Bachelor of Science

in Mathematics

Linear algebra I

Author:

Daniel Pérez

Contents

Note of the author

This lecture notes correspond to the spring term of the module *Algebra* MAT00010C. Please note that these notes have not been endorsed by the lecturers, and I have made several modifications to them (often substantial) after the lectures. They should not be considered as accurate representations of the actual lecture content, and it is highly likely that any errors present are solely my responsibility.

CHAPTER 1

Introduction to complex numbers

Definition 1.0.1 (Complex number)**.** *A complex number is an object of the form* $a+bi$*, where* a *and* b *are real numbers and* $i = \sqrt{-1}$*. The set of complex numbers is denoted* $\mathbb{C}$*.*

Famously, the square of any real number is non-negative: that is, $x^2 \geq 0$ for all real x. We define $i = \sqrt{-1}$ to be an object with the property that $i^2 = -1$. One consequence of extending our horizons to include complex numbers is that we can suddenly solve many equations than we could previously.

We have

$$\begin{aligned}
z_1 \pm z_2 &= (a_1 + ib_1) \pm (a_2 + ib_2) \\
&= (a_1 \pm a_2) + i(b_1 \pm b_2) \\
z_1 z_2 &= (a_1 + ib_1)(a_2 + ib_2) \\
&= (a_1 a_2 - b_1 b_2) + i(b_1 a_2 + a_1 b_2) \\
z^{-1} &= \frac{1}{a + ib} \\
&= \frac{a - ib}{a^2 + b^2}
\end{aligned}$$

Definition 1.0.2 (Complex conjugate)**.** *The complex conjugate of $z = a + ib$ is $a - ib$. It is written as $\overline{z}$ or z^*.*

Exercise 1.0.1. *Show that*

(i) $\overline{\overline{z}} = z$

(ii) $\overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}$

(iii) $\overline{z_1 z_2} = \overline{z_1 z_2}$

(iv) $\overline{z^{-1}} = \overline{z}^{-1}$

Definition 1.0.3 (Modulus)**.** *The modulus of $z = a + ib$, which is written as $|z|$, is defined as*

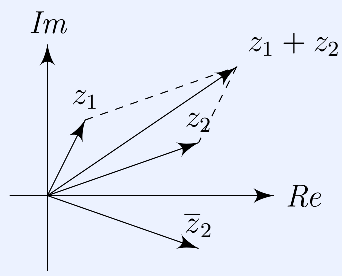
$$|z| = (a^2 + b^2)^{1/2} \tag{1.1}$$

Exercise 1.0.2. *Show that*

(i) $|z|^2 = z\overline{z}$

(ii) $z^{-1} = \frac{\overline{z}}{|z|^2}$

Definition 1.0.4 (Argand diagram)**.** *An Argand diagram is a diagram in which a complex number $z = x + iy$ is represented by a vector $\mathbf{p} = \begin{pmatrix} x \\ y \end{pmatrix}$. Addition of vectors corresponds to vector addition and $\overline{z}$ is the reflection of z in the x-axis.*



Definition 1.0.5 (Argument). *The* argument *is* $\theta = \arg z = \tan^{-1}(y/x)$. *The modulus is the length of the vector in the Argand diagram, and the argument is the angle between z and the real axis. We have*

$$z = r(\cos\theta + i\sin\theta)$$

Clearly the pair (r, θ) uniquely describes a complex number z, but each complex number $z \in \mathbb{C}$ can be described by many different θ since $\sin(2\pi + \theta) = \sin\theta$ *and* $\cos(2\pi + \theta) = \cos\theta$. *Often we take the principle value $\theta \in (-\pi, \pi]$.*

When writing $z_i = r_i(\cos\theta_i + i\sin\theta_i)$, we have

$$\begin{aligned} z_1 z_2 &= r_1 r_2[(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1)] \\ &= r_1 r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)] \end{aligned}$$

In other words, when multiplying complex numbers, the moduli multiply and the arguments add.

Theorem 1.0.1 (Triangle inequality). *For all $z_1, z_2 \in \mathbb{C}$, we have*

$$|z_1 + z_2| \leq |z_1| + |z_2|.$$

Alternatively, we have $|z_1 - z_2| \geq ||z_1| - |z_2||$.

Proof. Self-evident by geometry. Alternatively, by the cosine rule,

$$\begin{aligned} |z_1 + z_2| &= |z_1|^2 + |z_2|^2 - 2|z_1||z_2|\cos\phi \\ &\leq |z_1|^2 + |z_2|^2 + 2|z_1||z_2| \\ &= (|z_1| + |z_2|)^2 \end{aligned}$$

$\square$

1.1 Complex exponential function

Definition 1.1.1 (Exponential function)**.** *The exponential function is defined as*

$$\exp(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This series converges for all $x \in \mathbb{R}$ (see *Real Analysis I* course).

Theorem 1.1.1. $\exp(z_1)\exp(z_2) = \exp(z_1 + z_2)$

Proof.

$$\begin{aligned}
\exp(z_1)\exp(z_2) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{z_1^m}{m!}\frac{z_2^n}{n!} \\
&= \sum_{r=0}^{\infty}\sum_{m=0}^{r} \frac{z_1^{r-m}}{(r-m)!}\frac{z_2^m}{m!} \\
&= \sum_{r=0}^{\infty}\frac{1}{r!}\sum_{m=0}^{r} \frac{r!}{(r-m)!m!} z_1^{r-m} z_2^m \\
&= \sum_{r=0}^{\infty}\frac{(z_1+z_2)^r}{r!}
\end{aligned}$$

$\square$

Definition 1.1.2. *We write*

$$\exp 1 = e \tag{1.2}$$

Exercise 1.1.1. *Show for $n, p, q \in \mathbb{Z}$, where without loss of generality $q > 0$, that:*

$$e^n = \exp(n) \qquad e^{\frac{p}{q}} = \exp\left(\frac{p}{q}\right)$$

Proof. For $n = 1$ there is nothing to prove. For $n \geq 2$, and using the last theorem,

$$\exp(n) = \exp(1)\exp(n-1) = e\exp(n-1)$$

and thence by induction $\exp(n) = e^(n)$. From the power series definition with $n = 0$: $\exp(0) = 1 = e^0$. Therefore,

$$\exp(-1)\exp(1) = \exp(0)$$

and thence $\exp(-1) = \frac{1}{e} = e^{-1}$. For $n \leq 2$, proceed by induction as above. Next note that

$$\left(\exp\frac{p}{q}\right)^q = \exp(p) = e^p$$

Thence on taking the positive qth root,

$$\exp\left(\frac{p}{q}\right) = e^{\frac{p}{q}}$$

$\square$

Definition 1.1.3 (Complex xponential function). *The complex exponential function is defined as*

$$\exp(z) = e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

for $z \in \mathbb{C}$ and $z \notin \mathbb{R}$.

This automatically allows taking exponents of arbitrary complex numbers. Having defined exponentiation this way, we want to check that it satisfies the usual properties, such as $\exp(z + w) = \exp(z)\exp(w)$. To prove this, we will first need a helpful lemma.

Lemma 1.1.1.

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} = \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}$$

Proof.

$$\begin{aligned}
\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} &= a_{00} + a_{01} + a_{02} + \cdots \\
&+ a_{10} + a_{11} + a_{12} + \cdots \\
&+ a_{20} + a_{21} + a_{22} + \cdots \\
&= (a_{00}) + (a_{10} + a_{01}) + (a_{20} + a_{11} + a_{02}) + \cdots \\
&= \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}
\end{aligned}$$

□

This is not exactly a rigorous proof, since we should not hand-wave about infinite sums so casually. But in fact, we did not even show that the definition of $\exp(z)$ is well defined for all numbers z, since the sum might diverge. All these will be done in the *Real Analysis I* course.

1.2 The complex trigonometric functions

Definition 1.2.1 (Sine and cosine functions)**.** *Define, for all $z \in \mathbb{C}$,*

$$\begin{aligned}
\sin z &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} &= z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdots \\
\cos z &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} &= 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots
\end{aligned}$$

One very important result is the relationship between exp, sin and cos.

Theorem 1.2.1. *$e^{iz} = \cos z + i \sin z$.*

Alternatively, since $\sin(-z) = -\sin z$ and $\cos(-z) = \cos z$, we have

$$\begin{aligned}
\cos z &= \frac{e^{iz} + e^{-iz}}{2}, \\
\sin z &= \frac{e^{iz} - e^{-iz}}{2i}.
\end{aligned}$$

Proof.

$$\begin{aligned} e^{iz} &= \sum_{n=0}^{\infty} \frac{i^n}{n!} z^n \\ &= \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} z^{2n} + \sum_{n=0}^{\infty} \frac{i^{2n+1}}{(2n+1)!} z^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\ &= \cos z + i \sin z \end{aligned}$$

□

Thus we can write $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$.

1.3 Roots of unity

Definition 1.3.1 (Roots of unity). *A root of unity is a solution of $z^n = 1$, with $z \in \mathbb{C}$ and n a positive integer.*

Theorem 1.3.1. *There are n solutions of $z^n = 1$*

Proof. One solution is $z = 1$. Seek more general solutions of the form $z = re^{i\theta}$, with the restriction that $0 \leq \theta < 2\pi$ so that θ is not multi-valued. Then,

$$(re^{i\theta})^n = r^n (e^{i\theta})^n = r^n e^{in\theta} = 1$$

hence, $r^n = 1$ and $n\theta = 2k\pi$ with $k \in \mathbb{Z}$. We conclude that within the requirement that $0 \leq \theta < 2\pi$, there are n distinct roots given by $r = 1$ and $\theta = \frac{2k\pi}{n}$ with $k = 0, 1, \ldots, n-1$. □

Proposition 1.3.1. *If $\omega = \exp\left(\frac{2\pi i}{n}\right)$, then $1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$*

Proof. Two proofs are provided:

1. Consider the equation $z^n = 1$. The coefficient of z^{n-1} is the sum of all roots. Since the coefficient of z^{n-1} is 0, then the sum of all roots $= 1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$.

2. Since $\omega^n - 1 = (\omega - 1)(1 + \omega + \cdots + \omega^{n-1})$ and $\omega \neq 1$, dividing by $(\omega - 1)$, we have $1 + \omega + \cdots + \omega^{n-1} = (\omega^n - 1)/(\omega - 1) = 0$. □

1.4 Logarithm and complex powers

We know already that if $x \in \mathbb{R}$ and $x > 0$, the equation $e^y = x$ has a unique real solution, namely $y = \log x$.

Definition 1.4.1 (Complex logarithm)**.** *The* complex logarithm $w = \log z$ *is a solution to* $e^\omega = z$*, i.e.* $\omega = \log z$*. Writing* $z = re^{i\theta}$*, we have* $\log z = \log(re^{i\theta}) = \log r + i\theta$*. This can be multi-valued for different values of* θ *and, as above, we should select the* θ *that satisfies* $-\pi < \theta \leq \pi$*.*

(i) By definition

$$\exp(\log(z)) = z \tag{1.3}$$

(ii) Let $y = \log(z)$ and take the logarithm of both sides of the equation above to conclude that

$$\begin{aligned}\log(\exp(y)) &= \log(\exp(\log(z)))\\ &= \log(z)\\ &= y\end{aligned}$$

To understand the nature of the complex logarithm let $w = u + iv$ with $u, v \in \mathbb{R}$. Then, $e^{u+iv} = e^w = z = re^{i\theta}$, and hence $e^u = |z| = r$ and $v = \arg z = \theta + 2k\pi$ for any $k \in \mathbb{Z}$. Thus

$$\log(z) = w = u + iv = \log|z| + i \arg z \tag{1.4}$$

Since $\arg z$ is a multi-valued function, so is $\log z$.

Example 1.4.1. $\log 2i = \log 2 + i\frac{\pi}{2}$

Definition 1.4.2 (Principal value)**.** *The principal value of* $\log z$ *is such that*

$$-\pi < \arg(z) = \Im(\log z) \leq \pi \tag{1.5}$$

Example 1.4.2. *If $z = -x$ with $x \in \mathbb{R}$ and $x > 0$, then*

$$\begin{aligned}\log(z) &= \log|-x| + i\arg(-x)\\ &= \log|x| + (2k+1)i\pi\end{aligned}$$

for any $k \in \mathbb{Z}$. The principal value of $\log(-x)$ is $\log|x| + i\pi$.

Definition 1.4.3 (Complex power)**.** *The complex power z^α for $z, \alpha \in \mathbb{C}$ is defined as $z^\alpha = e^{\alpha \log z}$. This, again, can be multi-valued, as $z^\alpha = e^{\alpha\log|z|}e^{i\alpha\theta}e^{2in\pi\alpha}$ (there are finitely many values if $\alpha \in \mathbb{Q}$, infinitely many otherwise). Nevertheless, we make z^α single-valued by insisting $-\pi < \theta \leq \pi$.*

Example 1.4.3. *(i) For $a, b \in \mathbb{C}$, it follows that*

$$z^{ab} = \exp(ab\log(z)) = \exp(b(a\log(z)) = y^b$$

where $\log(y) = a\log(z)$. But from the definition of the logarithm, $\exp(\log(z)) = z$. Hence,

$$y = \exp(a\log(z)) = z^a$$

and thus

$$z^{ab} = (z^a)^b = (z^b)^a$$

(ii) The value of i^i is given by

$$\begin{aligned}i^i &= e^{i\log(i)}\\ &= e^{i(\log|i| + i\arg i)}\\ &= e^{i(\log 1 + 2ki\pi + i\pi/2)}\\ &= e^{-(2k+\frac{1}{2})\pi} \in \mathbb{R}\end{aligned}$$

1.5 De Moivre's theorem

Theorem 1.5.1 (De Moivre's theorem)**.**

$$\cos n\theta + i \sin n\theta = (\cos\theta + i\sin\theta)^n.$$

Proof. First prove for the $n \geq 0$ case by induction. The $n = 0$ case is true since it merely reads $1 = 1$. We then have

$$\begin{aligned}
(\cos\theta + i\sin\theta)^{n+1} &= (\cos\theta + i\sin\theta)^n(\cos\theta + i\sin\theta) \\
&= (\cos n\theta + i\sin n\theta)(\cos\theta + i\sin\theta) \\
&= \cos(n+1)\theta + i\sin(n+1)\theta
\end{aligned}$$

If $n < 0$, let $m = -n$. Then $m > 0$ and

$$\begin{aligned}
(cos\theta + i\sin\theta)^{-m} &= (\cos m\theta + i\sin m\theta)^{-1} \\
&= \frac{\cos m\theta - i\sin m\theta}{(\cos m\theta + i\sin m\theta)(\cos m\theta - i\sin m\theta)} \\
&= \frac{\cos(-m\theta) + i\sin(-m\theta)}{\cos^2 m\theta + \sin^2 m\theta} \\
&= \cos(-m\theta) + i\sin(-m\theta) \\
&= \cos n\theta + i\sin n\theta
\end{aligned}$$

□

Note that "$\cos n\theta + i\sin n\theta = e^{in\theta} = (e^{i\theta})^n = (\cos\theta + i\sin\theta)^n$" is not a valid proof of De Moivre's theorem, since we do not know yet that $e^{in\theta} = (e^{i\theta})^n$. In fact, De Moivre's theorem tells us that this is a valid rule to apply.

Example 1.5.1. *We have* $\cos 5\theta + i\sin 5\theta = (\cos\theta + i\sin\theta)^5$*. By binomial expansion of the RHS and taking real and imaginary parts, we have*

$$\cos 5\theta = 5\cos\theta - 20\cos^3\theta + 16\cos^5\theta$$
$$\sin 5\theta = 5\sin\theta - 20\sin^3\theta + 16\sin^5\theta$$

1.6 Lines and circles in the complex plane

Suppose that we want to represent a straight line through $z_0 \in \mathbb{C}$ parallel to $w \in \mathbb{C}$. The obvious way to do so is to let $z = z_0 + \lambda w$ where λ can take any real value. However, this is not an optimal way of doing so, since we are not using the power of complex numbers fully. This is just the same as the vector equation for straight lines.

Instead, we arrange the equation to give $\lambda = \frac{z-z_0}{w}$. We take the complex conjugate of this expression to obtain $\overline{\lambda} = \frac{\overline{z}-\overline{z_0}}{\overline{w}}$. The trick here is to realise that λ is a real number. So we must have $\lambda = \overline{\lambda}$. This means that we must have

$$\frac{z - z_0}{w} = \frac{\overline{z} - \overline{z_0}}{\overline{w}}$$

$$z\overline{w} - \overline{z}w = z_0\overline{w} - \overline{z}_0 w.$$

Theorem 1.6.1 (Equation of straight line). *The equation of a straight line through z_0 and parallel to w is given by*

$$z\overline{w} - \overline{z}w = z_0\overline{w} - \overline{z}_0 w.$$

The equation of a circle, on the other hand, is rather straightforward. Suppose that we want a circle with centre $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}^+$. By definition of a circle, a point z is on the circle iff its distance to c is ρ, i.e. $|z - c| = \rho$. Recalling that $|z|^2 = z\overline{z}$, we obtain,

$$\begin{aligned}
|z - c| &= \rho \\
|z - c|^2 &= \rho^2 \\
(z - c)(\overline{z} - \overline{c}) &= \rho^2 \\
z\overline{z} - \overline{c}z - c\overline{z} &= \rho^2 - c\overline{c}
\end{aligned}$$

Theorem 1.6.2 (Equation of a circle). *The general equation of a circle with centre $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}^+$ can be given by*

$$z\overline{z} - \overline{c}z - c\overline{z} = \rho^2 - c\overline{c}.$$

CHAPTER 2

Vector spaces

Definition 2.0.1 (Vector)**.** *A quantity that is specified by a [positive] magnitude and a direction in space.*

Definition 2.0.2 (Vector space)**.** *A vector space over $\mathbb{R}$ or $\mathbb{C}$ is a collection of vectors $\mathbf{v} \in V$, together with two operations: addition of two vectors and multiplication of a vector with a scalar (i.e. a number from $\mathbb{R}$ or $\mathbb{C}$, respectively).* Vector addition *has to satisfy the following axioms:*

(i) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ *(commutativity)*

(ii) $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ *(associativity)*

(iii) There is a vector $\mathbf{0}$ *such that* $\mathbf{a} + \mathbf{0} = \mathbf{a}$. *(identity)*

(iv) For all vectors $\mathbf{a}$*, there is a vector* $(-\mathbf{a})$ *such that* $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ *(inverse)*

Scalar multiplication *has to satisfy the following axioms:*

(i) $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

(ii) $(\lambda + \mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$.

(iii) $\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}$.

(iv) $1\mathbf{a} = \mathbf{a}$.

Note that $\lambda\mathbf{a}$ is either parallel ($\lambda \geq 0$) to or anti-parallel ($\lambda \leq 0$) to $\mathbf{a}$.

The magnitude of a vector $\mathbf{v}$ is written $|\mathbf{v}|$ or v. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are equal if they have the same magnitude, i.e. $|\mathbf{u}| = |\mathbf{v}|$, and they are in the same direction, i.e. $\mathbf{u}$ is parallel to $\mathbf{v}$ and in both vectors are in the same direction/sense.

Example 2.0.1. *Every complex number corresponds to a unique point in the complex plane, and hence to the position vector of that point.*

Definition 2.0.3 (Unit vector)**.** *A* unit vector *is a vector with length 1. We write a unit vector as* $\hat{\mathbf{v}}$.

Example 2.0.2. $\mathbb{R}^n$ *is a vector space with component-wise addition and scalar multiplication. Note that the vector space* $\mathbb{R}$ *is a line, but not all lines are vector spaces. For example,* $x + y = 1$ *is not a vector space since it does not contain* $\mathbf{0}$.

2.1 Scalar product

In a vector space, we can define the scalar product of two vectors, which returns a scalar (i.e. a real or complex number). We will first look at the usual scalar product defined for $\mathbb{R}^n$, and then define the scalar product axiomatically.

Definition 2.1.1 (Scalar product)**.** *The scalar product of two vectors* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *is defined geometrically to be the real (scalar) number*

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}| \cos\theta \tag{2.1}$$

where $0 \leq \theta \leq \pi$ *is the non-reflex angle between* $\boldsymbol{a}$ *and* $\boldsymbol{b}$.

The scalar product satisfies the following properties:

(i) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$

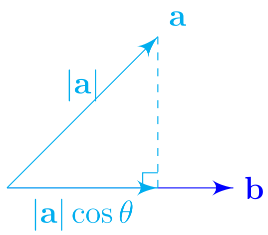
(ii) $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2 \geq 0$

(iii) $\mathbf{a} \cdot \mathbf{a} = 0$ iff $\mathbf{a} = \mathbf{0}$

(iv) If $0 \leq \theta < \frac{1}{2}\pi$, then $\mathbf{a} \cdot \mathbf{b} > 0$, while if $\frac{1}{2}\pi < \theta \leq \pi$, then $\mathbf{a} \cdot \mathbf{b} < 0$

(v) If $\mathbf{a} \cdot \mathbf{b} = 0$ and $\mathbf{a}, \mathbf{b} \neq \mathbf{0}$, then $\mathbf{a}$ and $\mathbf{b}$ are perpendicular.

Intuitively, this is the product of the parts of $\mathbf{a}$ and $\mathbf{b}$ that are parallel.



Using the dot product, we can write the projection of $\mathbf{b}$ onto $\mathbf{a}$ as $(|\mathbf{b}| \cos \theta)\hat{\mathbf{a}} = (\hat{\mathbf{a}} \cdot \mathbf{b})\hat{\mathbf{a}}$. Note that the scalar product is distributive over vector addition.

Example 2.1.1 (The cosine rule)**.**

$$
\begin{aligned}
|\overrightarrow{BC}|^2 &= |\overrightarrow{AC} - \overrightarrow{AB}|^2 \\
&= (\overrightarrow{AC} - \overrightarrow{AB}) \cdot (\overrightarrow{AC} - \overrightarrow{AB}) \\
&= |\overrightarrow{AB}|^2 + |\overrightarrow{AC}|^2 - 2|\overrightarrow{AB}||\overrightarrow{AC}| \cos \theta
\end{aligned}
$$

Definition 2.1.2 (Algebraic definition of a scalar product)**.** *In a real vector space V, an inner product or scalar product is a map $V \times V \to \mathbb{R}$ that satisfies the following axioms. It is written as $\mathbf{x} \cdot \mathbf{y}$ or $\langle \mathbf{x} \mid \mathbf{y} \rangle$.*

(i) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ *(symmetry)*

(ii) $\mathbf{x} \cdot (\lambda \mathbf{y} + \mu \mathbf{z}) = \lambda \mathbf{x} \cdot \mathbf{y} + \mu \mathbf{x} \cdot \mathbf{z}$ *(linearity in 2nd argument)*

(iii) $\mathbf{x} \cdot \mathbf{x} \geq 0$ *with equality iff* $\mathbf{x} = \mathbf{0}$ *(positive definite)*

Note that this is a definition only for real vector spaces, where the scalars are real. In particular, here we can use (i) and (ii) together to show linearity in 1st argument. However, this is generally not true for complex vector spaces.

Definition 2.1.3. *The norm of a vector, written as* $|\mathbf{a}|$ *or* $\|\mathbf{a}\|$*, is defined as*

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

Example 2.1.2. *Instead of the usual* $\mathbb{R}^n$ *vector space, we can consider the set of all real (integrable) functions as a vector space. We can define the following inner product:*

$$\langle f \mid g \rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x.$$

Theorem 2.1.1 (Cauchy-Schwarz inequality)**.** *For all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*,*

$$|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|.$$

Proof. Consider the expression $|\mathbf{x} - \lambda\mathbf{y}|^2$. We must have

$$\begin{aligned} |\mathbf{x} - \lambda\mathbf{y}|^2 &\geq 0 \\ (\mathbf{x} - \lambda\mathbf{y}) \cdot (\mathbf{x} - \lambda\mathbf{y}) &\geq 0 \\ \lambda^2|\mathbf{y}|^2 - \lambda(2\mathbf{x} \cdot \mathbf{y}) + |\mathbf{x}|^2 &\geq 0. \end{aligned}$$

Viewing this as a quadratic in λ, we see that the quadratic is non-negative and thus cannot have 2 real roots. Thus the discriminant $\Delta \leq 0$. So

$$\begin{aligned} 4(\mathbf{x} \cdot \mathbf{y})^2 &\leq 4|\mathbf{y}|^2|\mathbf{x}|^2 \\ (\mathbf{x} \cdot \mathbf{y})^2 &\leq |\mathbf{x}|^2|\mathbf{y}|^2 \\ |\mathbf{x} \cdot \mathbf{y}| &\leq |\mathbf{x}||\mathbf{y}|. \end{aligned}$$

□

Note that we proved this using the axioms of the scalar product. So this result holds for *all* possible scalar products on *any* (real) vector space.

Example 2.1.3. *Let* $\mathbf{x} = (\alpha, \beta, \gamma)$ *and* $\mathbf{y} = (1, 1, 1)$*. Then by the Cauchy-*

Schwarz inequality, we have

$$\alpha + \beta + \gamma \leq \sqrt{3}\sqrt{\alpha^2 + \beta^2 + \gamma^2}$$
$$\alpha^2 + \beta^2 + \gamma^2 \geq \alpha\beta + \beta\gamma + \gamma\alpha,$$

with equality if $\alpha = \beta = \gamma$*.*

Corollary 2.1.1 (Triangle inequality)**.**

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$

Proof.

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) \\ &= |\mathbf{x}|^2 + 2\mathbf{x} \cdot \mathbf{y} + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2. \end{aligned}$$

So

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$

□

2.2 Vector product

Definition 2.2.1. *From a geometric standpoint, the vector product* $\boldsymbol{a} \times \boldsymbol{b}$ *of an ordered pair a, b is a vector such that*

(i)

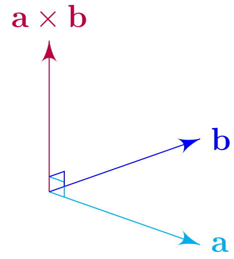
$$|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}| \sin\theta \tag{2.2}$$

with $0 \leq \theta \leq \pi$ *defined as before.*

(ii) $\boldsymbol{a} \times \boldsymbol{b}$ *is perpendicular or orthogonal to both* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *(if* $\boldsymbol{a} \times \boldsymbol{b} \neq 0$*.*

(iii) $\boldsymbol{a} \times \boldsymbol{b}$ *has the sense or direction defined by the 'right-hand rule', i.e. take a right hand, point the index finger in the direction of* $\boldsymbol{a}$*, the second finger*

in the direction of $\boldsymbol{b}$, *and then* $\boldsymbol{a} \times \boldsymbol{b}$ *is in the direction of the thumb.*



The vector product satisfies the following properties:

(i) $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

(ii) $\mathbf{a} \times \mathbf{a} = \mathbf{0}$.

(iii) $\mathbf{a} \times \mathbf{b} = \mathbf{0} \Rightarrow \mathbf{a} = \lambda\mathbf{b}$ for some $\lambda \in \mathbb{R}$ (or $\mathbf{b} = \mathbf{0}$).

(iv) $\mathbf{a} \times (\lambda\mathbf{b}) = \lambda(\mathbf{a} \times \mathbf{b})$.

(v) $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.

If we have a triangle OAB, its area is given by $\frac{1}{2}|\overrightarrow{OA}||\overrightarrow{OB}|\sin\theta = \frac{1}{2}|\overrightarrow{OA} \times \overrightarrow{OB}|$. We define the vector area as $\frac{1}{2}\overrightarrow{OA} \times \overrightarrow{OB}$.

Proposition 2.2.1.

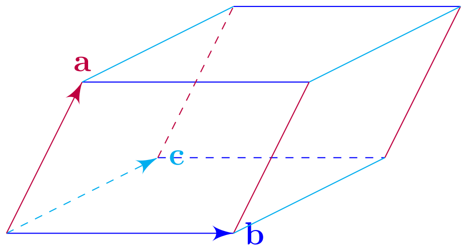
$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= (a_1\hat{\mathbf{i}} + a_2\hat{\mathbf{j}} + a_3\hat{\mathbf{k}}) \times (b_1\hat{\mathbf{i}} + b_2\hat{\mathbf{j}} + b_3\hat{\mathbf{k}}) \\
&= (a_2b_3 - a_3b_2)\hat{\mathbf{i}} + \cdots \\
&= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}
\end{aligned}$$

2.3 Scalar triple product

Definition 2.3.1 (Scalar triple product)**.** *The scalar triple product is defined as*

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

Proposition 2.3.1. *If a parallelepiped has sides represented by vectors* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *that form a right-handed system, then the volume of the parallelepiped is given by* $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$.



Proof. The area of the base of the parallelepiped is given by $|\mathbf{b}||\mathbf{c}|\sin\theta = |\mathbf{b} \times \mathbf{c}|$. Thus the volume$= |\mathbf{b} \times \mathbf{c}||\mathbf{a}|\cos\phi = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$, where ϕ is the angle between $\mathbf{a}$ and the normal to $\mathbf{b}$ and $\mathbf{c}$. However, since $\mathbf{a}, \mathbf{b}, \mathbf{c}$ form a right-handed system, we have $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \geq 0$. Therefore the volume is $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. □

Since the order of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ doesn't affect the volume, we know that

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

Theorem 2.3.1. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.

Proof. Let $\mathbf{d} = \mathbf{a} \times (\mathbf{b} + \mathbf{c}) - \mathbf{a} \times \mathbf{b} - \mathbf{a} \times \mathbf{c}$. We have

$$\begin{aligned}
\mathbf{d} \cdot \mathbf{d} &= \mathbf{d} \cdot [\mathbf{a} \times (\mathbf{b} + \mathbf{c})] - \mathbf{d} \cdot (\mathbf{a} \times \mathbf{b}) - \mathbf{d} \cdot (\mathbf{a} \times \mathbf{c}) \\
&= (\mathbf{b} + \mathbf{c}) \cdot (\mathbf{d} \times \mathbf{a}) - \mathbf{b} \cdot (\mathbf{d} \times \mathbf{a}) - \mathbf{c} \cdot (\mathbf{d} \times \mathbf{a}) \\
&= 0
\end{aligned}$$

Thus $\mathbf{d} = \mathbf{0}$. □

2.4 Bases

Let's consider 2D space, an origin O, and two non-zero and non-parallel vectors $\mathbf{a}$ and $\mathbf{b}$.

Definition 2.4.1 (Spanning set)**.** *A set of vectors* $\{\mathbf{a}, \mathbf{b}\}$ *spans* $\mathbb{R}^2$ *if for all vectors* $\mathbf{r} \in \mathbb{R}^2$*, there exist some* $\lambda, \mu \in \mathbb{R}$ *such that* $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b}$*.*

In $\mathbb{R}^2$, two vectors span the space if $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$.

Theorem 2.4.1. *The coefficients* λ, μ *are unique.*

Proof. Suppose that $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} = \lambda'\mathbf{a} + \mu'\mathbf{b}$. Take the vector product with $\mathbf{a}$ on both sides to get $(\mu - \mu')\mathbf{a} \times \mathbf{b} = \mathbf{0}$. Since $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$, then $\mu = \mu'$. Similarly, $\lambda = \lambda'$. □

Definition 2.4.2. *We refer to* (λ, μ) *as the components of* $\mathbf{r}$ *with respect to the ordered pair of vectors* $\mathbf{a}$ *and* $\mathbf{b}$*.*

Definition 2.4.3 (Linear independence)**.** *Two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are linearly independent if for* $\alpha, \beta \in \mathbb{R}$, $\alpha\mathbf{a} + \beta\mathbf{b} = \mathbf{0}$ *iff* $\alpha = \beta = 0$*. In* $\mathbb{R}^2$, $\mathbf{a}$ *and* $\mathbf{b}$ *are linearly independent if* $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$*.*

Definition 2.4.4 (Basis)**.** *A set of vectors is a basis of* $\mathbb{R}^2$ *if it spans* $\mathbb{R}^2$ *and are linearly independent.*

Example 2.4.1. $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}\} = \{(1, 0), (0, 1)\}$ *is a basis of* $\mathbb{R}^2$*. They are the standard basis of* $\mathbb{R}^2$*.*

Note that any two non-parallel vectors are linearly independent; the set $\{\mathbf{a}, \mathbf{b}\}$ does not have to be orthogonal to be a basis; and that in 2D space a basis always consists of two vectors.

Now consider 3D space, an origin O, and three non-zero and non-coplanar vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ (i.e. $[\mathbf{a}, \mathbf{b}, \mathbf{c}] \neq 0$). Then the position vector $\mathbf{r}$ of any point P in space can be expressed as

$$\mathbf{r} = \mathbf{OP} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} \tag{2.3}$$

for suitable and unique real scalars λ, μ and ν.

Theorem 2.4.2. *The coefficients λ, μ, ν are unique.*

Proof. We can show this by construction. Suppose that $\mathbf{r}$ is given by 2.3 and consider

$$\begin{aligned}\mathbf{r} \cdot (\mathbf{b} \times \mathbf{c}) &= (\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}) \cdot (\mathbf{b} \times \mathbf{c}) \\ &= \lambda\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) + \mu\mathbf{b} \cdot (\mathbf{b} \times \mathbf{c}) + \nu\mathbf{c}) \cdot (\mathbf{b} \times \mathbf{c}) \\ &= \lambda\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})\end{aligned}$$

since $\mathbf{b} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{c} \cdot (\mathbf{b} \times \mathbf{c}) = 0$. Hence, and similarly or by permutation,

$$\lambda = \frac{[\mathbf{r}, \mathbf{b}, \mathbf{c}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]} \qquad \mu = \frac{[\mathbf{r}, \mathbf{c}, \mathbf{a}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]} \qquad \nu = \frac{[\mathbf{r}, \mathbf{a}, \mathbf{b}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]}$$

□

Definition 2.4.5 (Linear independence)**.** *Two vectors* $\mathbf{a}$*,* $\mathbf{b}$ *and* $\mathbf{c}$ *and* α*,* β*,* $\gamma \in \mathbb{R}$*,*

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} + \gamma\boldsymbol{c} = \boldsymbol{0} \Rightarrow \alpha = \beta = \gamma = 0 \tag{2.4}$$

then we say that $\boldsymbol{a}$*,* $\boldsymbol{b}$ *and* $\boldsymbol{c}$ *are linearly independent.*

If $[\mathbf{a}, \mathbf{b}, \mathbf{c}] \neq 0$ the uniqueness of λ, μ and ν means that since $(0, 0, 0)$ is a solution to

$$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} = \mathbf{0}$$

it is also the unique solution, and hence the set $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is linearly independent. If $[\mathbf{a}, \mathbf{b}, \mathbf{c}] \neq 0$ the set $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ both spans 3D space and is linearly independent, it is hence a basis for 3D space. $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ do not have to be mutually orthogonal to be a basis.

2.5 Higher dimensional spaces

2.5.1 $\mathbb{R}^n$

Definition 2.5.1 (Linear independence). *A set of vectors* $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \cdots \mathbf{v}_m\}$ *is linearly independent if*

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i = \mathbf{0} \Rightarrow (\forall i)\, \lambda_i = 0.$$

Definition 2.5.2 (Spanning set). *A set of vectors* $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \cdots \mathbf{u}_m\} \subseteq \mathbb{R}^n$ *is a spanning set of* $\mathbb{R}^n$ *if*

$$(\forall \mathbf{x} \in \mathbb{R}^n)(\exists \lambda_i) \sum_{i=1}^{m} \lambda_i \mathbf{u}_i = \mathbf{x}$$

Definition 2.5.3 (Basis vectors). *A basis of* $\mathbb{R}^n$ *is a linearly independent spanning set. The standard basis of* $\mathbb{R}^n$ *is* $\mathbf{e}_1 = (1, 0, 0, \cdots 0), \mathbf{e}_2 = (0, 1, 0, \cdots 0), \cdots \mathbf{e}_n = (0, 0, 0, \cdots, 1)$.

Definition 2.5.4 (Orthonormal basis). *A basis* $\{\mathbf{e}_i\}$ *is orthonormal if* $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ *if* $i \neq j$ *and* $\mathbf{e}_i \cdot \mathbf{e}_i = 1$ *for all* i, j.

Using the Kronecker Delta symbol, which we will define later, we can write this condition as $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$.

Definition 2.5.5 (Dimension of vector space). *The dimension of a vector space is the number of vectors in its basis. (Exercise: show that this is well-defined)*

Definition 2.5.6 (Scalar product). *The* scalar product *of* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *is defined as* $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$.

The reader should check that this definition coincides with the $|\mathbf{x}||\mathbf{y}|\cos\theta$ definition in the case of $\mathbb{R}^2$ and $\mathbb{R}^3$.

2.5.2 $\mathbb{C}^n$

$\mathbb{C}^n$ is very similar to $\mathbb{R}^n$, except that we have complex numbers. As a result, we need a different definition of the scalar product. If we still defined $\mathbf{u} \cdot \mathbf{v} = \sum u_i v_i$, then if we let $\mathbf{u} = (0, i)$, then $\mathbf{u} \cdot \mathbf{u} = -1 < 0$. This would be bad if we want to use the scalar product to define a norm.

Definition 2.5.7 **($\mathbb{C}^n$).** $\mathbb{C}^n = \{(z_1, z_2, \cdots, z_n) : z_i \in \mathbb{C}\}$. *It has the same standard basis as* $\mathbb{R}^n$ *but the scalar product is defined differently. For* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$*,* $\mathbf{u} \cdot \mathbf{v} = \sum u_i^* v_i$*. The scalar product has the following properties:*

1. $\mathbf{u} \cdot \mathbf{v} = (\mathbf{v} \cdot \mathbf{u})^*$
2. $\mathbf{u} \cdot (\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda(\mathbf{u} \cdot \mathbf{v}) + \mu(\mathbf{u} \cdot \mathbf{w})$
3. $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *iff* $\mathbf{u} = \mathbf{0}$

Instead of linearity in the first argument, here we have $(\lambda \mathbf{u} + \mu \mathbf{v}) \cdot \mathbf{w} = \lambda^* \mathbf{u} \cdot \mathbf{w} + \mu^* \mathbf{v} \cdot \mathbf{w}$.

Example 2.5.1.

$$
\begin{aligned}
&\sum_{k=1}^{4} (-i)^k |\mathbf{x} + i^k \mathbf{y}|^2 \\
&= \sum (-i)^k \langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{x} + i^k \mathbf{y} \rangle \\
&= \sum (-i)^k (\langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{x} \rangle + i^k \langle \mathbf{x} + i^k \mathbf{y} \mid \mathbf{y} \rangle) \\
&= \sum (-i)^k (\langle \mathbf{x} \mid \mathbf{x} \rangle + (-i)^k \langle \mathbf{y} \mid \mathbf{x} \rangle + i^k \langle \mathbf{x} \mid \mathbf{y} \rangle + i^k (-i)^k \langle \mathbf{y} \mid \mathbf{y} \rangle) \\
&= \sum (-i)^k [(|\mathbf{x}|^2 + |\mathbf{y}|^2) + (-1)^k \langle \mathbf{y} \mid \mathbf{x} \rangle + \langle \mathbf{x} \mid \mathbf{y} \rangle] \\
&= (|\mathbf{x}|^2 + |\mathbf{y}|^2) \sum (-i)^k + \langle \mathbf{y} \mid \mathbf{x} \rangle \sum (-1)^k + \langle \mathbf{x} \mid \mathbf{y} \rangle \sum 1 \\
&= 4 \langle \mathbf{x} \mid \mathbf{y} \rangle.
\end{aligned}
$$

We can prove the Cauchy-Schwartz inequality for complex vector spaces using the same proof as the real case, except that this time we have to first multiply $\mathbf{y}$ by some $e^{i\theta}$ so that $\mathbf{x} \cdot (e^{i\theta} \mathbf{y})$ is a real number. The factor of $e^{i\theta}$ will drop off at the end when we take the modulus signs.

Definition 2.5.8 (Vector subspace). *A* vector subspace *of a vector space V is a subset of V that is also a vector space under the same operations. Both V and $\{\mathbf{0}\}$ are subspaces of V. All others are proper subspaces.*

A useful criterion is that a subset $U \subseteq V$ is a subspace iff

1. $\mathbf{x}, \mathbf{y} \in U \Rightarrow (\mathbf{x} + \mathbf{y}) \in U$.
2. $\mathbf{x} \in U \Rightarrow \lambda\mathbf{x} \in U$ *for all scalars* λ.
3. $\mathbf{0} \in U$.

This can be more concisely written as "U is non-empty and for all $\mathbf{x}, \mathbf{y} \in U$, $(\lambda\mathbf{x} + \mu\mathbf{y}) \in U$".

Example 2.5.2.

1. *If $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis of $\mathbb{R}^3$, then $\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}$ is a basis of a 2D subspace.*

   Suppose $\mathbf{x}, \mathbf{y} \in \operatorname{span}\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}$. Let

   $$\mathbf{x} = \alpha_1(\mathbf{a}+\mathbf{c}) + \beta_1(\mathbf{b}+\mathbf{c});$$
   $$\mathbf{y} = \alpha_2(\mathbf{a}+\mathbf{c}) + \beta_2(\mathbf{b}+\mathbf{c}).$$

   Then

   $$\lambda\mathbf{x} + \mu\mathbf{y} = (\lambda\alpha_1 + \mu\alpha_2)(\mathbf{a}+\mathbf{c}) + (\lambda\beta_1 + \mu\beta_2)(\mathbf{b}+\mathbf{c}) \in \operatorname{span}\{\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}\}.$$

   Thus this is a subspace of $\mathbb{R}^3$.

   Now check that $\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}$ is a basis. We only need to check linear independence. If $\alpha(\mathbf{a}+\mathbf{c}) + \beta(\mathbf{b}+\mathbf{c}) = \mathbf{0}$, then $\alpha\mathbf{a} + \beta\mathbf{b} + (\alpha+\beta)\mathbf{c} = \mathbf{0}$. Since $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis of $\mathbb{R}^3$, therefore $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent and $\alpha = \beta = 0$. Therefore $\mathbf{a}+\mathbf{c}, \mathbf{b}+\mathbf{c}$ is a basis and the subspace has dimension 2.

2. *Given a set of numbers α_i, let $U = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \alpha_i x_i = 0\}$. We show that this is a vector subspace of $\mathbb{R}^n$: Take $\mathbf{x}, \mathbf{y} \in U$, then consider $\lambda\mathbf{x} + \mu\mathbf{y}$. We have $\sum \alpha_i(\lambda x_i + \mu y_i) = \lambda \sum \alpha_i x_i + \mu \sum \alpha_i y_i = 0$. Thus $\lambda\mathbf{x} + \mu\mathbf{y} \in U$.*

   The dimension of the subspace is $n-1$ as we can freely choose x_i for $i = 1, \cdots, n-1$ and then x_n is uniquely determined by the previous x_i's.

3. *Let $W = \{\mathbf{x} \in \mathbb{R}^n : \sum \alpha_i x_i = 1\}$. Then $\sum \alpha_i(\lambda x_i + \mu y_i) = \lambda + \mu \neq 1$. Therefore W is not a vector subspace.*

2.6 Suffix notation

Let $\mathbf{v} \in \mathbb{R}^3$. We can write $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 = (v_1, v_2, v_3)$. So in general, the ith component of $\mathbf{v}$ is written as v_i. We can thus write vector equations in component form. For example, $\mathbf{a} = \mathbf{b} \to a_i = b_i$ or $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b} \to c_i = \alpha a_i + \beta b_i$. A vector has one free suffix, i, while a scalar has none.

Notation 2.6.1 (Einstein's summation convention)**.** *Consider a sum* $\mathbf{x} \cdot \mathbf{y} = \sum x_i y_i$. *The summation convention says that we can drop the* $\sum$ *symbol and simply write* $\mathbf{x} \cdot \mathbf{y} = x_i y_i$. *If suffixes are repeated once, summation is understood.*

Note that i is a dummy suffix and doesn't matter what it's called, i.e. $x_i y_i = x_j y_j = x_k y_k$ *etc.*

The rules of this convention are:

1. *Suffix appears once in a term: free suffix*
2. *Suffix appears twice in a term: dummy suffix and is summed over*
3. *Suffix appears three times or more: WRONG!*

Example 2.6.1. $[(\mathbf{a} \cdot \mathbf{b})\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b}]_i = a_j b_j c_i - a_j c_j b_i$ *summing over* j *understood.*

Definition 2.6.1 (Kronecker delta)**.**

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

We have

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{I}.$$

So the Kronecker delta represents an identity matrix.

Example 2.6.2. *1.* $a_i\delta_{i1} = a_1$*. In general,* $a_i\delta_{ij} = a_j$ *(*i *is dummy,* j *is free).*

2. $\delta_{ij}\delta_{jk} = \delta_{ik}$

3. $\delta_{ii} = n$ *if we are in* $\mathbb{R}^n$*.*

4. $a_p\delta_{pq}b_q = a_pb_p$ *with* p, q *both dummy suffices and summed over.*

Definition 2.6.2 (Alternating symbol ε_{ijk})**.** *Consider rearrangements of* $1, 2, 3$*. We can divide them into even and odd permutations. Even permutations include* $(1, 2, 3)$*,* $(2, 3, 1)$ *and* $(3, 1, 2)$*. These are permutations obtained by performing two (or no) swaps of the elements of* $(1, 2, 3)$*. (Alternatively, it is any "rotation" of* $(1, 2, 3)$*)*

The odd permutations are $(2, 1, 3)$*,* $(1, 3, 2)$ *and* $(3, 2, 1)$*. They are the permutations obtained by one swap only.*

Define

$$\varepsilon_{ijk} = \begin{cases} +1 & ijk \text{ is even permutation} \\ -1 & ijk \text{ is odd permutation} \\ 0 & \text{otherwise (i.e. repeated suffices)} \end{cases}$$

ε_{ijk} *has 3 free suffices.*

We have $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$ *and* $\varepsilon_{213} = \varepsilon_{132} = \varepsilon_{321} = -1$*.* $\varepsilon_{112} = \varepsilon_{111} = \cdots = 0$*.*

We have

1. $\varepsilon_{ijk}\delta_{jk} = \varepsilon_{ijj} = 0$

2. If $a_{jk} = a_{kj}$ (i.e. a_{ij} is symmetric), then $\varepsilon_{ijk}a_{jk} = \varepsilon_{ijk}a_{kj} = -\varepsilon_{ikj}a_{kj}$. Since $\varepsilon_{ijk}a_{jk} = \varepsilon_{ikj}a_{kj}$ (we simply renamed dummy suffices), we have $\varepsilon_{ijk}a_{jk} = 0$.

Proposition 2.6.1. $(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk}a_jb_k$

Proof. By expansion of formula $\square$

Theorem 2.6.1. $\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}$

Proof. Proof by exhaustion:

$$\text{RHS} = \begin{cases} +1 & \text{if } j = p \text{ and } k = q \\ -1 & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise} \end{cases}$$

LHS: Summing over i, the only non-zero terms are when $j, k \neq i$ and $p, q \neq i$. If $j = p$ and $k = q$, LHS is $(-1)^2$ or $(+1)^2 = 1$. If $j = q$ and $k = p$, LHS is $(+1)(-1)$ or $(-1)(+1) = -1$. All other possibilities result in 0. □

Equally, we have $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{jp}\delta_{iq}$ and $\varepsilon_{ijk}\varepsilon_{pjq} = \delta_{ip}\delta_{kq} - \delta_{iq}\delta_{kp}$.

Proposition 2.6.2.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$$

Proof. In suffix notation, we have

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_i(\mathbf{b} \times \mathbf{c})_i = \varepsilon_{ijk}b_jc_ka_i = \varepsilon_{jki}b_jc_ka_i = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}).$$

□

Theorem 2.6.2 (Vector triple product)**.**

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

Proof.

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i &= \varepsilon_{ijk}a_j(b \times c)_k \\ &= \varepsilon_{ijk}\varepsilon_{kpq}a_jb_pc_q \\ &= \varepsilon_{ijk}\varepsilon_{pqk}a_jb_pc_q \\ &= (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})a_jb_pc_q \\ &= a_jb_ic_j - a_jc_ib_j \\ &= (\mathbf{a} \cdot \mathbf{c})b_i - (\mathbf{a} \cdot \mathbf{b})c_i \end{aligned}$$

□

Similarly, $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}$.

2.7 Vector equations

Lines

Any line through $\mathbf{a}$ and parallel to $\mathbf{t}$ can be written as

$$\mathbf{x} = \mathbf{a} + \lambda \mathbf{t}.$$

By crossing both sides of the equation with $\mathbf{t}$, we have

Theorem 2.7.1. *The equation of a straight line through $\mathbf{a}$ and parallel to $\mathbf{t}$ is*

$$(\mathbf{x} - \mathbf{a}) \times \mathbf{t} = \mathbf{0} \text{ or } \mathbf{x} \times \mathbf{t} = \mathbf{a} \times \mathbf{t}.$$

Plane

To define a plane Π, we need a normal $\mathbf{n}$ to the plane and a fixed point $\mathbf{b}$. For any $\mathbf{x} \in \Pi$, the vector $\mathbf{x} - \mathbf{b}$ is contained in the plane and is thus normal to $\mathbf{n}$, i.e. $(\mathbf{x} - \mathbf{b}) \cdot \mathbf{n} = 0$.

Theorem 2.7.2. *The equation of a plane through $\mathbf{b}$ with normal $\mathbf{n}$ is given by*

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}.$$

If $\mathbf{n} = \hat{\mathbf{n}}$ is a unit normal, then $d = \mathbf{x} \cdot \hat{\mathbf{n}} = \mathbf{b} \cdot \hat{\mathbf{n}}$ is the perpendicular distance from the origin to Π.

Alternatively, if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ lie in the plane, then the equation of the plane is

$$(\mathbf{x} - \mathbf{a}) \cdot [(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})] = 0.$$

Example 2.7.1.

1. *Consider the intersection between a line $\mathbf{x} \times \mathbf{t} = \mathbf{a} \times \mathbf{t}$ with the plane $\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}$. Cross $\mathbf{n}$ on the right with the line equation to obtain*

$$(\mathbf{x} \cdot \mathbf{n})\mathbf{t} - (\mathbf{t} \cdot \mathbf{n})\mathbf{x} = (\mathbf{a} \times \mathbf{t}) \times \mathbf{n}$$

Eliminate $\mathbf{x} \cdot \mathbf{n}$ using $\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}$

$$(\mathbf{t} \cdot \mathbf{n})\mathbf{x} = (\mathbf{b} \cdot \mathbf{n})\mathbf{t} - (\mathbf{a} \times \mathbf{t}) \times \mathbf{n}$$

Provided $\mathbf{t}\cdot\mathbf{n}$ *is non-zero, the point of intersection is*

$$\mathbf{x} = \frac{(\mathbf{b}\cdot\mathbf{n})\mathbf{t} - (\mathbf{a}\times\mathbf{t})\times\mathbf{n}}{\mathbf{t}\cdot\mathbf{n}}.$$

Exercise: what if $\mathbf{t}\cdot\mathbf{n} = 0$*?*

2. *Shortest distance between two lines. Let* L_1 *be* $(\mathbf{x}-\mathbf{a}_1)\times\mathbf{t}_1 = \mathbf{0}$ *and* L_2 *be* $(\mathbf{x}-\mathbf{a}_2)\times\mathbf{t}_2 = \mathbf{0}$.

   The distance of closest approach s *is along a line perpendicular to both* L_1 *and* L_2*, i.e. the line of closest approach is perpendicular to both lines and thus parallel to* $\mathbf{t}_1\times\mathbf{t}_2$*. The distance* s *can then be found by projecting* $\mathbf{a}_1-\mathbf{a}_2$ *onto* $\mathbf{t}_1\times\mathbf{t}_2$*. Thus* $s = \left|(\mathbf{a}_1-\mathbf{a}_2)\cdot\frac{\mathbf{t}_1\times\mathbf{t}_2}{|\mathbf{t}_1\times\mathbf{t}_2|}\right|$.

Example 2.7.2. $\mathbf{x} - (\mathbf{x}\times\mathbf{a})\times\mathbf{b} = \mathbf{c}$. *Strategy: take the dot or cross of the equation with suitable vectors. The equation can be expanded to form*

$$\mathbf{x} - (\mathbf{x}\cdot\mathbf{b})\mathbf{a} + (\mathbf{a}\cdot\mathbf{b})\mathbf{x} = \mathbf{c}.$$

Dot this with $\mathbf{b}$ *to obtain*

$$\begin{aligned}\mathbf{x}\cdot\mathbf{b} - (\mathbf{x}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{b}) + (\mathbf{a}\cdot\mathbf{b})(\mathbf{x}\cdot\mathbf{b}) &= \mathbf{c}\cdot\mathbf{b}\\ \mathbf{x}\cdot\mathbf{b} &= \mathbf{c}\cdot\mathbf{b}.\end{aligned}$$

Substituting this into the original equation, we have

$$\mathbf{x}(1+\mathbf{a}\cdot\mathbf{b}) = \mathbf{c} + (\mathbf{c}\cdot\mathbf{b})\mathbf{a}$$

If $(1+\mathbf{a}\cdot\mathbf{b})$ *is non-zero, then*

$$\mathbf{x} = \frac{\mathbf{c} + (\mathbf{c}\cdot\mathbf{b})\mathbf{a}}{1+\mathbf{a}\cdot\mathbf{b}}$$

Otherwise, when $(1+\mathbf{a}\cdot\mathbf{b}) = 0$*, if* $\mathbf{c} + (\mathbf{c}\cdot\mathbf{b})\mathbf{a} \neq \mathbf{0}$*, then a contradiction is reached. Otherwise,* $\mathbf{x}\cdot\mathbf{b} = \mathbf{c}\cdot\mathbf{b}$ *is the most general solution, which is a plane of solutions.*

CHAPTER 3

Linear maps and matrices

Matrices are rectangular tables of elements that are used for describing linear equations and maps. They have a wide range of applications since many real-world problems are linear. For instance, electromagnetic waves follow linear equations, and most sounds are also considered "linear." Many computational approaches to nonlinear problems involve linearizations, which are good for computers. This section aims to introduce matrices and their properties.

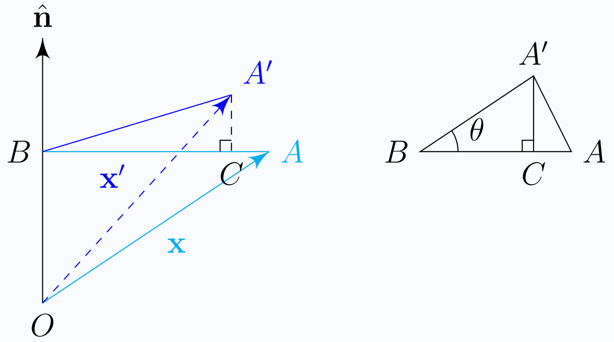
Example 3.0.1 (Rotation in $\mathbb{R}^3$)**.** *In $\mathbb{R}^3$, first consider the simple cases where we rotate about the z axis by θ. We call this rotation R and write $\mathbf{x}' = R(\mathbf{x})$. Suppose that initially, $\mathbf{x} = (x, y, z) = (r\cos\phi, r\sin\phi, z)$. Then after a rotation by θ, we get*

$$\begin{aligned}\mathbf{x}' &= (r\cos(\phi+\theta), r\sin(\phi+\theta), z)\\ &= (r\cos\phi\cos\theta - r\sin\phi\sin\theta, r\sin\phi\cos\theta + r\cos\phi\sin\theta, z)\\ &= (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z).\end{aligned}$$

We can represent this by a matrix R such that $x_i' = R_{ij}x_j$. Using our formula above, we obtain

$$R = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now consider the general case where we rotate by θ about $\hat{\mathbf{n}}$.



We have $\mathbf{x}' = \overrightarrow{OB} + \overrightarrow{BC} + \overrightarrow{CA'}$. We know that

$$\begin{aligned} \overrightarrow{OB} &= (\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} \\ \overrightarrow{BC} &= \overrightarrow{BA}\cos\theta \\ &= (\overrightarrow{BO} + \overrightarrow{OA})\cos\theta \\ &= (-(\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} + \mathbf{x})\cos\theta \end{aligned}$$

Finally, to get $\overrightarrow{CA}$, we know that $|\overrightarrow{CA'}| = |\overrightarrow{BA'}|\sin\theta = |\overrightarrow{BA}|\sin\theta = |\hat{\mathbf{n}}\times\mathbf{x}|\sin\theta$. Also, $\overrightarrow{CA'}$ is parallel to $\hat{\mathbf{n}}\times\mathbf{x}$. So we must have $\overrightarrow{CA'} = (\hat{\mathbf{n}}\times\mathbf{x})\sin\theta$.

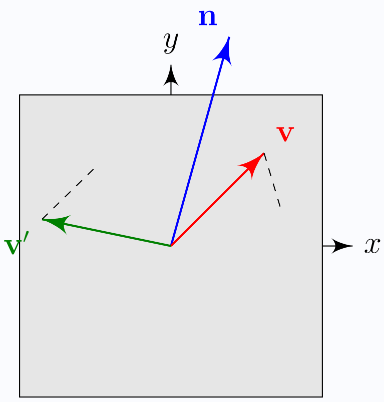
Thus $\mathbf{x}' = \mathbf{x}\cos\theta + (1-\cos\theta)(\hat{\mathbf{n}}\cdot\mathbf{x})\hat{\mathbf{n}} + \hat{\mathbf{n}}\times\mathbf{x}\sin\theta$. In components,

$$x_i' = x_i\cos\theta + (1-\cos\theta)n_jx_jn_i - \varepsilon_{ijk}x_jn_k\sin\theta.$$

We want to find an R such that $x_i' = R_{ij}x_j$. So

$$R_{ij} = \delta_{ij}\cos\theta + (1-\cos\theta)n_in_j - \varepsilon_{ijk}n_k\sin\theta.$$

Example 3.0.2 (Reflection in $\mathbb{R}^3$). *Suppose we want to reflect through a plane through O with normal $\hat{\mathbf{n}}$. First of all the projection of $\mathbf{v}$ onto $\hat{\mathbf{n}}$ is given by $(\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$. So we get $\mathbf{v}' = \mathbf{v} - 2(\mathbf{v}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$. In suffix notation, we have $v_i' = v_i - 2x_jn_jn_i$. So our reflection matrix is $R_{ij} = \delta_{ij} - 2n_in_j$.*



3.1 Linear maps

Definition 3.1.1 (Domain, codomain and image of map)**.** *Consider sets A and B and mapping $T : A \to B$ such that each $x \in A$ is mapped into a unique $x' = T(x) \in B$. A is the domain of T and B is the co-domain of T. Typically, we have $T : \mathbb{R}^n \to \mathbb{R}^m$ or $T : \mathbb{C}^n \to \mathbb{C}^m$.*

Definition 3.1.2 (Linear map)**.** *Let V, W be real (or complex) vector spaces, and $T : V \to W$. Then T is a linear map if*

1. *$T(\mathbf{a} + \mathbf{b}) = T(\mathbf{a}) + T(\mathbf{b})$ for all $\mathbf{a}, \mathbf{b} \in V$.*
2. *$T(\lambda\mathbf{a}) = \lambda T(\mathbf{a})$ for all $\lambda \in \mathbb{R}$ (or $\mathbb{C}$).*

Equivalently, we have $T(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda T(\mathbf{a}) + \mu T(\mathbf{b})$.

Example 3.1.1.
1. *Consider a translation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with $T(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ for some fixed, given $\mathbf{a}$. This is* not *a linear map since $T(\lambda\mathbf{x} + \mu\mathbf{y}) \neq \lambda\mathbf{x} + \mu\mathbf{y} + (\lambda + \mu)\mathbf{a}$.*
2. *Rotation, reflection and projection are linear transformations.*

3.2 Rank, kernel and nullity

Let $T : V \to W$ be a linear map (say with $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$). Recall that $T(V)$ is the image of V under T, and that $T(V)$ is a subspace of W.

Definition 3.2.1 (Rank)**.** *The rank is the dimension of the range of T, also known as the image of T, denoted by* $\text{rank}(T) = \dim(\text{range}(T)) = \dim(V)$.

Definition 3.2.2 (Kernel)**.** *The kernel of a linear map $T : V \to W$ is the set of all vectors in V that are mapped to zero in W, denoted by* $\ker(T) = \mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}$.

Claim: The kernel of a linear map $T : V \to W$ is a subspace of V.

Proof. Let $\mathbf{u}, \mathbf{v} \in \ker(T)$ and let $c \in \mathbb{R}$. Then we have:

$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so $\mathbf{u} + \mathbf{v} \in \ker(T)$. $T(c\mathbf{u}) = cT(\mathbf{u}) = c\mathbf{0} = \mathbf{0}$, so $c\mathbf{u} \in \ker(T)$. Therefore, $\ker(T)$ is closed under vector addition and scalar multiplication, and hence is a subspace of V. $\square$

Definition 3.2.3 (Nullity)**.** *The nullity of a linear map $T : V \to W$ is the dimension of the kernel of T, denoted by* $\text{nullity}(T) = \dim(\ker(T))$.

Theorem 3.2.1 (Rank-Nullity Theorem)**.** *The rank-nullity theorem states that for any linear map $T : V \to W$, we have:*

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

In other words, the sum of the dimension of the image of T and the dimension of the kernel of T is equal to the dimension of the domain of T.

Proof. (Non-examinable) Write $\dim(U) = n$ and $n(f) = m$. If $m = n$, then f is the zero map, and the proof is trivial, since $r(f) = 0$. Otherwise, assume $m < n$.

Suppose $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m\}$ is a basis of $\ker f$, Extend this to a basis of the whole of U to get $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m, \mathbf{e}_{m+1}, \cdots, \mathbf{e}_n\}$. To prove the theorem, we need to prove that $\{f(\mathbf{e}_{m+1}), f(\mathbf{e}_{m+2}), \cdots f(\mathbf{e}_n)\}$ is a basis of $\text{im}(f)$.

1. First show that it spans $\text{im}(f)$. Take $\mathbf{y} \in \text{im}(f)$. Thus $\exists \mathbf{x} \in U$ such that

$\mathbf{y} = f(\mathbf{x})$. Then

$$\mathbf{y} = f(\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_n \mathbf{e}_n),$$

since $\mathbf{e}_1, \cdots \mathbf{e}_n$ is a basis of U. Thus

$$\mathbf{y} = \alpha_1 f(\mathbf{e}_1) + \alpha_2 f(\mathbf{e}_2) + \cdots + \alpha_m f(\mathbf{e}_m) + \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$

The first m terms map to $\mathbf{0}$, since $\mathbf{e_1}, \cdots \mathbf{e_m}$ is the basis of the kernel of f. Thus

$$\mathbf{y} = \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$

2. To show that they are linearly independent, suppose

$$\alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}.$$

Then

$$f(\alpha_{m+1} \mathbf{e}_{m+1} + \cdots + \alpha_n \mathbf{e}_n) = \mathbf{0}.$$

Thus $\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n \mathbf{e}_n \in \ker(f)$. Since $\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$ span $\ker(f)$, there exist some $\alpha_1, \alpha_2, \cdots \alpha_m$ such that

$$\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n \mathbf{e}_n = \alpha_1 \mathbf{e_1} + \cdots + \alpha_m \mathbf{e}_m.$$

But $\mathbf{e}_1 \cdots \mathbf{e}_n$ is a basis of U and are linearly independent. So $\alpha_i = 0$ for all i. Then the only solution to the equation $\alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}$ is $\alpha_i = 0$, and they are linearly independent by definition. □

Example 3.2.1. *Calculate the kernel and image of* $f : \mathbb{R}^3 \to \mathbb{R}^3$, *defined by* $f(x, y, z) = (x + y + z, 2x - y + 5z, x + 2z)$.
First find the kernel: we've got the system of equations:

$$\begin{aligned} x + y + z &= 0 \\ 2x - y + 5z &= 0 \\ x + 2z &= 0 \end{aligned}$$

Note that the first and second equation add to give $3x + 6z = 0$, *which is identical to the third. Then using the first and third equation, we have* $y = -x - z = z$. *So the kernel is any vector in the form* $(-2z, z, z)$ *and is the span of* $(-2, 1, 1)$.

To find the image, extend the basis of $\ker(f)$ *to a basis of the whole of* $\mathbb{R}^3$: $\{(-2,1,1),(0,1,0),(0,0,1)\}$. *Apply* f *to this basis to obtain* $(0,0,0),(1,-1,0)$ *and* $(1,5,2)$. *From the proof of the rank-nullity theorem, we know that* $f(0,1,0)$ *and* $f(0,0,1)$ *is a basis of the image.*

To get the standard form of the image, we know that the normal to the plane is parallel to $(1,-1,0)\times(1,5,2) \parallel (1,1,-3)$. *Since* $\mathbf{0}\in \operatorname{im}(f)$, *the equation of the plane is* $x+y-3z=0$.

3.2.1 Composition of maps

The composition of linear maps is a way of combining two linear maps $T : U \to V$ and $S : V \to W$ to form a new linear map $S \circ T : U \to W$. The composition of T and S is sometimes denoted by ST, although the order is important, so we use $S \circ T$ to make it clear which map is applied first.

To define the composition of linear maps T and S, we need to make sure that the range of T is contained in the domain of S. In other words, $T(U) \subseteq V$ so that S can be applied to the output of T.

More formally, the composition of T and S is defined as:

Definition 3.2.4 (Composition of maps)**.**

$$(S\circ T)(\mathbf{u}) = S(T(\mathbf{u})) \quad \textit{for all } \mathbf{u}\in U \tag{3.1}$$

Note that the resulting map $S\circ T$ is also a linear map, since it preserves vector addition and scalar multiplication. Here are some examples of composition of linear maps:

Example 3.2.2. *Let* $T : \mathbb{R}^2 \to \mathbb{R}^3$ *be the linear map defined by* $T(\begin{pmatrix} x & y \end{pmatrix}) = \begin{pmatrix} x & y & 0 \end{pmatrix}$, *and let* $S : \mathbb{R}^3 \to \mathbb{R}^2$ *be the linear map defined by* $S(\begin{pmatrix} x & y & z \end{pmatrix}) = \begin{pmatrix} x+y & z \end{pmatrix}$. *Then the composition* $S\circ T : \mathbb{R}^2 \to \mathbb{R}^2$ *is given by:*

$$(S\circ T)(\begin{pmatrix} x & y \end{pmatrix}) = S(T(\begin{pmatrix} x & y \end{pmatrix})) = S(\begin{pmatrix} x & y & 0 \end{pmatrix}) = \begin{pmatrix} x+y & 0 \end{pmatrix}$$

Example 3.2.3. *Let* $T : \mathbb{R}^3 \to \mathbb{R}^2$ *be the linear map defined by* $T(\begin{pmatrix} x & y & z \end{pmatrix}) = \begin{pmatrix} x+y & z \end{pmatrix}$, *and let* $S : \mathbb{R}^2 \to \mathbb{R}$ *be the linear map defined by* $S(\begin{pmatrix} x & y \end{pmatrix}) = x - y$. *Then the composition* $S \circ T : \mathbb{R}^3 \to \mathbb{R}$ *is given by:*

$$(S \circ T)(\begin{pmatrix} x & y & z \end{pmatrix}) = S(T(\begin{pmatrix} x & y & z \end{pmatrix})) = S(\begin{pmatrix} x+y & z \end{pmatrix}) = (x+y) - z$$

3.3 Bases and the matrix description of maps

In the examples above, we have represented our linear maps by some object R such that $x'_i = R_{ij}x_j$. We call R the matrix for the linear map. In general, let $\alpha : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, and $\mathbf{x}' = \alpha(\mathbf{x})$.

Let $\{\mathbf{e}_i\}$ be a basis of $\mathbb{R}^n$. Then $\mathbf{x} = x_j\mathbf{e}_j$ for some x_j. Then we get

$$\mathbf{x}' = \alpha(x_j\mathbf{e}_j) = x_j\alpha(\mathbf{e}_j).$$

So we get that

$$x'_i = [\alpha(\mathbf{e}_j)]_i x_j.$$

We now define $A_{ij} = [\alpha(\mathbf{e}_j)]_i$. Then $x'_i = A_{ij}x_j$. We write

$$A = \{A_{ij}\} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & A_{ij} & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix}$$

Here A_{ij} is the entry in the ith row of the jth column. We say that A is an $m \times n$ matrix, and write $\mathbf{x}' = A\mathbf{x}$.

We see that the columns of the matrix are the images of the standard basis vectors under the mapping α.

Example 3.3.1. *In* $\mathbb{R}^2$, *consider a reflection in a line with an angle* θ *to the* x *axis. We know that* $\hat{\mathbf{i}} \mapsto \cos 2\theta\hat{\mathbf{i}} + \sin 2\theta\hat{\mathbf{j}}$, *with* $\hat{\mathbf{j}} \mapsto -\cos 2\theta\hat{\mathbf{j}} + \sin 2\theta\hat{\mathbf{i}}$. *Then the matrix is* $\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$.

Example 3.3.2. *In $\mathbb{R}^3$, as we've previously seen, a rotation by θ about the z axis is given by*

$$R = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

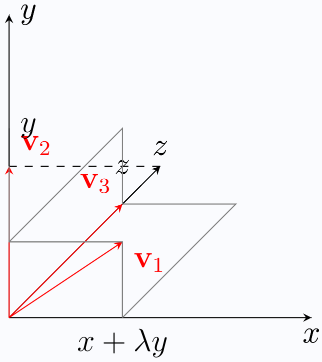
Example 3.3.3. *In $\mathbb{R}^3$, a reflection in plane with normal $\hat{\mathbf{n}}$ is given by $R_{ij} = \delta_{ij} - 2\hat{n}_i\hat{n}_j$. Written as a matrix, we have*

$$\begin{pmatrix} 1-2\hat{n}_1^2 & -2\hat{n}_1\hat{n}_2 & -2\hat{n}_1\hat{n}_3 \\ -2\hat{n}_2\hat{n}_1 & 1-2\hat{n}_2^2 & -2\hat{n}_2\hat{n}_3 \\ -2\hat{n}_3\hat{n}_1 & -2\hat{n}_3\hat{n}_2 & 1-2\hat{n}_3^2 \end{pmatrix}$$

Example 3.3.4. *Dilation ("stretching") $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ is given by a map $(x,y,z) \mapsto (\lambda x, \mu y, \nu z)$ for some λ, μ, ν. The matrix is*

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{pmatrix}$$

Example 3.3.5. *Shear: Consider $S : \mathbb{R}^3 \to \mathbb{R}^3$ that sheers in the x direction:*



We have $(x,y,z) \mapsto (x + \lambda y, y, z)$. Then

$$S = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3.4 Algebra of matrices

Definition 3.4.1 (Addition of matrices)**.** *Consider two linear maps* $\alpha, \beta : \mathbb{R}^n \to \mathbb{R}^m$*. The sum of* α *and* β *is defined by*

$$(\alpha + \beta)(\mathbf{x}) = \alpha(\mathbf{x}) + \beta(\mathbf{x})$$

In terms of the matrix, we have

$$(A + B)_{ij} x_j = A_{ij} x_j + B_{ij} x_j,$$

or

$$(A + B)_{ij} = A_{ij} + B_{ij}.$$

Definition 3.4.2 (Scalar multiplication of matrices)**.** *Define* $(\lambda\alpha)\mathbf{x} = \lambda[\alpha(\mathbf{x})]$*. So* $(\lambda A)_{ij} = \lambda A_{ij}$*.*

Definition 3.4.3 (Matrix multiplication)**.** *Consider maps* $\alpha : \mathbb{R}^\ell \to \mathbb{R}^n$ *and* $\beta : \mathbb{R}^n \to \mathbb{R}^m$*. The composition is* $\beta\alpha : \mathbb{R}^\ell \to \mathbb{R}^m$*. Take* $\mathbf{x} \in \mathbb{R}^\ell \mapsto \mathbf{x}'' \in \mathbb{R}^m$*. Then* $\mathbf{x}'' = (BA)\mathbf{x} = B\mathbf{x}'$*, where* $\mathbf{x}' = A\mathbf{x}$*. Using suffix notation, we have* $x_i'' = (B\mathbf{x}')_i = b_{ik} x_k' = B_{ik} A_{kj} x_j$*. But* $x_i'' = (BA)_{ij} x_j$*. So*

$$(BA)_{ij} = B_{ik} A_{kj}.$$

Generally, an $m \times n$ *matrix multiplied by an* $n \times \ell$ *matrix gives an* $m \times \ell$ *matrix.* $(BA)_{ij}$ *is given by the* i*th row of* B *dotted with the* j*th column of* A*.*

Note that the number of columns of B has to be equal to the number of rows of A for multiplication to be defined. If $\ell = m$ as well, then both BA and AB make sense, but $AB \neq BA$ in general. In fact, they don't even have to have the same dimensions.

Also, since function composition is associative, we get $A(BC) = (AB)C$.

Definition 3.4.4 (Transpose of matrix)**.** *If A is an $m \times n$ matrix, the* transpose A^T *is an $n \times m$ matrix defined by $(A^T)_{ij} = A_{ji}$.*

Proposition 3.4.1. *1.* $(A^T)^T = A$.

2. If $\mathbf{x}$ is a column vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{x}^T$ *is a row vector* $(x_1\ x_2 \cdots x_n)$.

3. $(AB)^T = B^T A^T$ *since* $(AB)^T_{ij} = (AB)_{ji} = A_{jk}B_{ki} = B_{ki}A_{jk}$
$= (B^T)_{ik}(A^T)_{kj} = (B^T A^T)_{ij}$.

3.4.1 Symmetric and hermitian matrices

Definition 3.4.5 (Hermitian conjugate)**.** *The Hermitian Conjugate of an $m \times n$ matrix A with complex entries is the $n \times m$ matrix denoted by $A^\dagger$ (or sometimes A^*) and defined by $A^\dagger = (A^T)^*$. Similarly, $(AB)^\dagger = B^\dagger A^\dagger$.*

Definition 3.4.6 (Hermitian matrix)**.** *A square matrix A is Hermitian if it is equal to its Hermitian conjugate, that is, if $A^\dagger = A$.*

Definition 3.4.7 (Symmetric matrix)**.** *A matrix is symmetric if $A^T = A$.*

Definition 3.4.8 (Anti/skew symmetric matrix)**.** *A matrix is anti-symmetric or skew symmetric if $A^T = -A$. The diagonals are all zero.*

Definition 3.4.9 (Skew-Hermitian matrix)**.** *A matrix is skew-Hermitian if $A^\dagger = -A$. The diagonals are pure imaginary.*

Definition 3.4.10 (Trace of matrix)**.** *The* trace *of an $n \times n$ matrix A is the sum of the diagonal.* $\mathrm{tr}(A) = A_{ii}$.

Proposition 3.4.2. $\operatorname{tr}(BC) = \operatorname{tr}(CB)$

Proof. $\operatorname{tr}(BC) = B_{ik}C_{ki} = C_{ki}B_{ik} = (CB)_{kk} = \operatorname{tr}(CB)$ □

Definition 3.4.11 (Identity matrix)**.** $I = \delta_{ij}$.

3.4.2 Decomposition of an $n \times n$ matrix

Any $n \times n$ matrix B can be split as a sum of symmetric and antisymmetric parts. Write

$$B_{ij} = \underbrace{\frac{1}{2}(B_{ij} + B_{ji})}_{S_{ij}} + \underbrace{\frac{1}{2}(B_{ij} - B_{ji})}_{A_{ij}}.$$

We have $S_{ij} = S_{ji}$, so S is symmetric, while $A_{ji} = -A_{ij}$, and A is antisymmetric. So $B = S + A$.

Furthermore , we can decompose S into an isotropic part (a scalar multiple of the identity) plus a trace-less part (i.e. sum of diagonal $= 0$). Write

$$S_{ij} = \underbrace{\frac{1}{n}\operatorname{tr}(S)\delta_{ij}}_{\text{isotropic part}} + \underbrace{(S_{ij} - \frac{1}{n}\operatorname{tr}(S)\delta_{ij})}_{T_{ij}}.$$

We have $\operatorname{tr}(T) = T_{ii} = S_{ii} - \frac{1}{n}\operatorname{tr}(S)\delta_{ii} = \operatorname{tr}(S) - \frac{1}{n}\operatorname{tr}(S)(n) = 0$.

Putting all these together,

$$B = \frac{1}{n}\operatorname{tr}(B)I + \left\{\frac{1}{2}(B + B^T) - \frac{1}{n}\operatorname{tr}(B)I\right\} + \frac{1}{2}(B - B^T).$$

In three dimensions, we can write the antisymmetric part A in terms of a single vector: we have

$$A = \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix}$$

and we can consider

$$\varepsilon_{ijk}\omega_k = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}$$

So if we have $\omega = (c, b, a)$, then $A_{ij} = \varepsilon_{ijk}\omega_k$.

This decomposition can be useful in certain physical applications. For example, if the matrix represents the stress of a system, different parts of the decomposition will correspond to different types of stresses.

Definition 3.4.12 (Inverse of matrix). *Consider an $m \times n$ matrix A and $n \times m$ matrices B and C. If $BA = I$, then we say B is the* left inverse *of A. If $AC = I$, then we say C is the right inverse of A. If A is square ($n \times n$), then $B = B(AC) = (BA)C = C$, i.e. the left and right inverses coincide. Both are denoted by A^{-1}, the inverse of A. Therefore we have*

$$AA^{-1} = A^{-1}A = I.$$

Note that not all square matrices have inverses. For example, the zero matrix clearly has no inverse.

Definition 3.4.13 (Invertible matrix). *If A has an inverse, then A is invertible.*

Proposition 3.4.3. $(AB)^{-1} = B^{-1}A^{-1}$

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

Definition 3.4.14 (Orthogonal and unitary matrices). *A real $n \times n$ matrix is orthogonal if $A^TA = AA^T = I$, i.e. $A^T = A^{-1}$. A complex $n \times n$ matrix is unitary if $U^\dagger U = UU^\dagger = I$, i.e. $U^\dagger = U^{-1}$.*

Note that an orthogonal matrix A satisfies $A_{ik}(A^T_{kj}) = \delta_{ij}$, i.e. $A_{ik}A_{jk} = \delta_{ij}$. We can see this as saying "the scalar product of two distinct rows is 0, and the scalar product of a row with itself is 1". Alternatively, the rows (and columns — by considering A^T) of an orthogonal matrix form an orthonormal set.

Similarly, for a unitary matrix, $U_{ik}U^\dagger_{kj} = \delta_{ij}$, i.e. $u_{ik}u^*_{jk} = u^*_{ik}u_{jk} = \delta_{ij}$. i.e. the rows are orthonormal, using the definition of complex scalar product.

Example 3.4.1. *1. The reflection in a plane is an orthogonal matrix. Since $R_{ij} = \delta_{ij} - 2n_i n_j$, We have*

$$\begin{aligned} R_{ik}R_{jk} &= (\delta_{ik} - 2n_i n_k)(\delta_{jk} - 2n_j n_k) \\ &= \delta_{ik}\delta_{jk} - 2\delta_{jk}n_i n_k - 2\delta_{ik}n_j n_k + 2n_i n_k n_j n_k \\ &= \delta_{ij} - 2n_i n_j - 2n_j n_i + 4n_i n_j(n_k n_k) \\ &= \delta_{ij} \end{aligned}$$

2. The rotation is an orthogonal matrix. We could multiply out using suffix notation, but it would be cumbersome to do so. Alternatively, denote rotation matrix by θ about $\hat{\mathbf{n}}$ as $R(\theta, \hat{\mathbf{n}})$. Clearly, $R(\theta, \hat{\mathbf{n}})^{-1} = R(-\theta, \hat{\mathbf{n}})$. We have

$$\begin{aligned} R_{ij}(-\theta, \hat{\mathbf{n}}) &= (\cos\theta)\delta_{ij} + n_i n_j(1 - \cos\theta) + \varepsilon_{ijk}n_k \sin\theta \\ &= (\cos\theta)\delta_{ji} + n_j n_i(1 - \cos\theta) - \varepsilon_{jik}n_k \sin\theta \\ &= R_{ji}(\theta, \hat{\mathbf{n}}) \end{aligned}$$

In other words, $R(-\theta, \hat{\mathbf{n}}) = R(\theta, \hat{\mathbf{n}})^T$. So $R(\theta, \hat{\mathbf{n}})^{-1} = R(\theta, \hat{\mathbf{n}})^T$.

3.5 Determinants

Consider a linear map $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$. The standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is mapped to $\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'$ with $\mathbf{e}_i' = A\mathbf{e}_i$. Thus the unit cube formed by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is mapped to the parallelepiped with volume

$$\begin{aligned} [\mathbf{e}_1', \mathbf{e}_2', \mathbf{e}_3'] &= \varepsilon_{ijk}(e_1')_i(e_2')_j(e_3')_k \\ &= \varepsilon_{ijk}A_{i\ell}\underbrace{(e_1)_\ell}_{\delta_{1\ell}} A_{jm}\underbrace{(e_2)_m}_{\delta_{2m}} A_{kn}\underbrace{(e_3)_n}_{\delta_{3n}} \\ &= \varepsilon_{ijk}A_{i1}A_{j2}A_{k3} \end{aligned}$$

We call this the determinant and write as

$$\det(A) = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

3.5.1 Permutations

To define the determinant for square matrices of arbitrary size, we first have to consider permutations.

Definition 3.5.1 (Permutation)**.** *A permutation of a set S is a bijection $\varepsilon : S \to S$.*

Notation 3.5.1. *Consider the set S_n of all permutations of $1, 2, 3, \cdots, n$. S_n contains $n!$ elements. Consider $\rho \in S_n$ with $i \mapsto \rho(i)$. We write*

$$\rho = \begin{pmatrix} 1 & 2 & \cdots & n \\ \rho(1) & \rho(2) & \cdots & \rho(n) \end{pmatrix}.$$

Definition 3.5.2 (Fixed point)**.** *A fixed point of ρ is a k such that $\rho(k) = k$. e.g. in $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}$, 3 is the fixed point. By convention, we can omit the fixed point and write as $\begin{pmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \end{pmatrix}$.*

Definition 3.5.3 (Disjoint permutation)**.** *Two permutations are* disjoint *if numbers moved by one are fixed by the other, and vice versa. e.g. $\begin{pmatrix} 1 & 2 & 4 & 5 & 6 \\ 5 & 6 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix} \begin{pmatrix} 1 & 4 & 5 \\ 5 & 1 & 4 \end{pmatrix}$, and the two cycles on the right hand side are disjoint. Disjoint permutations commute, but in general non-disjoint permutations do not.*

Definition 3.5.4 (Transposition and k-cycle)**.** *$\begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix}$ is a* 2-cycle *or a* transposition*, and we can simply write* (2 6)*. $\begin{pmatrix} 1 & 4 & 5 \\ 5 & 1 & 4 \end{pmatrix}$ is a 3-cycle, and we can simply write* (1 5 4)*. (1 is mapped to 5; 5 is mapped to 4; 4 is mapped to 1)*

Proposition 3.5.1. *Any q-cycle can be written as a product of 2-cycles.*

Proof. $(1\ 2\ 3\ \cdots\ n) = (1\ 2)(2\ 3)(3\ 4)\cdots(n-1\ n)$. □

Definition 3.5.5 (Sign of permutation). *The sign of a permutation $\varepsilon(\rho)$ is $(-1)^r$, where r is the number of 2-cycles when ρ is written as a product of 2-cycles. If $\varepsilon(\rho) = +1$, it is an even permutation. Otherwise, it is an odd permutation. Note that $\varepsilon(\rho\sigma) = \varepsilon(\rho)\varepsilon(\sigma)$ and $\varepsilon(\rho^{-1}) = \varepsilon(\rho)$.*

The proof that this is well-defined can be found in IA Groups.

Definition 3.5.6 (Levi-Civita symbol). *The Levi-Civita symbol is defined by*

$$\varepsilon_{j_1 j_2 \cdots j_n} = \begin{cases} +1 & \text{if } j_1 j_2 j_3 \cdots j_n \text{ is an even permutation of } 1, 2, \cdots n \\ -1 & \text{if it is an odd permutation} \\ 0 & \text{if any 2 of them are equal} \end{cases}$$

Clearly, $\varepsilon_{\rho(1)\rho(2)\cdots\rho(n)} = \varepsilon(\rho)$.

Definition 3.5.7 (Determinant). *The* determinant *of an $n \times n$ matrix A is defined as:*

$$\det(A) = \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n},$$

or equivalently,

$$\det(A) = \varepsilon_{j_1 j_2 \cdots j_n} A_{j_1 1} A_{j_2 2} \cdots A_{j_n n}.$$

Proposition 3.5.2.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

3.5.2 Properties of determinants

Proposition 3.5.3. $\det(A) = \det(A^T)$.

Proof. Take a single term $A_{\sigma(1)1} A_{\sigma(2)2} \cdots A_{\sigma(n)n}$ and let ρ be another permutation in

S_n. We have

$$A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n} = A_{\sigma(\rho(1))\rho(1)}A_{\sigma(\rho(2))\rho(2)}\cdots A_{\sigma(\rho(n))\rho(n)}$$

since the right hand side is just re-ordering the order of multiplication. Choose $\rho = \sigma^{-1}$ and note that $\varepsilon(\sigma) = \varepsilon(\rho)$. Then

$$\det(A) = \sum_{\rho\in S_n} \varepsilon(\rho)A_{1\rho(1)}A_{2\rho(2)}\cdots A_{n\rho(n)} = \det(A^T).$$

□

Proposition 3.5.4. *If matrix B is formed by multiplying every element in a single row of A by a scalar λ, then $\det(B) = \lambda\det(A)$. Consequently, $\det(\lambda A) = \lambda^n\det(A)$.*

Proof. Each term in the sum is multiplied by λ, so the whole sum is multiplied by λ^n. □

Proposition 3.5.5. *If 2 rows (or 2 columns) of A are identical, the determinant is 0.*

Proof. wlog, suppose columns 1 and 2 are the same. Then

$$\det(A) = \sum_{\sigma\in S_n} \varepsilon(\sigma)A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n}.$$

Now write an arbitrary σ in the form $\sigma = \rho(1\ 2)$. Then $\varepsilon(\sigma) = \varepsilon(\rho)\varepsilon((1\ 2)) = -\varepsilon(\rho)$. So

$$\det(A) = \sum_{\rho\in S_n} -\varepsilon(\rho)A_{\rho(2)1}A_{\rho(1)2}A_{\rho(3)3}\cdots A_{\rho(n)n}.$$

But columns 1 and 2 are identical, so $A_{\rho(2)1} = A_{\rho(2)2}$ and $A_{\rho(1)2} = A_{\rho(1)1}$. So $\det(A) = -\det(A)$ and $\det(A) = 0$. □

Proposition 3.5.6. *If 2 rows or 2 columns of a matrix are linearly dependent, then the determinant is zero.*

Proof. Suppose in A, (column r) + λ(column s) = 0. Define

$$B_{ij} = \begin{cases} A_{ij} & j \neq r \\ A_{ij} + \lambda A_{is} & j = r \end{cases}.$$

Then $\det(B) = \det(A) + \lambda \det(\text{matrix with column } r = \text{column } s) = \det(A)$. Then we can see that the rth column of B is all zeroes. So each term in the sum contains one zero and $\det(A) = \det(B) = 0$. □

Even if we don't have linearly dependent rows or columns, we can still run the exact same proof as above, and still get that $\det(B) = \det(A)$. Linear dependence is only required to show that $\det(B) = 0$. So in general, we can add a linear multiple of a column (or row) onto another column (or row) without changing the determinant.

Proposition 3.5.7. *Given a matrix A, if B is a matrix obtained by adding a multiple of a column (or row) of A to another column (or row) of A, then* $\det A = \det B$.

Corollary 3.5.1. *Swapping two rows or columns of a matrix negates the determinant.*

Proof. We do the column case only. Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n)$. Then

$$\begin{aligned}
\det(\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n) &= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j - (\mathbf{a}_i + \mathbf{a}_j) \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n) \\
&= \det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n) \\
&= -\det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n)
\end{aligned}$$

Alternatively, we can prove this from the definition directly, using the fact that the sign of a transposition is -1 (and that the sign is multiplicative). □

Proposition 3.5.8. $\det(AB) = \det(A)\det(B)$.

Proof. First note that $\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)} A_{\sigma(2)\rho(2)} = \varepsilon(\rho)\det(A)$, i.e. swapping columns (or rows) an even/odd number of times gives a factor ± 1 respectively. We can prove this by writing $\sigma = \mu\rho$.

Now

$$\begin{aligned}
\det AB &= \sum_\sigma \varepsilon(\sigma)(AB)_{\sigma(1)1}(AB)_{\sigma(2)2}\cdots(AB)_{\sigma(n)n} \\
&= \sum_\sigma \varepsilon(\sigma) \sum_{k_1,k_2,\cdots,k_n}^{n} A_{\sigma(1)k_1}B_{k_1 1}\cdots A_{\sigma(n)k_n}B_{k_n n} \\
&= \sum_{k_1,\cdots,k_n} B_{k_1 1}\cdots B_{k_n n} \underbrace{\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)k_1}A_{\sigma(2)k_2}\cdots A_{\sigma(n)k_n}}_{S}
\end{aligned}$$

Now consider the many different S's. If in S, two of k_1 and k_n are equal, then S is a determinant of a matrix with two columns the same, i.e. $S = 0$. So we only have to consider the sum over distinct k_is. Thus the k_is are are a permutation of $1, \cdots n$, say $k_i = \rho(i)$. Then we can write

$$\begin{aligned}
\det AB &= \sum_\rho B_{\rho(1)1}\cdots B_{\rho(n)n} \sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)}\cdots A_{\sigma(n)\rho(n)} \\
&= \sum_\rho B_{\rho(1)1}\cdots B_{\rho(n)n}(\varepsilon(\rho)\det A) \\
&= \det A \sum_\rho \varepsilon(\rho) B_{\rho(1)1}\cdots B_{\rho(n)n} \\
&= \det A \det B
\end{aligned}$$

□

Corollary 3.5.2. *If A is orthogonal,* $\det A = \pm 1$.

Proof.

$$\begin{aligned}
AA^T &= I \\
\det AA^T &= \det I \\
\det A \det A^T &= 1 \\
(\det A)^2 &= 1 \\
\det A &= \pm 1
\end{aligned}$$

□

Corollary 3.5.3. *If U is unitary, $|\det U| = 1$.*

Proof. We have $\det U^\dagger = (\det U^T)^* = \det(U)^*$. Since $UU^\dagger = I$, we have $\det(U)\det(U)^* = 1$. □

Proposition 3.5.9. *In $\mathbb{R}^3$, orthogonal matrices represent either a rotation ($\det = 1$) or a reflection ($\det = -1$).*

3.5.3 Minors and Cofactors

Definition 3.5.8 (Minor and cofactor)**.** *For an $n \times n$ matrix A, define A^{ij} to be the $(n-1)\times(n-1)$ matrix in which row i and column j of A have been removed.*
The minor of the ijth element of A is $M_{ij} = \det A^{ij}$
The cofactor of the ijth element of A is $\Delta_{ij} = (-1)^{i+j} M_{ij}$.

Notation 3.5.2. *We use $\overline{}$ to denote a symbol which has been missed out of a natural sequence.*

Example 3.5.1. $1, 2, 3, 5 = 1, 2, 3, \overline{4}, 5$.

The significance of these definitions is that we can use them to provide a systematic way of evaluating determinants. We will also use them to find inverses of matrices.

Theorem 3.5.1 (Laplace expansion formula)**.** *For any particular fixed i,*

$$\det A = \sum_{j=1}^{n} A_{ji}\Delta_{ji}.$$

Proof.

$$\det A = \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1,\cdots,\overline{j_i},\cdots j_n}^{n} \varepsilon_{j_1 j_2 \cdots j_n} A_{j_1 1} A_{j_2 2} \cdots \overline{A_{j_i i}} \cdots A_{j_n n}$$

Let $\sigma \in S_n$ be the permutation which moves j_i to the ith position, and leave everything else in its natural order, i.e.

$$\sigma = \begin{pmatrix} 1 & \cdots & i & i+1 & i+2 & \cdots & j_i-1 & j_i & j_i+1 & \cdots & n \\ 1 & \cdots & j_i & i & i+1 & \cdots & j_i-2 & j_i-1 & j_i+1 & \cdots & n \end{pmatrix}$$

if $j_i > i$, and similarly for other cases. To perform this permutation, $|i - j_i|$ transpositions are made. So $\varepsilon(\sigma) = (-1)^{i-j_i}$.

Now consider the permutation $\rho \in S_n$

$$\rho = \begin{pmatrix} 1 & \cdots & \cdots & \overline{j_i} & \cdots & n \\ j_1 & \cdots & \overline{j_i} & \cdots & \cdots & j_n \end{pmatrix}$$

The composition $\rho\sigma$ reorders $(1, \cdots, n)$ to $(j_1, j_2, \cdots, j_n)$. So $\varepsilon(\rho\sigma) = \varepsilon_{j_1\cdots j_n} = \varepsilon(\rho)\varepsilon(\sigma) = (-1)^{i-j_i}\varepsilon_{j_1\cdots\overline{j}_i\cdots j_n}$. Hence the original equation becomes

$$\begin{aligned} \det A &= \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1\cdots\overline{j}_i\cdots j_n} (-1)^{i-j_i}\varepsilon_{j_1\cdots\overline{j}_i\cdots j_n} A_{j_1 1}\cdots\overline{A_{j_i i}}\cdots A_{j_n n} \\ &= \sum_{j_i=1}^{n} A_{j_i i}(-1)^{i-j_i} M_{j_i i} \\ &= \sum_{j_i=1}^{n} A_{j_i i}\Delta_{j_i i} \\ &= \sum_{j=1}^{n} A_{ji}\Delta_{ji} \end{aligned}$$

$\square$

Example 3.5.2. $\det A = \begin{vmatrix} 2 & 4 & 2 \\ 3 & 2 & 1 \\ 2 & 0 & 1 \end{vmatrix}$. *We can pick the first row and have*

$$\begin{aligned} \det A &= 2\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} - 4\begin{vmatrix} 3 & 1 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 3 & 2 \\ 2 & 0 \end{vmatrix} \\ &= 2(2-0) - 4(3-2) + 2(0-4) \\ &= -8. \end{aligned}$$

Alternatively, we can pick the second column and have

$$\begin{aligned}\det A &= -4\begin{vmatrix}3 & 1\\ 2 & 1\end{vmatrix} + 2\begin{vmatrix}2 & 2\\ 2 & 1\end{vmatrix} - 0\begin{vmatrix}2 & 2\\ 3 & 1\end{vmatrix}\\ &= -4(3-2) + 2(2-4) - 0\\ &= -8.\end{aligned}$$

In practical terms, we use a combination of properties of determinants with a sensible choice of i to evaluate $\det(A)$.

Example 3.5.3. *Consider* $\begin{vmatrix}1 & a & a^2\\ 1 & b & b^2\\ 1 & c & c^2\end{vmatrix}$. *row 1 - row 2 gives*

$$\begin{vmatrix}0 & a-b & a^2-b^2\\ 1 & b & b^2\\ 1 & c & c^2\end{vmatrix} = (a-b)\begin{vmatrix}0 & 1 & a+b\\ 1 & b & b^2\\ 1 & c & c^2\end{vmatrix}.$$

Do row 2 - row 3. We obtain

$$(a-b)(b-c)\begin{vmatrix}0 & 1 & a+b\\ 0 & 1 & b+c\\ 1 & c & c^2\end{vmatrix}.$$

row 1 - row 2 gives

$$(a-b)(b-c)(a-c)\begin{vmatrix}0 & 0 & 1\\ 0 & 1 & b+c\\ 1 & c & c^2\end{vmatrix} = (a-b)(b-c)(a-c).$$

CHAPTER 4

Linear equations

The 'real' world can often be described by equations of various sorts. Some models result in linear equations of the type studied here. However, even when the real world results in more complicated models, the solution of these more complicated models often involves the solution of linear equations.

4.1 Inverse of an $n \times n$ matrix

Recall that the determinant of an $n \times n$ matrix A can be expressed in terms of the Laplace expansion formula

$$\det A = \sum_{k=1}^{n} A_{Ik}\Delta_{Ik} \qquad \text{for any } 1 \leq I \leq n \tag{4.1}$$

$$= \sum_{k=1}^{n} A_{kJ}\Delta_{kJ} \qquad \text{for any } 1 \leq J \leq n \tag{4.2}$$

Lemma 4.1.1.

$$\sum A_{ik}\Delta_{jk} = \delta_{ij}\det A$$

Proof. If $i \neq j$, then consider an $n \times n$ matrix B, which is identical to A except the jth row is replaced by the ith row of A. So Δ_{jk} of $B = \Delta_{jk}$ of A, since Δ_{jk} does not depend on the elements in row j. Since B has a duplicate row, we know that

$$0 = \det B = \sum_{k=1}^{n} B_{jk}\Delta_{jk} = \sum_{k=1}^{n} A_{ik}\Delta_{jk}.$$

If $i = j$, then the expression is $\det A$ by the Laplace expansion formula. ☐

Theorem 4.1.1. *If* $\det A \neq 0$*, then* A^{-1} *exists and is given by*

$$(A^{-1})_{ij} = \frac{\Delta_{ji}}{\det A}.$$

Proof.

$$(A^{-1})_{ik}A_{kj} = \frac{\Delta_{ki}}{\det A}A_{kj} = \frac{\delta_{ij}\det A}{\det A} = \delta_{ij}.$$

So $A^{-1}A = I$. ☐

The other direction is easy to prove. If $\det A = 0$, then it has no inverse, since for any matrix B, $\det AB = 0$, and hence AB cannot be the identity.

Example 4.1.1. *Consider the shear matrix* $S_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. *We have* $\det S_\lambda =$

1. *The cofactors are*

$$\begin{aligned} &\Delta_{11} = 1 \quad && \Delta_{12} = 0 \quad && \Delta_{13} = 0 \\ &\Delta_{21} - \lambda \quad && \Delta_{22} = 1 \quad && \Delta_{23} = 0 \\ &\Delta_{31} = 0 \quad && \Delta_{32} = 0 \quad && \Delta_{33} = 1 \end{aligned}$$

So $S_\lambda^{-1} = \begin{pmatrix} 1 & -\lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

How many arithmetic operations does it take to calculate the inverse of an $n \times n$ matrix? Let's count only the multiplication operations, as they are the most time-consuming.

Suppose calculating $\det A$ takes f_n multiplications. This involves n $(n-1) \times (n-1)$ determinants, and you need n more multiplications to put them together. Therefore, $f_n = nf_{n-1} + n$. Hence, $f_n = O(n!)$ (in fact, $f_n \approx (1+e)n!$).

To find the inverse, we need to calculate n^2 cofactors. Each cofactor is a $(n-1)$-determinant, and each takes $O((n-1)!)$. So the time complexity is $O(n^2(n-1)!) = O(n \cdot n!)$.

This is incredibly slow. Hence, while it is theoretically possible to solve systems of linear equations by inverting a matrix, most people do not do so in practice. Instead, we develop better methods to solve the equations. In fact, the "usual" method people use to solve equations by hand only has complexity $O(n^3)$, which is much faster.

4.2 Solving linear equations

Consider $A\mathbf{x} = \mathbf{d}$ where A is an $m \times n$ matrix, $\mathbf{x}$ and $\mathbf{d}$ are $m \times 1$ column vectors.

Definition 4.2.1 (Inhomogeneous equation)**.** *If $\mathbf{d} \neq \mathbf{0}$ then the system of equations is said to be a system of inhomogeneous equations.*

Definition 4.2.2 (Homogeneous equation)**.** *If $\mathbf{d} = \mathbf{0}$, then the system is homogeneous.*

Suppose $\det A \neq 0$. Then there exists a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ ($\mathbf{x} = \mathbf{0}$ for the homogeneous case).

To understand this result, recall that $\det A \neq 0$ means that the columns of A are linearly independent. These columns are the images of the standard basis vectors, $\mathbf{e}'_i = A\mathbf{e_i}$. Therefore, $\det A \neq 0$ implies that $\mathbf{e}'_i$ are linearly independent and form a basis of $\mathbb{R}^n$. This means that the image of A is the entire $\mathbb{R}^n$, so there exists a solution to the equation $A\mathbf{x} = \mathbf{b}$.

To show that the solution is unique, suppose $\mathbf{x}$ and $\mathbf{x}'$ are both solutions. Then

$A\mathbf{x} = A\mathbf{x}' = \mathbf{b}$. So $A(\mathbf{x} - \mathbf{x}') = \mathbf{0}$. This implies that $\mathbf{x} - \mathbf{x}'$ is in the kernel of A. Since the rank of A is n, the rank-nullity theorem tells us that the nullity is 0, which means the kernel is trivial. Therefore, we have $\mathbf{x} - \mathbf{x}' = \mathbf{0}$, i.e., $\mathbf{x} = \mathbf{x}'$.

4.2.1 Gaussian elimination

Consider the following general solution:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= d_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= d_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= d_m \end{aligned}$$

This system consists of m equations and n unknowns.

Assuming $A_{11} \neq 0$ (if it is, we can reorder the equations), we can eliminate x_1 from the remaining $(m-1)$ equations using the first equation. Then we can use the second equation to eliminate x_2 from the remaining $(m-2)$ equations. This process can be repeated until we are left with the following system:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1n}x_n &= d_1 \\ A_{22}^{(2)}x_2 + A_{23}^{(2)}x_3 + \cdots + A_{2n}^{(2)}x_n &= d_2 \\ &\vdots \\ A_{rr}^{(r)}x_r + \cdots + A_{rn}^{(r)}x_n &= d_r \\ 0 &= d_{r+1}^{(r)} \\ &\vdots \\ 0 &= d_m^{(r)} \end{aligned}$$

Here, $A_{ii}^{(i)} \neq 0$ (which we can achieve by reordering), and the superscript (i) refers to the "version number" of the coefficient, e.g., $A_{22}^{(2)}$ is the second version of the coefficient of x_2 in the second row.

Now let's consider the different possibilities for this system:

1. If $r < m$ and at least one of $d^{(r)} * r + 1, \cdots d_m^{(r)} \neq 0$, then the system is overdetermined and inconsistent with no solution.

Example 4.2.1. *Consider the following system:*

$$3x_1 + 2x_2 + x_3 = 3$$
$$6x_1 + 3x_2 + 3x_3 = 0$$
$$6x_1 + 2x_2 + 4x_3 = 6$$

After eliminating x_1, we are left with:

$$3x_1 + 2x_2 + x_3 = 3$$
$$0 - x_2 + x_3 = -6$$
$$0 - 2x_2 + 2x_3 = 0$$

Further elimination leads to:

$$3x_1 + 2x_2 + x_3 = 3$$
$$0 - x_2 + x_3 = -6$$
$$0 = 12$$

Since $d_3^{(3)} = 12 = 0$, there is no solution to this system.

2. If $r = n \leq m$, and all $d * r + i^{(r)} = 0$, then there is a unique solution for $x_n = d_n^{(n)}/A_{nn}^{(n)}$ and hence for all x_i by back substitution. This system is *determined.*

Example 4.2.2. *Consider the following system:*

$$2x_1 + 5x_2 = 2$$
$$4x_1 + 3x_2 = 11$$

After elimination, we are left with:

$$2x_1 + 5x_2 = 2$$
$$-7x_2 = 7$$

So $x_2 = -1$ and $x_1 = 7/2$, giving us a unique solution.

3. If $r < n$ and $d_{r+i}^{(r)} = 0$, then $x_{r+1}, \cdots x_n$ can be freely chosen, and there are infinitely many solutions. This system is underdetermined. For example:

$$x_1 + x_2 = 1$$
$$2x_1 + 2x_2 = 2$$

After elimination, we are left with:

$$x_1 + x_2 = 1$$
$$0 = 0$$

So $x_1 = 1 - x_2$ is a solution for any x_2.

In the case where $m = n$ and A is square, the determinant can be used to solve the system. Since row operations do not change the determinant and swapping rows gives a factor of (-1), we can write:

$$\det A = (-1)^k \begin{vmatrix} A_{11} & A_{12} & \cdots & \cdots & \cdots & A_{1n} \\ 0 & A_{22}^{(2)} & \cdots & \cdots & \cdots & A_{2n}^{(n)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{rr}^{(r)} & \cdots & A_{rn}^{(n)} \\ 0 & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

This determinant is upper triangular (all elements below diagonal are 0) and the determinant is the product of its diagonal elements.

If $r < n$ (and $d_i^{(r)} = 0$ for $i > r$), then we have case (ii) and $\det A = 0$. If $r = n$, then $\det A = (-1)^k A_{11} A_{22}^{(2)} \cdots A_{nn}^{(n)} \neq 0$.

In the case where $n = m$, this method takes $O(n^3)$ operations, which is much less than inverting the matrix. Therefore, this is an efficient way of solving equations.

4.3 The rank of a matrix

Consider a linear map $\alpha : \mathbb{R}^n \to \mathbb{R}^m$. Recall the rank $r(\alpha)$ is the dimension of the image. Suppose that the matrix A is associated with the linear map. We also call $r(A)$ the rank of A.

Recall that if the standard basis is $\mathbf{e}_1, \cdots \mathbf{e}_n$, then $A\mathbf{e}_1, \cdots, A\mathbf{e}_n$ span the image (but not necessarily linearly independent). Further, $A\mathbf{e}_1, \cdots, A\mathbf{e}_n$ are the columns of the matrix A. Hence $r(A)$ is the number of linearly independent columns.

Definition 4.3.1 (Column and row rank of linear map). *The column rank of a matrix is the maximum number of linearly independent columns.*

The row rank of a matrix is the maximum number of linearly independent rows.

Theorem 4.3.1. *The column rank and row rank are equal for any $m \times n$ matrix.*

Proof. Let r be the row rank of A. Write the biggest set of linearly independent rows as $\mathbf{v}_1^T, \mathbf{v}_2^T, \cdots \mathbf{v}_r^T$ or in component form $\mathbf{v}_k^T = (v_{k1}, v_{k2}, \cdots, v_{kn})$ for $k = 1, 2, \cdots, r$.

Now denote the ith row of A as $\mathbf{r}_i^T = (A_{i1}, A_{i2}, \cdots A_{in})$.

Note that every row of A can be written as a linear combination of the $\mathbf{v}$'s. (If $\mathbf{r_i}$ cannot be written as a linear combination of the $\mathbf{v}$'s, then it is independent of the $\mathbf{v}$'s and $\mathbf{v}$ is not the maximum collection of linearly independent rows) Write

$$\mathbf{r}_i^T = \sum_{k=1}^{r} C_{ik}\mathbf{v}_k^T.$$

For some coefficients C_{ik} with $1 \le i \le m$ and $1 \le k \le r$.

Now the elements of A are

$$A_{ij} = (\mathbf{r}_i)_j^T = \sum_{k=1}^{r} C_{ik}(\mathbf{v}_k)_j,$$

or

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix} = \sum_{k=1}^{r} \mathbf{v}_{kj} \begin{pmatrix} C_{1k} \\ C_{2k} \\ \vdots \\ C_{mk} \end{pmatrix}$$

So every column of A can be written as a linear combination of the r column vectors $\mathbf{c}_k$. Then the column rank of $A \le r$, the row rank of A.

Apply the same argument to A^T to see that the row rank is $\le$ the column rank. $\square$

4.4 The homogeneous problems $A\mathbf{x} = 0$

We will restrict our attention to the square case where the number of unknowns is equal to the number of equations, i.e., A is an $n \times n$ matrix. Our goal is to solve $A\mathbf{x} = \mathbf{0}$.

Firstly, if $\det A \neq 0$, then A^{-1} exists and $\mathbf{x}^{-1} = A^{-1}\mathbf{0} = \mathbf{0}$, which is the unique solution. Therefore, if $A\mathbf{x} = \mathbf{0}$ with $\mathbf{x} \neq \mathbf{0}$, then $\det A = 0$.

4.4.1 Geometrical interpretation of $A\mathbf{x} = \mathbf{0}$

We will consider a 3×3 matrix

$$A = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{pmatrix}$$

$A\mathbf{x} = \mathbf{0}$ means that $\mathbf{r}_i \cdot \mathbf{x} = 0$ for all i. Each equation $\mathbf{r}_i \cdot \mathbf{x} = 0$ represents a plane through the origin. Therefore, the solution is the intersection of the three planes.

There are three possibilities:

1. If $\det A = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3] \neq 0$, then $\text{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = \mathbb{R}^3$ and thus $r(A) = 3$. By the rank-nullity theorem, $n(A) = 0$ and the kernel is $\{\mathbf{0}\}$. Therefore, $\mathbf{x} = \mathbf{0}$ is the unique solution.

2. If $\det A = 0$, then $\dim(\text{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}) = 1$ or 2.

   (a) If rank $= 2$, without loss of generality, assume $\mathbf{r}_1, \mathbf{r}_2$ are linearly independent. Therefore, $\mathbf{x}$ lies on the intersection of two planes $\mathbf{x} \cdot \mathbf{r}_1 = 0$ and $\mathbf{x} \cdot \mathbf{r}_2 = 0$, which is the line $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda \mathbf{r}_1 \times \mathbf{r}_2\}$ (Since $\mathbf{x}$ lies on the intersection of the two planes, it has to be normal to the normals of both planes). All such points on this line also satisfy $\mathbf{x} \cdot \mathbf{r}_3 = 0$ since $\mathbf{r}_3$ is a linear combination of $\mathbf{r}_1$ and $\mathbf{r}_2$. The kernel is a line, $n(A) = 1$.

   (b) If rank $= 1$, then $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ are parallel. Therefore, $\mathbf{x} \cdot \mathbf{r}_1 = 0 \Rightarrow \mathbf{x} \cdot \mathbf{r}_2 = \mathbf{x} \cdot \mathbf{r}_3 = 0$. All $\mathbf{x}$ that satisfy $\mathbf{x} \cdot \mathbf{r}_1 = 0$ are in the kernel, and the kernel is now a plane. $n(A) = 2$.

(We also have the trivial case where $r(A) = 0$, we have the zero mapping and the kernel is $\mathbb{R}^3$)

4.4.2 Linear mapping view of $A\mathbf{x} = \mathbf{0}$

In the general case, consider a linear map $\alpha : \mathbb{R}^n \to \mathbb{R}^n$ $\mathbf{x} \mapsto \mathbf{x}' = A\mathbf{x}$. The kernel $k(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$ has dimension $n(A)$.

1. If $n(A) = 0$, then $A(\mathbf{e}_1), A(\mathbf{e}_2), \cdots, A(\mathbf{e}_n)$ is a linearly independent set, and $r(A) = n$.

2. If $n(A) > 0$, then the image is not the whole of $\mathbb{R}^n$. Let $\{\mathbf{u}_i\}, i = 1, \cdots, n(A)$ be a basis of the kernel, i.e. so given any solution to $A\mathbf{x} = \mathbf{0}$, $\mathbf{x} = \sum_{i=1}^{n(A)} \lambda_i \mathbf{u}_i$ for some λ_i. Extend $\{\mathbf{u}_i\}$ to be a basis of $\mathbb{R}^n$ by introducing extra vectors $\mathbf{u}_i$ for $i = n(A) + 1, \cdots, n$. The vectors $A(\mathbf{u}_i)$ for $i = n(A) + 1, \cdots, n$ form a basis of the image.

4.5 General solution of $A\mathbf{x} = \mathbf{d}$

Finally consider the general equation $A\mathbf{x} = \mathbf{d}$, where A is an $n \times n$ matrix and $\mathbf{x}, \mathbf{d}$ are $n \times 1$ column vectors. We can separate into two main cases.

1. $\det(A) \neq 0$. So A^{-1} exists and $n(A) = 0$, $r(A) = n$. Then for any $\mathbf{d} \in \mathbb{R}^n$, a unique solution must exists and it is $\mathbf{x} = A^{-1}\mathbf{d}$.

2. $\det(A) = 0$. Then A^{-1} does not exist, and $n(A) > 0$, $r(A) < n$. So the image of A is not the whole of $\mathbb{R}^n$.

   (a) If $\mathbf{d} \notin \operatorname{im} A$, then there is no solution (by definition of the image)

   (b) If $\mathbf{d} \in \operatorname{im} A$, then by definition there exists at least one $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{d}$. The general solution of $A\mathbf{x} = \mathbf{d}$ can be written as $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$, where $\mathbf{x}_0$ is a particular solution (i.e. $A\mathbf{x}_0 = \mathbf{d}$), and $\mathbf{y}$ is any vector in $\ker A$ (i.e. $A\mathbf{y} = \mathbf{0}$). (cf. Isomorphism theorem)

   If $n(A) = 0$, then $\mathbf{y} = \mathbf{0}$ only, and then the solution is unique (i.e. case (i)).

If $n(A) > 0$, then $\{\mathbf{u}_i\}, i = 1, \cdots, n(A)$ is a basis of the kernel. Hence

$$\mathbf{y} = \sum_{j=1}^{n(A)} \mu_j \mathbf{u}_j,$$

so

$$\mathbf{x} = \mathbf{x}_0 + \sum_{j=1}^{n(A)} \mu_j \mathbf{u}_j$$

for any μ_j, i.e. there are infinitely many solutions.

Example 4.5.1.

$$\begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ b \end{pmatrix}$$

We have $\det A = 1 - a$. *If* $a \neq 1$, *then* A^{-1} *exists and*

$$A^{-1} = \frac{1}{1-a} = \frac{1}{1-a} \begin{pmatrix} 1 & -1 \\ -a & 1 \end{pmatrix}.$$

Then

$$\mathbf{x} = \frac{1}{1-a} \begin{pmatrix} 1-b \\ -a+b \end{pmatrix}.$$

If $a = 1$, *then*

$$A\mathbf{x} = \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix} = (x_1 + x_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So $\operatorname{im} A = \operatorname{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ *and* $\ker A = \operatorname{span} \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$. *If* $b \neq 1$, *then* $\begin{pmatrix} 1 \\ b \end{pmatrix} \notin \operatorname{im} A$ *and there is no solution. If* $b = 1$, *then* $\begin{pmatrix} 1 \\ b \end{pmatrix} \in \operatorname{im} A$.

We find a particular solution of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. *So The general solution is*

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Example 4.5.2. *Find the general solution of*

$$\begin{pmatrix} a & a & b \\ b & a & a \\ a & b & a \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$$

We have $\det A = (a-b)^2(2a+b)$. *If* $a \neq b$ *and* $b \neq -2a$, *then the inverse exists and there is a unique solution for any* c. *Otherwise, the possible cases are*

1. $a = b, b \neq -2a$. *So* $a \neq 0$. *The kernel is the plane* $x + y + z = 0$ *which is* $\operatorname{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$ *We extend this basis to* $\mathbb{R}^3$ *by adding* $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

   So the image is the span of $\begin{pmatrix} a \\ a \\ a \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. *Hence if* $c \neq 1$, *then* $\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$ *is not in the image and there is no solution. If* $c = 1$, *then a particular solution is* $\begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix}$ *and the general solution is*

   $$\mathbf{x} = \begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

2. *If* $a \neq b$ *and* $b = -2a$, *then* $a \neq 0$. *The kernel satisfies*

   $$\begin{aligned} x + y - 2z &= 0 \\ -2x + y + z &= 0 \\ x - 2y + z &= 0 \end{aligned}$$

   This can be solved to give $x = y = z$, *and the kernel is* $\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$. *We add* $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ *to form a basis of* $\mathbb{R}^3$. *So the image is the span of*

$$\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

If $\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$ *is in the image, then*

$$\begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

Then the only solution is $\mu = 0, \lambda = 1, c = -2$. *Thus there is no solution if* $c \neq -2$, *and when* $c = -2$, *pick a particular solution* $\begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix}$ *and the general solution is*

$$\mathbf{x} = \begin{pmatrix} \frac{1}{a} \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

3. *If* $a = b$ *and* $b = -2a$, *then* $a = b = 0$ *and* $\ker A = \mathbb{R}^3$. *So there is no solution for any* c.

CHAPTER 5

Eigenvalues and eigenvectors

The energy levels in quantum mechanics are eigenvalues, ionization potentials in Hartree-Fock theory are eigenvalues, the principal moments of inertia of an inertia tensor are eigenvalues, the resonance frequencies in mechanical systems (like a violin string, or a strand of DNA held fixed by optical tweezers, or the Tacoma Narrows bridge are eigenvalues, etc.

5.1 Motivation and basic definitions

Theorem 5.1.1 (Fundamental theorem of algebra). *Let $p(z)$ be a polynomial of degree (or order) $m \geq 1$, i.e.*

$$p(z) = \sum_{j=0}^{m} c_j z^j$$

where $c_j \in \mathbb{C}$ and $c_m \neq 0$. Then the Fundamental Theorem of Algebra states that *the equation*

$$p(z) = \sum_{j=0}^{m} c_j z^j = 0 \tag{5.1}$$

has a solution in $\mathbb{C}$. *In other words,* $p(z) = 0$ *has precisely* m *(not necessarily distinct) roots in the complex plane, accounting for multiplicity.*

This will be proved when you come to study complex variable theory. Note that we have the disclaimer "accounting for multiplicity". For example, $x^2 - 2x + 1 = 0$ has only one distinct root, 1, but we say that this root has multiplicity 2, and is thus counted twice.

Corollary 5.1.1. *If* $m \geq 1$, $c_j \in \mathbb{C}$ $(j = 0, 1, \ldots, m)$ *and* $c_m \neq 0$, *then we can find* $\omega_1, \omega_2, \ldots, \omega_m \in \mathbb{C}$ *such that*

$$p(z) = \sum_{j=0}^{m} c_j z^j = c_m \prod_{j=1}^{m} (z - \omega_j) \tag{5.2}$$

Formally, multiplicity is defined as follows:

Definition 5.1.1 (Multiplicity of root)**.** *The root* $z = \omega$ *has multiplicity* k *if* $(z - \omega)^k$ *is a factor of* $p(z)$ *but* $(z - \omega)^{k+1}$ *is not.*

A complex polynomial of degree m has precisely m roots (each counted with its multiplicity).

Example 5.1.1. *Let* $p(z) = z^3 - z^2 - z + 1 = (z - 1)^2(z + 1)$. *So* $p(z) = 0$ *has roots* $1, 1, -1$, *where* $z = 1$ *has multiplicity* 2.

Eigenvalues and eigenvectors of maps

Definition 5.1.2 (Eigenvector and eigenvalue)**.** *Let* $\alpha : \mathbb{C}^n \to \mathbb{C}^n$ *be a linear map with associated matrix* A. *Then* $\mathbf{x} \neq \mathbf{0}$ *is an eigenvector of* A *if*

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

for some λ. λ *is the associated eigenvalue. This means that the direction of the* *eigenvector is preserved by the mapping, but is scaled up by λ.*

The following it is noteworthy:

1. Let ℓ be the subspace (or line in $\mathbb{R}^n$) defined by span $\{\mathbf{x}\}$. Then $\mathcal{A}(\ell) \subseteq \ell$, i.e., ℓ is an invariant subspace under $\mathcal{A}$.

2. If $\mathcal{A}(\mathbf{x}) = \lambda\mathbf{x}$ then

$$\mathcal{A}^m(\mathbf{x}) = \lambda^m\mathbf{x}$$

and

$$(c_0 + c_1\mathcal{A} + \ldots c_m\mathcal{A}^m)\mathbf{x} = (c_0 + c_1\lambda + \ldots + c_m\lambda^m)\mathbf{x}$$

where $c_j \in \mathbb{F}$ $(j = 0, 1, \ldots, m)$. Let $p(z)$ be the polynomial

$$p(z) = \sum_{j=0}^{m} c_j z^j$$

Then for any polynomial p of $\mathcal{A}$, and an eigenvector $\mathbf{x}$ of $\mathcal{A}$,

$$p(\mathcal{A})\mathbf{x} = p(\lambda)\mathbf{x}$$

Eigenvalues and eigenvectors of matrices

Suppose that A is the square $n \times n$ matrix associated with the map $\mathcal{A}$ for a given basis. Then consistent with the definition for maps, if $A\mathbf{x} = \lambda\mathbf{x}$ for some non-zero vector $\mathbf{x} \in \mathbb{F}^n$ and $\lambda \in \mathbb{F}$, we say that $\mathbf{x}$ is an eigenvector of A with eigenvalue λ. This equation can be written as

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{5.3}$$

Theorem 5.1.2. *λ is an eigenvalue of A iff*

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

Proof. $(\Rightarrow)$ Suppose that λ is an eigenvalue and $\mathbf{x}$ is the associated eigenvector. We can rearrange the equation in the definition above to

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

and thus

$$\mathbf{x} \in \ker(A - \lambda I)$$

But $\mathbf{x} \neq \mathbf{0}$. So $\ker(A - \lambda I)$ is non-trivial and $\det(A - \lambda I) = 0$. The $(\Leftarrow)$ direction is similar. $\square$

Definition 5.1.3 (Characteristic equation of matrix)**.** *The* characteristic equation *of A is*

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

Definition 5.1.4 (Characteristic polynomial of matrix)**.** *The* characteristic polynomial *of A is*

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

From the definition of the determinant,

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \\ &= \varepsilon_{j_1 j_2 \cdots j_n}(A_{j_1 1} - \lambda\delta_{j_1 1}) \cdots (A_{j_n n} - \lambda\delta_{j_n n}) \\ &= c_0 + c_1\lambda + \cdots + c_n\lambda^n \end{aligned}$$

for some constants $c_0, \cdots, c_n$. From this, we see that

1. $p_A(\lambda)$ has degree n and has n roots. So an $n \times n$ matrix has n eigenvalues (accounting for multiplicity).

2. If A is real, then all $c_i \in \mathbb{R}$. So eigenvalues are either real or come in complex conjugate pairs.

3. $c_n = (-1)^n$ and $c_{n-1} = (-1)^{n-1}(A_{11} + A_{22} + \cdots + A_{nn}) = (-1)^{n-1}\operatorname{tr}(A)$. But c_{n-1} is the sum of roots, i.e. $c_{n-1} = (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots \lambda_n)$, so

   $$\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

   Finally, $c_0 = p_A(0) = \det(A)$. Also c_0 is the product of all roots, i.e. $c_0 = \lambda_1\lambda_2 \cdots \lambda_n$. So

   $$\det A = \lambda_1\lambda_2 \cdots \lambda_n.$$

The kernel of the matrix $A - \lambda I$ is the set $\{\mathbf{x} : A\mathbf{x} = \lambda\mathbf{x}\}$. This is a vector subspace because the kernel of any map is always a subspace.

5.2 Eigenspaces and diagonal matrices

5.2.1 Eigenspaces and multiplicity

Definition 5.2.1 (Eigenspace)**.** *The eigenspace denoted by E_λ is the kernel of the matrix $A - \lambda I$, i.e. the set of eigenvectors with eigenvalue λ.*

Definition 5.2.2 (Algebraic multiplicity of eigenvalue)**.** *The algebraic multiplicity $M(\lambda)$ or M_λ of an eigenvalue λ is the multiplicity of λ in $p_A(\lambda) = 0$. By the fundamental theorem of algebra,*

$$\sum_\lambda M(\lambda) = n.$$

If $M(\lambda) > 1$, then the eigenvalue is degenerate.

Definition 5.2.3 (Geometric multiplicity of eigenvalue)**.** *The geometric multiplicity $m(\lambda)$ or m_λ of an eigenvalue λ is the dimension of the eigenspace, i.e. the maximum number of linearly independent eigenvectors with eigenvalue λ.*

Definition 5.2.4 (Defect of eigenvalue)**.** *The defect Δ_λ of eigenvalue λ is*

$$\Delta_\lambda = M(\lambda) - m(\lambda).$$

It can be proven that $\Delta_\lambda \geq 0$, i.e. the geometric multiplicity is never greater than the algebraic multiplicity.

5.2.2 Linearly independent eigenvectors

Theorem 5.2.1. *Suppose $n \times n$ matrix A has distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then the corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly independent.*

Proof. Proof by contradiction: Suppose $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly dependent. Then we can find non-zero constants d_i for $i = 1, 2, \cdots, r$, such that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_r\mathbf{x}_r = \mathbf{0}.$$

Suppose that this is the shortest non-trivial linear combination that gives $\mathbf{0}$ (we may need to re-order $\mathbf{x}_i$).

Now apply $(A - \lambda_1 I)$ to the whole equation to obtain

$$d_1(\lambda_1 - \lambda_1)\mathbf{x}_1 + d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0}.$$

We know that the first term is $\mathbf{0}$, while the others are not (since we assumed $\lambda_i \neq \lambda_j$ for $i \neq j$). So

$$d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0},$$

and we have found a shorter linear combination that gives $\mathbf{0}$. Contradiction. □

Example 5.2.1.

1. $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. *Then* $p_A(\lambda) = \lambda^2 + 1 = 0$. *So* $\lambda_1 = i$ *and* $\lambda_2 = -i$.

   To solve $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$, *we obtain*

$$\begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{0}.$$

   So we obtain

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

   to be an eigenvector. Clearly any scalar multiple of $\begin{pmatrix} 1 \\ i \end{pmatrix}$ *is also a solution, but still in the same eigenspace* $E_i = \operatorname{span}\begin{pmatrix} 1 \\ i \end{pmatrix}$

Solving $(A - \lambda_2 I)\mathbf{x} = \mathbf{0}$ *gives*

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

So $E_{-i} = \operatorname{span}\begin{pmatrix} 1 \\ -i \end{pmatrix}$.

Note that $M(\pm i) = m(\pm i) = 1$*, so* $\Delta_{\pm i} = 0$*. Also note that the two eigenvectors are linearly independent and form a basis of* $\mathbb{C}^2$*.*

2. *Consider*

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$$

Then $\det(A - \lambda I) = 0$ *gives* $45 + 21\lambda - \lambda^2 - \lambda^3$*. So* $\lambda_1 = 5, \lambda_2 = \lambda_3 = -3$*. The eigenvector with eigenvalue* 5 *is*

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

We can find that the eigenvectors with eigenvalue -3 *are*

$$\mathbf{x} = \begin{pmatrix} -2x_2 + 3x_3 \\ x_2 \\ x_3 \end{pmatrix}$$

for any x_2, x_3*. This gives two linearly independent eigenvectors, say* $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$.

So $M(5) = m(5) = 1$ *and* $M(-3) = m(-3) = 2$*, and there is no defect for both of them. Note that these three eigenvectors form a basis of* $\mathbb{C}^3$*.*

3. *Let*

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

Then $0 = p_A(\lambda) = -(\lambda+2)^4$. *So* $\lambda = -2, -2, -2$. *To find the eigenvectors, we have*

$$(A + 2I)\mathbf{x} = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ -2 & -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

The general solution is thus $x_1 + x_2 - x_3 = 0$, *and the general solution is thus* $x = \begin{pmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{pmatrix}$. *The eigenspace* $E_{-2} = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\}$.

Hence $M(-2) = 3$ *and* $m(-2) = 2$. *Thus the defect* $\Delta_{-2} = 1$. *So the eigenvectors do not form a basis of* $\mathbb{C}^3$.

4. *Consider the reflection* R *in the plane with normal* $\mathbf{n}$. *Clearly* $R\mathbf{n} = -\mathbf{n}$. *The eigenvalue is* -1 *and the eigenvector is* $\mathbf{n}$. *Then* $E_1 = \operatorname{span}\{\mathbf{n}\}$. *So* $M(-1) = m(-1) = 1$.

   If $\mathbf{p}$ *is any vector in the plane,* $R\mathbf{p} = \mathbf{p}$. *So this has an eigenvalue of* 1 *and eigenvectors being any vector in the plane. So* $M(1) = m(1) = 2$.

   So the eigenvectors form a basis of $\mathbb{R}^3$.

5. *Consider a rotation* R *by* θ *about* $\mathbf{n}$. *Since* $R\mathbf{n} = \mathbf{n}$, *we have an eigenvalue of* 1 *and eigenspace* $E_1 = \operatorname{span}\{\mathbf{n}\}$.

   We know that there are no other real eigenvalues since rotation changes the direction of any other vector. The other eigenvalues turn out to be $e^{\pm i\theta}$. *If* $\theta \neq 0$, *there are 3 distinct eigenvalues and the eigenvectors form a basis of* $\mathbb{C}^3$.

6. *Consider a shear*

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

   The characteristic equation is $(1-\lambda)^2 = 0$ *and* $\lambda = 1$. *The eigenvectors corresponding to* $\lambda = 1$ *is* $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. *We have* $M(1) = 2$ *and* $m(1) = 1$. *So* $\Delta_1 = 1$.

5.2.3 Diagonal matrices

Definition 5.2.5 (Diagonal matrix)**.** *If $n \times n$ matrix A has n distinct eigenvalues, and hence has n linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n$, then with respect to this eigenvector basis, A is diagonal.*

In this basis, $v_1 = (1, 0, \cdots, 0)$ etc. We know that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ (no summation). So the image of the ith basis vector is λ_i times the ith basis. Since the columns of A are simply the images of the basis,

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

The fact that A can be diagonalized by changing the basis is an important observation. We will now look at how we can change bases and see how we can make use of this.

5.3 Change of basis

We have now identified as least two types of 'nice' bases, i.e. orthonormal bases and bases of eigenvectors. A linear map A does not change if we change basis, but the matrix representing it does. The aim of this section, which is somewhat of a diversion from our study of eigenvalues and eigenvectors, is to work out how the elements of a matrix transform under a change of basis.

5.3.1 Transformation matrices

Let $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ and $\{\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_n\}$ be 2 different bases of $\mathbb{R}^n$ or $\mathbb{C}^n$. Then we can write

$$\tilde{\mathbf{e}_j} = \sum_{i=1}^{n} P_{ij}\mathbf{e}_i$$

i.e. P_{ij} is the ith component of $\tilde{\mathbf{e}_j}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$. Note that the sum is made as $P_{ij}\mathbf{e}_i$, not $P_{ij}\mathbf{e}_j$. This is different from the formula for matrix multiplication.

Matrix P has as its columns the vectors $\tilde{\mathbf{e}_j}$ relative to $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$. So $P = (\tilde{\mathbf{e}_1}\ \tilde{\mathbf{e}_2}\ \cdots\ \tilde{\mathbf{e}_n})$ and

$$P(\mathbf{e}_i) = \tilde{\mathbf{e}_i}$$

Similarly, we can write

$$\mathbf{e}_i = \sum_{k=1}^{n} Q_{ki} \tilde{\mathbf{e}_k}$$

with $Q = (\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_n)$.

5.3.2 Properties of transformation matrices

Substituting this into the equation for $\tilde{\mathbf{e}_j}$, we have

$$\begin{aligned}\tilde{\mathbf{e}_j} &= \sum_{i=1}^{n} \left(\sum_{k=1}^{n} Q_{ki} \tilde{\mathbf{e}_k} \right) P_{ij} \\ &= \sum_{k=1}^{n} \tilde{\mathbf{e}_k} \left(\sum_{i=1}^{n} Q_{ki} P_{ij} \right)\end{aligned}$$

But $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \cdots, \tilde{\mathbf{e}}_n$ are linearly independent, so this is only possible if

$$\sum_{i=1}^{n} Q_{ki} P_{ij} = \delta_{kj},$$

which is just a fancy way of saying $QP = I$, or $Q = P^{-1}$.

5.3.3 Transformation law for vectors

Consider a vector $\mathbf{u}$, and suppose that in terms of the $\{\mathbf{e}_i\}$ basis it can be written in component form as

$$\mathbf{u} = \sum_{i=1}^{n} u_i \mathbf{e}_i$$

Similarly, in the $\{\tilde{\mathbf{e}_i}\}$ basis suppose that $\mathbf{u}$ can be written in component form as

$$\mathbf{u} = \sum_{i=1}^{n} \tilde{u_i} \tilde{\mathbf{e}_i}$$

Note that this is the same vector $\mathbf{u}$ but has different components with respect to different bases. Using the transformation matrix above for the basis, we have

$$\begin{aligned}\mathbf{u} &= \sum_{j=1}^{n} \tilde{u_j} \sum_{i=1}^{n} P_{ij} \mathbf{e}_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij} \tilde{u_j} \right) \mathbf{e}_i\end{aligned}$$

By comparison, we know that

$$u_i = \sum_{j=1}^{n} P_{ij}\tilde{u}_j$$

Theorem 5.3.1. *Denote vector as* $\mathbf{u}$ *with respect to* $\{\mathbf{e}_i\}$ *and* $\tilde{\mathbf{u}}$ *with respect to* $\{\tilde{\mathbf{e}_i}\}$. *Then*

$$\mathbf{u} = P\tilde{\mathbf{u}} \text{ and } \tilde{\mathbf{u}} = P^{-1}\mathbf{u}$$

Example 5.3.1. *Take the first basis as* $\{\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1)\}$ *and the second as* $\{\tilde{\mathbf{e}_1} = (1,1), \tilde{\mathbf{e}_2} = (-1,1)\}$.

So $\tilde{\mathbf{e}_1} = \mathbf{e}_1 + \mathbf{e}_2$ *and* $\tilde{\mathbf{e}_2} = -\mathbf{e}_1 + \mathbf{e}_2$. *We have*

$$\mathbf{P} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Then for an arbitrary vector $\mathbf{u}$, *we have*

$$\begin{aligned}
\mathbf{u} &= u_1\mathbf{e}_1 + u_2\mathbf{e}_2 \\
&= u_1\frac{1}{2}(\tilde{\mathbf{e}_1} - \tilde{\mathbf{e}_2}) + u_2\frac{1}{2}(\tilde{\mathbf{e}_1} + \tilde{\mathbf{e}_2}) \\
&= \frac{1}{2}(u_1 + u_2)\tilde{\mathbf{e}_1} + \frac{1}{2}(-u_1 + u_2)\tilde{\mathbf{e}_2}.
\end{aligned}$$

Alternatively, using the formula above, we obtain

$$\begin{aligned}
\tilde{\mathbf{u}} &= P^{-1}\mathbf{u} \\
&= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}(u_1 + u_2) \\ \frac{1}{2}(-u_1 + u_2) \end{pmatrix}
\end{aligned}$$

Which agrees with the above direct expansion.

Transformation law for matrices representing linear maps

Consider a linear map $\alpha : \mathbb{C}^n \to \mathbb{C}^n$ with associated $n \times n$ matrix A. We have

$$\mathbf{u}' = \alpha(\mathbf{u}) = A\mathbf{u}.$$

Denote $\mathbf{u}$ and $\mathbf{u}'$ as being with respect to basis $\{\mathbf{e}_i\}$ (i.e. same basis in both spaces), and $\tilde{\mathbf{u}}, \tilde{\mathbf{u}}'$ with respect to $\{\tilde{\mathbf{e}}_i\}$.

Using what we've got above, we have

$$\begin{aligned}
\mathbf{u}' &= A\mathbf{u} \\
P\tilde{\mathbf{u}}' &= AP\tilde{\mathbf{u}} \\
\tilde{\mathbf{u}}' &= P^{-1}AP\tilde{\mathbf{u}} \\
&= \tilde{A}\tilde{\mathbf{u}}
\end{aligned}$$

So

Theorem 5.3.2.

$$\tilde{A} = P^{-1}AP.$$

Example 5.3.2. *Consider the shear* $S_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ *with respect to the standard basis. Choose a new set of basis vectors by rotating by* θ *about the* $\mathbf{e}_3$ *axis:*

$$\begin{aligned}
\tilde{\mathbf{e}_1} &= \cos\theta \mathbf{e}_1 + \sin\theta \mathbf{e}_2 \\
\tilde{\mathbf{e}_2} &= -\sin\theta \mathbf{e}_1 + \cos\theta \mathbf{e}_2 \\
\tilde{\mathbf{e}_3} &= \mathbf{e}_3
\end{aligned}$$

So we have

$$P = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, P^{-1} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Now use the basis transformation laws to obtain

$$\tilde{S}_\lambda = \begin{pmatrix} 1 + \lambda\sin\theta\cos\theta & \lambda\cos^2\theta & 0 \\ -\lambda\sin^2\theta & 1 - \lambda\sin\theta\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Clearly this is much more complicated than our original basis. This shows that choosing a sensible basis is important.

More generally, given $\alpha : \mathbb{C}^m \to \mathbb{C}^n$, given $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{x}' \in \mathbb{C}^n$ with $\mathbf{x}' = A\mathbf{x}$. We know that A is an $n \times m$ matrix.

Suppose $\mathbb{C}^m$ has a basis $\{\mathbf{e}_i\}$ and $\mathbb{C}^n$ has a basis $\{\mathbf{f}_i\}$. Now change bases to $\{\tilde{\mathbf{e}_i}\}$ and $\{\tilde{\mathbf{f}_i}\}$.

We know that $\mathbf{x} = P\tilde{\mathbf{x}}$ with P being an $m \times m$ matrix, with $\mathbf{x}' = R\tilde{\mathbf{x}}'$ with R being an $n \times n$ matrix.

Combining both of these, we have

$$\begin{aligned} R\tilde{\mathbf{x}}' &= AP\tilde{\mathbf{x}} \\ \tilde{\mathbf{x}}' &= R^{-1}AP\tilde{\mathbf{x}} \end{aligned}$$

Therefore $\tilde{A} = R^{-1}AP$.

Example 5.3.3. *Consider $\alpha : \mathbb{R}^3 \to \mathbb{R}^2$, with respect to the standard bases in both spaces,*

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix}$$

Use a new basis $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ in $\mathbb{R}^2$ and keep the standard basis in $\mathbb{R}^3$. The basis change matrix in $\mathbb{R}^3$ is simply I, while

$$R = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}, R^{-1} = \frac{1}{9}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}$$

is the transformation matrix for $\mathbb{R}^2$. So

$$\begin{aligned} \tilde{A} &= \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix} I \\ &= \frac{1}{9}\begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 17/9 \\ 0 & 1 & 2/9 \end{pmatrix} \end{aligned}$$

We can alternatively do it this way: we know that $\tilde{\mathbf{f}_1} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \tilde{\mathbf{f}_2} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ Then *we know that*

$$\tilde{\mathbf{e}_1} = \mathbf{e}_1 \mapsto 2\mathbf{f}_1 + \mathbf{f}_2 = \mathbf{f}_1$$
$$\tilde{\mathbf{e}_2} = \mathbf{e}_2 \mapsto 3\mathbf{f}_1 + 6\mathbf{f}_2 = \tilde{\mathbf{f}_1} + \tilde{\mathbf{f}_2}$$
$$\tilde{\mathbf{e}_3} = \mathbf{e}_3 \mapsto 4\mathbf{f}_1 + 3\mathbf{f}_2 = \frac{17}{9}\tilde{\mathbf{f}_1} + \frac{2}{9}\tilde{\mathbf{f}_2}$$

and we can construct the matrix correspondingly.

5.4 Similar matrices

Definition 5.4.1 (Similar matrices)**.** *Two $n \times n$ matrices A and B are similar if there exists an invertible matrix P such that*

$$B = P^{-1}AP,$$

i.e. they represent the same map under different bases. Alternatively, we say that they are in the same conjugacy class.

Note that

1. Similarity is an equivalence relation.
2. A map from A to $\mathrm{P}^{-1}\mathrm{AP}$ is sometimes known as a similarity transformation.
3. The matrices representing a map $\mathcal{A}$ with respect to different bases are similar.
4. he identity matrix (or a multiple of it) is similar only with itself (or a multiple of it) since $\mathrm{P}^{-1}\mathrm{IP}=\mathrm{I}$

Exercise 5.4.1. *Show that a $n \times n$ matrix with a unique eigenvalue (i.e. an eigenvalue with an algebraic multiplicity of n), and with n linearly independent eigenvectors, has to be a multiple of the identity matrix.*

Proposition 5.4.1. *Similar matrices have the following properties:*

1. *Similar matrices have the same determinant.*
2. *Similar matrices have the same trace.*

3. Similar matrices have the same characteristic polynomial.

Note that (iii) implies (i) and (ii) since the determinant and trace are the coefficients of the characteristic polynomial

Proof. They are proven as follows:

1. $\det B = \det(P^{-1}AP) = (\det A)(\det P)^{-1}(\det P) = \det A$

2.

$$\begin{aligned}
\operatorname{tr} B &= B_{ii} \\
&= P^{-1}_{ij} A_{jk} P_{ki} \\
&= A_{jk} P_{ki} P^{-1}_{ij} \\
&= A_{jk} (PP^{-1})_{kj} \\
&= A_{jk} \delta_{kj} \\
&= A_{jj} \\
&= \operatorname{tr} A
\end{aligned}$$

3.

$$\begin{aligned}
p_B(\lambda) &= \det(B - \lambda I) \\
&= \det(P^{-1}AP - \lambda I) \\
&= \det(P^{-1}AP - \lambda P^{-1} I P) \\
&= \det(P^{-1}(A - \lambda I)P) \\
&= \det(A - \lambda I) \\
&= p_A(\lambda)
\end{aligned}$$

□

5.5 Diagonalizable maps and matrices

Definition 5.5.1. *A linear map $\mathcal{A} : \mathbb{F}^n \to \mathbb{F}^n$ is said to be diagonalizable if $\mathbb{F}^n$ has a basis consisting of eigenvectors of $\mathcal{A}$.*

Further, we have shown that a map $\mathcal{A}$ with n distinct eigenvalues has a basis of eigenvectors. It follows that if a matrix A has n distinct eigenvalues, then it is diagonalizable by means of a similarity transformation using the transformation matrix that changes to a basis of eigenvectors.

Definition 5.5.2 (Diagonalizable matrices)**.** *An $n \times n$ matrix A is diagonalizable if it is similar to a diagonal matrix. We showed above that this is equivalent to saying the eigenvectors form a basis of $\mathbb{C}^n$.*

Proposition 5.5.1. *For an $n \times n$ matrix A acting on $V = \mathbb{R}^n$ or $\mathbb{C}^n$, the following conditions are equivalent:*

1. *there exists a basis of eigenvectors of A for V, named $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ which $A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$ for each i*
2. *there exists an $n \times n$ invertible matrix P with the property that*

$$P^{-1}AP = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

If either of these conditions hold, then A is diagonalizable.

Proof. Note that for any matrix P, AP has columns $A\mathbf{C}_i(P)$, and PD has columns $\lambda_i \mathbf{C}_i(P)$. Then (i) and (ii) are related by choosing $\mathbf{v}_i = \mathbf{C}_i(P)$. Then $P^{-1}AP = D \iff AP = PD \iff A\mathbf{v}_i = \lambda_i \mathbf{v}_i$.

In essence, given a basis of eigenvectors as in (i), the relation above defines P, and if the eigenvectors are linearly independent then P is invertible. Conversely, given a matrix P as in (ii), its columns are a basis of eigenvectors. $\square$

5.5.1 Criteria for diagonalizability

Proposition 5.5.2. *Consider an $n \times n$ matrix A.*

1. *A is diagonalisable if it has n distinct eigenvalues (sufficient condition).*
2. *A is diagonalisable if and only if for every eigenvalue λ, $M_\lambda = m_\lambda$ (necessary and sufficient condition).*

Proof. Use the proposition and corollary above.

1. If we have n distinct eigenvalues, then we have n linearly independent eigenvectors. Hence they form a basis.
2. If λ_i are all the distinct eigenvalues, then $\mathcal{B}_{\lambda_1} \cup \cdots \cup \mathcal{B}_{\lambda_r}$ are linearly independent. The number of elements in this new basis is $\sum_i m_{\lambda_i} = \sum_i M_{\lambda_i} = n$ which is the degree of the characteristic polynomial. So we have a basis.

Note that case (i) is just a specialisation of case (ii) where both multiplicities are 1. $\square$

Let us consider some examples.

1. Let
$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix} \implies \lambda = 5, -3; \quad M_5 = m_5 = 1; \quad M_{-3} = m_{-3} = 2$$
So A is diagonalisable by case (ii) above, and moreover
$$P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}; \quad P^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 2 & -3 \\ -2 & 4 & 6 \\ 1 & 2 & 5 \end{pmatrix} \implies P^{-1}AP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$
2. Let
$$A = \begin{pmatrix} -3 & -1 & 1 \\ -1 & -3 & 1 \\ -2 & 2 & 0 \end{pmatrix} \implies \lambda = -2; \quad M_{-2} = 3 > m_{-2} = 2$$
So A is not diagonalizable. As a check, if it were diagonalizable, then there would be some matrix P such that $P^{-1}AP = -2I \Rightarrow A = P(-2I)P^{-1} = -2I$ This is a contradiction.

Consider the second example in Section 5.2,

$$A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$$

We found three linear eigenvectors

$$\tilde{\mathbf{e}_1} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \tilde{\mathbf{e}_2} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \tilde{\mathbf{e}_3} = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

If we let

$$P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, P^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 2 & -3 \\ -2 & 4 & 6 \\ 1 & 2 & 5 \end{pmatrix},$$

then

$$\tilde{A} = P^{-1}AP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix},$$

so A is diagonalizable.

Theorem 5.5.1. *Let $\lambda_1, \lambda_2, \cdots, \lambda_r$, with $r \leq n$ be the distinct eigenvalues of A. Let $B_1, B_2, \cdots B_r$ be the bases of the eigenspaces $E_{\lambda_1}, E_{\lambda_2}, \cdots, E_{\lambda_r}$ correspondingly. Then the set $B = \bigcup_{i=1}^{r} B_i$ is linearly independent.*

This is similar to the proof we had for the case where the eigenvalues are distinct. However, we are going to do it much concisely, and the actual meat of the proof is actually just a single line.

Proof. Write $B_1 = \{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \cdots \mathbf{x}_{m(\lambda_1)}^{(1)}\}$. Then $m(\lambda_1) = \dim(E_{\lambda_1})$, and similarly for all B_i.

Consider the following general linear combination of all elements in B. Consider the equation

$$\sum_{i=1}^{r} \sum_{j=1}^{m(\lambda_i)} \alpha_{ij}\mathbf{x}_j^{(i)} = 0.$$

The first sum is summing over all eigenspaces, and the second sum sums over the basis vectors in B_i. Now apply the matrix

$$\prod_{k=1,2,\cdots,\overline{K},\cdots,r}(A-\lambda_k I)$$

to the above sum, for some arbitrary K. We obtain

$$\sum_{j=1}^{m(\lambda_K)}\alpha_{Kj}\left[\prod_{k=1,2,\cdots,\overline{K},\cdots,r}(\lambda_K-\lambda_k)\right]\mathbf{x}_j^{(K)}=0.$$

Since the $\mathbf{x}_j^{(K)}$ are linearly independent (B_K is a basis), $\alpha_{Kj}=0$ for all j. Since K was arbitrary, all α_{ij} must be zero. So B is linearly independent. □

Proposition 5.5.3. *A is diagonalizable iff all its eigenvalues have zero defect.*

5.5.2 Canonical form of complex matrices

Given a matrix A, if its eigenvalues all have non-zero defect, then we can find a basis in which it is diagonal. However, if some eigenvalue does have defect, we can still put it into an almost-diagonal form. This is known as the Jordan normal form.

Theorem 5.5.2. *Any 2×2 complex matrix A is similar to exactly one of*

$$\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix},\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix},\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$$

Proof. For each case:

1. If A has two distinct eigenvalues, then eigenvectors are linearly independent. Then we can use P formed from eigenvectors as its columns

2. If $\lambda_1=\lambda_2=\lambda$ and $\dim E_\lambda=2$, then write $E_\lambda=\operatorname{span}\{\mathbf{u},\mathbf{v}\}$, with $\mathbf{u},\mathbf{v}$ linearly independent. Now use $\{\mathbf{u},\mathbf{v}\}$ as a new basis of $\mathbb{C}^2$ and $\tilde{A}=P^{-1}AP=\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}=\lambda I$

   Note that since $P^{-1}AP=\lambda I$, we have $A=P(\lambda I)P^{-1}=\lambda I$. So A is *isotropic*, i.e. the same with respect to any basis.

3. If $\lambda_1 = \lambda_2 = \lambda$ and $\dim(E_\lambda) = 1$, then $E_\lambda = \operatorname{span}\{\mathbf{v}\}$. Now choose basis of $\mathbb{C}^2$ as $\{\mathbf{v}, \mathbf{w}\}$, where $\mathbf{w} \in \mathbb{C}^2 \setminus E_\lambda$.

   We know that $A\mathbf{w} \in \mathbb{C}^2$. So $A\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{w}$. Hence, if we change basis to $\{\mathbf{v}, \mathbf{w}\}$, then $\tilde{A} = P^{-1}AP = \begin{pmatrix} \lambda & \alpha \\ 0 & \beta \end{pmatrix}$.

   However, A and $\tilde{A}$ both have eigenvalue λ with algebraic multiplicity 2. So we must have $\beta = \lambda$. To make $\alpha = 1$, let $\mathbf{u} = (\tilde{A} - \lambda I)\mathbf{w}$. We know $\mathbf{u} \neq \mathbf{0}$ since $\mathbf{w}$ is not in the eigenspace. Then

   $$(\tilde{A} - \lambda I)\mathbf{u} = (\tilde{A} - \lambda I)^2\mathbf{w} = \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}\mathbf{w} = \mathbf{0}.$$

   So $\mathbf{u}$ is an eigenvector of $\tilde{A}$ with eigenvalue λ.

   We have $\mathbf{u} = \tilde{A}\mathbf{w} - \lambda\mathbf{w}$. So $\tilde{A}\mathbf{w} = \mathbf{u} + \lambda\mathbf{w}$.

   Change basis to $\{\mathbf{u}, \mathbf{w}\}$. Then A with respect to this basis is $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

   This is a two-stage process: P sends basis to $\{\mathbf{v}, \mathbf{w}\}$ and then matrix Q sends to basis $\{\mathbf{u}, \mathbf{w}\}$. So the similarity transformation is

   $$Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ)$$

   .

□

Proposition 5.5.4. *The canonical form, or Jordan normal form, exists for any $n \times n$ matrix A. Specifically, there exists a similarity transform such that A is similar to a matrix to $\tilde{A}$ that satisfies the following properties:*

1. $\tilde{A}_{\alpha\alpha} = \lambda_\alpha$*, i.e. the diagonal composes of the eigenvalues.*
2. $\tilde{A}_{\alpha,\alpha+1} = 0$ *or* 1.
3. $\tilde{A}_{ij} = 0$ *otherwise.*

The actual theorem is actually stronger than this, and the Jordan normal form satisfies some additional properties in addition to the above. However, we shall not go into details, and this is left for the IB Linear Algebra course.

Example 5.5.1. *Let*

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

The eigenvalues are $-2, -2, -2$ *and the eigenvectors are* $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. *Pick* $\mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. *Write* $\mathbf{u} = (A - \lambda I)\mathbf{w} = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ -2 & -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -2 \end{pmatrix}$. *Note that* $A\mathbf{u} = -2\mathbf{u}$. *We also have* $A\mathbf{w} = \mathbf{u} - 2\mathbf{w}$. *Form a basis* $\{\mathbf{u}, \mathbf{w}, \mathbf{v}\}$, *where* $\mathbf{v}$ *is another eigenvector linearly independent from* $\mathbf{u}$, *say* $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

Now change to this basis with $P = \begin{pmatrix} -1 & 1 & 1 \\ -1 & 0 & 0 \\ -2 & 0 & 1 \end{pmatrix}$. *Then the Jordan normal form is* $P^{-1}AP = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}$

5.6 Cayley-Hamilton theorem

5.6.1 Matrix polynomials

If A is an $n \times n$ complex matrix and

$$p(t) = c_0 + c_1 t + c_2^2 + \cdots + c_k t^k$$

is a polynomial, then

$$p(A) = c_0 I + c_1 A + c_2 A^2 + \cdots + c_k A^k$$

We can also define power series on matrices (subject to convergence). For example, the exponential series which always converges:

$$\exp(A) = I + A + \frac{1}{2}A^2 + \cdots + \frac{1}{r!}A^r + \ldots$$

For a diagonal matrix, polynomials and power series can be computed easily since the power of a diagonal matrix just involves raising its diagonal elements to said power. Therefore,

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \implies p(D) = \begin{pmatrix} p(\lambda_1) & 0 & \cdots & 0 \\ 0 & p(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(\lambda_n) \end{pmatrix}$$

Therefore,

$$\exp(D) = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix}$$

If $B = P^{-1}AP$ (similar to A) where P is an $n \times n$ invertible matrix, then

$$B^r = P^{-1}A^rP$$

Therefore,

$$p(B) = p(P^{-1}AP) = P^{-1}p(A)P$$

Of special interest is the characteristic polynomial,

$$\chi_A(t) = \det(A - tI) = c_0 + c_1t + c_2t^2 + \cdots + c_nt^n$$

where $c_0 = \det A$, and $c_n = (-1)^n$.

Theorem 5.6.1 (Cayley-Hamilton Theorem)**.**

$$\chi_A(A) = c_0I + c_1A + c_2A^2 + \cdots + c_nA^n = 0$$

Less formally, a matrix satisfies its own characteristic equation.

Note that we can find an expression for the matrix inverse.

$$-c_0 I = A(c_1 + c_2 A + \cdots + c_n A^{n-1})$$

If $c_0 = \det A \neq 0$, then

$$A^{-1} = \frac{-1}{c_0}(c_1 + c_2 A + \cdots + c_n A^{n-1})$$

5.6.2 Proofs of special cases of Cayley-Hamilton theorem

Proof for a 2×2 matrix. Let A be a general 2×2 matrix.

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \chi_A(t) = t^2 - (a+d)t + (ad - bc)$$

We can check the theorem by substitution.

$$\chi_A(A) = A^2 - (a+d)A - (ad - bc)I$$

This is shown on the last example sheet. □

Proof for diagonalisable $n \times n$ matrices. Consider A with eigenvalues λ_i, and an invertible matrix P such that $P^{-1}AP = D$, where D is diagonal.

$$\chi_A(D) = \begin{pmatrix} \chi_A(\lambda_1) & 0 & \cdots & 0 \\ 0 & \chi_A(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_A(\lambda_n) \end{pmatrix} = 0$$

since the λ_i are eigenvalues. Then

$$\chi_A(A) = \chi_A(PDP^{-1}) = P\chi_A(D)P^{-1} = 0$$

□

5.6.3 Proof in general case (non-examinable)

Proof. Let $M = A - tI$. Then $\det M = \det(A - tI) = \chi_A(t) = \sum_{r=0} c_r t^r$. We can construct the adjugate matrix.

$$\operatorname{adj} M = \sum_{r=0}^{n-1} B_r t^r$$

Therefore,

$$\begin{aligned}\operatorname{adj} M M = (\det M) I &= \left(\sum_{r=0}^{n-1} B_r t^r\right)(A - tI) \\ &= B_0 A + (B_1 A - B_0)t + (B_2 A - B_1)t^2 + \cdots + \\ &+ (B_{n-1}A - B_{n-2})t^{n-1} - B_{n-1}t\end{aligned}$$

Now by comparing coefficients,

$$\begin{aligned} C_0 I &= B_0 A \\ C_1 I &= B_1 A - B_0 \\ &\vdots \\ C_{n-1} I &= B_{n-1} A - B_{n-2} \\ C_n I &= -B_{n-1} \end{aligned}$$

Summing all of these coefficients, multiplying by the relevant powers,

$$\begin{aligned} & C_0 I + C_1 A + C_2 A^2 + \cdots + C_n A^n \\ = & B_0 A + (B_1 A^2 - B_0 A) + (B_2 A^3 - B_1 A^2) + \cdots + (B_{n-1}A^n - B_{n-2}A^{n-1}) - B_{n-1}A^n \\ = & 0 \end{aligned}$$

□

5.7 Eigenvalues and eigenvectors of a Hermitian matrix

Theorem 5.7.1. *The eigenvalues of a Hermitian matrix H are real.*

Proof. Suppose that H has eigenvalue λ with eigenvector $\mathbf{v} \neq 0$. Then

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

We pre-multiply by $\mathbf{v}^\dagger$, a $1 \times n$ row vector, to obtain

$$\mathbf{v}^\dagger H \mathbf{v} = \lambda \mathbf{v}^\dagger \mathbf{v} \tag{$*$}$$

We take the Hermitian conjugate of both sides. The left hand side is

$$(\mathbf{v}^\dagger H\mathbf{v})^\dagger = \mathbf{v}^\dagger H^\dagger \mathbf{v} = \mathbf{v}^\dagger H\mathbf{v}$$

since H is Hermitian. The right hand side is

$$(\lambda\mathbf{v}^\dagger\mathbf{v})^\dagger = \lambda^*\mathbf{v}^\dagger\mathbf{v}$$

So we have

$$\mathbf{v}^\dagger H\mathbf{v} = \lambda^*\mathbf{v}^\dagger\mathbf{v}.$$

From $(*)$, we know that $\lambda\mathbf{v}^\dagger\mathbf{v} = \lambda^*\mathbf{v}^\dagger\mathbf{v}$. Since $\mathbf{v} \neq 0$, we know that $\mathbf{v}^\dagger\mathbf{v} = \mathbf{v}\cdot\mathbf{v} \neq 0$. So $\lambda = \lambda^*$ and λ is real. □

Theorem 5.7.2. *The eigenvectors of a Hermitian matrix H corresponding to distinct eigenvalues are orthogonal.*

Proof. Let

$$H\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{i}$$

$$H\mathbf{v}_j = \lambda_j\mathbf{v}_j. \tag{ii}$$

Pre-multiply (i) by $\mathbf{v}_j^\dagger$ to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_i\mathbf{v}_j^\dagger\mathbf{v}_i. \tag{iii}$$

Pre-multiply (ii) by $\mathbf{v}_i^\dagger$ and take the Hermitian conjugate to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_j\mathbf{v}_j^\dagger\mathbf{v}_i. \tag{iv}$$

Equating (iii) and (iv) yields

$$\lambda_i\mathbf{v}_j^\dagger\mathbf{v}_i = \lambda_j\mathbf{v}_j^\dagger\mathbf{v}_i.$$

Since $\lambda_i \neq \lambda_j$, we must have $\mathbf{v}_j^\dagger\mathbf{v}_i = 0$. So their inner product is zero and are orthogonal. □

So we know that if a Hermitian matrix has n distinct eigenvalues, then the eigenvectors form an orthonormal basis. However, if there are degenerate eigenvalues, it is more difficult, and requires the Gram-Schmidt process.

5.7.1 Gram-Schmidt orthogonalization (non-examinable)

Suppose we have a set $B = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r\}$ of linearly independent vectors. We want to find an orthogonal set $\tilde{\mathbf{B}} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$.

Define the projection of $\mathbf{w}$ onto $\mathbf{v}$ by $\mathcal{P}_{\mathbf{v}}(\mathbf{w}) = \frac{\langle \mathbf{v}|\mathbf{w}\rangle}{\langle \mathbf{v}|\mathbf{v}\rangle}\mathbf{v}$. Now construct $\tilde{\mathbf{B}}$ iteratively:

1. $\mathbf{v}_1 = \mathbf{w}_1$
2. $\mathbf{v}_2 = \mathbf{w}_2 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w})$

   Then we get that $\langle \mathbf{v}_1 \mid \mathbf{v}_2\rangle = \langle \mathbf{v}_1 \mid \mathbf{w}_2\rangle - \left(\frac{\langle \mathbf{v}_1|\mathbf{w}_2\rangle}{\langle \mathbf{v}_1|\mathbf{v}_1\rangle}\right)\langle \mathbf{v}_1 \mid \mathbf{v}_1\rangle = 0$
3. $\mathbf{v}_3 = \mathbf{w}_3 - \mathcal{P}_{\mathbf{v}_1}(\mathbf{w}_3) - \mathcal{P}_{\mathbf{v}_2}(\mathbf{w}_3)$
4. $\vdots$
5. $\mathbf{v}_r = \mathbf{w}_r - \sum_{j=1}^{r-1} \mathcal{P}_{\mathbf{v}_j}(\mathbf{w}_r)$

At each step, we subtract out the components of $\mathbf{v}_i$ that belong to the space of $\{\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}\}$. This ensures that all the vectors are orthogonal. Finally, we normalize each basis vector individually to obtain an orthonormal basis.

5.7.2 Unitary transformation

Suppose U is the transformation between one orthonormal basis and a new orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$, i.e. $\langle \mathbf{u}_i \mid \mathbf{u}_j\rangle = \delta_{ij}$. Then

$$U = \begin{pmatrix} (\mathbf{u}_1)_1 & (\mathbf{u}_2)_1 & \cdots & (\mathbf{u}_n)_1 \\ (\mathbf{u}_1)_2 & (\mathbf{u}_2)_2 & \cdots & (\mathbf{u}_n)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{u}_1)_n & (\mathbf{u}_2)_n & \cdots & (\mathbf{u}_n)_n \end{pmatrix}$$

Then

$$\begin{aligned} (U^\dagger U)_{ij} &= (U^\dagger)_{ik} U_{kj} \\ &= U^*_{ki} U_{kj} \\ &= (\mathbf{u}_i)^*_k (\mathbf{u}_j)_k \\ &= \langle \mathbf{u}_i \mid \mathbf{u}_j\rangle \\ &= \delta_{ij} \end{aligned}$$

So U is a unitary matrix.

5.7.3 Diagonalization of $n \times n$ Hermitian matrices

Theorem 5.7.3. *An $n \times n$ Hermitian matrix has precisely n orthogonal eigenvectors.*

Proof. (Non-examinable) Let $\lambda_1, \lambda_2, \cdots, \lambda_r$ be the distinct eigenvalues of H ($r \leq n$), with a set of corresponding orthonormal eigenvectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$. Extend to a basis of the whole of $\mathbb{C}^n$

$$B' = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{n-r}\}$$

Now use Gram-Schmidt to create an orthonormal basis

$$\tilde{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-r}\}.$$

Now write

$$P = \begin{pmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r & \mathbf{u}_1 & \cdots & \mathbf{u}_{n-r} \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

We have shown above that this is a unitary matrix, i.e. $P^{-1} = P^\dagger$. So if we change basis, we have

$$\begin{aligned} P^{-1}HP &= P^\dagger HP \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & c_{11} & c_{12} & \cdots & c_{1,n-r} \\ 0 & 0 & \cdots & 0 & c_{21} & c_{22} & \cdots & c_{2,n-r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & c_{n-r,1} & c_{n-r,2} & \cdots & c_{n-r,n-r} \end{pmatrix} \end{aligned}$$

Here C is an $(n-r)\times(n-r)$ Hermitian matrix. The eigenvalues of C are also eigenvalues of H because $\det(H - \lambda I) = \det(P^\dagger HP - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_r - \lambda)\det(C - \lambda I)$. So the eigenvalues of C are the eigenvalues of H.

We can keep repeating the process on C until we finish all rows. For example, if the eigenvalues of C are all distinct, there are $n-r$ orthonormal eigenvectors $\mathbf{w}_j$ (for $j = r+1, \cdots, n$) of C. Let

$$Q = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & \uparrow & \uparrow & & \uparrow \\ & & & & \mathbf{w}_{r+1} & \mathbf{w}_{r+2} & \cdots & \mathbf{w}_n \\ & & & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

with other entries 0. (where we have a $r \times r$ identity matrix block on the top left corner and a $(n-r) \times (n-r)$ with columns formed by $\mathbf{w}_j$)

Since the columns of Q are orthonormal, Q is unitary. So $Q^\dagger P^\dagger HPQ = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_n)$, where the first r λs are distinct and the remaining ones are copies of previous ones.

The n linearly-independent eigenvectors are the columns of PQ.

□

So it now follows that H is diagonalizable via transformation $U(= PQ)$. U is a unitary matrix because P and Q are. We have

$$D = U^\dagger HU$$
$$H = UDU^\dagger$$

Note that a real symmetric matrix S is a special case of Hermitian matrices. So we have

$$D = Q^T SQ$$
$$S = QDQ^T$$

Example 5.7.1. *Find the orthogonal matrix which diagonalizes the following* *real symmetric matrix:* $S = \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix}$ *with* $\beta \neq 0 \in \mathbb{R}$.

We find the eigenvalues by solving the characteristic equation: $\det(S - \lambda I) = 0$, *and obtain* $\lambda = 1 \pm \beta$.

The corresponding eigenvectors satisfy $(S - \lambda I)\mathbf{x} = 0$, *which gives* $\mathbf{x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}$

We change the basis from the standard basis to $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ *(which is just a rotation by* $\pi/4$*).*

The transformation matrix is $Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$. *Then we know that* $S = QDQ^T$ *with* $D = \mathrm{diag}(1, -1)$

5.7.4 Normal matrices

We have seen that the eigenvalues and eigenvectors of Hermitian matrices satisfy some nice properties. More generally, we can define the following:

Definition 5.7.1 (Normal matrix)**.** *A* normal matrix *as a matrix that commutes with its own Hermitian conjugate, i.e.*

$$NN^\dagger = N^\dagger N$$

Hermitian, real symmetric, skew-Hermitian, real anti-symmetric, orthogonal, unitary matrices are all special cases of normal matrices.

It can be shown that:

Proposition 5.7.1.

1. *If* λ *is an eigenvalue of* N, *then* λ^* *is an eigenvalue of* $N^\dagger$.
2. *The eigenvectors of distinct eigenvalues are orthogonal.*
3. *A normal matrix can always be diagonalized with an orthonormal basis of eigenvectors.*

CHAPTER 6

Quadratic forms and conics

We want to study quantities like $x_1^2 + x_2^2$ and $3x_1^2 + 2x_1x_2 + 4x_2^2$. For example, conic sections generally take this form. The common characteristic of these is that each term has degree 2. Consequently, we can write it in the form $\mathbf{x}^\dagger A\mathbf{x}$ for some matrix A.

Definition 6.0.1. *A sesquilinear form is a quantity $F = \mathbf{x}^\dagger A\mathbf{x} = x_i^* A_{ij} x_j$. If A is Hermitian, then F is a Hermitian form. If A is real symmetric, then F is a quadratic form.*

Theorem 6.0.1. *Hermitian forms are real.*

Proof. $(\mathbf{x}^\dagger H\mathbf{x})^* = (\mathbf{x}^\dagger H\mathbf{x})^\dagger = \mathbf{x}^\dagger H^\dagger \mathbf{x} = \mathbf{x}^\dagger H\mathbf{x}$. So $(\mathbf{x}^\dagger H\mathbf{x})^* = \mathbf{x}^\dagger H\mathbf{x}$ and it is real. □

We know that any Hermitian matrix can be diagonalized with a unitary transformation. So $F(\mathbf{x}) = \mathbf{x}^\dagger H\mathbf{x} = \mathbf{x}^\dagger UDU^\dagger \mathbf{x}$. Write $\mathbf{x}' = U^\dagger \mathbf{x}$. So $F = (\mathbf{x}')^\dagger D\mathbf{x}'$, where $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

We know that $\mathbf{x}'$ is the vector $\mathbf{x}$ relative to the eigenvector basis. So

$$F(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i |x_i'|^2$$

The eigenvectors are known as the principal axes.

Example 6.0.1. *Take* $F = 2x^2 - 4xy + 5y^2 = \mathbf{x}^T S\mathbf{x}$, *where* $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ *and* $S = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix}$.

Note that we can always choose the matrix to be symmetric. This is since for any antisymmetric A, *we have* $\mathbf{x}^\dagger A\mathbf{x} = 0$. *So we can just take the symmetric part.*

The eigenvalues are $1, 6$ *with corresponding eigenvectors* $\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

Now change basis with

$$Q = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

Then $\mathbf{x}' = Q^T\mathbf{x} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2x + y \\ x - 2y \end{pmatrix}$. *Then* $F = (x')^2 + 6(y')^2$.

So $F = c$ *is an ellipse.*

6.1 Quadrics

Definition 6.1.1 (Quadric). *A quadric is an* n*-dimensional surface defined by the zero of a real quadratic polynomial, i.e.*

$$\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0,$$

where A *is a real* $n \times n$ *matrix,* $\mathbf{x}, \mathbf{b}$ *are* n*-dimensional column vectors and* c *is a constant scalar.*

As noted in example, anti-symmetric matrix has $\mathbf{x}^T A\mathbf{x} = 0$, so for any A, we can split it into symmetric and anti-symmetric parts, and just retain the symmetric part

$S = (A + A^T)/2$. So we can have

$$\mathbf{x}^T S\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0$$

with S symmetric.

Since S is real and symmetric, we can diagonalize it using $S = QDQ^T$ with D diagonal. We write $\mathbf{x}' = Q^T\mathbf{x}$ and $\mathbf{b}' = Q^T\mathbf{b}$. So we have

$$(\mathbf{x}')^T D\mathbf{x}' + (\mathbf{b}')^T\mathbf{x}' + c = 0.$$

If S is invertible, i.e. with no zero eigenvalues, then write $\mathbf{x}'' = \mathbf{x}' + \frac{1}{2}D^{-1}\mathbf{b}'$ which shifts the origin to eliminate the linear term $(\mathbf{b}')^T\mathbf{x}'$ and finally have (dropping the prime superfixes)

$$\mathbf{x}^T D\mathbf{x} = k.$$

So through two transformations, we have ended up with a simple quadratic form.

6.1.1 Conic sections $(n = 2)$

From the equation above, we obtain

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = k.$$

We have the following cases:

1. $\lambda_1\lambda_2 > 0$: we have ellipses with axes coinciding with eigenvectors of S. (We require $\operatorname{sgn}(k) = \operatorname{sgn}(\lambda_1, \lambda_2)$, or else we would have no solutions at all)

2. $\lambda_1\lambda_2 < 0$: say $\lambda_1 = k/a^2 > 0$, $\lambda_2 = -k/b^2 < 0$. So we obtain

   $$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1,$$

   which is a hyperbola.

3. $\lambda_1\lambda_2 = 0$: Say $\lambda_2 = 0$, $\lambda_1 \neq 0$. Note that in this case, our symmetric matrix S is not invertible and we cannot shift our origin using as above.

   From our initial equation, we have

   $$\lambda_1(x_1')^2 + b_1'x_1' + b_2'x_2' + c = 0.$$

We perform the coordinate transform (which is simply completing the square!)

$$\begin{aligned}
x_1'' &= x_1' + \frac{b_1'}{2\lambda_1}\\
x_2'' &= x_2' + \frac{c}{b_2'} - \frac{(b_1')^2}{4\lambda_1 b_2'}
\end{aligned}$$

to remove the x_1' and constant term. Dropping the primes, we have

$$\lambda_1 x_1^2 + b_2 x_2 = 0,$$

which is a parabola.

Note that above we assumed $b_2' \neq 0$. If $b_2' = 0$, we have $\lambda_1(x_1')^2 + b_1' x_1' + c = 0$. If we solve this quadratic for x_1', we obtain 0, 1 or 2 solutions for x_1 (and x_2 can be any value). So we have 0, 1 or 2 straight lines.

These are known as conic sections. As you will see in IA Dynamics and Relativity, this are the trajectories of planets under the influence of gravity.

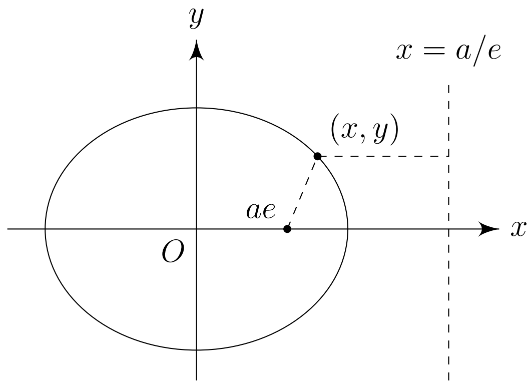
6.2 Focus-directrix property

Conic sections can be defined in a different way, in terms of

Definition 6.2.1 (Conic sections)**.** *The eccentricity and scale are properties of a conic section that satisfy the following:*
Let the foci of a conic section be $(\pm ae, 0)$ and the directrices be $x = \pm a/e$.
A conic section is the set of points whose distance from focus is $e\times$ distance from directrix which is closer to that of focus (unless $e = 1$, where we take the distance to the other directrix).

Now consider the different cases of e:

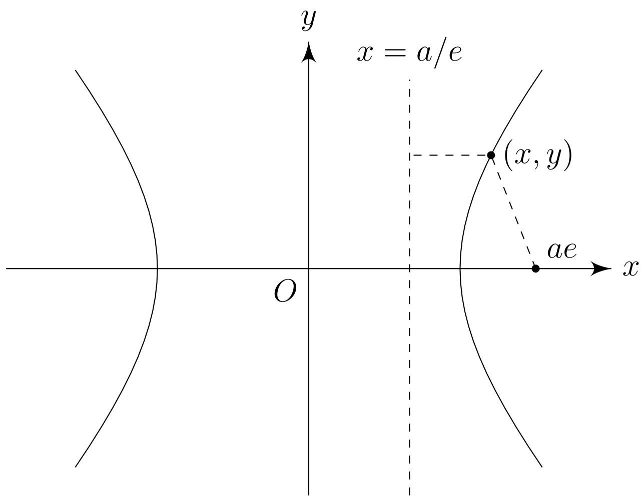
1. $e < 1$. By definition,



$$\sqrt{(x-ae)^2+y^2} = e\left(\frac{a}{e} - x\right)$$
$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1$$

Which is an ellipse with semi-major axis a and semi-minor axis $a\sqrt{1-e^2}$. (if $e = 0$, then we have a circle)

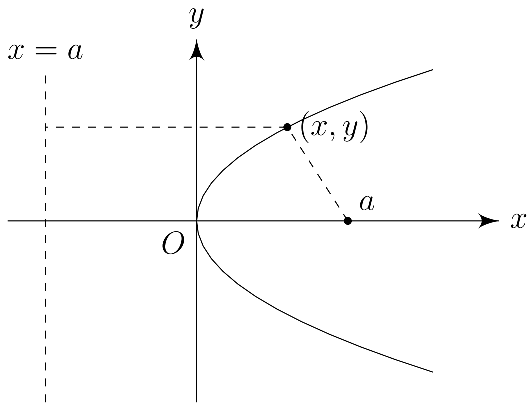
2. $e > 1$. So



$$\sqrt{(x-ae)^2+y^2} = e\left(x - \frac{a}{e}\right)$$
$$\frac{x^2}{a^2} - \frac{y^2}{a^2(e^2-1)} = 1$$

and we have a hyperbola.

3. $e = 1$: Then



$$\sqrt{(x-a)^2 + y^2} = (x+1)$$
$$y^2 = 4ax$$

and we have a parabola.

Conics also work in polar coordinates. We introduce a new parameter l such that l/e is the distance from the focus to the directrix. So

$$l = a|1 - e^2|.$$

We use polar coordinates (r, θ) centered on a focus. So the focus-directrix property is

$$r = e\left(\frac{l}{e} - r\cos\theta\right)$$
$$r = \frac{l}{1 + e\cos\theta}$$

We see that $r \to \infty$ if $\theta \to \cos^{-1}(-1/e)$, which is only possible if $e \geq 1$, i.e. hyperbola or parabola. But ellipses have $e < 1$. So r is bounded, as expected.

CHAPTER 7

Transformation groups

We have previously seen that orthogonal matrices are used to transform between orthonormal bases. Alternatively, we can see them as transformations of space itself that preserve distances, which is something we will prove shortly.

Using this as the definition of an orthogonal matrix, we see that our definition of orthogonal matrices is dependent on our choice of the notion of distance, or metric. In special relativity, we will need to use a different metric, which will lead to the Lorentz matrices, the matrices that conserve distances in special relativity. We will have a brief look at these as well.

7.1 Groups of orthogonal matrices

We know that if a matrix R is orthogonal, we have $R^T R = I \iff (R\mathbf{x}) \cdot (R\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \iff$ the rows or columns are orthonormal. The set of $n \times n$ matrices R forms the orthogonal group $O_n = O(n)$. If $R \in O(n)$ then $\det R = \pm 1$. $SO_n = SO(n)$ is the special orthogonal group, which is the subgroup of $O(n)$ defined by $\det R = 1$. If some matrix R is an element of $O(n)$, then R preserves the modulus of n-dimensional volume. If $R \in SO(n)$, then R preserves not only the modulus but also the sign of such a volume.

$SO(n)$ consists precisely of all rotations in $\mathbb{R}^n$. $O(n) \setminus SO(n)$ consists of all reflections. For some specific $H \in O(n) \setminus SO(n)$, any element of $O(n)$ can be written as a product of H with some element in $SO(n)$, i.e. R or RH with $R \in SO(n)$. For example, if n is odd, we can choose $H = -I$.

Now, we can consider the transformation $x'_i = R_{ij}x_j$ under two distinct points of view.

- (active) The rotation R acts on the vector $\mathbf{x}$ and yields a new vector $\mathbf{x}'$. The x'_i are components of the transformed vector in terms of the standard basis vectors.
- (passive) The x'_i are components of the same vector $\mathbf{x}$ but with respect to new orthonormal basis vectors $\mathbf{u}_i$. In general, $\mathbf{x} = \sum_i x_i \mathbf{e}_i = \sum_i x'_i \mathbf{u}_i$ which is true where $\mathbf{u}_i = \sum_j R_{ij}\mathbf{e}_j = \sum_j \mathbf{e}_j P_{ji}$. So $P = R^{-1} = R^T$ where P is the change of basis matrix.

Proposition 7.1.1. *The set of all $n \times n$ orthogonal matrices P forms a group under matrix multiplication.*

Proof.

0. If P, Q are orthogonal, then consider $R = PQ$. $RR^T = (PQ)(PQ)^T = P(QQ^T)P^T = PP^T = I$. So R is orthogonal.
1. I satisfies $II^T = I$. So I is orthogonal and is an identity of the group.
2. Inverse: if P is orthogonal, then $P^{-1} = P^T$ by definition, which is also orthogonal.
3. Matrix multiplication is associative since function composition is associative.

$\square$

Definition 7.1.1 (Orthogonal group)**.** *The orthogonal group $O(n)$ is the group of orthogonal matrices.*

Definition 7.1.2 (Special orthogonal group). *The special orthogonal group is the subgroup of $O(n)$ that consists of all orthogonal matrices with determinant* 1.

In general, we can show that any matrix in $O(2)$ is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

7.2 Motivation and basic definitions

7.2.1 Length preserving matrices

Theorem 7.2.1. *Let $P \in O(n)$. Then the following are equivalent:*

1. *P is orthogonal*
2. $|P\mathbf{x}| = |\mathbf{x}|$
3. $(P\mathbf{x})^T(P\mathbf{y}) = \mathbf{x}^T\mathbf{y}$, *i.e.* $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.
4. *If* $(\mathbf{v}_1, \mathbf{v_2}, \cdots, \mathbf{v}_n)$ *are orthonormal, so are* $(P\mathbf{v}_1, P\mathbf{v}_2, \cdots, P\mathbf{v}_n)$
5. *The columns of P are orthonormal.*

Proof. We do them one by one:

1. $\Rightarrow$ (ii): $|P\mathbf{x}|^2 = (P\mathbf{x})^T(P\mathbf{x}) = \mathbf{x}^T P^T P\mathbf{x} = \mathbf{x}^T\mathbf{x} = |\mathbf{x}|^2$
2. $\Rightarrow$ (iii): $|P(\mathbf{x}+\mathbf{y})|^2 = |\mathbf{x}+\mathbf{y}|^2$. The right hand side is

   $$(\mathbf{x}^T + \mathbf{y}^T)(\mathbf{x}+\mathbf{y}) = \mathbf{x}^T\mathbf{x} + y^T\mathbf{y} + \mathbf{y}^T\mathbf{x} + \mathbf{x}^T\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2\mathbf{x}^T\mathbf{y}.$$

   Similarly, the left hand side is

   $$|P\mathbf{x} + P\mathbf{y}|^2 = |P\mathbf{x}|^2 + |P\mathbf{y}| + 2(P\mathbf{x})^T P\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2(P\mathbf{x})^T P\mathbf{y}.$$

   So $(P\mathbf{x})^T P\mathbf{y} = \mathbf{x}^T\mathbf{y}$.
3. $\Rightarrow$ (iv): $(P\mathbf{v}_i)^T P\mathbf{v}_j = \mathbf{v}_i^T\mathbf{v}_j = \delta_{ij}$. So $P\mathbf{v}_i$'s are also orthonormal.
4. $\Rightarrow$ (v): Take the $\mathbf{v}_i$'s to be the standard basis. So the columns of P, being $P\mathbf{e}_i$, are orthonormal.

5. $\Rightarrow$ (i): The columns of P are orthonormal. Then $(PP^T)_{ij} = P_{ik}P_{jk} = (P_i)\cdot(P_j) = \delta_{ij}$, viewing P_i as the ith column of P. So $PP^T = I$. $\square$

Therefore the set of length-preserving matrices is precisely $O(n)$.

7.3 Minkowski space

Consider a new 'inner product' on $\mathbb{R}^2$ given by

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T J \mathbf{y}; \quad J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\therefore \left(\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}, \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \right) = x_0 y_0 - x_1 y_1$$

We start indexing these vectors from zero, not one. Here are some important properties.

- This 'inner product' is not positive definite. In fact, $(\mathbf{x}, \mathbf{x}) = x_0^2 - x_1^2$. (This is a quadratic form for $\mathbf{x}$ with eigenvalues ± 1.)
- It is bilinear and symmetric.
- Defining $\mathbf{e}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, they obey

  $$(\mathbf{e}_0, \mathbf{e}_0) = -(\mathbf{e}_1, \mathbf{e}_1) = 1; \quad (\mathbf{e}_0, \mathbf{e}_1) = 0$$

  This is similar to orthonormality, in this generalised sense.

This inner product is known as the Minkowski metric on $\mathbb{R}^2$. $\mathbb{R}^2$ with this metric is called Minkowski space.

7.4 Lorentz transformations

Consider the Minkowski $1 + 1$ dimension spacetime (i.e. 1 space dimension and 1 time dimension)

Definition 7.4.1 (Minkowski inner product)**.** *The* Minkowski *inner product of 2 vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *is*

$$\langle \mathbf{x} \mid \mathbf{y} \rangle = \mathbf{x}^T J \mathbf{y},$$

where

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

Then $\langle \mathbf{x} \mid \mathbf{y} \rangle = x_1 y_1 - x_2 y_2$.

This is to be compared to the usual Euclidean inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, given by

$$\langle \mathbf{x} \mid \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{x}^T I \mathbf{y} = x_1 y_1 + x_2 y_2.$$

Definition 7.4.2 (Preservation of inner product)**.** *A transformation matrix* M *preserves the Minkowski inner product if*

$$\langle \mathbf{x} | \mathbf{y} \rangle = \langle M\mathbf{x} | M\mathbf{y} \rangle$$

for all $\mathbf{x}, \mathbf{y}$.

We know that $\mathbf{x}^T J \mathbf{y} = (M\mathbf{x})^T J M \mathbf{y} = \mathbf{x}^T M^T J M \mathbf{y}$. Since this has to be true for all $\mathbf{x}$ and $\mathbf{y}$, we must have

$$J = M^T J M.$$

We can show that M takes the form of

$$H_\alpha = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} \text{ or } K_{\alpha/2} = \begin{pmatrix} \cosh\alpha & -\sinh\alpha \\ \sinh\alpha & -\cosh\alpha \end{pmatrix}$$

where H_α is a hyperbolic rotation, and $K_{\alpha/2}$ is a hyperbolic reflection.

This is technically all matrices that preserve the metric, since these only include matrices with $M_{11} > 0$. In physics, these are the matrices we want, since $M_{11} < 0$ corresponds to inverting time, which is frowned upon.

Definition 7.4.3 (Lorentz matrix)**.** *A* Lorentz matrix *or a Lorentz boost is a* *matrix in the form*

$$B_v = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}.$$

Here $|v| < 1$*, where we have chosen units in which the speed of light is equal to* 1*. We have* $B_v = H_{\tanh^{-1} v}$

Definition 7.4.4 (Lorentz group)**.** *The* Lorentz group *is a group of all Lorentz matrices under matrix multiplication.*

It is easy to prove that this is a group. For the closure axiom, we have $B_{v_1}B_{v_2} = B_{v_3}$, where

$$v_3 = \tanh(\tanh^{-1} v_1 + \tanh^{-1} v_2) = \frac{v_1 + v_2}{1 + v_1 v_2}$$

The set of all B_v is a group of transformations which preserve the Minkowski inner product.