a vector, say $\mathbf{u}$, such that $\mathbf{u}$ is not in the set of their combinations (so we cannot make $\mathbf{u}$ from the vectors in the set) then the vectors that span this are linearly independent to $\mathbf{u}$.

**Proposition 5.3.4.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ linear independent vectors in $\mathbb{R}^n$ and let $\mathbf{u} \in \mathbb{R}^n$ such that*

$$\mathbf{u} \notin \operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k).$$

*Then*

$$\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u} \text{ are linear independent.}$$

*Proof.* Consider the equation

$$\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k + \alpha \mathbf{u} = \mathbf{0}.$$

We need to show $\lambda_1 = \cdots = \lambda_k = \alpha = 0$ is the only solution. We have two cases:

1) $\alpha = 0$. But then we have $\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k = \mathbf{0}$, which has only the trivial solution since the $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linear independent.

2) $\alpha \neq 0$. But then we can write $\mathbf{u} = -\frac{1}{\alpha}(\lambda_1 \mathbf{u}_1 + \cdots + \lambda_k \mathbf{u}_k)$. But this is impossible since $\mathbf{u}$ is not in $\operatorname{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

$\square$

# 6 Real vector spaces

## 6.1 Vector spaces and subspaces

A (real) vector space is a non-empty set[5] $V$ together with two operations:

(i) Addition: This is a function

$$V \times V \to V$$

(Given any $\mathbf{u}$ and $\mathbf{v}$ in $V$, we write their sum as $\mathbf{u} + \mathbf{v}$);

(ii) Scalar multiplication: This is a function

$$\mathbb{R} \times V \to V$$

(Given $\mathbf{u} \in V$ and $\lambda \in \mathbb{R}$ we write the multiplication of $\mathbf{u}$ by $\lambda$ as $\lambda\mathbf{u}$).

These operations must additionally satisfy the axioms of a real vector space:

Axioms for addition:

(i) Existence of additive identity: there is a vector $\mathbf{0}$ in $V$ so that $\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}$ for all $\mathbf{v}$ in $V$;

(ii) Commutativity: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ for all $\mathbf{v}$ and $\mathbf{w}$ in $V$;

(iii) Existence of additive inverses: for each $\mathbf{v}$ in $V$ there is an element $-\mathbf{v}$ in $V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0} = (-\mathbf{v}) + \mathbf{v}$;

(iv) Associativity: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

Axioms for scalar multiplication:

(i) $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v}$ in $V$;

(ii) $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$ in $V$;

(iii) Associativity: $(\lambda\mu)\mathbf{v} = \lambda(\mu\mathbf{v})$ for all $\lambda, \mu$ in $\mathbb{R}$ and for all $\mathbf{v}$ in $V$;

(iv) Distributivity: For all $\lambda, \mu \in \mathbb{R}$ and for all $\mathbf{v}, \mathbf{w} \in V$ we have

  a) $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$;

  b) $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}$.

We refer to elements of a vector space as vectors[6].

**Remark 6.1.1.** *Instead of $\mathbb{R}$ one can also allow other scalars such that complex numbers $\mathbb{C}$ or the rationals $\mathbb{Q}$. Then, for example, $\mathbb{C}^n$ is a vector space over $\mathbb{C}$ and $\mathbb{Q}^n$ is a vector space over $\mathbb{Q}$. In fact (almost) everything in these notes holds if we replace $\mathbb{R}$ with $\mathbb{Q}$ or $\mathbb{C}$. In general, one considers vector spaces over a field $F$, where $F$ plays the role of $\mathbb{R}, \mathbb{C}$ or $\mathbb{Q}$. (Very roughly, a field is a set of scalars, where we have nice addition and multiplication, and in particular every non-zero scalar has a multiplicative inverse).*

**Definition 6.1.2** *(Vector subspace).* A vector subspace of a vector space $V$ is a non-empty subset that is itself a vector space with respect to the same addition and multiplication operations of $V$.

---

[5]This could be for any type, not just vectors or reals, I know, it's a bit bloody confusing.
[6]Big shock.

This is just a subset of a vector space $V$, don't forget that this can still be for any type of object not just our typical $\mathbb{R}^n$ vectors[7].

**Proposition 6.1.3** (Subspace criterion)**.** *A non-empty subset $U$ of a vector space $V$ is a vector subspace of $V$ if and only if it satisfies the following three conditions:*

    *(i) closure under addition: if $\mathbf{u}_1$ and $\mathbf{u}_2$ are in $U$, then $\mathbf{u}_1 + \mathbf{u}_2$ is also in $U$;*

    *(ii) closure under scalar multiplication: if $\mathbf{u}$ is in $U$ and $\lambda$ is in $\mathbb{R}$, then $\lambda \mathbf{u}$ is in $U$;*

    *(iii) existence of an origin: the vector $\mathbf{0}$ is in $U$.*

## 6.2   Linear independence, spanning sets, bases, and dimension

The notations for linear combination, linear span, spanning sets, and linear independence for a general vector space $V$ are defined exactly in the same way as for $\mathbb{R}^n$.

**Definition 6.2.1** *(Basis)***.** A (finite) basis for a vector space $V$ is a finite subset of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $V$ such that

    (i) the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent, and

    (ii) $V = \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$.

**Assorted Lemmas**:

**Lemma 6.2.2.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ vectors in a vector space $V$. Then $\mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is a vector subspace of $V$.*

**Lemma 6.2.3.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be vectors in a vector space $V$ and assume that one of the vectors, say $\mathbf{u}_1$, can be expressed as a linear combination of the others. Then*

$$\mathrm{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = \mathrm{span}(\mathbf{u}_2, \ldots, \mathbf{u}_k).$$

**Lemma 6.2.4.** *Suppose that the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is not linearly independent. Then for some $1 \leq i \leq k$, we have that*

$$\mathbf{u}_i \in \mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_k).$$

*Proof.* Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ are not linearly independent there is some linear dependence relation

$$\sum_{j=1}^{k} \lambda_j \mathbf{u}_j = \mathbf{0}$$

with at least one of the $\lambda_j$ non zero. Say that $\lambda_i \neq 0$. Then the linear dependence relation may be rearranged to read

$$\mathbf{u}_i = \sum_{j=1}^{i-1} \frac{\lambda_j}{-\lambda_i} \mathbf{u}_j + \sum_{j=i+1}^{k} \frac{\lambda_j}{-\lambda_i} \mathbf{u}_j.$$

The equivalence of the spans then follows from Lemma 6.2.3.       □

---

[7]Why would we make it so easy to understand by just thinking about $\mathbb{R}^n$.

**Lemma 6.2.5.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ linearly independent vectors in a vector space $V$ and let $\mathbf{u} \in V$ such that*

$$\mathbf{u} \notin \mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_k).$$

*Then*

$$\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u} \text{ are linearly independent.}$$

**Theorem 6.2.6** *(Existence of finite basis)***.** *Every vector space that is spanned by finitely many vectors has a basis.*

**Theorem 6.2.7** *("What is a basis?")***.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be vectors in a vector space $V$. Then the following statements are (pairwise) equivalent:*

(i) *The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form a basis of $V$.*

(ii) *The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form a maximal linear independent set in $V$, i.e., one cannot add another vector in $V$ to the $\mathbf{u}_i$, and still obtain a linearly independent set.*

(iii) *Any vector $\mathbf{u}$ in $V$ can be written uniquely as*

$$\mathbf{u} = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k.$$

(iv) *The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_j$ form a minimal spanning set of $V$; i.e., one cannot remove any of the $\mathbf{u}_i$ and still have a spanning set of $V$.*

*Proof.* **Add proof.** $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We call a vector space finite dimensional if it has a finite basis.

For example, we have seen that $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathbb{R}[x]_n$ and is therefore finite dimensional.

**Theorem 6.2.8** *(Steinitz Exchange Theorem)***.** *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a spanning set of a vector space $V = \mathrm{span}(S)$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\} \subseteq V$ be a linearly independent set.*

*Then there exist $\ell$ distinct vectors $\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_\ell} \in S$ such that for all $j$ with $0 \le j \le \ell$ we have*

$$(S \setminus \{\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_\ell}\}) \cup \{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$$

*is a spanning set for $V$.*

*Proof.* **Add proof.** $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Corollary 6.2.9.** *In the situation above $k \ge \ell$. In other words, a spanning set for a vector space is always at least as big as any linearly independent set within the vector space.*

**Corollary 6.2.10.** *Let $V$ be a finite-dimensional vector space. Then every basis for $V$ has the same number of elements.*

*Proof.* **Add proof.** $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Definition 6.2.11** *(Dimension).* Let $V$ be a vector space that is generated by finitely many vectors. The dimension, $\dim(V)$, of $V$ is defined to be the cardinality of any basis for $V$.

This definition is best understood through the following examples.

**Example 6.2.12.** The vector space $\mathbb{R}[x]_n$ has basis $1, x, \ldots, x^n$. Hence

$$\dim \mathbb{R}[x]_n = n + 1.$$

Additionally, the vector space $M_{m,n}(\mathbb{R})$ has a basis of matrices $E_{rs}$ (for $1 \leq r \leq m$ and $1 \leq s \leq n$) and so

$$\dim M_{m,n}(\mathbb{R}) = m \cdot n.$$

**Theorem 6.2.13.** *Let $V$ be a finite dimensional vector space and $U$ a vector subspace of $V$. Then*

*(i) $U$ is also finite-dimensional;*

*(ii) $\dim(U) \leq \dim(V)$;*

*(iii) $\dim(U) = \dim(V) \iff U = V$.*

*Proof.* **Add proof**. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Remark 6.2.14.** *Not every vector space is finite-dimensional.*

This remark is shown through the following example.

**Example 6.2.15.** The vector space $\mathbb{R}[x]$ of all polynomials has the infinite basis $S = \{1, x, x^2, x^3, \ldots\}$. Every element in $\mathbb{R}[x]$ is the unique linear combination of finitely many elements in $S$.

The following theorem provides us with a list of useful observations.

**Theorem 6.2.16.** *Let $V$ be a vector space of dimension $k$ and let $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ be $\ell$ vectors in $V$. Then*

*(i) If $\ell > k$, then the $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ are linearly dependent.*

*(ii) If $\ell < k$, then the $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ do not span $V$.*

*(iii) If $k = \ell$, then the following statements are (pairwise) equivalent*

  *(a) The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form a basis of $V$.*

  *(b) The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linear independent.*

  *(c) The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span $V$.*

*Proof.* **Insert proof**. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

## 6.3 Bases and dimension in $\mathbb{R}^n$

**Remark 6.3.1.** *Let $U$ be a vector subspace of $\mathbb{R}^n$. Then, of course, a basis for $U$ is any linearly independent set of vectors that span $U$.*

**Theorem 6.3.2.** *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be $k$ vectors in $\mathbb{R}^n$. Then*

  *(i) If $k > n$, then the $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are linearly dependent.*

  *(ii) If $k < n$, then the $\mathbf{u}_1, \ldots, \mathbf{u}_k$ do not span $\mathbb{R}^n$.*

  *(iii) If $k = n$, then*

$$\mathbf{u}_1, \ldots, \mathbf{u}_n \text{ form a basis of } \mathbb{R}^n \iff \text{ The matrix } (\mathbf{u}_1, \ldots, \mathbf{u}_n) \text{ is non singular}$$
$$\iff \det(\mathbf{u}_1, \ldots, \mathbf{u}_n) \neq 0.$$

*So, the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ form a basis for $\mathbb{R}^n$ if and only if $k = n$ and the square matrix $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is non-singular.*

*Proof.* **Add proof.** $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Definition 6.3.3.** Let $A \in M_{n \times k}(\mathbb{R})$ be a $n \times k$ matrix.

  (i) The column space of $A$ is the span of the columns of $A$. The column rank of $A$ is the dimension of the column space of $A$.

    The column space is the subspace of all $\mathbf{b} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$ has a solution.

  (ii) The row space of $A$ is the span of the rows of $A$. The row rank of $A$ is the dimension of the row space of $A$.

    The row space is the subspace of all $c \in \mathbb{R}^k$ such that $A^t\mathbf{y} = \mathbf{c}$ has a solution.

  (iii) The nullspace (or kernal) of $A$ is the solution space to the homogeneous system $A\mathbf{x} = \mathbf{0}$. The nullity of $A$ is the dimension of the nullspace.

**Lemma 6.3.4.** *Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be linear independent vectors in $\mathbb{R}^n$ and let $P \in M_n(\mathbb{R})$ be an invertible matrix. Then the vectors $P\mathbf{v}_1, \ldots, P\mathbf{v}_k$ are linear independent.*

*Proof.* Consider the equation $\lambda_1(P\mathbf{v}_1) + \cdots + \lambda_k(P\mathbf{v}_k) = \mathbf{0}$. By linearity of matrix multiplication this is equivalent to the equation $P(\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k) = \mathbf{0}$, and multiplying with $P^{-1}$ gives $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k = P^{-1}\mathbf{0} = \mathbf{0}$. But the $\mathbf{v}_i$ are linear independent, so we have $\lambda_1 = \cdots = \lambda_k = 0$. Thus the vectors $P\mathbf{v}_1, \ldots, P\mathbf{v}_k$ are linear independent. $\qquad\qquad$ $\square$

**Proposition 6.3.5.** *Let $A \in M_{n \times k}(\mathbb{R})$ and let $P \in M_n(\mathbb{R})$ be an invertible matrix. Then $A$ and $PA$ have the same column rank, row rank, and nullity. In particular, elementary row operations do not change column rank, row rank, and nullity. More precisely,*

  *1. $A$ and $PA$ have the same row space. In particular, ERO's do not change the row space.*

2. *PA may have a different column space than A, but the dimension stays the same. In particular, ERO's do not change the dimension of the column space.*

3. *A and PA have the same nullsapce. In particular, ERO's do not change the nullspace.*

*Proof.* Since every invertible matrix is a product of elementary matrices, it suffices to consider an elementary matrix $E$.

(i) After applying an ERO the new rows are a linear combination of the old ones (think about what each ERO does). This could make the row space smaller (conceivably). but not bigger. However, the process is reversible (instead consider the same argument with $E^{-1}$), so it must be that the space didn't change.

(ii) Denote the column space of $A$ by $U$ and let $\mathbf{u}_1, \ldots, \mathbf{u}_\ell$ be a basis of $U$. But $A\mathbf{x} = \mathbf{b}$ if and only if $(EA)\mathbf{x} = E\mathbf{b}$ (since applying an ERO does not change the solution set as per Lemma 3.5.1). This shows that the column space of $EA$ is equal to $EU := \{E\mathbf{u} : \mathbf{u} \in U\}$.

We claim $E\mathbf{u}_1, \ldots, E\mathbf{u}_\ell$ is a basis of $EU$, and this will show the dimension of the two spaces $U$ and $EU$ is the same. To do this we need to show the vectors form a linearly independent spanning set of $EU$. By Lemma 6.3.4 we know $E\mathbf{u}_1, \ldots, E\mathbf{v}_\ell$ are linearly independent. For spanning, let $\mathbf{v} \in EU$ be given by $\mathbf{v} = E\mathbf{u}$ for some $\mathbf{u} \in U$. We can write $\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_\ell \mathbf{u}_\ell$ for some $\lambda_i \in \mathbb{R}$. But then $\mathbf{v} = E\mathbf{u} = E(\lambda_1 \mathbf{u}_1 + \cdots + \lambda_\ell \mathbf{u}_\ell) = \lambda_1(E\mathbf{u}_1) + \cdots + \lambda_\ell(E\mathbf{u}_\ell)$.

(iii) This follows immediately from Lemma 3.5.1.

$\square$

**Theorem 6.3.6** *(Rank Theorem). The row rank of a matrix is equal to its column rank.*

**Definition 6.3.7** *(Rank).* We define the rank of a matrix $A$ to be its row rank $=$ column rank. We write $\text{rank}(A) = \text{rk}(A)$.

**Theorem 6.3.8** *(Rank-Nullity Theorem). Let $A \in M_{n \times k}(\mathbb{R})$ be an $n \times k$ matrix. Then*

$$\text{rk}(A) + \text{nullity}(A) = k.$$

*Proof.* We prove Theorem 6.3.6 and this theorem together. By Proposition 6.3.5 we can switch to the RREF $A'$ of $A$. After switching some columns if necessary we can assume it looks as follows:

$$A' = \left( \begin{array}{c|c} I_r & C \\ \hline \mathbf{0}_{(n-r) \times r} & \mathbf{0}_{(n-r) \times (k-r)} \end{array} \right).$$

For some $r \times (k-r)$ matrix $C$. But the first $r$ columns are linear independent (they are simply the first $r$ standard basis vectors of $\mathbb{R}^n$), and the other columns are in their span. Thus, span(column space of $A'$) $=$ span$\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ and

$$\text{column rank of } A' = r.$$

But also
$$\text{row rank of } A' = r,$$
since the first $r$ rows are linearly independent in $\mathbb{R}^k$ (since the 1's in the $I_r$ are in different slots), and the other rows are zero. Finally, upon taking $x_{r+1}, \ldots, x_k$ as free variables, we see that a general vector in the nullspace of $A'$ (that is, a solution $\mathbf{x}$ to the system $A'\mathbf{x} = \mathbf{0}$) is given by

$$\mathbf{x} = x_{r+1}\begin{pmatrix} -\mathbf{c}_1 \\ --- \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_{r+2}\begin{pmatrix} -\mathbf{c}_2 \\ --- \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_k\begin{pmatrix} -\mathbf{c}_{k-r} \\ --- \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

where $\mathbf{c}_i \in \mathbb{R}^r$ denotes the $i$-th column of $C$. The spanning vectors are clearly linear independent and so

$$\text{nullity}(A') = k - r.$$

$\square$

**Corollary 6.3.9.** *For any matrix $A \in M_{n \times k}(\mathbb{R})$ we have*

(i) *The rank of $A$ is the number of leading 1's in the RREF of $A$.*

(ii) *The nullity of $A$ is the number columns in the RREF of $A$ without a leading 1; i.e., the number of free variables in the solution set to $A\mathbf{x} = \mathbf{0}$.*

## 6.4 Sums and intersections of vector subspaces

Let $V$ be a vector space and let $U, W$ be subspaces. Then define

$$U + W := \{\mathbf{v} \in V : \mathbf{v} = \mathbf{u} + \mathbf{w} \text{ for some } \mathbf{u} \in U, \mathbf{w} \in W\},$$
$$U \cap W := \{\mathbf{v} \in V : \mathbf{v} \in U \text{ and } \mathbf{v} \in W\}.$$

Both $U + W$ and $U \cap W$ are vector subspaces of $V$.

**Proposition 6.4.1.** *Let $U$ and $W$ be finite-dimensional vector subspaces of a vector space $V$. Then*

$$\dim(U) + \dim(W) = \dim(U + W) + \dim(U \cap W).$$

*Proof.* Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for $U \cap W$ (we know such a basis exists by Theorem 6.2.13). Using Lemma 6.2.5 multiple times, we may repeatedly add vectors in $U$ to this set (keeping it linearly independent) until we span $U$. Do the same for $W$. Then, we have extended the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ to bases of $U$ and $W$ given by

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_s\}, \quad \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_t\}$$

respectively. Then we claim that

$$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_s, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_t\}$$

is a basis for $U + W$. If this claim is true we have

$$\dim(U) + \dim(W) = s + t = (s + t - r) + r = \dim(U + W) + \dim(U \cap W).$$

In order to prove this claim, observe that $B$ spans $U + W$, all we need to do now is establish linear independence. If $s = 0$ or $t = 0$ it is trivial. Assume $s \geq 1$ and $t \geq 1$. Now consider

$$\mathbf{0} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r + \mu_{r+1} \mathbf{u}_{r+1} + \cdots + \mu_s \mathbf{u}_s + \gamma_{r+1} \mathbf{w}_{r+1} + \cdots + \gamma_t \mathbf{w}_t.$$

Suppose at least one of the $\gamma_j$ is non-zero. This gives

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r + \mu_{r+1} \mathbf{u}_{r+1} + \cdots + \mu_s \mathbf{u}_s + = -(\gamma_{r+1} \mathbf{w}_{r+1} + \cdots + \gamma_t \mathbf{w}_t).$$

The left hand side is in $U$ and the right hand side is in $W \setminus (U \cap W)$. But, the intersection of these two sets is empty, so we have a contradiction. Thus $\gamma_{r+1} = \cdots = \gamma_t = 0$. Since the vectors

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_s\}$$

are linearly independent we must also have $\lambda_1 = \cdots = \lambda_r = \mu_{r+1} = \cdots = \mu_s = 0$. Thus the vectors $B$ are linearly independent. $\qquad \square$

**Definition 6.4.2** *(Direct Sum).* Suppose that $U$ and $W$ are vector subspaces of $V$ with

(i) $U + W = V$;

(ii) $U \cap W = \{\mathbf{0}\}$.

Then, we say $V$ is the direct sum of $U$ and $W$ and write $V = U \oplus W$.

**Proposition 6.4.3.** *Let $U$ and $W$ be two subspaces of a vector space $V$. Then the following statements are pairwise equivalent:*

(i) $V = U \oplus W$;

(ii) *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$ is a basis for $U$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$ is a basis for $W$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_s, \mathbf{w}_1, \ldots, \mathbf{w}_t\}$ is a basis for $V$;*

(iii) $V = U + W$ *and* $\dim V = \dim U + \dim W$;

(iv) *Every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \mathbf{u} + \mathbf{w}$ where $\mathbf{u} \in U$ and $\mathbf{w} \in W$.*

*Proof.* $(i) \iff (ii) \iff (iii)$ follows from Proposition 6.4.1 and its proof. For $(i) \iff (iv)$ see problem sheet 9. $\qquad \square$

## 6.5 Coordinates

If $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for $V$ and if $\mathbf{u}$ is any point of $V$ then (by Theorem 6.2.7) we can uniquely express $\mathbf{u}$ as

$$\mathbf{u} = x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n.$$

We call the real numbers $(x_1, \ldots, x_n)$ the coordinates of $\mathbf{u}$ with respect to this basis.

The coordinates are simply a collection of real numbers, if we wished we could write them as a column vector. We obtain a map, which we call the coordinate map,

$$\Phi = \Phi_B : V \to \mathbb{R}^n$$

$$\mathbf{u} \mapsto \Phi(\mathbf{u}) := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It is clear that the coordinate map depends on the choice of the basis $B$.

For a column vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, we have that $(x_1, \ldots, x_n)$ are the coordinates of $\mathbf{x}$ with respect to the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

**Lemma 6.5.1.** *Consider a basis* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *of* $\mathbb{R}^n$, *and let* $P = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ *be the associated matrix of column vectors. Then the coordinates* $(x_1, \ldots, x_n)$ *of* $\mathbf{u}$ *with respect to* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *are given by*

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = P^{-1}\mathbf{u}.$$

*Note that if* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ *then* $P = I_n$.

# 7  Linear maps

## 7.1  Definition and examples of linear maps

**Definition 7.1.1** *(Linear map).* Let $V$ and $W$ be two vector spaces, with addition denoted by $+_V$ and $+_W$ respectively. A linear map $T : V \to W$ is a function satisfying the following properties

  (i) for all $\mathbf{u}$ and $\mathbf{v}$ in $V$ we have $T(\mathbf{u} +_V \mathbf{v}) = T(\mathbf{u}) +_W T(\mathbf{v})$;

  (ii) for all $\mathbf{v}$ in $V$ and all $\lambda \in \mathbb{R}$ we have $T(\lambda \mathbf{v}) = \lambda T(\mathbf{v})$.

A linear map is sometimes referred to as a linear transformation or a linear mapping.

**Example 7.1.2.** The map $T$ given by

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3x - y + z \\ z \\ 0 \end{pmatrix} = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

is a linear map $\mathbb{R}^3 \to \mathbb{R}^3$

**Example 7.1.3.** The map $T$ given by

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

is not linear, since $T \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \neq \mathbf{0}$.

It is however called an affine map, it is of the form $A\mathbf{x} + \mathbf{b}$.

**Lemma 7.1.4.** *Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of a vector space $V$. Then*

  *(i) Any linear map $T : V \to W$ of vector spaces is determined by its values $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ on the basis vectors of $V$.*

  *(ii) Conversely, given arbitrary vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$ in $W$, then there exists a linear map $T : V \to W$ such that*

$$T(\mathbf{v}_1) = \mathbf{w}_1, \ldots, T(\mathbf{v}_n) = \mathbf{w}_n.$$

*Proof.* □

**Lemma 7.1.5.** *A map $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there is a $m \times n$ matrix $A$ such that*

$$T(\mathbf{x}) = A\mathbf{x}.$$

*Moreover, $A$ is given by*

$$A = (T(\mathbf{e}_1), T(\mathbf{e}_2), \cdots, T(\mathbf{e}_n)).$$

*So the ith column is the image of the standard basis vector $\mathbf{e}_i$.*

We can see that the only linear maps between Euclidean spaces are the matrices.

## 7.2 Properties of linear maps

**Lemma 7.2.1.** *Let $T : V \to W$ and $S : W \to U$ be two linear maps of vector spaces. Then the composition $S \circ T : V \to U$ is a linear map. In particular, if $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^k$ are linear maps given by $T(\mathbf{x}) = A\mathbf{x}$ and $S(\mathbf{y}) = B\mathbf{y}$, then the composition $S \circ T : \mathbb{R}^n \to \mathbb{R}^k$ is again linear, and moreover we have*

$$(S \circ T)(\mathbf{x}) = (BA)\mathbf{x}.$$

*So the composition of linear maps corresponds to matrix multiplication.*

**Definition 7.2.2** *(Isomorphism).* Let $T : V \to W$ be a bijective linear map of vector spaces (that is, a linear map which is one-to-one and onto). Then we call $T$ an isomorphism and call $V$ and $W$ isomorphic. We write $V \cong W$.

**Lemma 7.2.3.** *Let $T : V \to W$ be an isomorphism of vector spaces. Then the inverse map $T^{-1} : W \to V$ is also a linear map, in fact an isomorphism. For the linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ given by the matrix $A \in M_n(\mathbb{R})$, we have that $T$ is an isomorphism if and only if $A$ is invertible. In that case, we have*

$$T^{-1}(\mathbf{y}) = A^{-1}\mathbf{y}.$$

**Lemma 7.2.4.** *Let $V$ b a vector space of dimension $n$. Then the coordinate map $\Phi : V \to \mathbb{R}^n$ is an isomorphism. In particular, every finite dimensional vector space is isomorphic to the Euclidean space of the same dimension.*

**Theorem 7.2.5.** *Let $V$ be a vector space of dimension $n$ and let $W$ be a vector space of dimension $m$. Let $\phi : V \to \mathbb{R}^n$ and $\psi : W \to \mathbb{R}^m$ be two coordinate maps with respect to the choice of some bases $V$ and $W$ respectively. Let $T : V \to W$ be a linear map. Then there exists a $m \times n$ matrix $A$ so that for all $\mathbf{v} \in V$ we have*

$$\phi(T(\mathbf{v})) = A\phi(\mathbf{v}).$$

*We say $T$ is represented by $A$. (The matrix $A$ depends on the choice of the bases made.)*

## 7.3 Image, kernel, rank and nullity

**Definition 7.3.1** *(Image and kernel).* Let $V$ and $W$ be vector spaces and $T : V \to W$ be a linear map. We define the image of $T$ by

$$\operatorname{im}(T) = \{\mathbf{w} \in W : \text{there exists } \mathbf{v} \in V \text{ so that } \mathbf{w} = T(\mathbf{v})\}.$$

We define the kernel of $T$ by

$$\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}.$$

**Lemma 7.3.2.** *Let $T : V \to W$ be a linear map of vector spaces. Then $\operatorname{im}(T)$ is a vector subspace of $W$ and $\ker(T)$ is a vector subspace of $V$.*

**Lemma 7.3.3.** *Let $T : V \to W$ be a linear map of vector spaces, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $V$. Then the image $\operatorname{im}(T)$ is spanned by $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$.*

*Proof.* Let $\mathbf{w} = T(\mathbf{u})$ be in the image of $T$. We can write $\mathbf{u} = \lambda_1 \mathbf{V}_1 + \cdots + \lambda_n \mathbf{v}_n$, and then we see immediately

$$\mathbf{w} = T(\mathbf{u}) = T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n).$$

So $\mathbf{w}$ is in the span of $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$. $\qquad\square$

**Proposition 7.3.4.** *Let $T : V \to W$ be a linear map of vector spaces. Then the following statements are (pairwise) equivalent:*

(i) $\ker(T) = \{\mathbf{0}\}$;

(ii) *T is one-to-one (that is, it is injective);*

(iii) *If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linear independent vectors in $V$, then the images $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)$ are linearly independent in $W$.*

**Corollary 7.3.5.** *Let $T : V \to W$ be a linear map of vector spaces with $\ker(T) = \{\mathbf{0}\}$. Assume $\dim V = n$ with a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Then the images $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ are a basis for $\mathrm{im}(T)$. In particular, we have $\dim(\mathrm{im}(T)) = \dim(V)$.*

**Definition 7.3.6** *(Rank and nullity of a linear map).* Let $V$ and $W$ be vector spaces and $T : V \to W$ be a linear map. We define the rank of $T$ to be

$$\mathrm{rk}(T) = \dim(\mathrm{im}(T)).$$

We define the nullity of $T$ to be

$$n(T) = \dim(\ker(T)).$$

**Proposition 7.3.7.** *Let $T : V \to W$ be a linear map of finite-dimensional vector spaces and let $\phi : V' \to V$ and $\psi : W \to W'$ be two isomorphisms. Then*

$$\mathrm{rk}(\psi \circ T \circ \phi) = \mathrm{rk}(T) \quad and \quad n(\psi \circ T \circ \phi) = n(T).$$

**Theorem 7.3.8** *(Rank-Nullity Theorem for linear maps).* *Let $V$ and $W$ be two finite-dimensional vector spaces and $T : V \to W$ be a linear map. Then*

$$\dim(V) = \mathrm{rk}(T) + n(T).$$

**Remark 7.3.9.** *Alternatively, we could argue directly. Let $\mathbf{u}_1, \ldots, \mathbf{u}_r$ be a basis for $\ker(T)$ and let $\mathbf{w}_1, \ldots, \mathbf{w}_s$ be a basis for $\mathrm{im}(T)$. By the definition of $\mathrm{im}(T)$ there exist $\mathbf{v}_1, \ldots, \mathbf{v}_s$ with $T(\mathbf{v}_i) = \mathbf{w}_i$ for $1 \le i \le s$. Now we claim that $\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{v}_1, \ldots, \mathbf{v}_s$ is a basis for $V$, which will prove the result.*

**Theorem 7.3.10.** *For two finite-dimensional vector spaces $V$ and $W$ we have*

$$V \cong W \iff \dim V = \dim W.$$

## 7.4 Change of basis and of coordinates

**Theorem 7.4.1.** *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Suppose that with respect to the standard bases $T$ is represented by the matrix $A$. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be any basis of $\mathbb{R}^n$ and let $P = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$. Let $S' = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be any basis of $\mathbb{R}^m$ and let $Q = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$. Then with respect to the bases $S$ of $\mathbb{R}^n$ and $S'$ of $\mathbb{R}^m$ the transformation $T$ is represented by the matrix $B = Q^{-1}AP$.*

**Corollary 7.4.2.** *Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ are any bases of $\mathbb{R}^n$. Let $P$ be a matrix so that $\mathbf{v}'_j = P\mathbf{v}_j$. Suppose that $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ and $\{\mathbf{w}'_1, \ldots, \mathbf{w}'_m\}$ are any bases of $\mathbb{R}^m$. Let $Q$ be a matrix so that $\mathbf{w}'_j = Q\mathbf{w}_j$.*

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Suppose that with respect to the bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ the map $T$ is represented by the matrix $A$. Then with respect to the bases $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ and $\{\mathbf{w}'_1, \ldots, \mathbf{w}'_m\}$ the map $T$ is represented by the matrix $B = Q^{-1}AP$.

# 8 Eigenvalues, Eigenvectors, and Diagonalisation

## 8.1 Prologue

## 8.2 Eigenvalues and eigenvectors

**Definition 8.2.1.** Let $A$ be a square matrix. A non-zero $n$-vector is an eigen-vector of $A$ with eigenvalue $\lambda \in \mathbb{C}$, if

$$A\mathbf{v} = \lambda\mathbf{v}.$$

From this we can see

$$\mathbf{0} = A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda I)\mathbf{v}.$$

By the fundamental lemma of invertible matrices we have the existence for such a $\mathbf{v}$ for a given $\lambda$ is equivalent to the matrix $A - \lambda I$ being non-invertible. Hence the set of eigenvalues $\lambda$ is the set of solutions to

$$\det(A - \lambda I) = 0.$$

We define $p_A(t) = \det(A - tI)$ to be the characteristic polynomial of $A$.

**Example 8.2.2.** If $A$ is a diagonal matrix, then the eigenvalues of $A$ are precisely the entries on the diagonal. For example,

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

has characteristic polynomial $p_A(t) = -(t+2)(t-3)(t-1)$, so the eigenvalues are $\lambda = -2, 3, 1$.

To find the corresponding **eigenvectors**, we solve

$$(A - \lambda I)\mathbf{v} = \mathbf{0}.$$

**Multiple eigenvalues.** If we have a polynomial $p(t)$ with degree $N$, the following decomposition holds:

$$p(t) = c(t - \lambda_1)^{k_1} \cdot (t - \lambda_2)^{k_2} \cdot \ldots \cdot (t - \lambda_p)^{k_p},$$

with $c \in \mathbb{C}$ is non-zero. The roots, $\lambda \in \mathbb{C}$, are all distinct, i.e. $\lambda_i \neq \lambda_j$ for all $i \neq j$.

**Definition 8.2.3.** The algebraic multiplicity of the root $\lambda_i$ is given by $k_i \in \mathbb{N}$, with $k_i \neq 0$, and $\sum_{i=1}^{p} k_i = N$.

**Remark 8.2.4.** *If an eigenvalue $\lambda$ has multiplicity $k \neq 0 \in \mathbb{N}$ it means that the characteristic polynomial takes the form*

$$p_A(t) = (t - \lambda)^k Q(t),$$

*where the polynomial $Q(t)$ has degree $N - k$ and it is non-vanishing on $\lambda$, i.e. $Q(\lambda) \neq 0$.*

**Proposition 8.2.5.** *There are $p$ linearly independent eigenvectors corresponding to $\lambda$, where $1 \le p \le k$.*

**Definition 8.2.6.** The eigenspace $V_\lambda$ is the $p$-dimensional vector space spanned by these eigenvectors. Since the eigenspace $V_\lambda$ is generated by all the eigenvectors associated with the eigenvalue $\lambda$ we can also write it as

$$V_\lambda = \ker(A - \lambda I),$$

where the kernel, ker, or a linear operator $B$ is a subspace of the full vector space $V$ given by

$$\ker B = \{\mathbf{v} \in V \mid B\mathbf{v} = 0\}.$$

**Definition 8.2.7.** We call $p = \dim V_\lambda$ the geometric multiplicity of $\lambda$.

## 8.3 The Cayley-Hamilton theorem

**Theorem 8.3.1** *(Cayley-Hamilton).* *Let $B$ be a square matrix with characteristic polynomial $p_B(t)$. Then $p_B(B) = 0$, as a matrix equation.*

*Partial proof.* Let us show this first for $2 \times 2$ matrices by direct calculation. Let

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then $p_B(t) = t^2 - (a+d)t + (ad - bc)$, and

$$p_B(B) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

$\square$

*Proof.* The matrix $B$ defines a linear transformation $T$. Let us change basis to a flag basis for $T$, i.e. an ordered basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ such that $B$ in this basis becomes upper triangular

$$A = P^{-1}BP = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ 0 & a_{22} & \cdots & a_{2N} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{NN} \end{pmatrix}$$

where $P$ is the change of basis matrix. We use the result that for any linear transformation $T$ we can always find such a change of basis.

The characteristic polynomial does not change under changes of basis (see the proposition below), so the characteristic polynomial of the matrix in the new basis, which takes the form

$$p_A(t) = (a_{11} - t)(a_{22} - t) \cdots (a_{NN} - t),$$

is the same as the original characteristic polynomial $p_B(t)$. Also,

$$\begin{aligned}
p_A(A) &= (a_{11} - A)(a_{22} - A)\cdots(a_{NN} - A) \\
&= (P^{-1}a_{11}P - P^{-1}BP)(P^{-1}a_{22}P - P^{-1}BP)\cdots(P^{-1}a_{NN}P - P^{-1}BP) \\
&= P^{-1}(a_{11} - B)(a_{22} - B)\cdots(a_{NN} - B)P \\
&= P^{-1}p_B(B)P
\end{aligned}$$

so $p_A(A)$ will vanish if and only if $p_B(B)$ vanishes. We will now show that $p_A(A) = 0$. To prove this matrix equation we will show for all $1 \leq i \leq N$ that $\forall \mathbf{v} \in V_i := \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_i)$ we have $p_i(A)\mathbf{v} = 0$, with

$$p_i(A) = (A - a_{11}I)(A - a_{22}I)\cdots(A - a_{ii}I).$$

The case $i = N$ is the statement we are after, since in this case $p_N(A) = p_A(A)$ and $V_N$ is the full vector space. We do this using induction on $i$. For $i = 1$ we have that $V_1 = \text{span}(\mathbf{v}_1)$ and $p_1 = (A - a_{11}I)$. Clearly, from the upper triangular form of $A$,

$$A\mathbf{v}_1 = a_{11}\mathbf{v}_1$$

so for $i = 1$ indeed $p_1\mathbf{v} = 0$ for all $\mathbf{v} \in V_1$. Let us assume that the inductive hypothesis is valid up to $i - 1$, we will show that it also holds for $i$. To prove it we note that $\forall \mathbf{v} \in V_i$, we can write

$$\mathbf{v} = \alpha\mathbf{v}_i + \mathbf{w},$$

with $\alpha \in \mathbb{C}$ and $\mathbf{w} \in V_{i-1}$. By linearity (and noting that $p_i = p_{i-1}(A - a_{ii}I) = (A - a_{ii})p_{i-1}$)

$$p_i\mathbf{v} = \alpha p_{i-1}(A - a_{ii}I)\mathbf{v}_i + (A - a_{ii})p_{i-1}\mathbf{w}$$

but $p_{i-1}\mathbf{w} = 0$ by the inductive hypothesis, so

$$p_i\mathbf{v} = \alpha p_{i-1}(A - a_{ii}I)\mathbf{v}_i.$$

Now, because of the upper triangular form, if we apply $A$ to $\mathbf{v}_i$ we obtain

$$A\mathbf{v}_i = a_{ii}\mathbf{v}_i$$

with $\mathbf{w}_{i-1} \in V_{i-1}$. Therefore

$$p_{i-1}(A - a_{ii}I)\mathbf{v}_i = p_{i-1}\mathbf{w}_{i-1} = 0$$

concluding our theorem. $\qquad\qquad\square$

## 8.4   The similarity relation

**Definition 8.4.1.** An equivalence relation, $\sim$, on a set $X$, is a binary relation with the following properties:

(i) $a \sim a$ (reflexivity).

(ii) $a \sim b \implies b \sim a$ (symmetry).

(iii) If $a \sim b$ and $b \sim c \implies a \sim c$ (transitivity).

$\forall a, b, c \in X$.

**Definition 8.4.2.** We say that two square matrices $A$ and $B$ (of the same size) are similar, written $A \sim B$, if there exists an invertible matrix $M$ such that $A = MBM^{-1}$.

**Proposition 8.4.3.** *Similarity is an equivalence relation.*

The set of all $n \times n$ matrices gets partitioned into equivalence classes

$$[A] = \left\{ B \,\middle|\, B \sim A, \text{ equivalently } B = MAM^{-1} \text{ for some } M \in \mathrm{GL}(n, \mathbb{R}) \right\}.$$

We have that $\mathrm{GL}(n, \mathbb{R})$ denotes the set of invertible $n \times n$ matrices with real coefficients, i.e.

$$\mathrm{GL}(n, \mathbb{R}) = \{ M \in M_n(\mathbb{R}) \text{ s.t. } \det M \neq 0 \}.$$

This set is usually called General Linear group. Changes of basis matrices are elements of the general linear group.

**Proposition 8.4.4.** *Similar matrices have the same eigenvalues.*

## 8.5 Diagonalisation by change of basis

**Definition 8.5.1.** A square matrix $A$ is said to be diagonalisable if $A$ is similar to a diagonal matrix.

**Proposition 8.5.2.** *The $n \times n$ matrix $A$ is diagonalisable if and only if it has $n$ linearly independent eigenvectors.*

**Example 8.5.3.** If possible, diagonalise the matrix

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

*Solution.* $p(t) = t^2 - 4t + 3 = (t-1)(t-3)$, so we get eigenvalues $\lambda_1 = 1, \lambda_2 = 3$.

Now we need to get a similar matrix, this will have the same eigenvalues. Thus $\lambda = 1$

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0} = (A - \lambda I)\mathbf{v} = (A - I)\mathbf{v} = \begin{pmatrix} -2 & 4 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

solving this gives the condition $-2v_1 + 4v_2 = 0$, taking $v_2 = 1$ gives us $v_1 = 2$, so

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

$\lambda = 3$

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0} = (A - \lambda I)\mathbf{v} = (A - 3I)\mathbf{v} = \begin{pmatrix} -4 & 4 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

which gives $v_1 = v_2$, taking $v_2 = 1$ gives $v_1 = 1$, hence

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We can set

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad M^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix},$$

then

$$M^{-1}AM = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

$\triangle$

**Remark 8.5.4.** *We can actually check our answer without computing $M^{-1}$, we just have to check that $AM = MD$, with $D$ being the diagonal matrix of eigenvalues (in the right order).*

**Example 8.5.5.** Solve the system of ODEs

$$\dot{x} = -x + 4y,$$
$$\dot{y} = -2x + 5y,$$

where $x = x(t)$, $y = y(t)$ and $\dot{x} = \frac{dx(t)}{dt}$, similarly $\dot{y} = \frac{dy(t)}{dt}$.

*Solution.* If we diagonalise the system this is the same thing as changing basis. If $\mathbf{v}(t)$ denotes the column vector with entries $(x(t), y(t))$, put $\mathbf{w}(t) = M^{-1}\mathbf{v}(t)$. Then $\mathbf{w}(t)$ satisfies the system $\dot{\mathbf{w}}(t) = D\mathbf{w}(t)$, where $D = \text{diag}(1, 3)$.

The solution is $w_1(t) = c_1 e^t$, $w_2(t) = c_2 e^{3t}$, with $c_1$ and $c_2$ constant; and then using $\mathbf{v}(t) = M\mathbf{w}(t)$ gives

$$x(t) = 2c_1 e^t + c_2 e^{3t},$$
$$y(t) = c_1 e^t + c_{,} e^{3t}.$$

$\triangle$

**Proposition 8.5.6.** *Eigenvectors that correspond to distinct eigenvalues of $A$ are linearly independent. In particular, if all eigenvalues of $A$ are distinct, then $A$ is diagonalisable.*

**Example 8.5.7.** Let us try to diagonalise

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

*Solution.* $p(t) = t^2 + 1$, so the eigenvalues are $\lambda_1 = i, \lambda_2 = -i$. Then we get

*Using the fact that $A$ is similar to a diagonal matrix and that similar matrices have the same eigenvalues, so $A \sim B$ with $B = MAM^{-1}$ is equivalent to $M^{-1}BM = A$.*

$$M = \begin{pmatrix} -2 + i & -2 - i \\ 5 & 5 \end{pmatrix}, \quad M^{-1} = \frac{1}{10}\begin{pmatrix} -5i & 1 - 2i \\ 5i & 1 + 2i \end{pmatrix},$$

and

$$M^{-1}AM = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

$\triangle$

**Example 8.5.8.** If

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

derive a formula for $A^n$.

*Solution.* Note that $A = MDM^{-1}$, $D$ is the diagonal matrix, so after we get that $p_A(t) = (t-3)(t-1)$ and $\lambda_1 = 1, \lambda_2 = 3$, we can see that

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix},$$

Finding the eigenvectors using the usual method we get

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad M^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

From here we can see that

$$
\begin{aligned}
A = MDM^{-1} \iff A^n &= (MDM^{-1})(MDM^{-1})\cdots(MDM^{-1}) \\
&= (M)D(M^{-1}M)D(M^{-1}M)\cdots(M^{-1}M)D(M^{-1}) \\
&= MD^n M^{-1} \\
&= \begin{pmatrix} 2 - 3^n & 2 \cdot 3^n - 2 \\ 1 - 3^n & 2 \cdot 3^n - 1 \end{pmatrix}.
\end{aligned}
$$

$\triangle$

# 9 Inner-product spaces

## 9.1 Real bilinear forms and inner products

The dot product is linear in both "entries" (we will say bilinear), i.e.

$$(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) \cdot \mathbf{w} = \alpha \mathbf{v}_1 \cdot \mathbf{w} + \beta \mathbf{v}_2 \cdot \mathbf{w}$$
$$\mathbf{v} \cdot (\alpha \mathbf{w}_1 + \beta \mathbf{w}_2) = \alpha \mathbf{v} \cdot \mathbf{w}_1 + \beta \mathbf{v} \cdot \mathbf{w}_2.$$

**Definition 9.1.1.** A bilinear form $B$ on a real vector space $V$ is a mapping $B : V \times V \to \mathbb{R}$ assigning to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a real number $B(\mathbf{u}, \mathbf{v})$ satisfying the following for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\lambda \in \mathbb{R}$:

(i) $B(\mathbf{u} + \mathbf{v}, \mathbf{w}) = B(\mathbf{u}, \mathbf{w}) + B(\mathbf{v}, \mathbf{w})$,

(ii) $B(\mathbf{w}, \mathbf{u} + \mathbf{v}) = B(\mathbf{w}, \mathbf{u}) + B(\mathbf{w}, \mathbf{v})$,

(iii) $B(\lambda \mathbf{u}, \mathbf{v}) = B(\mathbf{u}, \lambda \mathbf{v}) = \lambda B(\mathbf{u}, \mathbf{v})$.

**Definition 9.1.2.** An inner product on a real vector space $V$ is a map $(\cdot, \cdot) :$ $V \times V \mapsto \mathbb{R}$ which assigns to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a real number $(\mathbf{u}, \mathbf{v})$, satisfying the following: for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\lambda \in \mathbb{R}$

(i) $(\mathbf{u} + \mathbf{v}, \mathbf{w}) = (\mathbf{u}, \mathbf{w}) + (\mathbf{v}, \mathbf{w})$,

(ii) $(\lambda \mathbf{u}, \mathbf{v}) = \lambda (\mathbf{u}, \mathbf{v})$;

(iii) $(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})$ (symmetry);

(iv) $(\mathbf{v}, \mathbf{v}) \geq 0$ with $(\mathbf{v}, \mathbf{v}) = 0 \iff \mathbf{v} = 0$ (positivity).

**Definition 9.1.3.** A real inner-product space is a real vector space equipped with an inner product.

**Example 9.1.4.** $V = \mathbb{R}^n$ with the standard Euclidean inner product (dot product)

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n.$$

**Example 9.1.5.** On $V = \mathbb{R}^2$, is the expression $(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_1 y_2 + x_2 y_2$ an inner product?

*Solution.* No as $\mathbf{x} = (1, 1)$ and $\mathbf{y} = (1, 0)$, then $(\mathbf{x}, \mathbf{y}) = 1$, and $(\mathbf{y}, \mathbf{x}) = 2$. Hence this does not satisfy symmetry. $\triangle$

**Example 9.1.6.** On $V = \mathbb{R}^2$, is the expression $(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_1 y_2 + x_2 y_1 + x_2 y_2$ an inner product?

*Solution.* Both symmetry and bilinearity hold, however positivity does not. Look at $(\mathbf{x}, \mathbf{x}) = (x_1 + x_2)^2$, so we can just have $x_1 = -x_2$ or vice versa, for example $\mathbf{x} = (2, -2)$, $(\mathbf{x}, \mathbf{x}) = (2 - 2)^2 = 0$ but $\mathbf{x} \neq 0$. $\triangle$

**Definition 9.1.7.** We refer to the matrix $A_{ij} := B(\mathbf{v}_j, \mathbf{v}_i)$ as the matrix form of $B$ in the basis $\{\mathbf{v}_i\}$.

**Definition 9.1.8.** We say that an $n \times n$ matrix $B$ is symmetric if

$$B^t = B,$$

and anti-symmetric (or skew-symmetric) if

$$B^t = -B.$$

**Definition 9.1.9.** Every matrix $B$ can be decomposed into its symmetric part $B_+$ and anti-symmetric part $B_-$ simply via

$$B = \frac{B + B^t}{2} + \frac{B - B^t}{2} = B_+ + B_-,$$

where $B_+ := \frac{B + B^t}{2}$ and $B_- := \frac{B - B^t}{2}$. Clearly $B_+$ is a symmetric matrix and $B_-$ is an anti-symmetric matrix.

**Definition 9.1.10.** A symmetric $n \times n$ matrix $B$ is positive-definite if

$$\mathbf{v}^t B \mathbf{v} > 0$$

for all non-zero vectors $\mathbf{v} \in \mathbb{R}^n$.

**Proposition 9.1.11.** *A symmetric matrix $B$ is positive-definite if and only if its eigenvalues are all positive.*

*Proof.* " $\implies$ ". For each eigenvalue $\lambda_i$ choose an eigenvector $\mathbf{v}_i$ with eigenvalue $\lambda_i$. For a positive-definite matrix we then have

$$0 < \mathbf{v}_i^t B \mathbf{v}_i = \lambda_i \mathbf{v}_i^t \mathbf{v}_i = \lambda_i \mathbf{v}_i \cdot \mathbf{v}_i.$$

But $\mathbf{v}_i \cdot \mathbf{v}_i > 0$, so we conclude that $\lambda_i > 0$.

" $\impliedby$ ". Given a real $n \times n$ symmetric matrix $B$, we can form a basis for $\mathbb{R}^n$ formed by eigenvectors $\mathbf{v}_i$ of $B$, with eigenvalue $\lambda_i$, and such that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$. If we express an arbitrary non-zero vector $\mathbf{v}$ in this basis as $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{v}_i$ (with at least one $c_i \neq 0$) we have

$$
\begin{aligned}
\mathbf{v}^t B \mathbf{v} &= \left( \sum_{i=1}^n c_i \mathbf{v}_i \right) \cdot \left( B \sum_{j=1}^n c_j \mathbf{v}_j \right) \\
&= \left( \sum_{i=1}^n c_i \mathbf{v}_i \right) \cdot \left( \sum_{j=1}^n c_j \lambda_j \mathbf{v}_j \right) \\
&= \sum_{i=1}^n c_i^2 (\mathbf{v}_i \cdot \mathbf{v}_i) \lambda_i.
\end{aligned}
$$

where in going from the second to the third lines we have used $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$. The last line is greater than 0 if all the eigenvalues are positive. $\square$

**Proposition 9.1.12.** *An alternative, equivalent, criterion for a real symmetric matrix $B$ to be positive-definite is that all $n$ of its upper-left square submatrices,*

*i.e. the upper left* 1*-by-*1 *corner of B, the upper left* 2*-by-*2 *corner of B, etc, B itself, have positive determinant.*

*This is called Sylvester's criterion, and these subdeterminants are called the leading principle minors of B.*

**Example 9.1.13.** We have

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

is this positive-definite?

*Solution.* No, the subdeterminants are

$$\det \begin{pmatrix} 1 \end{pmatrix} = 1 \qquad \text{and} \qquad \det B = \det \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 1 - 1 = 0$$

so $B$ is not positive-definite by the previous proposition. $\triangle$

**Proposition 9.1.14.** *Consider a bilinear form B on $\mathbb{R}^n$, and denote by A its matrix form for some given choice of basis for $\mathbb{R}^n$. B defines an inner product on $\mathbb{R}^n$ if and only if its matrix representation A is a positive-definite symmetric matrix.*

*Proof.* $\square$

## 9.2 Hermitian inner products

**Definition 9.2.1.** A complex (or hermitian) inner product on a complex vector space $V$ assigns to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a complex number $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfying the following for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\lambda \in \mathbb{C}$:

(i) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$;

(ii) $\langle \lambda \mathbf{u}, \mathbf{v} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle$;

(iii) $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ (hermiticity);

(iv) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ with $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \iff \mathbf{v} = 0$ (positivity).

**Remark 9.2.2.** *We will use $\overline{z}$ to denote the complex conjugate of the complex number $z$.*

**Example 9.2.3.** $V = \mathbb{C}^2$ and $\langle \mathbf{z}, \mathbf{w} \rangle = 3z_1\overline{w}_1 + 2z_2\overline{w}_2 + iz_1\overline{w}_2 - iz_2\overline{w}_1$. Hermiticity is satisfied, the linearity properties are satisfied since the right hand side is linear in $z_1, z_2$. For positivity, we have

$$\begin{aligned} \langle \mathbf{z}, \mathbf{z} \rangle &= 3z_1\overline{z}_1 + 2z_2\overline{z}_2 + iz_1\overline{z}_2 - iz_2\overline{z}_1 \\ &= 2|z_1|^2 + |z_2|^2 + |z_1 - iz_2|^2 \\ &\geq 0, \end{aligned}$$

with equality if and only if $z_1 = z_2 = 0$. So positivity is satisfied as well.

**Example 9.2.4.** $V = \mathbb{C}^2$ and $\langle \mathbf{z}, \mathbf{w} \rangle = z_1 \overline{w}_1 + i z_1 \overline{w}_2 - i z_2 \overline{w}_1 + z_2 \overline{w}_2$ satisfies linearity and hermiticity, but positivity fails, since $\langle \mathbf{z}, \mathbf{z} \rangle = |z_1 - i z_2|^2$ which is zero for $\mathbf{z} = (1, -i)$.

**Matrix version of the hermitian product**

**Definition 9.2.5.** The Hermitian transpose (or complex-conjugate transpose) $B^*$ of a matrix $B$ is the matrix with components

$$(B^*)_{ij} = \overline{B_{ji}}.$$

**Remark 9.2.6.** *If we denote by $\overline{B}$ the matrix such that*

$$(\overline{B})_{ij} = \overline{B_{ij}}$$

*(that is, the matrix whose entries are the complex conjugates of those of $B$) then*

$$B^* = (\overline{B})^t = \overline{B^t}.$$

**Remark 9.2.7.** *If $A$ and $B$ are matrices, then $(AB)^* = \overline{(AB)^t} = \overline{B^t A^t} = B^* A^*$.*

**Proposition 9.2.8.** *If $V = \mathbb{C}^n$ with basis $\{\mathbf{v}_i\}$, then an inner product on $V$ is realised by a $n \times n$ matrix $B$ as*

$$\langle \mathbf{u}, \mathbf{w} \rangle = \mathbf{y}^* B \mathbf{x}$$

*where $\mathbf{x}$ and $\mathbf{y}$ are the vector of components of $\mathbf{u}$ and $\mathbf{w}$ in the $\{\mathbf{v}_i\}$ basis (so $\mathbf{u} = x_1 \mathbf{v}_1 + \ldots + x_n \mathbf{v}_n$ and $\mathbf{w} = y_1 \mathbf{v}_1 + \ldots + y_n \mathbf{v}_n$) and $B$ is the matrix with components*

$$B_{ij} = \langle \mathbf{v}_j, \mathbf{v}_i \rangle.$$

*Proof.* The argument is analogous to the one in the real case. That $B$ has the components given in the proposition follows from expanding $\mathbf{u}, \mathbf{w} \in V$ in the $\{\mathbf{v}_i\}$ basis, and using (sesqui)linearity. That the resulting matrix $B$ is hermitian and positive-definite follows from the definitions below. $\qquad \square$

**Definition 9.2.9.** We say that a matrix $B$ is hermitian if $B^* = B$. Similarly $B$ is an anti-hermitian matrix if $B^* = -B$.

**Proposition 9.2.10.** *$A$ is Hermitian if and only if $\mathbf{y}^* A \mathbf{x} = (A\mathbf{y}) * \mathbf{x}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. If we choose the standard inner product on $\mathbb{C}^n$, this can be written as $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.*

*Proof.* This follows by choosing $\mathbf{x}$ and $\mathbf{y}$ the standard basis vectors of $\mathbb{C}^n$. For instance, choosing $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$ we have

$$\langle \mathbf{e}_i, A\mathbf{e}_j \rangle = \langle \mathbf{e}_i, \sum_{k=1}^{n} A_{kj} \mathbf{e}_k \rangle = \overline{A_{ij}}$$

and similarly $\langle A\mathbf{e}_i, \mathbf{e}_j \rangle = A_{ji}$. $\qquad \square$

**Definition 9.2.11.** We say that a hermitian matrix $B$ is positive-definite if

$$\mathbf{v}^* B \mathbf{v} > 0$$

for all non-zero $\mathbf{v} \in \mathbb{C}^n$.

**Proposition 9.2.12.** *A hermitian matrix has real eigenvalues.*

*Proof.* □

**Proposition 9.2.13.** *A hermitian matrix is positive-definite if and only if its eigenvalues are all positive.*

*Proof.* The proof is entirely analogous to the one in the real case, just by replacing $\mathbf{v}^t$ by $\mathbf{v}^*$ everywhere. □

**Proposition 9.2.14** (Sylvester's criterion). *A hermitian matrix $B$ is positive-definite if and only if the determinants of all of its leading minors are positive.*

## 9.3 Norms, orthogonality and the Cauchy-Schwarz inequality

**Definition 9.3.1.** Given a vector space $V$ over $\mathbb{R}$ (or $\mathbb{C}$), a norm on $V$ is a function usually denoted by $\| \cdot \| : V \to \mathbb{R}$ with the following properties

(i) $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$ (absolute homogeneity),

(ii) $\|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{v}\| + \|\mathbf{u}\|$ (triangle inequality or subadditivity),

(iii) If $\|\mathbf{v}\| = 0$ then $\mathbf{v} = \mathbf{0}$ (separation of points),

for all $a \in \mathbb{R}$ (or $\mathbb{C}$) and all $\mathbf{v}, \mathbf{u} \in V$.

**Definition 9.3.2.** If $(V, (\,,\,))$ is a real inner product space, then we define the norm $\|\mathbf{v}\|$ of a vector $\mathbf{v} \in V$ by

$$\|\mathbf{v}\| := \sqrt{(\mathbf{v}, \mathbf{v})}.$$

**Definition 9.3.3.** A vector $\mathbf{v} \in V$ is a unit vector if $\|\mathbf{v}\| = 1$.

**Definition 9.3.4.** Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $(\mathbf{u}, \mathbf{v}) = 0$.

**Definition 9.3.5.** An orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $V$ is a basis consisting of mutually orthogonal unit vectors. For the bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ to be orthonormal we then must have

$$(\mathbf{u}_i, \mathbf{u}_i) = \|\mathbf{u}_i\|^2 = 1, \text{ for } i = 1, \ldots, n, \qquad (\mathbf{u}_i, \mathbf{u}_j) = 0, \text{ when } i \neq j,$$

these conditions can be written in the compact form

$$(\mathbf{u}_i, \mathbf{u}_j) = \delta_{ij},$$

with $\delta_{ij}$ denoting Kronecker's delta.

**Theorem 9.3.6** *(The real Cauchy-Schwarz inequality).* *We have*

$$(\mathbf{u}, \mathbf{v})^2 \leq (\mathbf{u}, \mathbf{u})(\mathbf{v}, \mathbf{v})$$

*with equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. Equivalently, $|(\mathbf{u}, \mathbf{v})| \leq \|\mathbf{u}\|\|\mathbf{v}\|$.*

*Proof.* For $\mathbf{u} = \mathbf{0}$ the proof is trivial, suppose $\mathbf{u} \neq \mathbf{0}$. For all real numbers $x$ we have

$$\|x\mathbf{u} - \mathbf{v}\|^2 = x^2\|\mathbf{u}\|^2 - 2x(\mathbf{u}, \mathbf{v}) + \|\mathbf{v}\|^2 \geq 0,$$

with equality if and only if $x\mathbf{u} - \mathbf{v} = 0$.

If $\mathbf{u}, \mathbf{v}$ are linearly independent then there is no $x$ such that $x\mathbf{u} = \mathbf{v}$ then the quadratic equation in $x$ we just formed would have a discriminant less than zero

$$(\mathbf{u}, \mathbf{v})^2 - \|\mathbf{u}\|\|\mathbf{v}\|^2 < 0.$$

Alternatively, if they are linearly dependent, from linearity of the inner product we can see that

$$(\mathbf{u}, \mathbf{v})^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2 = 0.$$

$\square$

**Corollary 9.3.7** (The triangle inequality). *For all $\mathbf{u}, \mathbf{v} \in V$, we have*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

*Proof.*

$$\begin{aligned}
(\|\mathbf{u}\| + \|\mathbf{v}\|)^2 - \|\mathbf{u} + \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 - \|\mathbf{u}\|^2 - 2(\mathbf{u}, \mathbf{v}) - \|\mathbf{v}\|^2 \\
&= 2(\|\mathbf{u}\|\|\mathbf{v}\| - (\mathbf{u}, \mathbf{v})) \\
&\geq 0.
\end{aligned}$$

$\square$

**Corollary 9.3.8.** *For a real inner-product space, the angle $\theta \in [0, \pi]$ between $\mathbf{u}$ and $\mathbf{v}$, given by $(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$, is well-defined.*

*Proof.* By Cauchy-Schwarz (Theorem 9.3.6) we know that $-\|\mathbf{u}\|\|\mathbf{v}\| \leq (\mathbf{u}, \mathbf{v}) \leq \|\mathbf{u}\|\|\mathbf{v}\|$. Additionally, if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal then $(\mathbf{u}, \mathbf{v}) = 0$ so the angle between them is
$theta = pm\pi/2$. If they are parallel, linearly dependent with $\mathbf{v} = x\mathbf{u}$ with $x > 0$, the angle is $\theta = 0$, if they are anti-parallel, linearly dependent with $\mathbf{v} = x\mathbf{u}$ with $x < 0$, the angle is $\theta = \pi$. $\square$

**The norm induced by a Hermitian inner product**

**Definition 9.3.9.** Let $(V, \langle\,,\,\rangle)$ be a Hermitian inner product space, then we define the norm $\|\mathbf{v}\|$ of a vector $\mathbf{v} \in V$ by

$$\|\mathbf{v}\| \coloneqq \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

**Definition 9.3.10.** A vector $\mathbf{v} \in V$ is a unit vector if $\|\mathbf{v}\| = 1$.

**Definition 9.3.11.** Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

**Theorem 9.3.12** *(The complex Cauchy-Schwarz inequality). Let $(V, \langle\,,\,\rangle)$ be a Hermitian inner product space, then for all vectors $\mathbf{u}, \mathbf{v} \in V$ we have*

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2,$$

*with equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.*

*Proof.* Assume $\mathbf{u} \neq 0$ and $\langle \mathbf{u}, \mathbf{v} \rangle \neq 0$, otherwise the statement is trivial and let us consider the following quadratic expression in $x \in \mathbb{C}$

$$\|x\mathbf{u} - \mathbf{v}\|^2 = |x|^2 \|\mathbf{u}\|^2 - 2\mathrm{Re}(x\langle \mathbf{u}, \mathbf{v} \rangle) + \|\mathbf{v}\|^2 \geq 0,$$

with equality if and only if $x\mathbf{u} - \mathbf{v} = 0$.

Since this expression must be non-negative for all $x \in \mathbb{C}$ we can choose in particular

$$x = \lambda \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{\langle \mathbf{u}, \mathbf{v} \rangle},$$

with $\lambda \in \mathbb{R}$.

The above inequality becomes

$$\lambda^2 \|\mathbf{u}\|^2 - 2\lambda |\langle \mathbf{u}, \mathbf{v} \rangle| + \|\mathbf{v}\|^2 \geq 0,$$

at this point we can apply the same argument used in the real Cauchy-Schwarz inequality (Theorem 9.3.6) $\qquad\square$

## 9.4 The Gram-Schmidt procedure

Our goal of this procedure is to form an orthonormal set of vectors in an inner-product space. Given any set of linearly independent vectors we can form an orthonormal set of vectors. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be linearly independent vectors in an inner-product space $V$. these span a $k$-dimensional subspace $U$ of $V$. Our task is to construct an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_k$ for $U$, The procedure is as follows.

**The Gram-Schmidt procedure**:

(i) Define $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. Then $\mathbf{u}_1$ is a unit vector, and $\mathrm{span}\{\mathbf{u}_1\} = \mathrm{span}\{\mathbf{v}_1\}$.

(ii) Define $\bar{\mathbf{v}}_2 = \mathbf{v}_2 - (\mathbf{v}_2, \mathbf{u}_1)\mathbf{u}_1$. Then $(\mathbf{u}_1, \bar{\mathbf{v}}_2) = 0$ and $\bar{\mathbf{v}}_2 \neq 0$, since $\mathbf{v}_2 \notin \mathrm{span}\{\mathbf{u}_1\}$. Now define $\mathbf{u}_2 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. Then $\mathbf{u}_1, \mathbf{u}_2$ are mutually orthogonal unit vectors, and $\mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ (because clearly $u_1, u_2 \in \mathrm{span}(v_1, v_2)$ and also $v_1, v_2 \in \mathrm{span}(u_1, u_2)$).

(iii) Suppose now that mutually orthogonal unit vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ have been found with $\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ for some $r$ with $1 \leq r \leq k$. If $r < k$, then define

$$\mathbf{v}_{r+1}^- = \mathbf{v}_{r+1} - (\mathbf{v}_{r+1}, \mathbf{u}_1)\mathbf{u}_1 - \ldots - (\mathbf{v}_{r+1}, \mathbf{u}_r)\mathbf{u}_r.$$

Then

$$(\mathbf{v}_{r+1}^-, \mathbf{u}_1) = \cdots = (\mathbf{v}_{r+1}^-, \mathbf{u}_r) = 0,$$

and $\mathbf{v}_{r+1}^- \neq 0$ since $\mathbf{v}_{r+1} \notin \mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. Define $\mathbf{u}_{r+1} = \frac{\mathbf{v}_{r+1}^-}{\|\mathbf{v}_{r+1}^-\|}$. Then $\mathbf{u}_1, \ldots, \mathbf{u}_{r+1}$ are mutually orthogonal unit vectors, and $\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{r+1}\} = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}\}$

**Example 9.4.1.** In $\mathbb{R}^3$ with the standard inner product, apply Gram-Schmidt to

$$\mathbf{v}_1 = (1, 1, 0) \qquad \mathbf{v}_2 = (1, 0, 1) \qquad \mathbf{v}_3 = (0, 1, 1).$$

*Solution.*

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\bar{\mathbf{v}}_2 = \mathbf{v}_2 - (\mathbf{v}_2, \mathbf{u}_1)\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

$$\mathbf{u}_2 = \frac{1}{6} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \bar{\mathbf{v}}_3 = \mathbf{v}_3 - (\mathbf{v}_3,$$

$$= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \frac{2}{3} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},$$

$$\mathbf{u}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}.$$

$\triangle$

## 9.5    Orthogonal complement and orthogonal projection

Let us explore some consequences of orthogonality between vectors in a vector space $V$.

**Definition 9.5.1.** Let $U$ be a vector subspace of an inner product space $V$. Then the orthogonal complement of $U$ is the set

$$U^{\perp} = \{\mathbf{v} \in V \; : \; (\mathbf{u}, \mathbf{v}) = 0 \text{ for all } \mathbf{u} \in U\},$$

of vectors in $V$ which are orthogonal to all the vectors in $U$.

**Remark 9.5.2.** $U^{\perp}$ *is a vector subspace of $V$.*

**Definition 9.5.3.** Given any vector $\mathbf{v} \in V$, there is a unique vector $\mathbf{u} \in U$ called the orthogonal projection of $\mathbf{v}$ onto $U$ (at least if $U$ is finite-dimensional), and one may find this $\mathbf{u}$ explicitly as follows.

**Proposition 9.5.4.** *If $U$ is finite-dimensional, then there always exists a unique decomposition $\mathbf{v} = \mathbf{u} + \tilde{\mathbf{u}}$, where $\mathbf{u} \in U$ and $\tilde{\mathbf{u}} \in U^{\perp}$. so we can write $V = U \oplus U^{\perp}$.*

*Proof.* Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthonormal basis of $U$, and set

$$\mathbf{u} = (\mathbf{v}, \mathbf{u}_1)\mathbf{u}_1 + \ldots + (\mathbf{v}, \mathbf{u}_k)\mathbf{u}_k, \qquad \tilde{\mathbf{u}} = \mathbf{v} - \mathbf{u}.$$

Clearly $\mathbf{u} \in U$; and for each $\mathbf{u}_i$ note that $(\mathbf{u}, \mathbf{u}_i) = (\mathbf{v}, \mathbf{u}_i) \cdot (\mathbf{u}_i, \mathbf{u}_i) = (\mathbf{v}, \mathbf{u}_i)$. So for each $\mathbf{u}_i$ we have

$$\begin{aligned}
(\tilde{\mathbf{u}}, \mathbf{u}_i) &= (\mathbf{v} - \mathbf{u}, \mathbf{u}_i) \\
&= (\mathbf{v}, \mathbf{u}_i) - (\mathbf{u}, \mathbf{u}_i) \\
&= (\mathbf{v}, \mathbf{u}_i) - (\mathbf{v}, \mathbf{u}_i) \\
&= 0.
\end{aligned}$$

Hence $\tilde{\mathbf{u}} \in U^\perp$. This shows that $V = U + U^\perp$. To show that $V = U \oplus U^\perp$ we need to show additionally that $U \cap U^\perp = \{\mathbf{0}\}$. Take $\mathbf{w} = w_1\mathbf{u}_1 + \ldots + w_k\mathbf{u}_k$. Clearly $\mathbf{w} \in U$, and let us assume that also $\mathbf{w} \in U^\perp$. In this case, by the definition of $U^\perp$, we must have that $(\mathbf{w}, \mathbf{u}_i) = 0$ for all $i \in 1, \ldots, k$. Since the $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ form an orthonormal basis, this implies (by linearity of the inner product) $w_i = 0$ for all $i$, and therefore $\mathbf{w} = \mathbf{0}$. To show uniqueness, assume that there is a second decomposition $\mathbf{v} = \mathbf{u} + \tilde{\mathbf{u}} = \mathbf{u}' + \tilde{\mathbf{u}}'$, with $\mathbf{u}' \in U$ and $\tilde{\mathbf{u}} \in U^\perp$. This implies $\mathbf{u}' - \mathbf{u} = \tilde{\mathbf{u}} - \tilde{\mathbf{u}}'$. The left hand side of this equality lives in $U$, while the right hand side lives in $U^\perp$. Since $U \cap U^\perp = \{\mathbf{0}\}$ this implies $\mathbf{u}' - \mathbf{u} = \mathbf{0}$ and $\tilde{\mathbf{u}} - \tilde{\mathbf{u}}' = \mathbf{0}$, so the decomposition is indeed unique. $\qquad\square$

**Definition 9.5.5.** Given a finite-dimensional vector subspace $U$ of the inner product space $(V, (\, , \,))$ we can consider the orthogonal projection operator $P$ onto $U$, defined by

$$P(\mathbf{v}) := (\mathbf{v}, \mathbf{u}_1)\mathbf{u}_1 + \ldots + (\mathbf{v}, \mathbf{u}_k)\mathbf{u}_k,$$

where $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis of $U$.