Vectors and Matrices

November 30, 2021

Contents

0 Introduction

This course covers "linear algebra", topics in *algebra* & *geometry*
It involves approaches that are

concrete & abstract
computational & conceptual

The key ideas to develop / build on are:

- Elementary geometry (Euclidean): points, lines, planes in 2d or 3d; length, angles
- Points described by coordinates
- Points described by vectors; what is a vector?
- Simple transformations e.g. rotations & reflections $\to$ linear maps.

0.1 Plan

1. Complex Numbers
2. Vectors in 3 dimensions
3. Vectors in General, $\mathbb{R}^n$ & $\mathbb{C}^n$
4. Matrices & Linear Maps
5. Determinants & Inverses
6. Eigenvalues & Eigenvectors
7. Changing Bases, Canonical Forms & Symmetries

1 Complex Numbers

1.1 Basic Definitions

The following terms will not be defined here but assumed to be understood:

- $\mathbb{C}$, $+$, $\times$
- conjugate, modulus, argument
- complex plane / Argand diagram

Construct $\mathbb{C}$ by adding an element i to real numbers $\mathbb{R}$, with

$$i^2 = -1.$$

Any complex number $z \in \mathbb{C}$ has the form

$$z = x + iy \qquad \text{with } x, y \in \mathbb{R};$$

$x = \mathrm{Re}(z)$ is the *real part*; $y = \mathrm{Im}(z)$ is the *imaginary part.*
$\mathbb{R} \subset \mathbb{C}$ consisting of elements $x = i0 = x$.
In following, use notation above &

$$z_1 = x_1 + iy_1, \quad z_2 = x_2 + iy_2 \quad \text{etc.}$$

1. Addition (& subtraction). Define

   $$z_1 \pm z_2 = (x_1 \pm x_2) + i(y_1 \pm y_2)$$

2. Multiplication. Define

   $$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)$$

   If $z \neq 0$, note that

   $$z^{-1} = \frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}$$

   satisfies $zz^{-1} = 1$.

3. Complex conjugate Define

   $$\overline{z} = z^* = x - iy$$

   Then:

   $$\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z})$$

   and

   $$\mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z})$$

   $\overline{(\overline{z})} = z$ & further

   $$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$$
   $$\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$$

4. Modulus is defined by $r = |z|$, real & ≥ 0, with $r^2 = |z|^2 = z\overline{z} = x^2 + y^2$

5. Argument $\theta = \arg(z)$ real, defined for $z \neq 0$ by

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

for some real θ (this is known as *polar form*)

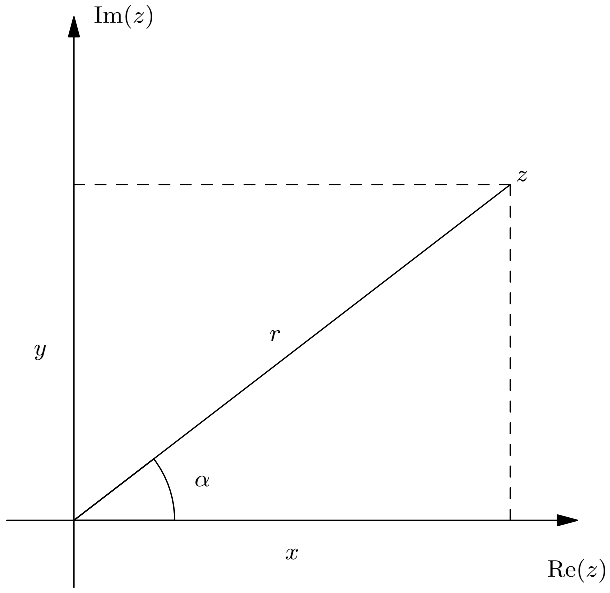
$$\cos\theta = \frac{x}{\sqrt{x^2+y^2}}, \quad \sin\theta = \frac{y}{\sqrt{x^2+y^2}}$$

$$\implies \tan\theta = \frac{y}{x}$$

$\arg(z)$ is determined only mod 2π i.e. can change $\theta \to \theta + 2n\pi$ for $n \in \mathbb{Z}$.
To make it unique we can restrict the range, e.g. the *principal value* defined by

$$-\pi < \theta \leq \pi$$

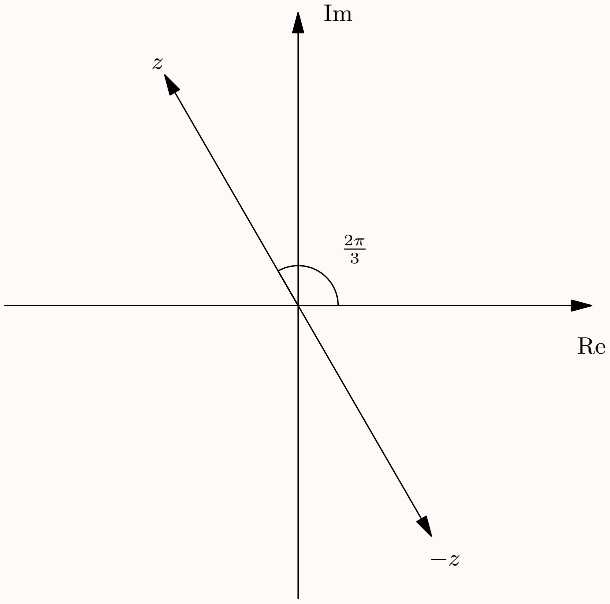
6. Argand diagram & Complex Plane Plot $\mathrm{Re}(z)$ & $\mathrm{Im}(z)$ on orthogonal axes, then $r = |z|$ & $\theta = \arg(z)$ are length & angle shown



Example. Consider

$$z = -1 + i\sqrt{3} = 2\left(-\frac{1}{2} + i\frac{\sqrt{3}}{2}\right)$$

here $z = 2$ & $\arg(z) = \frac{2\pi}{3} + 2n\pi$. Note $\tan\theta = -\sqrt{3} \implies \theta = \frac{2\pi}{3} + 2n\pi = \arg(z)$ *or* $\theta = -\frac{\pi}{3} + 2n\pi = \arg(-z)$.



1.2 Basic Properties & Consequences

Aside (motivating the definitions leading to $\mathbb{C}$)
Note that $\mathbb{Z}$ can be seen as a way to solve some equations involving $\mathbb{Z}$, for example $x+3 = 0$. Rational numbers can then be used to solve other equations such as $5x+1 = 0$, and real numbers are used to solve some quadratics and other higher degree polynomials, such as x^2-2. Finally, the complex numbers are used to allow us to solve more equations that we couldn't before, such as $x^2 + 4 = 0$. This leads to the fundamental theorem of algebra.

(i) $\mathbb{C}$ with operations $+, \times$ is a *field*.

i.e. $\mathbb{C}$ with $+$ is an abelian group & distributive laws hold, i.e.

$$z_1(z_2 + z_3) = z_1z_2 + z_1z_3.$$

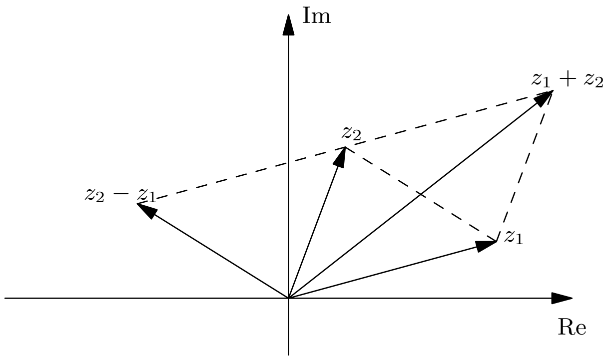
(ii) Fundamental Theorem of Algebra A polynomial of degree n with coefficients in $\mathbb{C}$ can be written as a product of n linear factors

$$P(z) = c_nz^n + \cdots + c_1z + c_0 \qquad c_i \in \mathbb{C}, c_n \neq 0$$

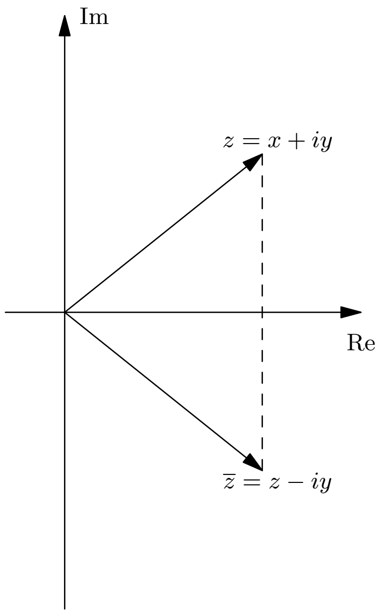
$$= c_n(z-\alpha_1)\cdots(z-\alpha_n) \qquad \alpha_i \in \mathbb{C}.$$

Hence $P(z) = 0$ has at least one root & n roots counted with multiplicity.

(iii) Addition & Subtraction as parallelogram constructions:



Complex conjugation is reflection in real axis



(iv) **Proposition** (Composition Property)**.** Modulus / length obeys

$$|z_1 z_2| = |z_1||z_2|$$

Proof. This result follows immediately by just expanding. □

Proposition (Triangle Inquality)**.**

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

Proof. Compare

$$LHS^2 = (z_1 + z_2)\overline{(z_1 + z_2)}$$
$$RHS^2 = |z_1|^2 + 2|z_1||z_2| + |z_2|^2$$

Compare "cross terms":

$$z_1\overline{z_2} + z_2\overline{z_1} \leq 2|z_1||z_2|$$
$$\iff \frac{1}{2}(z_1\overline{z_2} + \overline{(z_1\overline{z_2})}) \leq |z_1||\overline{z_2}|$$
$$\iff \text{Re}(z_1\overline{z_2}) \leq |z_1\overline{z_2}|$$

as desired. □

Proposition (Alternative form of triangle inequality)**.** Replace z_1 by $z_2 - z_1$ and rearrange to get

$$|z_2 - z_1| \geq |z_2| - |z_1|$$
$$\text{or} \geq |z_1| - |z_2$$

so

$$|z_2 - z_1| \geq ||z_2| - |z_1||$$

(v) **Proposition.** $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ implies that

$$z_1 z_2 = r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2))$$

Proof. Just expand and apply trig formulae. □

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

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \qquad \forall n \in \mathbb{Z}$$

(for $z \neq 0$, $z^0 = 1$ & $z^{-n} = (z^{-1})^n$ for $n > 0$.)

Proof. Use the proposition above and induct. □

1.3 Exponential & Trigonometric Functions

Define exp, cos, sin as functions on $\mathbb{C}$ by

$$\exp(z) = e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

$$\begin{aligned}\cos(z) &= \frac{1}{2}(e^{iz} + e^{-iz}) \\ &= 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots\end{aligned}$$

$$\begin{aligned}\sin(z) &= \frac{1}{2i}(e^{iz} - e^{-iz}) \\ &= z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdots\end{aligned}$$

These series converge $\forall z \in \mathbb{C}$ and such series can be multiplied, rearranged, and differentiated.
Furthermore

$$e^z e^w = e^{z+w}$$

From above

$$e^0 = 1 \qquad \text{and} \qquad (e^z)^n = e^{nz} \qquad n \in \mathbb{Z}$$

Proof. Induction for positive integers, and for negative integers use

$$e^z e^{-z} = 1 \implies e^{-z} = (e^z)^{-1}$$

□

Lemma. For $z = x + iy$

(i) $e^z = e^x(\cos y + i \sin y)$

(ii) exp on $\mathbb{C}$ takes all complex values except 0.

(iii) $e^z = 1 \iff z = 2n\pi i$, $n \in \mathbb{Z}$.

Proof.

(i) $e^{x+iy} = e^x e^{iy}$ but $e^{iy} = \cos y + i \sin y$.

(ii) $|e^z| = e^x$ takes all real values > 0. $\arg(e^z) = y$ taking all possible values.

(iii)

$$\begin{aligned}e^z = 1 &\iff e^x = 1, \cos y = 1, \sin y = 0 \\ &\iff x = 0 \text{ and } y = 2\pi n\end{aligned}$$

as required.

□

Returning to polar form or mod / arg form (Subsection 1.1 (v)), this can be written

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}$$

for $r = |z|$ and $\theta = \arg(z)$.
De Moivre's Theorem now follows from

$$(e^{i\theta})^n = e^{in\theta}.$$

Roots of unity

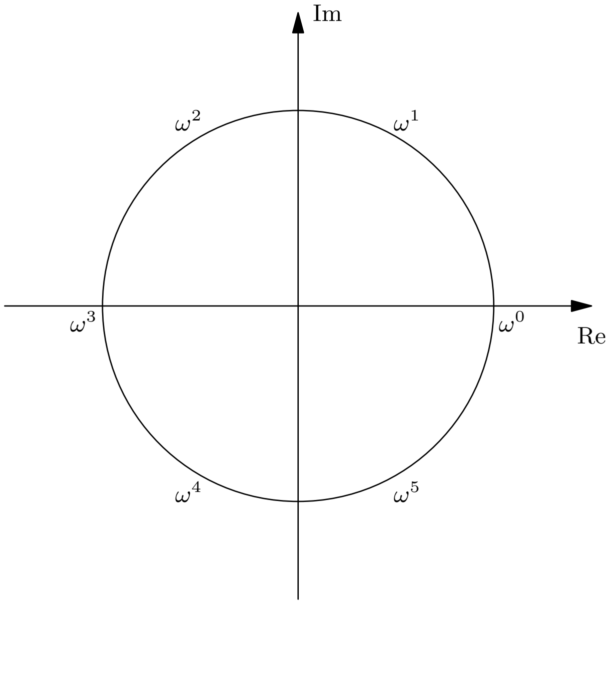
z is an N-th root of unity if $z^N = 1$. To find all solutions:

$$\begin{aligned} z = re^{i\theta} &\text{ satisfies } z^N = 1 \\ \iff& r^N e^{iN\theta} = 1 \\ \iff& r^N = 1 \text{ and } N\theta = 2n\pi \qquad n \in \mathbb{Z} \end{aligned}$$

This gives N distinct solutions:

$$\begin{aligned} z &= e^{2\pi n/N} \qquad n = 0, 1, \ldots, N-1 \\ &= \cos\frac{2\pi n}{N} + i\sin\frac{2\pi n}{N} \\ &= \omega^n \end{aligned}$$

where $\omega = e^{2\pi/N}$.

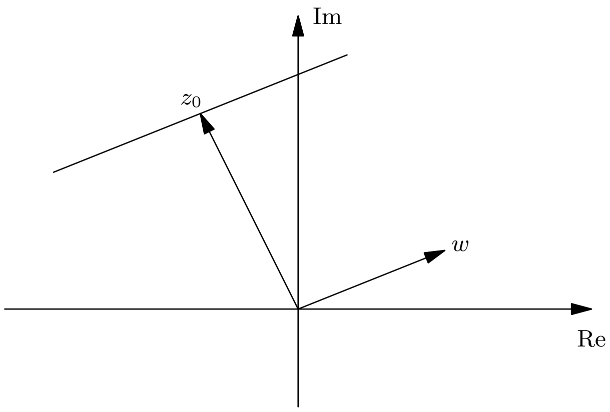


1.4 Transformations; lines & circles

Consider the following transformations on $\mathbb{C}$ (maps $\mathbb{C} \to \mathbb{C}$).

$z \mapsto z + a$	translation by $a \in \mathbb{C}$
$z \mapsto \lambda z$	scaling by $\lambda \in \mathbb{R}$
$z \mapsto e^{i\alpha} z$	rotation by $\alpha \in \mathbb{R}$
$z \mapsto \overline{z}$	reflection in real axis
$z \mapsto \dfrac{1}{z}$	inversion

Consider general point on a *line* in $\mathbb{C}$ through z_0 and parallel to $w \neq 0$ (fixed z_0, $w \in \mathbb{C}$):



$$z = z_0 + \lambda w$$

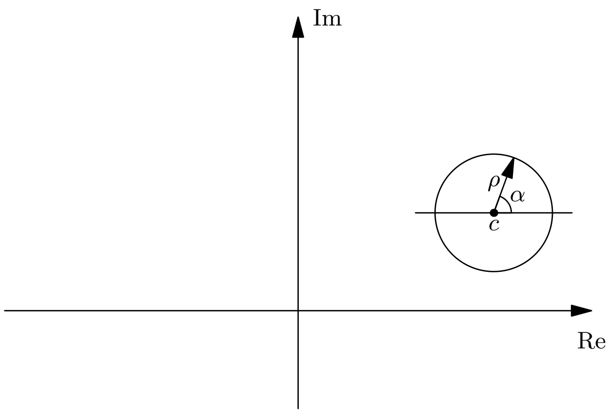
for any real parameter λ.
To eliminate λ, take conjugate

$$\overline{z} = \overline{z_0} + \lambda \overline{w}$$

and then combine

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

Consider general point on a *circle* with centre $c \in \mathbb{C}$ and radius ρ:



$$z = c + \rho e^{i\alpha} \qquad \text{for any real } \alpha$$

Equivalently

$$|z - c| = \rho$$

or $|z^2| - \bar{c}z - c\bar{z} = \rho^2 - |c|^2$. (squaring sides above). Möbius transformations are generated by translations, scalings, rotations and inversion. They can be viewed as acting on

$$\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$$

which is geometrically a sphere (see IA Groups).

1.5 Logarithms & Complex Powers

Define

$$w = \log z \qquad z \in \mathbb{C}, z \neq 0$$

by

$$e^w = \exp w = z$$

i.e. log is inverse of exp but exp is many-to-one ($e^z = e^{z+2n\pi i}$) and so log is *multi-valued*.

$$z = re^{i\theta} = e^{\log r} e^{i\theta} = e^{\log r + i\theta} \implies \log z = \log(r + i\theta) = \log|z| + i \arg|z|$$

Multiple values of arg and log are related:

$$\theta \to \theta + 2n\pi$$

$$\log z \to \log z + 2n\pi i$$

where $n \in \mathbb{Z}$. To make them single valued we can restrict e.g. $0 \leq \theta < 2\pi$ or $-\pi < \theta \leq \pi$ (called the *principal value*).

Example.

$$z = -3i = 3(-i) = e^{\log 3} e^{-i\pi/2 + 2n\pi i} = e^{\log 3 - i\pi/2 + 2n\pi i}$$

Hence

$$\log z = \log 3 - \frac{i\pi}{2} + 2n\pi i$$

$$\arg z = \begin{cases} 3\pi/2 & \text{if we use } 0 \leq \theta < 2\pi \\ -\pi/2 & \text{if we use } -\pi < \theta \leq \pi \end{cases}$$

We define *complex powers* by

$$z^\alpha = e^{\alpha \log z} \qquad z \in \mathbb{C}, z \neq 0 \& \alpha \in \mathbb{C}$$

This is multi-valued in general under the change $\arg z \to \arg z + 2n\pi$

$$z^\alpha \to z^\alpha e^{2\pi i n\alpha}$$

(i) If $\alpha = P \in \mathbb{Z}$ then $z^\alpha = z^p$ unique.

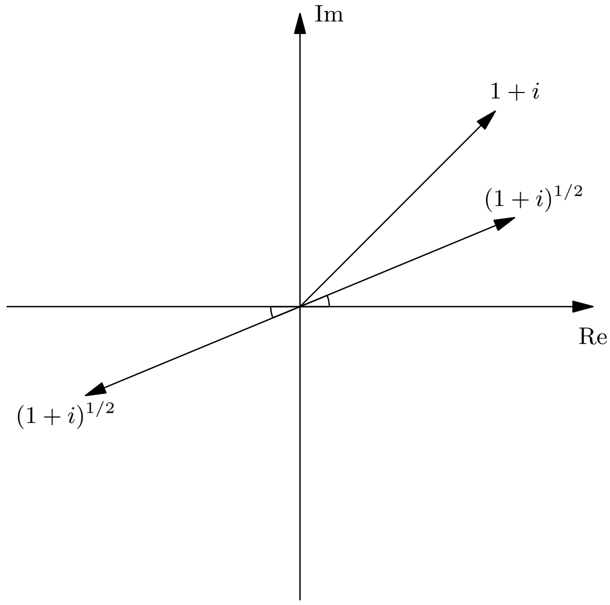
(ii) If $\alpha = \frac{p}{q} \in \mathbb{Q}$, then $z^\alpha = z^{p/q}$ takes finitely many values.

but in general we have *infinitely* many values.

Examples

- $(1+i)^{1/2}$: $1+i = \sqrt{2}e^{i\pi/4} = e^{\frac{1}{2}\log 2 + i\pi/4}$ Hence

$$\begin{aligned} \log(1+i) &= \frac{1}{2}\log 2 + \frac{i\pi}{4} + 2n\pi i \\ \implies (1+i)^{1/2} &= e^{\frac{1}{2}\log(1+i)} \\ &= e^{\frac{1}{4}\log 2 + i\pi/8 + n\pi i} \\ &= 2^{1/4}e^{i\pi/8}(-1)^n \end{aligned}$$



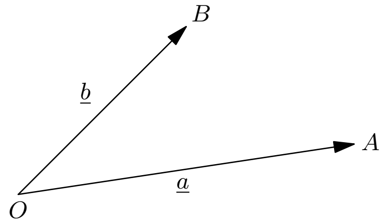
- $$\begin{aligned} (-3i)^i &= e^{i\log(-3i)} \\ &= e^{i(\log 3 - i\pi/2 + 2n\pi i)} \\ &= e^{i\log 3}e^{\pi/2 - 2n\pi} \qquad n \in \mathbb{Z} \end{aligned}$$

2 Vectors in 3 Dimensions

A vector is a quantity with magnitude and direction (e.g. force, electric and magnetic fields) - all examples modelled on *position*.

Take geometrical approach to position vectors in 3D space based on standard (Euclidean) notions of points, lines, planes, length, angle etc. Choose point O as the origin, then points A, B have position vectors

$$\underline{a} = \vec{OA}, \qquad \underline{b} = \vec{OB}$$



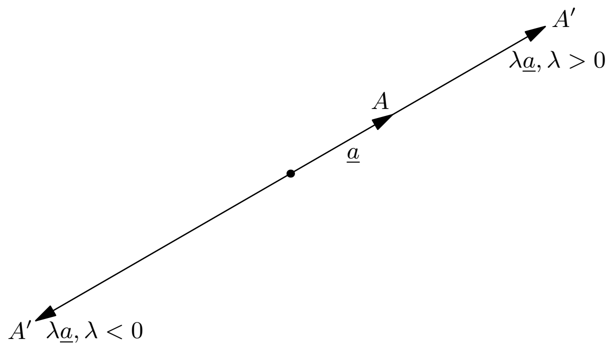
lengths denoted by $|\underline{a}| = |\vec{OA}|$. Also, $\underline{o}$ is the position vector for O.

2.1 Vector Addition and Scalar Multiplication

(i) Scalar Multiplication Given $\underline{a}$, position vector for A, and a *scalar* $\lambda \in \mathbb{R}$, $\lambda\underline{a}$ is position vector of point A' on OA with

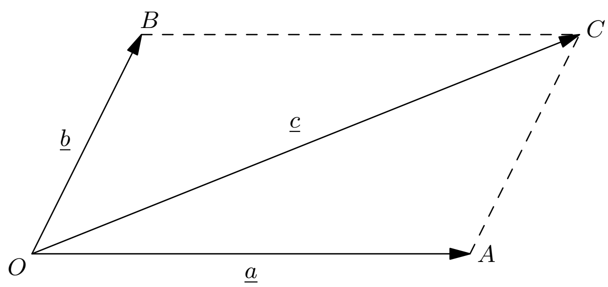
$$|\lambda\underline{a}| = |\vec{OA'}| = |\lambda||ula|$$

as shown



Say $\underline{a}$ and $\underline{b}$ are *parallel*, $\underline{a} \parallel \underline{b}$ iff $\underline{a} = \lambda\underline{b}$ or $\underline{b} = \lambda\underline{a}$. This definition allows $\lambda < 0$, and $\lambda = 0$ so $\underline{a} \parallel \underline{o}$ for any $\underline{a}$.

(ii) Given $\underline{a}$, $\underline{b}$ position vectors of A, B, construct a parallelogram $OACB$



and define $\underline{a} + \underline{b} = \underline{c}$, position vector of point C provided $\underline{a} \nparallel \underline{b}$; if $\underline{a} \parallel \underline{b}$ then we can write $\underline{a} = \alpha \underline{u}$, $\underline{b} = \beta \underline{u}$ for some $\underline{u}$, and then

$$\underline{a} + \underline{b} = (\alpha + \beta)\underline{u}$$

(iii) Properties For any vectors $\underline{a}, \underline{b}, \underline{c}$

$$\underline{a} + \underline{o} = \underline{o} + \underline{a} = \underline{a}$$

so $\underline{o}$ is the identity for $+$. We also have that there exists some $-\underline{a}$ such that

$$\underline{a} = (-\underline{a}) = (-\underline{a}) + \underline{a} = \underline{o}$$

so there exists an inverse of every vector. We also have

$$\underline{a} + \underline{b} = \underline{b} + \underline{a}$$

so $+$ is commutative. It is also associative, i.e.

$$\underline{a} + (\underline{b} + \underline{c}) = (\underline{a} + \underline{b}) + \underline{c}$$

We also have the following properties

$$\lambda(\underline{a} + \underline{b}) = \lambda\underline{a} + \lambda\underline{b}$$

$$(\lambda + \mu)\underline{a} = \lambda\underline{a} + \mu\underline{a}$$

$$\lambda(\mu\underline{a}) = (\lambda\mu)\underline{a}$$

All can be checked geometrically i.e. associativity from parallelepiped.

(iv) Linear Combinations and Span A *linear combination* of vectors $\underline{a}, \underline{b}, \dots, \underline{c}$ is an expression

$$\alpha\underline{a} + \beta\underline{b} + \cdots + \gamma\underline{c}$$

for some $\alpha, \beta, \dots, \gamma \in \mathbb{R}$. The *span* of a set of vectors is

$$\begin{aligned} &\text{span}\{\underline{a}, \underline{b}, \dots, \underline{c}\} \\ = \{\alpha\underline{a} + \beta\underline{b} + \cdots + \gamma\underline{c} &: \alpha, \beta, \dots, \gamma \in \mathbb{R}\} \end{aligned}$$

If $\underline{a} \neq \underline{a}$ then $\text{span}\{\underline{a}\} = \{\lambda\underline{a}\}$, i.e. the *line* through O and A. If $\underline{a} \nparallel \underline{b}$ then

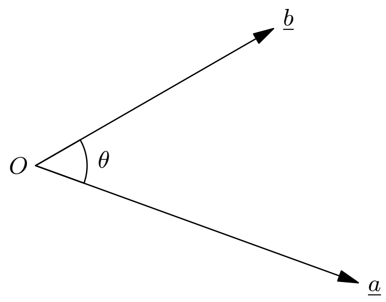
$$\text{span}\{\underline{a}, \underline{b}\} = \{\alpha\underline{a} + \beta\underline{b} : \alpha, \beta \in \mathbb{R}\}$$

i.e. the *plane* through O, A and B.

2.2 Scalar or Dot Product

(i) Definition: Given $\underline{a}$ and $\underline{b}$ let θ be the angle between them; then

$$\underline{a} \cdot \underline{b} = |\underline{a}||\underline{b}| \cos\theta$$



scalar or *dot product* or *inner product* (θ defined unless $|\underline{a}|$ or $|\underline{b}| = 0$ and then $\underline{a} \cdot \underline{b} = 0$.)

(ii) Properties

$$\underline{a} \cdot \underline{b} = \underline{b} \cdot \underline{a}$$

$$\underline{a} \cdot \underline{a} = |\underline{a}|^2 \geq 0 \& = 0 \text{ iff } \underline{a} = 0$$

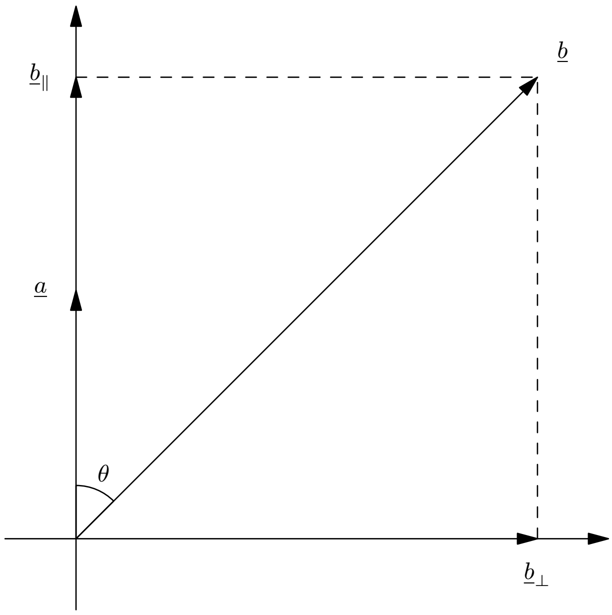
$$(\lambda \underline{a}) \cdot \underline{b} = \lambda(\underline{a} \cdot \underline{b}) = \underline{a} \cdot (\lambda \underline{b})$$

$$\underline{a} \cdot (\underline{b} + \underline{c}) = \underline{a} \cdot \underline{b} + \underline{a} \cdot \underline{c}$$

(iii) Interpretation For $\underline{a} \neq 0$, consider $\underline{u} = \frac{\underline{a}}{|\underline{a}|}$

$$\underline{u} \cdot \underline{b} = \frac{1}{\underline{a}} \underline{a} \cdot \underline{b} = |\underline{b}| \cos\theta$$

is *component* of $\underline{b}$ along $\underline{a}$.



We can *resolve* $\underline{b} = \underbrace{\underline{b}_{\parallel}}_{\parallel \underline{a}} + \underbrace{\underline{b}_{\perp}}_{\perp \underline{a}}$ where $\underline{a} \perp \underline{b}$ iff $\underline{a} \cdot \underline{b} = 0$. Note $\underline{a} \cdot \underline{b} = \underline{a} \cdot \underline{b}_{\parallel}$. (The expressions can be computed as $\underline{b}_{\parallel} = (\underline{b} \cdot \underline{u})\underline{u}$, $\underline{b}_{\perp} = \underline{b} - (\underline{b} \cdot \underline{u})\underline{u}$.

In general, vectors $\underline{a}$ and $\underline{b}$ are *orthogonal* or *perpendicular*, written

$$\underline{a} \perp \underline{b} \iff \underline{a} \cdot \underline{b} = 0$$

definition allows $\underline{a}$ or $\underline{b} = \underline{o}$; $\underline{o} \perp$ any vector.

2.3 Orthonormal Bases and Components

Choose vectors $\underline{e_1}$, $\underline{e_2}$, $\underline{e_3}$ that are *orthonormal* i.e. each of unit length and mutually perpendicular.

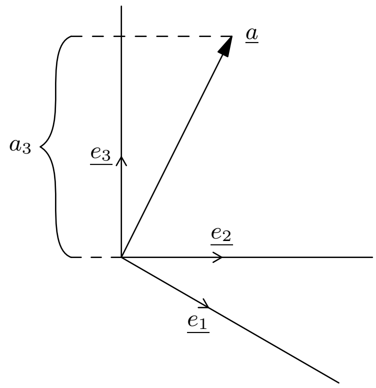
$$\underline{e_j} \cdot \underline{j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Equivalent to choosing Cartesian axes along these directions, $\{\underline{e_i}\}$ is a *basis*: any vector can be expressed

$$\underline{a} = \sum_i a_i \underline{e_i} = a_1 \underline{e_1} + a_2 \underline{e_2} = a_3 \underline{e_3}$$

and each component a_i is *uniquely* determined.

$$a_i = \underline{e_i} \cdot \underline{a}$$



Each $\underline{a}$ can now be identified with set of components in

$$\underbrace{(a_1, a_2, a_3)}_{\text{row vector}} \qquad \text{or} \qquad \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}}_{\text{column vector}}$$

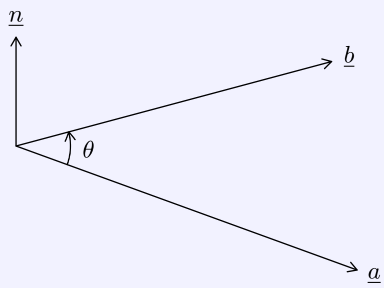
Note

$$\begin{aligned} \underline{a} \cdot \underline{b} &= asdf \left(\sum_i a_i \underline{e_i} \right) \cdot \left(\sum_j b_j \underline{e_j} \right) \\ &= a_1 b_1 + a_2 b_2 + a_3 b_3 \\ \text{and} |\underline{a}|^2 &= a_1^2 + a_2^2 + a_3^2 \qquad \text{Pythagoras} \end{aligned}$$

$\underline{e_1}$, $\underline{e_2}$, $\underline{e_3}$ are also often written $\underline{i}$, $\underline{j}$, $\underline{k}$.

2.4 Vector or Cross Product

Definition. Given $\underline{a}$ and $\underline{b}$, let θ be angle between them measured in sense shown relative to a unit normal $\underline{n}$ to the plan they span



"right-handed sense". (unit normal $\equiv$ unit vector $\perp$ plane);
then

$$\underline{a} \times \underline{b} = |\underline{a}||\underline{b}| \sin\theta \underline{n}$$

(sometimes $\wedge$ is used instead of $\times$) is *vector* or *cross* product.
Note n is defined up to a choice of sign if $\underline{a} \nparallel \underline{b}$, but changing sign of $\underline{n}$ means changing θ to $2\pi - \theta$ so definition is undefined; $\underline{n}$ is not defined it $\underline{a} \parallel \underline{b}$, and θ is not defined it $|\underline{a}|$ or $|\underline{b}| = 0$, but $\underline{a} \times \underline{b} = \underline{o}$ in these cases.

Properties

$$\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$$

$$(\lambda \underline{a}) \times \underline{b} = \lambda(\underline{a} \times \underline{b}) = \underline{a} \times (\lambda \underline{b})$$

$$\underline{a} \times (\underline{b} + \underline{c}) = \underline{a} \times \underline{b} + \underline{a} \times \underline{c}$$

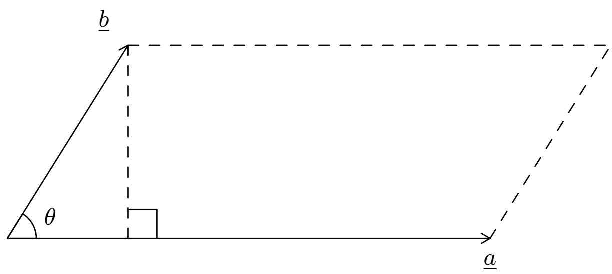
$$\underline{a} \times \underline{b} = \underline{o} \iff \underline{a} \parallel \underline{b}$$

$$\underline{a} \times \underline{b} \perp \underline{a} \& \underline{b}$$

$$\underline{a} \cdot (\underline{a} \times \underline{b}) = \underline{b} \cdot (\underline{a} \times \underline{b}) = 0.$$

Interpretations

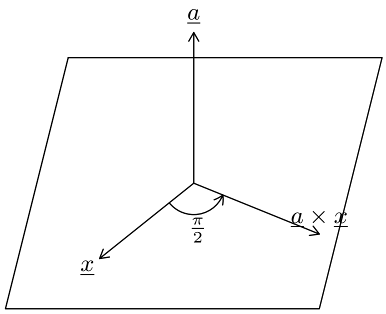
- $\underline{a} \times \underline{b}$ is the *vector* area of the parallelogram shown



$$\underline{a} \times \underline{b} = |\underline{a}||\underline{b}| \sin\theta \qquad \text{for } \sin\theta \geq 0 = \text{“base”} \times \text{“}\perp \text{ height”}$$

scalar area
Direction of normal $\underline{n}$ gives orientation of parallelogram in space.

- Fix $\underline{a}$ and consider $\underline{x} \perp \underline{a}$; then $\underline{x} \mapsto \underline{a} \times \underline{x}$ scales $|\underline{x}|$ by a factor of $|\underline{a}|$ and rotates $\underline{x}$ by $\pi/2$ in plane $\perp \underline{a}$ as shown.



Component Expressions

Consider $\underline{e_1}$, $\underline{e_2}$, $\underline{e_3}$ orthonormal basis as in section 2.3 but assume in addition that

$$\underline{e_1} \times \underline{e_2} = \underline{e_3} = -\underline{e_2} \times \underline{e_1}$$

$$\underline{e_1} \times \underline{e_3} = \underline{e_1} = -\underline{e_3} \times \underline{e_2}$$

$$\underline{e_3} \times \underline{e_1} = \underline{e_2} = -\underline{e_1} \times \underline{e_3}$$

(all equalities from any one) This is called a *right-handed* orthonormal basis. Now for

$$\underline{a} = \sum_i a_i \underline{e_i} = (a_1\underline{e_1} + a_2\underline{e_2} + a_3\underline{e_3})$$

and

$$\underline{b} = \sum_j b_j \underline{e_j} = (b_1\underline{e_1} + b_2\underline{e_2} + b_3\underline{e_3})$$

we get

$$\begin{aligned} \underline{a} \times \underline{b} = (a_2b_3 - a_3b_2)&\underline{e_1} \\ + (a_3b_1 - a_1b_3)&\underline{e_2} \\ + (a_1b_2 - a_2b_1)&\underline{e_3} \end{aligned}$$

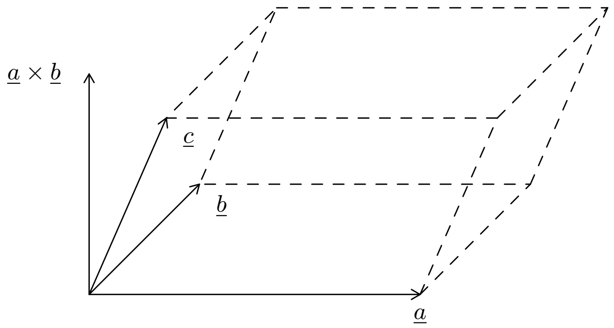
2.5 Triple Products

Scalar Triple Product

Notation. Define

$$\begin{aligned} &\underline{a}\cdot(\underline{b}\times\underline{c}) = \underline{b}\cdot(\underline{c}\times\underline{a}) = \underline{c}\cdot(\underline{a}\times\underline{b}) \\ &= -\underline{a}\cdot(\underline{c}\times\underline{b}) = -\underline{b}\cdot(\underline{a}\times\underline{c}) = -\underline{c}\cdot(\underline{b}\times\underline{a}) \\ &= [\underline{a},\underline{b},\underline{c}] \end{aligned}$$

Interpretation: $|\underline{c}\cdot(\underline{a}\times\underline{b})|$ is volume of parallelepiped shown = (area of parallelogram base) $\times$ ($\perp$ height) $= |\underline{a}\times\underline{b}||c||\cos\phi|$



$\underline{c}\cdot\underline{a}\times\underline{b}$ "signed volume"; if $\underline{c}\cdot\underline{a}\times\underline{b} > 0$ say $\underline{a}$, $\underline{b}$, $\underline{c}$ right-handed set.

Remark. $\underline{a}\cdot\underline{b}\times\underline{c} = 0$ if and only if $\underline{a}$, $\underline{b}$ and $\underline{c}$ are *co-planar* meaning one of them lies in plane spanned by other two. For example $\underline{c} = \alpha\underline{a} + \beta\underline{b}$ belonging to $\operatorname{span}\{\underline{a},\underline{b}\}$.

Example.

$$\underline{a} = (2, 0, -1) \qquad \underline{b} = (7, -3, 5)$$

$$\begin{aligned} \implies \underline{a}\times\underline{b} &= (0.5 - (-1)(-3))\underline{e_1} \\ &+ ((-1)\cdot 7 - 2.5)\underline{e_2} \\ &+ (2\cdot(-3) - 0.7)\underline{e_3} \\ &= (-3, -17, -6) \end{aligned}$$

Test whether $\underline{a}$, $\underline{b}$, $\underline{c}$ coplanar with $\underline{c} = (3, -3, 7)$

$$\underline{c}\cdot\underline{a}\times\underline{b} = 3(-3) + (-3)(-17) + 7(-6) = 0;$$

consistent with $\underline{c} = \underline{b} - 2\underline{a}$.

Vector Triple Product

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

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

Form of RHS is constrained by definitions above, or could check explicitly. Return to these formulas using index notation and summation convention.

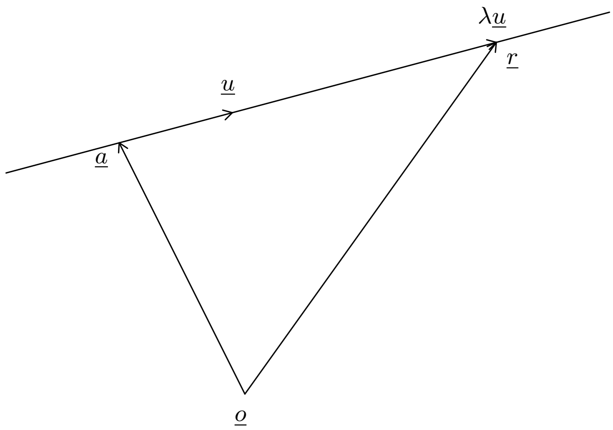
2.6 Lines, Planes and Other Vector Equations

(a) Lines

General point on a line through $\underline{a}$ with direction $\underline{u}(\neq \underline{o})$ has position vector

$$\underline{r} = \underline{a} = \lambda\underline{u} \qquad \lambda \in \mathbb{R}$$

parametric form



Alternative form without parameter λ obtained by crossing with $\underline{u}$:

$$\underline{u} \times \underline{r} = \underline{u} \times \underline{a}$$

Conversely

$$\underline{u} \times (\underline{r} - \underline{a}) = \underline{o}$$

and this holds if and only if

$$\underline{r} - \underline{a} = \lambda\underline{u}$$

for some real λ. Now consider

$$\underline{u} \times \underline{r} = \underline{C}$$

where $\underline{u}$, $\underline{c}$ are given vectors with $\underline{u} \neq \underline{o}$. Note that

$$\underline{u} \cdot (\underline{u} \times \underline{r}) = \underline{u} \cdot \underline{c} = 0$$

If $\underline{u} \cdot \underline{c} \neq 0$ then we hae a contradiction i.e. no solutions. If $\underline{u} \cdot \underline{c} = 0$, try a particular solution by considering

$$\underline{u} \times (\underline{u} \times \underline{c}) = (\underline{u} \cdot \underline{c})\underline{u} - (\underline{u} \cdot \underline{u})\underline{c} = -|\underline{u}|^2\underline{c}$$

Hence

$$\underline{a} = -\frac{1}{|\underline{u}|^2}(\underline{u} \times \underline{c})$$

is a solution. General solution (arguing as before) is

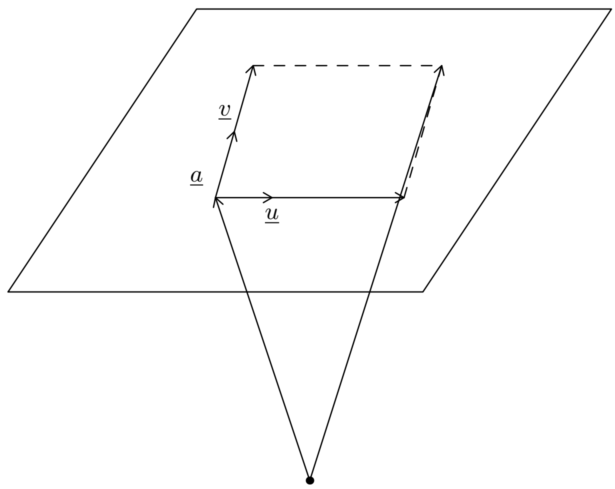
$$\underline{r} = \underline{a} + \lambda \underline{u}$$

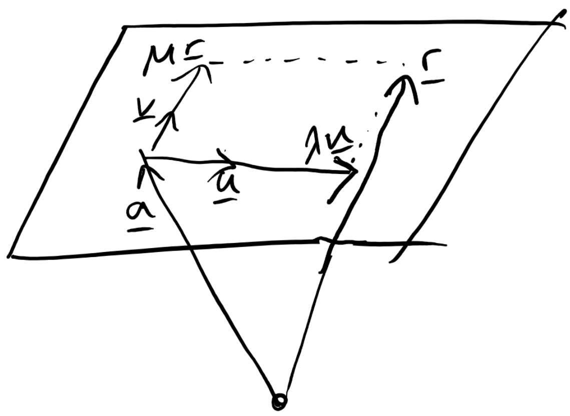
(b) Planes

General point on a plane through $\underline{a}$ with directions $\underline{u}$, $\underline{v}$ in plane ($\underline{u} \not\parallel \underline{v}$) has position vector

$$\underline{r} = \underline{a} + \lambda \underline{u} + \mu \underline{v} \qquad \lambda, \mu \in \mathbb{R}$$

parametric form





Alternative form without parameters obtained by dotting with normal

$$\underline{n} = \underline{u} \times \underline{v} \neq (\underline{o} \text{ since } \underline{u} \not\parallel \underline{v} \text{ but not necessarily a unit vector})$$

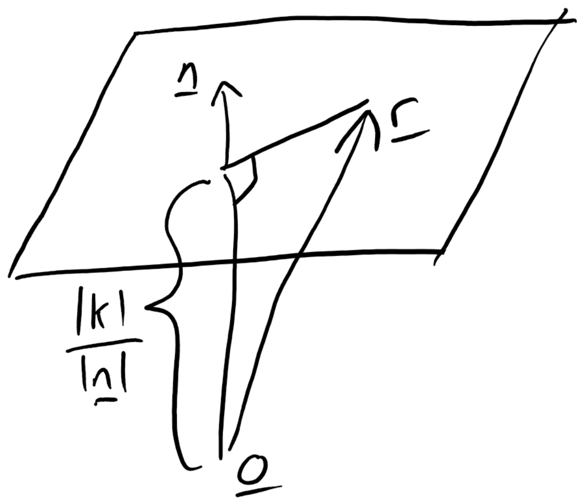
This gives

$$\underline{n} \cdot \underline{r} = \underline{n} \cdot \underline{a} = k$$

where k is a constant. Note component of $\underline{r}$ along $\underline{n}$ is

$$\frac{\underline{n} \cdot \underline{r}}{|\underline{n}|} = \frac{k}{|\underline{n}|} \qquad \text{(constant)}$$

is clearly a plane and moreover $\frac{|k|}{|\underline{n}|}$ is perpendicular distance of plane from $\underline{o}$.



(c) <u>Other Vector Equations</u>
Consider equations for $\underline{r}$ (unknown) written in vector notation with given (constant) vectors. Possible approaches:

- Can re-write and convert to some standard form, e.g.

$$|\underline{r}|^2 + \underline{r} \cdot \underline{a} = k, \qquad \text{constant}$$

  Then we can complete the square:

$$|\underline{r} + \frac{1}{2}\underline{a}|^2 = (\underline{r} + \frac{1}{2}\underline{a}) \cdot (\underline{r} + \frac{1}{2}\underline{a}) = k + \frac{1}{4}|\underline{a}|^2$$

  Equation of a sphere, centre $-\frac{1}{2}\underline{a}$ and radius $\sqrt{k + \frac{1}{4}|\underline{a}|^2}$, provided $k + \frac{1}{4}|\underline{a}|^2 > 0$. For equations linear in $\underline{r}$.
- Try dotting and crossing with constant vectors to learn more (see examples).
- Can try expressing

$$\underline{r} = \alpha\underline{a} + \beta\underline{b} + \gamma\underline{c}$$

  for some non-ω-planar $\underline{a}$, $\underline{b}$, $\underline{c}$ and solve for α, β, γ.
- Can choose basis and use index / matrix notation.

2.7 Index (suffix) Notation and the Summation Convention

(a) Components; δ and ε

Write vectors $\underline{a}$, $\underline{b}$, ... in terms of components. a_i, b_i, ..., with respect to an orthonormal, right-handed basis

$$\underline{e_1}, \underline{e_2}, \underline{e_3}$$

Indices or suffices $i, j, k, l, p, q, \ldots$ take values 1, 2, 3. Then

$$\underline{c} = \alpha\alpha\underline{a} + \beta\underline{b}$$

$$\iff c_i = [\alpha\underline{a} + \beta\underline{b}] = \alpha a_i \beta b_i$$

for $i = 1, 2, 3$ (*free index*)

$$\underline{a} \cdot \underline{b} = \sum_i a_i b_i = \sum_j a_j b_j$$

$$\underline{x} = \underline{a} = (\underline{b} \cdot \underline{c})\underline{d}$$

for $j = 1, 2, 3$ free index.

$$\iff x_j = a_j + \left(\sum_k b_k c_k\right) d_j$$

Definition (Kronecker Delta)**.**

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

$$\delta ij = \delta ji \qquad \text{(symmetric)}$$

As an asdfasdf matrix

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

Then

$$\underline{e_i} \cdot \underline{j} = \delta_{ij}$$

and

$$\begin{aligned} \underline{a} \cdot \underline{b} &= \left(\sum_i a_i \underline{e_i} \right) \cdot \left(\sum_j b_j \underline{e_j} \right) \\ &= \sum_{ij} a_i b_j \underline{e_i} \cdot \underline{e_j} \\ &= \sum_{ij} a_i b_j \delta_{ij} \\ &= \sum_i a_i b_i \end{aligned}$$

Definition (Levi-Civita Epsilon)**.**

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ odd permutation of } (1,2,3) \\ 0 & \text{else} \end{cases}$$

i.e.

$$\varepsilon_{123} = \varepsilon 231 = \varepsilon_{312} = +1$$

$$\varepsilon_{321} = \varepsilon_{213} = \varepsilon_{132} = -1$$

$$\varepsilon_{ijk} = 0 \qquad \text{if any two index values match}$$

Note that ε_{ijk} is *totally* antisymmetric: exchanging any pair of indices produces a change in sign. Then

$$\underline{e_i} \times \underline{e_j} = \sum_k \varepsilon_{ijk} \underline{e_k}$$

e.g.

$$\underline{e_2} \times \underline{e_1} = \sum_k \varepsilon_{21k} \underline{e_k} = \varepsilon_{213} \underline{e_3}$$

And

$$\begin{aligned} \underline{a} \times \underline{b} = \left(\sum_i a_i \underline{e_i} \right) \times \left(\sum_j b_j \underline{e_j} \right) \\ &= \sum_{ij} a_i b_j \underline{e_i} \times \underline{e_j} \\ &= \sum_{ij} a_i b_j \sum_k \varepsilon_{ijk} \underline{e_k} \\ &= \sum_k \left(\sum_{ij} \varepsilon_{ijk} a_i b_j \right) \underline{e_k} \end{aligned}$$

Hence

$$(\underline{a} \times \underline{b})_k = \sum_{ij} \varepsilon_{ijk} a_i b_j$$

e.g.

$$\begin{aligned} (\underline{a} \times \underline{b})_3 &= \sum_{ij} \varepsilon_{ij3} a_i b_j \\ &= \varepsilon_{123} a_1 b_2 + \varepsilon_{213} a_2 b_1 \\ &= a_1 b_2 - a_2 b_1 \end{aligned}$$

(b) Summation Convention

With component / index notation, we observe that indices that appear *twice* in a given term are (usually) summed over. In the summation convention we *omit* $\sum$ signs for repeated indices: the sum is understood.

Examples

(i)

$$\begin{aligned} & a_i\delta_{ij} \qquad \sum_i \text{ understood} \\ =& a_1\delta_{1j} + a_2\delta_{2j} + a_3\delta_{3j} \\ =& \begin{cases} a_1 & \text{if } j=1 \\ a_2 & \text{if } j=2 \\ a_3 & \text{if } j=3 \end{cases} \end{aligned}$$

or

$$a_i\delta_{ij} = a_j$$

true for $j = 1, 2, 3$.

(ii) Here on the first line we have $\sum_{i,j}$ is understood, and on the second line we have the $\sum_i$ is understood

$$\begin{aligned} \underline{a} \cdot \underline{b} &= \delta_{ij}a_ib_j \\ &= a_ib_i \end{aligned}$$

(iii) Here $\sum_{j,k}$ is understood

$$(\underline{a} \times \underline{b})_i = \varepsilon_{ijk}a_jb_k$$

(iv) Here $\sum_{ijk}$ is understood

$$\underline{a} \cdot \underline{b} \times \underline{c} = \varepsilon_{ijk}a_ib_jc_k$$

(v) Here $\sum_i$ is understood

$$\delta_{ii} = \delta_{11} + \delta 22 + \delta_{33} = 3$$

(vi) On the last line we have that $\sum_j$ is understood

$$\begin{aligned} [(\underline{a} \cdot \underline{c})\underline{b} - (\underline{a} \cdot \underline{b})\underline{c}]_i &= (\underline{a} \cdot \underline{c})b_i - (\underline{a} \cdot \underline{b})c_i \\ &= a_jc_jb_i - a_jb_jc_i \end{aligned}$$

Summation Convention Rules

(i) An index occurring exactly *once* in any term must appear once in *every* term and it can take any value - a *free* index.

(ii) An index occurring exactly *twice* in a given term is summed over - a *repeated* or *contracted* or *dummy* index.

(iii) No index can occur more than twice.

Application: proof of the vector triple product identity. Consider

$$\begin{aligned}[\underline{a} \times (\underline{b} \times \underline{c})]_i &= \varepsilon_{ijk} a_j (\underline{b} \times \underline{c})_k \\ &= \varepsilon_{ijk} a_j \varepsilon_{kpq} b_p c_q \\ &= \varepsilon_{ijk} \varepsilon_{pqk} a_j b_p c_q\end{aligned}$$

Now

$$\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$$

(see section (c) velow). Then

$$\begin{aligned}[\underline{a} \times (\underline{b} \times \underline{c})]_i &= (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}) a_j b_p c_q \\ &= a_j \delta_{ip} b_p \delta_{jq} c_q - a_j \delta_{jp} b_p \delta_{iq} c_q \\ &= a_j b_i c_j - a_j b_j c_i \\ &= (a_j c_j) b_i - (a_j b_j) c_i \\ &= (\underline{a} \cdot \underline{c}) b_i - (\underline{a} \cdot \underline{b}) c_i \\ &= [(\underline{a} \cdot \underline{c})\underline{b} - (\underline{a} \cdot \underline{b})\underline{c}]_i\end{aligned}$$

True for $i = 1, 2, 3$ hence

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

(c) $\varepsilon\ \varepsilon$ identities

- Expected to know this and quote it:

$$\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp} = \varepsilon_{kij}\varepsilon_{kpq}$$

Check: RHs and LHS are both antisymmetric (change sign) under

$$i \leftrightarrow j \qquad \text{or} \qquad p \leftrightarrow q$$

SO both sides vanish if i and j or p and q take same values. Now suffices to check

$$i = p = 1 \qquad \text{and} \qquad j = q = 2$$

$$LHS = \varepsilon_{123}\varepsilon_{123} = +1$$

$$RHS = \delta_{11}\delta_{22} - \delta_{12}\delta_{21} = +1$$

or $i = q = 1$ and $j = p = 2$

$$LHS = \varepsilon_{123}\varepsilon_{213} = (+1)(-1) = -1$$

$$RHS = \delta_{12}\delta 21 - \delta_{11}\delta_{22} = -1$$

All other index choices work similarly.

- $\varepsilon_{ijk}\varepsilon_{pjk} = 2\delta_{ip}$ contract result aove

$$\begin{aligned}\varepsilon_{ijk}\varepsilon_{pjk} &= \delta_{ip}\delta_{jj} - \delta_{ij}\delta_{jp} \\ &= 3\delta_{ip} - \delta_{ip} \\ &= 2\delta_{ip}\end{aligned}$$

- $\varepsilon_{ijk}\varepsilon_{ijk} = 6$.

3 Vectors in General; $\mathbb{R}^n$ and $\mathbb{C}^n$

3.1 Vectors in $\mathbb{R}^n$

(a) Definitions

If we regard vectors as sets of components, it is easy to generalise from 3 to n dimensions.

- Let $\mathbb{R}^n = \{\underline{x} = (x_1, \ldots, x_n : x_i \in \mathbb{R}\}$ and define
  (i) <u>addition</u>

$$\underline{x} + \underline{y} = (x_1 + y_1, \ldots, x_n + y_n)$$

  (ii) <u>scalar multiplication</u>

$$\lambda \underline{x} = (\lambda x_1, \ldots, \lambda x_n)$$

  for any $\underline{x}, \underline{y} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

- *Inner product* or *scalar product* on $\mathbb{R}^n$ is defined by

$$\underline{x} \cdot \underline{y} = \sum_i x_i y_i = x_1 y_1 + \cdots x_n y_n$$

  <u>Properties</u>

  (i) Symmetric $\underline{c} \cdot \underline{y} = \underline{y} \cdot \underline{x}$

  (ii) Bilinear (linear in each vector)

$$(\lambda \underline{X} + \lambda' \underline{x}') \cdot \underline{y} = \lambda(\underline{x} \cdot \underline{y}) + \lambda'(\underline{x}' \cdot \underline{y})$$

  and

$$\underline{x} \cdot (\mu \underline{y} + \mu' \underline{y}' = \mu(\underline{x} \cdot \underline{y}) + \mu'(\underline{x} \cdot \underline{y}')$$

  (iii) <u>Positive definite</u>

$$\underline{x} \cdot \underline{x} = \sum_i x_i^2 \geq 0$$

  and is equal to 0 if and only if $\underline{x} = \underline{o}$. The *length* or *norm* of vector $\underline{x}$ is $|\underline{x}|(\geq 0)$ defined by $|\underline{z}|^2 = \underline{x} \cdot \underline{x}$.

  (iv) For $\underline{x} \in \mathbb{R}^n$ we can write

$$\underline{x} = \sum_i x_i \underline{e_i}$$

  where

$$\begin{aligned} \underline{e_1} &= (1, 0, \ldots, 0) \\ \underline{e_1} &= (0, 1, \ldots, 0) \\ &\vdots \\ \underline{e_n} &= (0, 0, \ldots, 1) \end{aligned}$$

call $\{\underline{e_i}\}$ the *standard basis* for $\mathbb{R}^n$. Note that it is orthonormal:

$$\underline{e_i} \cdot \underline{e_j} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

(b) Cauchy-Schwarz and △ Inequalities Proposition

Proposition (Cauchy-Schwarz)**.**

$$|\underline{x} \cdot \underline{y}| \leq |\underline{x}||\underline{y}|$$

for $\underline{x}, \underline{y} \in \mathbb{R}^n$ and equality holds if and only if $\underline{x} = \lambda \underline{y}$ or $\underline{y} = \lambda \underline{x}$ ($\underline{x} \parallel \underline{y}$) for some $\lambda \in \mathbb{R}$.

Deductions reveal geometrical aspects of inner product:

(i) Set

$$\underline{x} \cdot \underline{y} = |\underline{x}||\underline{y}| \cos\theta$$

to define angle θ between $\underline{x}$ and $\underline{y}$

(ii) △ inequality holds

$$|\underline{x} + \underline{y}| \leq |\underline{x}| + |\underline{y}|$$

Now we present a proof of the Cauchy-Schwarz inequality

Proof. If $\underline{y} = \underline{o}$, result is immediate.
If $\underline{y} \neq \underline{o}$, consider

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

This is a quadratic in real λ with at most one real root, so discriminant satisfies

$$(-2\underline{x} \cdot \underline{y})^2 - 4|\underline{x}|^2|\underline{y}|^2 \leq 0$$

Equality holds if and only if $disc = 0$ which holds if and only if $\lambda\underline{y} = \underline{x}$ for some $\lambda \in \mathbb{R}$.
□ Now we present a proof of the △ inequality.

Proof.

$$LHS^2 = |\underline{x} + \underline{y}|^2 = |\underline{x}|^2 + 2\underline{x} \cdot \underline{y} + |\underline{y}|^2$$

$$RHS^2 = (|\underline{x}| + |\underline{y}|)^2 = |\underline{x}|^2 + 2|\underline{x}||\underline{y}| + |\underline{y}|^2$$

and compare using Cauchy-Schwarz. □

(c) Comments

Inner product on $\mathbb{R}^n$.

$$\underline{a} \cdot \underline{b} = \delta_{ij} a_i b_i$$

Component definition matches geometrical definition for $n = 3$ (section 2.2).
In $\mathbb{R}^3$ also have a cross product with component definition

$$(\underline{a} \times \underline{b})_i = \varepsilon_{ijk} a_j b_k$$

(geometrical definition given in section 2.4)
In $\mathbb{R}^n$ we have $\varepsilon_{ij\ldots l}$ totally antisymmetric. (see chapter 5). Cannot use this to define vector-valued product except in $n = 3$. But in $\mathbb{R}^2$ have ε_{ij} with

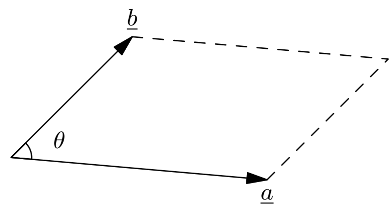
$$\varepsilon_{12} = -\varepsilon_{21} = 1$$

and can use this to define an additional scalar cross product in 2D.

$$\begin{aligned}[\underline{a}, \underline{b}] &= \varepsilon_{ij} a_i b_j \\ &= a_1 b_2 - a_2 b_1 \qquad \text{for } \underline{a}, \underline{b} \in \mathbb{R}^2\end{aligned}$$

Geometrically, this gives (signed) area of parallelogram

$$[\underline{a}, \underline{b}] = |\underline{a}||\underline{b}| \sin\theta$$



Compare with

$$[\underline{a}, \underline{b}, \underline{c}] = \underline{a} \times \underline{b} \times \underline{c} = \varepsilon_{ijk} a_i b_j c_k$$

(signed) volume of parallelepiped.

3.2 Vector Spaces

(a) Axioms; span; subspaces

Let V be a set of objects called *vectors* with operations

(i) $\underline{v} + \underline{w} \in V$

(ii) $\lambda \underline{v} \in V$

(the above expressions are defined $\forall \underline{v}, \underline{w} \in V$ and $\forall \lambda \in \mathbb{R}$)
Then V is a *real vector space* if V is an abelian group under $+$ and

$$\begin{aligned}\lambda(\underline{v}+\underline{w}) &= \lambda\underline{v}+\lambda\underline{w}\\ (\lambda+\mu)\underline{v} &= \lambda\underline{v}+\mu\underline{v}\\ \lambda(\mu\underline{v}) &= (\lambda\mu)\underline{v}\\ 1\underline{v} &= \underline{v}\end{aligned}$$

These axioms or key properties apply to geometrical vectors with V 3D space or to vectors in $V = \mathbb{R}^n$, as above, as well as other examples.
For vectors $\underline{v_1}, \underline{v_2}, \dots, \underline{v_r} \in V$ we can form a *linear combination*

$$\lambda_1\underline{v_1} + \lambda_2\underline{v_2} + \cdots + \lambda_r\underline{v_r} \in V$$

for any $\lambda_i \in \mathbb{R}$; the span is defined

$$\text{span}\{\underline{v_1}, \underline{v_2}, \dots, \underline{v_r}\} = \{\sum_i \lambda_i\underline{v_i} : \lambda_i \in \mathbb{R}\}$$

A *subspace* of V is a subset that is itself a vector space.
Note V and $\{\underline{o}\}$ are subspaces.

$$\text{span}\{\underline{v_1}, \underline{v_2}, \dots, \underline{v_r}\}$$

is a subspace for any vectors $\underline{v_1}, \dots, \underline{v_r}$. Note: a non-empty subset $U \subseteq V$ is a subspace if and only if

$$\underline{v}, \underline{w} \in U \implies \lambda\underline{v} + \mu\underline{w} \in U \;\; \forall \lambda, \mu \in \mathbb{R}$$

Example. In 3D or $\mathbb{R}^3$ a line or plane through $\underline{o}$ is a subspace but a line or plane that doesn't contain $\underline{o}$ is not a subspace. For example

$$\underline{v_1} = \begin{pmatrix}1\\0\\-1\end{pmatrix}, \underline{v_2} = \begin{pmatrix}1\\1\\-2\end{pmatrix}, \underline{n} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

$$\text{span}\{\underline{v_1}, \underline{v_2}\} = \{\underline{r} : \underline{n}\cdot\underline{r} = 0\}$$

which is a plane and subspace. But

$$\{ulr : \underline{n}\cdot\underline{r} = 1\}$$

is a plane but not a subspace ($\underline{r}$, $\underline{r}'$ on plane then $(\underline{r}+\underline{r}')\cdot\underline{n} = 2$)

(b) Linear Dependence and Independence

For vectors $\underline{v_1}, \underline{v_2}, \ldots, \underline{v_r} \in V$, with V a real vector space, consider the *linear relation*

$$\lambda_1 \underline{v_1} + \lambda_2 \underline{v_2} + \cdots + \lambda_r \underline{v_r} = \underline{o} \qquad (*)$$

If $(*) \implies \lambda_i = 0$ for every i then the vectors form a *linearly independent* set (they obey only the *trivial* linear relation with $\lambda_i = 0$).
If $(*)$ holds with at least one $\lambda_i \neq 0$ then the vectors form a *linearly dependent* set (they obey a *non-trivial* linear relation.)

Examples

- $$\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\2\end{pmatrix}$$

  is linearly dependent because

  $$0\begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix} + (-1)\begin{pmatrix}0\\2\end{pmatrix} = 0$$

  Note that we cannot express $\begin{pmatrix}1\\0\end{pmatrix}$ in terms of the others, but it is still linearly dependent.

- Any set containing $\underline{o}$ is linearly dependent. For example

  $$\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}\}$$

  we have

  $$0\begin{pmatrix}1\\0\end{pmatrix} + 412\begin{pmatrix}0\\0\end{pmatrix} = \underline{o}$$

  non-trivial linear relation.

- $\{\underline{a}, \underline{b}, \underline{c}\}$ in $\mathbb{R}^3$ linearly independent if $\underline{a} \cdot \underline{b} \times \underline{c} \neq 0$. Consider

  $$\alpha \underline{a} + \beta \underline{b} + \gamma \underline{c} = \underline{o}$$

  Take dot with $\underline{b} \times \underline{c}$ to get

  $$\alpha \underline{a} \cdot \underline{b} \times \underline{c} = 0 \implies \alpha = 0$$

  and we can get $\beta = \gamma = 0$ with a similar argument.

(c) Inner Product

This is an additional structure on a real vector space V, also characterised by axioms. For $\underline{v}, \underline{w} \in V$ write inner product $\underline{v} \cdot \underline{w}$ or $(\underline{v}, \underline{w}) \in \mathbb{R}$. This satisfies axioms corresponding to the properties in section 3.1(a)

(i) symmetric

(ii) bilinear

(iii) positive definite

Lemma. In a real vector space V with inner product, if $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are non-zero and orthogonal:

$$\underbrace{(\mathbf{v}_i, \mathbf{v}_i) \neq 0}_{\text{fixed } i} \qquad \text{and} \qquad \underbrace{(\mathbf{v}_i, \mathbf{v}_j) = 0}_{i \neq j}$$

then $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent.

Proof.

$$\begin{aligned} \sum_i \alpha_i \mathbf{v}_i &= \mathbf{0} \\ (\mathbf{v}_j, \sum_i \alpha_i \mathbf{v}_i) &= \sum_i \alpha_i (\mathbf{v}_j, \mathbf{v}_i) \\ &= \alpha_j (\mathbf{v}_j, \mathbf{v}_j) \\ &= 0 \\ \implies \alpha_j &= 0 \end{aligned}$$

as claimed. □

3.3 Bases and Dimension

For a vector space V, a basis is a set

$$\mathfrak{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$$

such that

(i) $\mathfrak{B}$ spans V, i.e. any $\mathbf{v} \in V$ can be written

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$$

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

Given (ii), the coefficients v_i in (i) are unique since

$$\sum_i v_i \mathbf{e}_i = \sum_i v_i' \mathbf{e}_i$$
$$\implies \sum_i (v_i - v_i')\mathbf{e}_i = \mathbf{0}$$
$$\implies v_i = v_i'$$

v_i are *components* of $\mathbf{v}$ with respect to $\mathfrak{B}$.

Examples

Standard basis for $\mathbb{R}^n$ consisting of

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

is a basis according to general definition.

(i)
$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$$

(ii) $\mathbf{x} = \mathbf{0}$ if and only if $x_1 = x_2 = \cdots = x_n = 0$.

Many other bases can be chosen, for example in $\mathbb{R}^2$ we have bases

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\},$$

or $\{\mathbf{a}, \mathbf{b}\}$ for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ with $\mathbf{a} \not\parallel \mathbf{b}$.

In $\mathbb{R}^3$, $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a basis if and only if

$$\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} \neq 0.$$

Consider previous example of plane through $\mathbf{0}$, subspace in $\mathbb{R}^3$

$$\mathbf{n} \cdot \mathbf{r} = 0 \qquad \text{with} \qquad \mathbf{n} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix};$$

we have $\{\mathbf{v}_1, \mathbf{v}_2\}$ basis with

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

not normalised or $\perp$ but could choose orthonormal basis

$$\{\mathbf{u}_1, \mathbf{u}_2\} \qquad \text{with} \qquad \mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \text{and} \qquad \mathbf{u}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}.$$

Theorem. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}$ are bases for a real vector space V, then

$$n = m.$$

Definition. The number of vectors in any basis is the *dimension* of V, dim V.

Note. $\mathbb{R}^n$ has dimension n (!)

Proof.

$$\mathbf{f}_a = \sum_i A_{ai}\mathbf{e}_i$$

and

$$\mathbf{e}_i = \sum_a B_{ia}\mathbf{f}_a$$

for constants A_{ai} and B_{ia} and we use ranges of indices $i, j = 1, \ldots, n$ and $a, b = 1, \ldots, m$ [since $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_a\}$ are bases]

$$\begin{aligned} \implies \mathbf{f}_a &= \sum_i A_{ai}\left(\sum_b B_{ib}\mathbf{f}_b\right) \\ &= \sum_b \left(\sum_i A_{ai}B_{ib}\mathbf{f}_b\right) \end{aligned}$$

But coefficients with respect to a basis are unique so

$$\sum_i A_{ai}B_{ib} = \delta_{ab}$$

Similarly

$$\mathbf{e}_i = \sum_j \left(\sum_a B_{ia}A_{aj}\right)\mathbf{e}_j$$

and hence

$$\sum_a B_{ia}A_{aj} = \delta_{ij}$$

Now

$$\begin{aligned} &\sum_{ia} A_{ai}B_{ia} = \sum_a \delta_{aa} = m \\ =&\sum_{ia} B_{ia}A_{ai} = \sum_i \delta_{ii} = n \\ \implies& m = n, \text{ as required.} \end{aligned}$$

□

The steps in the proof above are within the scope of the course; but the proof without prompts is *non-examinable.*

Note. By convention the vector space $\{\mathbf{0}\}$ has dimension $\mathbf{0}$. Not every vector space is finite dimensional!

Proposition. Let V be a vector space of dimension n (for example $\mathbb{R}^n$).

(i) If $Y = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ spans V, then $m \geq n$ and in the case where $m > n$, we can remove vectors can be removed from Y to get a basis.

(ii) If $X = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ are linearly independent then $k \leq n$ and in the case $k < n$ we can add vectors to X to get a basis.

3.4 Vectors in $\mathbb{C}^n$

(a) Definitions

Let $\mathbb{C}^n = \{\mathbf{z} = (z_1, \ldots, z_n) : z_j \in \mathbb{Z}\}$ and define:

- addition $\mathbf{z} + \mathbf{w} = (z_1 + w_1, \ldots, z_n + w_n)$
- scalar multiplication $\lambda\mathbf{z} = (\lambda z_1, \ldots, \lambda z_n)$ for any $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$.

Taking *real* scalars $\lambda, \mu \in \mathbb{R}$, $\mathbb{C}^n$ is a real vector space obeying axioms or key properties in section 3.2(a).
Taking *complex* scalars $\lambda, \mu \in \mathbb{C}$, $\mathbb{C}^n$ is a *complex* vector space - same axioms or key properties hold, and definitions of linear combinations, linear (in)dependence, span, bases, dimension all generalise to complex scalars.

The distinction matters, for example

$$\mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n$$

with $z_j = x_j + iy_j$, $x_j, y_j \in \mathbb{R}$ then

$$\mathbf{z} = \sum_j x_j \mathbf{e}_j + \sum_j y_j \mathbf{f}_j$$

(real linear combination) where

$$\mathbf{e}_j = \underbrace{(0, \ldots, 1, \ldots, 0)}_{\text{position } j}$$

$$\mathbf{f}_j = \underbrace{(0, \ldots, i, \ldots, 0)}_{\text{position } j}$$

therefore $\{\mathbf{e}_1, \ldots, \mathbf{e}_n, \mathbf{f}_1, \ldots, \mathbf{f}_n\}$ basis for $\mathbb{C}^n$ as a *real* vector space. So real dimension is $2n$. But

$$\mathbf{z} = \sum_j = z_j \mathbf{e}_j \qquad \text{and} \qquad \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$$

is a basis for $\mathbb{C}^n$ as a complex vector space, dimension n (over $\mathbb{C}$).

(b) Inner Product

Inner product or scalar product on $\mathbb{C}^n$ is defined by

$$(\mathbf{z}, \mathbf{w}) = \sum_j \overline{z_j} w_j = \overline{z_1} w_1 + \cdots + \overline{z_n} w_n$$

Properties

(i) hermitian $(\mathbf{w}, \mathbf{z}) = \overline{(\mathbf{z}, \mathbf{w})}$.

(ii) Linear / anti-linear

$$(\mathbf{z}, \lambda\mathbf{w} + \lambda'\mathbf{w}') = \lambda(\mathbf{z}, \mathbf{w}) + \lambda'(\mathbf{z}, \mathbf{w}')$$

$$(\mu\mathbf{z} + \mu'\mathbf{z}', \mathbf{w}) = \overline{\mu}(\mathbf{z}, \mathbf{w}) + \overline{\mu'}(\mathbf{z}', \mathbf{w})$$

(iii) positive definite
$(\mathbf{z}, \mathbf{z}) = \sum_i |z_i|^2$ is real and ≥ 0, and 0 if and only if $\mathbf{z} = \mathbf{0}$.

Defined *length* or *norm* of $\mathbf{z}$ to be $|\mathbf{z}| \geq 0$ with $|\mathbf{z}|^2 = (\mathbf{z}, \mathbf{z})$.
Define $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ to be *orthogonal* if $(\mathbf{z}, \mathbf{w}) = 0$.
Note: the standard basis $\{\mathbf{e_j}\}$ for $\mathbb{C}^n$ (see part (a)) is orthonormal

$$(\mathbf{e_i}, \mathbf{e_j}) = \delta_{ij}$$

Also, if $\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_k}$ are non-zero and orthogonal in sense above, then they are linearly independent over $\mathbb{C}$ (same argument as in real case).

Example. Complex inner product on $\mathbb{C}$ ($n = 1$) is

$$(z, w) = \overline{z}w$$

Let $z = a_1 + ia_2$ (real and imaginary part) and $w = b_1 + ib_2$. Then

$$\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2 \qquad \text{corresponding vectors.}$$

$$\begin{aligned} \overline{z}w &= (a_1b_1 + a_2b_2) + i(a_1b_2 - a_2b_1) \\ &= \mathbf{a} \cdot \mathbf{b} + i[\mathbf{a}, \mathbf{b}] \end{aligned}$$

recover scalar dot and cross product in $\mathbb{R}^2$.

4 Matrices and Linear Maps

4.1 Introduction

(a) Definitions

A *linear map* or *linear transformation* is a function

$$T : V \to W$$

between vector spaces V (dim n) and W (dim m) such that

$$T(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in V$ and $\lambda, \mu \in \mathbb{R}$ or $\mathbb{C}$ for V, W both real or complex vector spaces. [mostly concerned with $V = \mathbb{R}^n$ or $\mathbb{C}^n$, $W = \mathbb{R}^m$ or $\mathbb{C}^m$]

Note. A linear map is completely determined by its action on a basis $\{\mathbf{e}_n, \ldots, \mathbf{e}_n\}$ for V, since

$$T\left(\sum_i x_i \mathbf{e}_i\right) = \sum_i x_i T(\mathbf{e}_i)$$

—

$$\mathbf{x}' = T(\mathbf{x}) \in W$$

is the *image* of $\mathbf{x} \in V$

$$\mathrm{Im}(T) = \{\mathbf{x}' \in W : \mathbf{x}' = T(\mathbf{x}) \text{ for some } \mathbf{x} \in V\}$$

is the *image* of T.

Lemma. $\mathrm{Ker}(T)$ is a subspace of V and $\mathrm{Im}(T)$ is a subspace of W.

Check. $\mathbf{x}, \mathbf{y} \in \mathrm{Ker}(T) \implies T(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y}) = \mathbf{0}$ and $\mathbf{0} \in \mathrm{Ker}(T)$ so the result follows.
Also $\mathbf{0} \in \mathrm{Im}(T)$ and $\mathbf{x}', \mathbf{y}' \in \mathrm{Im}(T)$ then

$$T(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y}) = \lambda \mathbf{x}' + \mu \mathbf{y}' \in \mathrm{Im}(T)$$

for some $\mathbf{x}, \mathbf{y} \in V$. □

<u>Examples</u>

(i) zero linear map $T : V \to W$ is given by $T(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x} \in V$. Then $\mathrm{Im}\ T = \{\mathbf{0}\}$, $\mathrm{Ker}\ T = V$.

(ii) For $V = W$, the identity linear map $T : V \to V$ is given by

$$T(\mathbf{x}) = \mathbf{x} \qquad \forall \mathbf{x} \in V$$

then $\text{Im } T = V$, $\text{Ker } T = \{\mathbf{0}\}$.

(iii) $V = W = \mathbb{R}^3$, $\mathbf{x}' = T(\mathbf{x})$ given by

$$\mathbf{x}'_1 = 3x_1 + x_2 + 5x_3$$

$$\mathbf{x}'_2 = -x_1 - 2x_3$$

$$\mathbf{x}'_3 = 2x_1 + x_2 + 3x_3$$

$$\text{Ker}(T) = \left\{ \lambda \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \right\} \qquad \text{(dim 1)}$$

$$\text{Im}(T) = \left\{ \lambda \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\} \qquad \text{(dim 2)}$$

(b) Rank and Nullity

dim Im(T) is the *rank* of T $(\leq m)$ and dim Ker(T) is the *nullity* of T $(\leq n)$

Theorem (Rank-nullity)**.** For

$$T : V \to W$$

a linear map (as in (a) above)

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

Examples - refer to part (a) above

(i) $\text{null}(T) + \text{rank}(T) = n + 0 = n$

(ii) $\text{null}(T) + \text{rank}(T) = 0 + n = n$

(iii) $\text{null}(T) + \text{rank}(T) = 1 + 2 = 3$

Note that the following proof is *non-examinable.*
Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_k$ be a basis for Ker(T) so $T(\mathbf{e}_i) = \mathbf{0}$ for $i = 1, \ldots, k$. Extend by $\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n$ to get a basis for V. Claim

$$\mathfrak{B} = \{T(\mathbf{e}_{k+1}, \ldots, T(\mathbf{e}_n)\}$$

is a basis for Im(T). The result then follows since $\text{null}(T) = k$ and $\text{rank}(T) = n - k$, implying $\text{null}(T) + \text{rank}(T) = n$.
To check the claim:

- $\mathfrak{B}$ spans $\mathrm{Im}(T)$ since

$$\begin{aligned} \mathbf{x} &= \sum_{1}^{n} x_i \mathbf{e}_i \\ \implies T(\mathbf{x}) &= \sum_{i=k+1}^{n} x_i T(\mathbf{e}_i) \end{aligned}$$

- $\mathfrak{B}$ is linearly independent since

$$\begin{aligned} \sum_{i=k+1}^{n} \lambda_i T(\mathbf{e}_i) &= \mathbf{0} \\ \implies T\left(\sum_{i=k=1}^{n} \lambda_i \mathbf{e}_i\right) &= \mathbf{0} \\ \implies \sum_{i=k+1}^{n} \lambda_i \mathbf{e}_i &\in \mathrm{Ker}(T) \\ \implies \sum_{i=k+1}^{n} \lambda_i \mathbf{e}_i &= \sum_{i=1}^{k} \mu_i \mathbf{e}_i \end{aligned}$$

  But $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent in V

$$\begin{aligned} \implies \lambda_i = 0 \qquad (i = k+1, \ldots, n) \\ \implies \mu_i = 0 \qquad (i = 1, \ldots, k) \end{aligned}$$

  hence $\mathfrak{B}$ is linearly independent.

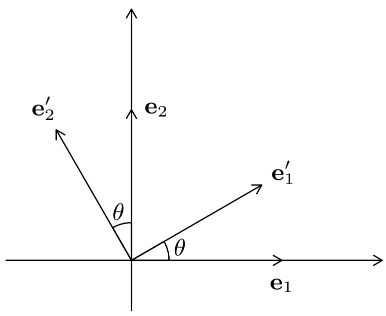
Therefore $\mathfrak{B}$ is a basis. □

4.2 Geometrical Examples

(a) Rotations

In $\mathbb{R}^2$, rotation about $\mathbf{0}$ through angle θ is defined by

$$\begin{aligned} \mathbf{e}_1 \mapsto \mathbf{e}_1' &= (\cos\theta)\mathbf{e}_1 + (\sin\theta)\mathbf{e}_2 \\ \mathbf{e}_2 \mapsto \mathbf{e}_2' &= -(\sin\theta)\mathbf{e}_1 + (\cos\theta)\mathbf{e}_2 \end{aligned}$$



In $\mathbb{R}^3$, rotation about axis given by $\mathbf{e}_3$ is defined as above, with

$$\mathbf{e}_3 \mapsto \mathbf{e}_3' = \mathbf{e}_3$$

Now consider rotation about axis $\mathbf{n}$ (unit vector).
Given $\mathbf{x}$, resolve $\|$ and $\perp$ to $\mathbf{n}$:

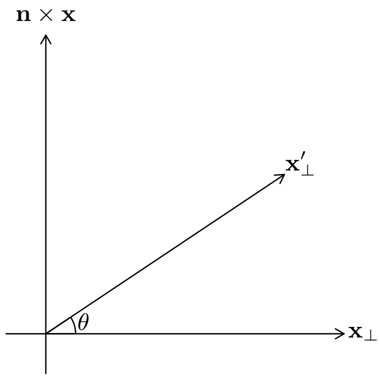
$$\mathbf{x} + \mathbf{x}_\| + \mathbf{x}_\perp$$

with $\mathbf{x}_\| = (\mathbf{x}\cdot\mathbf{n})\mathbf{n}$, and hence $\mathbf{n}\cdot\mathbf{x}_\perp = 0$. Under rotation

$$\mathbf{x}_\| \mapsto \mathbf{x}_\|' = \mathbf{x}_\|$$
$$\mathbf{x}_\perp \mapsto \mathbf{x}_\perp' = (\cos\theta)\mathbf{x}_\perp + (\sin\theta)\mathbf{n}\times\mathbf{x}$$

by considering plane $\perp$ $\mathbf{n}$, comparing to rotation in $\mathbb{R}^2$ and noting that

$$|\mathbf{x}_\perp = |\mathbf{n}\times\mathbf{n}\times\mathbf{x}|$$



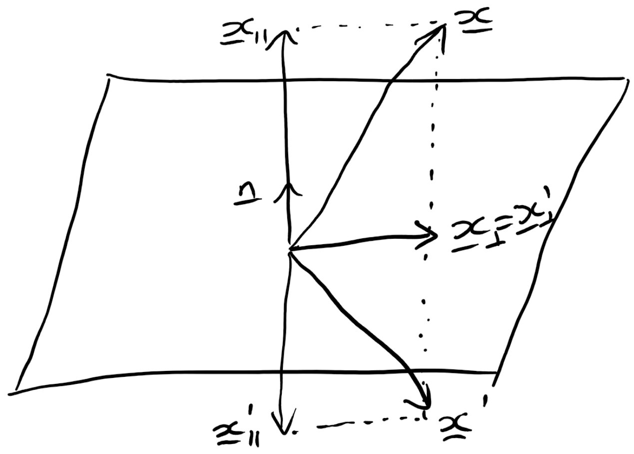
Re-assemble:

$$\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x}_\|' + \mathbf{x}_\perp' = (\cos\theta)\mathbf{X} + (1-\cos\theta)(\mathbf{n}\cdot\mathbf{x})\mathbf{n} + \sin\theta\mathbf{n}\times\mathbf{x}.$$

(b) Reflections

Consider *reflection* in plane in $\mathbb{R}^3$ (or line in $\mathbb{R}^2$ through $\mathbf{0}$ with unit normal $\mathbf{n}$.
Given $\mathbf{x}$, resolve $\|$ and $\perp$ to $\mathbf{n}$:

$$\mathbf{x}_\| \mapsto \mathbf{x}_\|' = -\mathbf{x}_\|$$
$$\mathbf{x}_\perp \mapsto \mathbf{x}_\perp' = \mathbf{x}_\perp$$



$$\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{n})\mathbf{n}$$

(c) Dilations

A dilation by scale factors α, β, γ (real, > 0) along axes $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ in $\mathbb{R}^3$ is defined by

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \mapsto \mathbf{x}' = \alpha x_1\mathbf{e}_1 + \beta x_2\mathbf{e}_2 + \gamma x_3\mathbf{e}_3$$

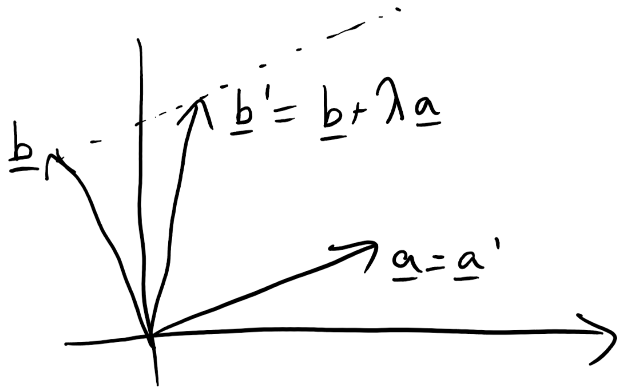
[unit cube $\to$ cuboid]

(d) Shears

Given $\mathbf{a}$, $\mathbf{b}$ orthogonal unit vectors ($|\mathbf{a}| = |\mathbf{b}| = 1$ and $\mathbf{a} \cdot \mathbf{b} = 0$) define a shear with parameter λ by

$$\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} + \lambda(\mathbf{x} \cdot \mathbf{b})\mathbf{a}$$

Definition applies in $\mathbb{R}^n$ and $\mathbf{u}' = \mathbf{u}$ for any vector $\mathbf{u} \perp \mathbf{b}$.



4.3 Matrices as Linear Maps $\mathbb{R}^n \to \mathbb{R}^n$

(a) Definitions

Consider a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ and standard bases $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_a\}$. Let $\mathbf{x}' = T(\mathbf{x})$ with

$$\mathbf{x} = \sum_i x_i \mathbf{e}_i = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{x}' = \sum_a x'_a \mathbf{f}_a = \begin{pmatrix} x'_1 \\ \vdots \\ x'_m \end{pmatrix}$$

Linearity implies T is determined by

$$T(\mathbf{e}_i) = \mathbf{e}'_i = \mathbf{C}_i \in \mathbb{R}^m (i = 1, \ldots, n);$$

take these as *columns* of an $m \times n$ *array* or *matrix* with *rows*

$$\mathbf{R}_a \in \mathbb{R}^n (a = 1, \ldots, m).$$

M has entries $M_{ai} \in \mathbb{R}$ where a labels rows and i labels columns.

$$\begin{pmatrix} \uparrow & & \uparrow \\ \mathbf{C}_1 & \cdots & \mathbf{C}_n \\ \downarrow & & \downarrow \end{pmatrix} = M = \begin{pmatrix} \leftarrow & \mathbf{R}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{R}_m & \rightarrow \end{pmatrix}$$

$$(\mathbf{C}_i)_a = M_{ai} = (\mathbf{R}_a)_i$$

Action of T is given by matrix M multiplying vector $\mathbf{x}$

$$\boxed{\mathbf{x}' = M\mathbf{x} \text{ defined by } \mathbf{x}'_a = M_{ai} x_i \ (\textstyle\sum \text{ convention})}$$

This follows from definitions above since

$$\mathbf{x}' = T\left(\sum_i x_i \mathbf{e}_i\right) = \sum_i x_i \mathbf{C}_i$$

$$\begin{aligned} \implies (\mathbf{x}')_a &= \sum_i x_i (\mathbf{C}_i)_a \\ &= \sum_i M_{ai} x_i \\ &= \sum_i (\mathbf{R}_a)_i x_i \\ &= \mathbf{R}_a \cdot \mathbf{x} \end{aligned}$$

Now regard properties of T as properties of M.

$$\mathrm{Im}(T) = \mathrm{Im}(M) = \mathrm{span}\{\mathbf{C}_1, \ldots, \mathbf{C}_n\}$$

image of M (or T) is span of column S.

$$\mathrm{Ker}(T) = \mathrm{Ker}(M) = \{\mathbf{x} : \mathbf{R}_a \cdot \mathbf{x} = 0\ \forall a\}$$

kernel of M is subspace $\perp$ all rows

(b) Examples

(Refer to sections 4.1 and 4.2)

(i) Zero map $\mathbb{R}^n \to \mathbb{R}^m$ corresponds to *zero matrix* $M = 0$ with $M_{ai} = 0$.

(ii) Identity map $\mathbb{R}^n \to \mathbb{R}^n$ corresponds to *identity matrix*

$$M = I = \begin{pmatrix} 1 & & & \\ & 1 & & 0 \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$$

with $I_{ij} = \delta_{ij}$.

(iii) $\mathbb{R}^3 \to \mathbb{R}^3$, $\mathbf{x}' = T(\mathbf{x}) = M\mathbf{x}$ with

$$M = \begin{pmatrix} 3 & 1 & 5 \\ -1 & 0 & -2 \\ 2 & 1 & 3 \end{pmatrix}, \mathbf{C}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \mathbf{C}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{C}_3 = \begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix}$$

$$\begin{aligned} \operatorname{Im}(T) &= \operatorname{Im}(M) \\ &= \operatorname{span}\{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\} \\ &= \operatorname{span}\{\mathbf{C}_1, \mathbf{C}_2\} \qquad \text{since } \mathbf{C}_3 = 2\mathbf{C}_1 - \mathbf{C}_2 \end{aligned}$$

$$\begin{aligned} \mathbf{R}_1 &= \begin{pmatrix} 3 & 1 & 5 \end{pmatrix} \\ \mathbf{R}_2 &= \begin{pmatrix} -1 & 0 & -2 \end{pmatrix} \\ \mathbf{R}_3 &= \begin{pmatrix} 2 & 1 & 3 \end{pmatrix} \\ \mathbf{R}_2 \times \mathbf{R}_3 &= \begin{pmatrix} 2 & -1 & -1 \end{pmatrix} = \mathbf{u} \end{aligned}$$

and we can notice that $\mathbf{u}$ is perpendicular to all rows. In fact

$$\operatorname{Ker}(T) = \operatorname{Ker}(M) = \{\lambda \mathbf{u}\}.$$

(iv) Rotation through θ about $\mathbf{0}$ in $\mathbb{R}^2$

$$\begin{aligned} \mathbf{e}_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \mathbf{C}_1 \\ \mathbf{e}_2 &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \mathbf{C}_2 \\ \implies M &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \end{aligned}$$

(v) Dilation $\mathbf{x}' = M\mathbf{x}$ with scale factors α, β, γ along axes in $\mathbb{R}^3$:

$$M = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix}$$

(vi) Reflection in plane $\perp \mathbf{n}$ (unit vector) matrix H:

$$\mathbf{x}' = H\mathbf{x} = \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{n})\mathbf{n}$$

$$\begin{aligned} \mathbf{x}'_i &= x_i - 2x_j n_j n_i \\ &= (\delta_{ij} - 2n_i n_j)x_j \\ H_{ij} &= \delta_{ij} - 2n_i n_j \end{aligned}$$

For example

$$\mathbf{n} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad\qquad n_i n_j = \frac{1}{3} \ \forall i, j$$

$$H = \frac{1}{3} \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix}$$

(vii) Shear $\mathbf{x}' = S\mathbf{x} = \mathbf{x} + \lambda(\mathbf{b} \cdot \mathbf{x})\mathbf{a}$

$$\mathbf{x}'_i = S_{ij} x_j$$

with

$$S_{ij} = \delta_{ij} + \lambda a_i b_j$$

for example in $\mathbb{R}^2$ with $\mathbf{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

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

(viii) Rotation win $\mathbb{R}^3$ with axis $\mathbf{n}$ and angle θ,

$$\mathbf{x}' = R\mathbf{x} \qquad x'_i = R_{ij} x_j$$

where $R_{ij} = \delta_{ij} \cos\theta + (1 - \cos\theta) n_i n_j - (\sin\theta) \varepsilon ijk n_k$ (see Example Sheet 2).

(c) Isometries, area and determinant in $\mathbb{R}^2$

Consider linear map $\mathbb{R}^2 \to \mathbb{R}^2$ given by a 2×2 matrix M:

$$\mathbf{x} \mapsto \mathbf{x}' = M\mathbf{x}$$

(i) When is M an *isometry* preserving lengths $|\mathbf{x}'| = |\mathbf{x}|$. This is equivalent to preserving inner products

$$\mathbf{x}' \cdot \mathbf{y}' = \mathbf{x} \cdot \mathbf{y}$$

(since $\mathbf{x} \cdot \mathbf{y} = \frac{1}{2}(|\mathbf{x}+\mathbf{y}|^2 - |\mathbf{x}|^2 - |\mathbf{y}|^2)$). Necessary conditions are

$$M\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \qquad \text{for some } \theta\text{; most general unit vector in } \mathbb{R}^2$$

$$M\begin{pmatrix}0\\1\end{pmatrix} = \pm\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix} \qquad \text{general unit vector perpendicular to other}$$

Simple to check that these conditions are also sufficient and have two cases:

$$M = R = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} \qquad \textit{rotation}$$

or

$$M = H = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix} \qquad \textit{reflection}$$

Compare with expression for reflection in Section 4.3(b)(vi)

$$H_{ij} = \delta_{ij} - 2n_i n_j$$

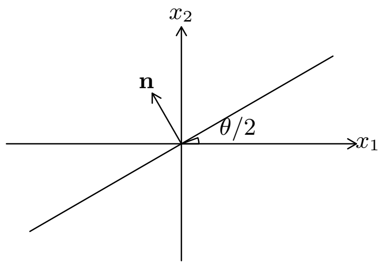
and note for

$$\mathbf{n} = \begin{pmatrix}n_1\\n_2\end{pmatrix} = \begin{pmatrix}-\sin\theta/2\\ \cos\theta/2\end{pmatrix}$$

we get

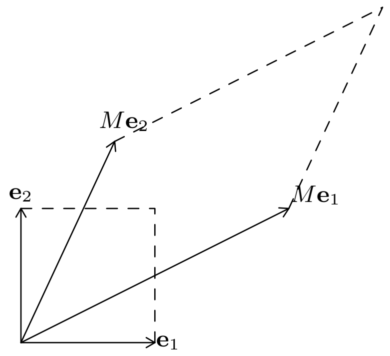
$$H = \begin{pmatrix}1-2\sin^2\theta/2 & 2\sin\theta/2\cos\theta/2\\ 2\sin\theta/2\cos\theta/2 & 1-2\cos^2\theta/2\end{pmatrix}$$

agreeing with H above. This is reflection in a line in $\mathbb{R}^2$ as shown



(ii) How does M change *areas* in $\mathbb{R}^2$ (in general)? Consider unit square in $\mathbb{R}^2$, mapped to parallelogram as shown, with area

$$[M\mathbf{e}_1, M\mathbf{e}_2]$$



"scalar cross product"

$$\left[\begin{pmatrix} M_{11} \\ M_{21} \end{pmatrix}, \begin{pmatrix} M_{12} \\ M_{22} \end{pmatrix}\right] = M_{11}M_{22} - M_{12}M_{21} = \det M$$

where $\det M$ is the *determinant* of 2×2 matrix

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

This is factor (with sign) by areas are scaled under M.
Now compare with (i):

$$\det R = +1, \qquad \det H = -1$$

In either case $|\det M| = +1$. Consider shear

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

this has $\det S = +1$ but it does not preserve lengths.

4.4 Matrices for Linear Maps in General

Consider a linear map

$$T : V \to W$$

between real or complex vector spaces of dimension n, m, respectively and choose bases $\{\mathbf{e}_i\}$ with $i = 1, \ldots, n$ for V and $\{\mathbf{f}_a\}$ with $a = 1, \ldots, m$ for W. The matrix M for T with respect to these bases is an $m \times n$ array with entries $M_{ai} \in \mathbb{R}$ or $\mathbb{C}$. It is defined by

$$T(\mathbf{e}_i) = \sum_a \mathbf{f}_a M_{ai}$$

note index positions. This is chosen to ensure that $T(\mathbf{x}) = \mathbf{x}'$ where

$$\mathbf{x} = \sum_i x_i \mathbf{e}_i$$

and

$$\mathbf{x}' = \sum_a x'_a \mathbf{f}_a$$

if and only if

$$x'_a = \sum_i M_{ai} x_i$$

i.e.

$$\begin{pmatrix} x'_1 \\ \vdots \\ x'_m \end{pmatrix} = \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{m1} & \cdots & M_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Moral. Given choice of bases $\{\mathbf{e}_i\}$ and $\{\mathbf{f}\}$

- V is identified with $\mathbb{R}^n$ (or $\mathbb{C}^n$)
- W is identified with $\mathbb{R}^m$ (or $\mathbb{C}^m$)
- T is identified with $m \times n$ matrix M

Note. There are natural ways to combine linear maps.
If $S : V \to W$ is also linear, then so is

$$\alpha T + \beta S : V \to W$$

defined by

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

Or if $S : U \to V$ is also linear, then so is

$$T \circ S : U \to W$$

composition of maps.

4.5 Matrix Algebra

(a) Linear Combinations

If M and N are $m \times n$ matrices, then $\alpha M + \beta N$ is an $m \times n$ matrix defined by

$$(\alpha M + \beta N)_{ai} = \alpha M_{ai} + \beta N_{ai}$$

($a = 1, \ldots, m$; $i = 1, \ldots, n$) [If M, N represent linear maps $T, S : V \to W$, then $\alpha M + \beta N$ represents $\alpha T + \beta S$, all with respect to same choice of bases.]

(b) Matrix Multiplication

If A is an $m \times n$ matrix, entries A_{ai} ($\in \mathbb{R}$ or $\mathbb{C}$) and B is an $n \times p$ matrix, entries B_{ir}, then AB is an $m \times p$ matrix defined by

$$(AB)_{ar} = A_{ai} B_{ir}$$

The product AB is not defined unless

$$\#\text{ cols of } A = \#\text{ rows of } B$$

$$a = 1, \ldots, m$$

$$i = 1, \ldots, n$$

$$r = 1, \ldots, p.$$

Matrix multiplication corresponds to composition of linear maps

$$[(AB)\mathbf{x}]_a = (AB_{ar})x_r$$

and compare

$$\begin{aligned} [A(B\mathbf{x})]_a &= A_{ai}(B\mathbf{x})_i \\ &= A_{ai}(B_{ir}x_r) \\ &= (A_{ai}B_{ir})x_r \end{aligned}$$

Example.

$$A = \begin{pmatrix} 1 & 3 \\ -5 & 0 \\ 2 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \end{pmatrix}$$

$$AB = \begin{pmatrix} 7 & -3 & 8 \\ -5 & 0 & 5 \\ 4 & -1 & 1 \end{pmatrix}$$

$$BA = \begin{pmatrix} -1 & 2 \\ 13 & 9 \end{pmatrix}$$

Helpful points of view

(i) Regarding $\mathbf{x} \in \mathbb{R}^n$ as a column vector or $n \times 1$ matrix, definition of matrix multiplying a matrix as vector agree.

(ii) For product AB (A is $m \times n$, B is $n \times p$) have columns

$$\mathbf{C}_r(B) \in \mathbb{R}^n$$

$$\mathbf{C}_r(AB) \in \mathbb{R}^m$$

related by

$$\mathbf{C}_r(AB) = A\mathbf{C}_r(B)$$

(iii)

$$AB = \begin{pmatrix} & \vdots & \\ \leftarrow & \mathbf{R}_a(A) & \rightarrow \\ & \vdots & \end{pmatrix} \begin{pmatrix} & \uparrow & \\ \cdots & \mathbf{C}_r(B) & \cdots \\ & \downarrow & \end{pmatrix}$$

$$\begin{aligned} A(B)_{ar} &= [\mathbf{R}_a(A)]_i [\mathbf{C}_r(B)]_i \\ &= \mathbf{R}_a(A) \cdot \mathbf{C}_r(B) \end{aligned}$$

dot product in $\mathbb{R}^n$ for real matrices.

Properties of matrix products

$$\begin{aligned} (\lambda M + \mu N)P &= \lambda(MP) + \mu(NP) \\ P(\lambda M + \mu N) &= \lambda(PM) + \mu(PN) \\ (MN)P &= M(NP) \end{aligned}$$

(c) Matrix Inverses

Consider a $m \times n$ matrix and B, C $n \times m$, B is a *left* inverse for A if

$$BA = I \qquad (n \times n);$$

C is a *right* inverse for A if

$$AC = I \qquad (m \times m).$$

If $m = n$, and A is *square*, one of these implies the other and $B = C = A^{-1}$ *the* inverse.

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

Not every matrix has an inverse; if it does it is called *invertible* or *non-singular*.
Consider map $\mathbb{R}^N \to \mathbb{R}^n$ given by real matrix M. If $\mathbf{x}' = M\mathbf{x}$ and M^{-1} exists then $\mathbf{x} = M^{-1}\mathbf{x}'$.
For $n = 2$,

$$\begin{aligned} x_1' &= M_{11}x_1 + M_{12}x_2 \\ x_2' &= M_{21}x_1 + M_{22}x_2 \\ \implies M_{22}x_1' - M_{12}x_2' &= (\det M)x_1 \\ \text{and } -M_{21}x_1' + M_{11}x_2' &= (\det M)x_2 \end{aligned}$$

So, if $\det M = M_{11}M_{22} - M_{12}M_{21} \neq 0$ then

$$M^{-1} = \frac{1}{\det M}\begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}$$

Examples

$$\begin{aligned} R(\theta) &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ R(\theta)^{-1} &= R(-\theta) \\ H(\theta) &= \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \\ H(\theta)^{-1} &= H(\theta) \\ S(\lambda) &= \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \\ S(\lambda)^{-1} &= S(-\lambda) \end{aligned}$$

(d) Transpose and Hermitian Conjugate

(i) If M is an $m \times n$ matrix, then transpose $M^\top$ is an $n \times m$ matrix defined by

$$(M^\top)_{ia} = M_{ai}$$

"exchange rows and columns"

$$a = 1, \ldots, m; i = 1, \ldots, n$$

Properties

$$\begin{aligned} (\alpha + \beta B)^\top &= \alpha A^\top + \beta B^\top (A, B \;\; m \times n) \\ (AB)^\top &= B^\top A^\top \end{aligned}$$

Check:

$$\begin{aligned} [(AB)^\top]_{ra} &= (AB)_{ar} \\ &= A_{ai}B_{ir} \\ &= (A^\top)_{ia}(B^\top)_{ri} \\ &= (B^\top)_{ri}(A^\top)_{ia} \\ &= (B^\top A^\top)_{ra} \qquad \text{as required.} \end{aligned}$$

Note.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{column vector, } n \times 1 \text{ matrix}$$

$$\implies \mathbf{x}^\top = (x_1, \dots, x_n) \qquad \text{row vector, } 1 \times n \text{ matrix}$$

Inner product on $\mathbb{R}^n$ is

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y} \qquad \text{scalar } 1 \times 1 \text{ matrix}$$

but $\mathbf{y}\mathbf{x}^\top = M$, $n \times n$ matrix with $M_{ij} = y_i x_j$.

(ii) If M is square, $n \times n$, then M is *symmetric* if and only if $M^\top = M$ or $M_{ij} = M_{ji}$ and *antisymmetric* if and only if $M^\top = -M$ or $M_{ij} = -M_{ji}$. Any square can be written as a sum of symmetric and antisymmetric parts:

$$M = S + A$$

where $S = \frac{1}{2}(M + M^\top)$ and $A = \frac{1}{2}(M - M^\top)$.

Example. If A is 3×3 antisymmetric, then it can be re-written in terms of vector $\mathbf{a}$

$$A = \begin{pmatrix} 0 & a_2 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{pmatrix}$$

$$A_{ij} = \varepsilon_{ijk} a_k \qquad \text{and} \qquad a_k = \frac{1}{2} \varepsilon_{kij} A_{ij}$$

Then

$$\begin{aligned} (A\mathbf{x})_i &= A_{ij} x_j \\ &= \varepsilon_{ijk} a_k x_j \\ &= (\mathbf{x} \times \mathbf{a})_i \end{aligned}$$

(iii) If M is $m \times n$ matrix the *hermitian conjugate* $M^\dagger$ is defined by

$$(M^\dagger)_{ia} = \overline{M}_{ai}$$

or

$$M^\dagger = \overline{M}^\top = \overline{(M^\top)}$$

<u>Properties</u>

$$\begin{aligned} (\alpha A + \beta B)^\dagger &= \overline{\alpha} A^\dagger + \overline{\beta} B^\dagger \\ (AB)^\dagger) &= B^\dagger A^\dagger \end{aligned}$$

Note.

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \qquad \text{column vector, } n \times 1 \text{ matrix}$$

$$\implies \mathbf{z}^\dagger = (\overline{z_1}, \ldots, \overline{z_n}) \qquad \text{row vector, } 1 \times n \text{ matrix}$$

Inner product on $\mathbb{C}^n$ is

$$(\mathbf{z}, \mathbf{w}) = \mathbf{z}^\dagger \mathbf{w} \qquad \text{scalar } 1 \times 1 \text{ matrix}$$

(iv) If M is *square* $n \times n$ then M is *hermitian* if $M^\dagger = M$ or $M_{ij} = \overline{M}_{ji}$ and *anti-hermitian* if $M^\dagger = -M$ or $M_{ij} = -\overline{M}_{ji}$.

(e) Trace

For any square $n \times n$ matrix M, the *trace* is defined by

$$\mathrm{Tr}(M) = M_{ii} = M_{11} + \cdots + Mnn \qquad \text{(sum of diagonal entries)}$$

Properties

$$\mathrm{tr}(\alpha M + \beta N) = \alpha \mathrm{tr}(M) + \beta \mathrm{tr}(N)$$

$$\mathrm{tr}(MN) = \mathrm{tr}(NM)$$

check:

$$\begin{aligned} (MN)_i i &= M_{ia} N_{ai} \\ &= N_{ai} M_{ia} \\ &= (NM)_{aa} \end{aligned}$$

$$\mathrm{tr}(M) = \mathrm{tr}(M^\top)$$

$$\mathrm{tr}(I) = n \qquad \text{for } I \;\; n \times n.$$

$$I_{ij} = \delta_{ij} \qquad \text{and} \qquad I_{ii} = \delta_{ii} = n$$

Previously decomposed

$$M = S + A \qquad \text{symmetric / antisymmetric parts}$$

Let $T = S - \frac{1}{n}(\mathrm{tr}(S))I$ or $T_{ij} = S_{ij} - \frac{1}{n}\mathrm{tr}(S)S_{ij}$, then $T_{ii} = \mathrm{tr}(T) = 0$; and note $\mathrm{tr}(M) = \mathrm{tr}(S)$ and $\mathrm{tr}(A) = 0$. So

$$M = \underbrace{T}_{\text{symm and traceless}} + \underbrace{A}_{\text{antisymm part}} + \frac{1}{n} \underbrace{\mathrm{tr}(M) I}_{\text{pure trace}}$$

Example.

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 3 & 2 \\ 3 & 5 & 4 \\ 2 & 4 & 3 \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ -1 & -2 & 0 \end{pmatrix}$$

$$\mathrm{tr}(S) = \mathrm{tr}(M) = 9$$

$$T = \begin{pmatrix} -2 & 3 & 2 \\ 3 & 2 & 4 \\ 2 & 4 & 0 \end{pmatrix}$$

$$M = T + A + 3I$$

Furthermore $A\mathbf{x} = \mathbf{x} \times \mathbf{a}$ where $\mathbf{a} = (2, -1, -1)$.

Orthogonal and Unitary Matrices

A real $n \times n$ matrix U is orthogonal if and only if

$$U^\top U = UU^\top = I$$

i.e.

$$U^\top = U^{-1}$$

These conditions can be written

$$U_{ki}U_{kj} = U_{ik}U_{jk} = \delta_{ij}$$

(the left implies the columns are orthonormal, and the middle implies that the rows are orthonormal). [recall $[\mathbf{C}_i(U)]_k = U_{ki} = [R_k(U)]_i$]

$$\underbrace{\begin{pmatrix} & \vdots & \\ \leftarrow & \mathbf{C}_i & \rightarrow \\ & \vdots & \end{pmatrix}}_{U^\top} \underbrace{\begin{pmatrix} & \uparrow & \\ \cdots & \mathbf{C}_j & \cdots \\ & \downarrow & \end{pmatrix}}_{U} = I$$

$$\mathbf{C}_i \cdot \mathbf{C}_j = \delta_{ij}$$

Equivalent definition U is orthogonal if and only if it preserves the inner product on $\mathbb{R}^n$

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

To check equivalence, write this as

$$(U\mathbf{x})^\top (U\mathbf{y}) = \mathbf{x}^\top \mathbf{y}$$

$$\begin{aligned} LHS &= (\mathbf{x}^\top U^\top)(U\mathbf{y}) \\ &= \mathbf{x}^\top (U^\top U)\mathbf{y} \\ &= RHS \ \forall \mathbf{x}, \mathbf{y} \end{aligned}$$

if and only if $U^\top U = I$. Note, since $\mathbf{C}_i = U\mathbf{e}_i$, columns are orthonormal is equivalent to

$$(U\mathbf{e}_i) \cdot (\mathbf{U}\mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

Examples
In $\mathbb{R}^2$ we found all orthogonal matrices (section 4.3(c)):

$$\text{rotations } R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$\text{and reflections } H(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

Clearly

$$R(\theta)^\top = R(-\theta) = R(\theta)^{-1}$$
$$H(\theta)^\top = H(\theta) = H(\theta)^{-1}$$

In $\mathbb{R}^3$ found matrix $R(\theta)$ for rotation through θ about axis $\mathbf{n}$

$$R(\theta)^\top = R(-\theta)$$

since

$$R(\theta)_{ij} = R(-\theta)_{ji}$$

and can check explicitly

$$R(\theta)^\top R(\theta) = R(-\theta)R(\theta) = I$$

or

$$R(\theta)_{ki}R(\theta)_{kj} = \delta_{ij}$$

A complex $n \times n$ matrix U is *unitary* if and only if

$$U^\dagger U = UU^\dagger = I$$

i.e.

$$U^\dagger = U^{-1}$$

Equivalent definition: U is unitary if and only if it preserves the inner product on $\mathbb{C}^n$

$$(U\mathbf{z}, U\mathbf{w}) = (\mathbf{z}, \mathbf{w}) \ \forall \mathbf{z}, \mathbf{w} \in \mathbb{C}^n$$

To check equivalence write this as

$$(U\mathbf{z})^\dagger (U\mathbf{w}) = \mathbf{z}^\dagger \mathbf{w}$$

$$\begin{aligned} LHS &= (\mathbf{z}^\dagger U^\dagger)(U\mathbf{w}) \\ &= \mathbf{z}^\dagger (U^\dagger U)\mathbf{w} \\ &= RHS \ \forall \mathbf{z}, \mathbf{w} \end{aligned}$$

if and only if $U^\dagger U = I$.

5 Determinants and Inverses

5.1 Introduction

Consider a linear map

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

If T is invertible then

$$\underbrace{\text{Ker } T = \{\mathbf{0}\}}_{\substack{\text{because } T \\ \text{one-to-one}}} \qquad \text{and} \qquad \underbrace{\text{Im } T = \mathbb{R}^n}_{T \text{ is onto}}$$

These conditions are equivalent by rank-nullity. Conversely, if these conditions hold, then

$$\mathbf{e}'_1 = T(\mathbf{e}_1), \ldots, \mathbf{e}'_n = T(\mathbf{e}_n)$$

is a basis (where $\{\mathbf{e}_i\}$ standard basis) and we can define a linear map T^{-1} by

$$T^{-1}(\mathbf{e}'_1) = \mathbf{e}_1, \ldots, T^{-1}(\mathbf{e}'_n) = \mathbf{e}_n$$

How can we test whether the conditions hold from matrix M representing T:

$$T(\mathbf{x}) = M\mathbf{x}$$

and how can we find M^{-1} when they do hold?
For any M $(n \times n)$ we will define a related matrix $\widetilde{M}$ $(n \times n)$ and a scalar, the *determinant* $\det(M)$ or $|M|$ such that

$$\widetilde{M}M = (\det M)I \qquad (*)$$

Then if $\det M \neq 0$, M is invertible with

$$M^{-1} = \frac{1}{\det M}\widetilde{M}$$

<u>For $n = 2$</u> we found in section 4.4(c) that $(*)$ holds with

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \qquad \text{and} \qquad \widetilde{M} = \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}$$

and

$$\begin{aligned} \det M &= \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} \\ &= M_{11}M_{22} - M_{12}M_{21} \\ &= [M\mathbf{e}_1, M\mathbf{e}_2] \\ &= [\mathbf{C}_1(M), \mathbf{C}_2(M)] \\ &= \varepsilon_{ij} M_{i1} M_{j2} \end{aligned}$$

Factor by which areas are scaled under M

$$\det M \neq 0 \iff \{M\mathbf{e}_1, M\mathbf{e}_2\} \text{ linearly independent } \iff \operatorname{Im}(M) = \mathbb{R}^2$$

<u>For $n = 3$</u> consider similarly

$$\begin{aligned}[M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3] &= [\mathbf{C}_1(M), \mathbf{C}_2(M), \mathbf{C}_3(M)] && \text{scalar triple product} \\ &= \varepsilon_{ijk} M_{i1} M_{j2} M_{k3} \\ &= \det M, \qquad \text{definition for } n = 3\end{aligned}$$

This is factor by which volumes are scaled under M and

$$\det M \neq 0 \iff \{M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3\} \text{ linearly independent } \iff \operatorname{Im}(M) = \mathbb{R}^3$$

Now define $\widetilde{M}$ from M using rows / column notation:

$$\mathbf{R}_1(\widetilde{M}) = \mathbf{C}_2(M) \times \mathbf{C}_3(M)$$

$$\mathbf{R}_2(\widetilde{M}) = \mathbf{C}_3(M) \times \mathbf{C}_1(M)$$

$$\mathbf{R}_3(\widetilde{M}) = \mathbf{C}_1(M) \times \mathbf{C}_2(M)$$

and note that

$$\begin{aligned}(\widetilde{M}M)_{ij} &= \mathbf{R}_i(\widetilde{M}) \cdot \mathbf{C}_j(M) \\ &= \underbrace{(\mathbf{C}_1(M) \cdot \mathbf{C}_2(M) \times \mathbf{C}_3(M))}_{\det M} \delta_{ij}\end{aligned}$$

as required.

Example.

$$M = \begin{pmatrix} 1 & 3 & 0 \\ 0 & -1 & 2 \\ 4 & 1 & -1 \end{pmatrix}$$

$$\mathbf{C}_2 \times \mathbf{C}_3 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 6 \end{pmatrix}$$

$$\mathbf{C}_3 \times \mathbf{C}_1 = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 8 \\ -1 \\ -2 \end{pmatrix}$$

$$\mathbf{C}_1 \times \mathbf{C}_2 = \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} \times \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 11 \\ -1 \end{pmatrix}$$

$$\widetilde{M} = \begin{pmatrix} -1 & 3 & 6 \\ 8 & -1 & -2 \\ 4 & 1 & -1 \end{pmatrix}$$

and $\widetilde{M}M = (\det M)I$ where

$$\det M = \mathbf{C}_1 \cdot \mathbf{C}_2 \times \mathbf{C}_3 = 23.$$

5.2 ε and Alternating Forms

(a) ε and Permutations

Recall: a permutation σ on the set $\{1, 2, \dots, n\}$ is a bijection from this set to itself, specified by list

$$\sigma(1), \sigma(2), \dots, \sigma(n)$$

Permutation σ form a group, the symmetric group S_n of order $n!$. The *sign* or *signature* $\varepsilon(\sigma) = (-1)^k$ where K is the number of transpositions (this is well-defined). The *alternating* or ε symbol in $\mathbb{R}^n$ or $\mathbb{C}^n$ is defined by

$$\varepsilon_{\underbrace{ij\dots}_{n \text{ indices}}} = \begin{cases} +1 & \text{if } i, j, \dots, l \text{ is an even permutation} \\ -1 & \text{if } i, j, \dots, l \text{ is an odd permutaiton} \\ 0 & \text{else} \end{cases}$$

If σ any permutation of $1, 2, \dots, n$ then

$$\varepsilon_{\sigma(1)\sigma(2)\cdots\sigma(n)} = \varepsilon(\sigma)$$

Lemma 1.

$$\sigma_{\sigma(i)\sigma(j)\cdots\sigma(l)} = \varepsilon(\sigma)\varepsilon_{ij\cdots l}$$

(ε totally antisymmetric is a corollary)

Proof. If $i, j, \ldots, l$ is *not* a permutation of $1, 2, \ldots, n$ then $RHS = LHS = 0$. If $i = \rho(1)$, $j = \rho(2)$, $\ldots$, $l = \rho(n)$ for some permutation ρ then

$$RHS = \varepsilon(\sigma)\varepsilon(\rho) = \varepsilon(\sigma\rho) = LHS$$

as required. □

(b) Alternating Forms and Linear (In)dependence

Given $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ or $\mathbb{C}^n$ the *alternating form* combines them to produce scalar, defined by

$$\begin{aligned}[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] &= \varepsilon_{ij\cdots l}(\mathbf{v}_1)_i(\mathbf{v}_2)_j \cdots (\mathbf{v}_n)_l \\ &= \sum_\sigma \varepsilon(\sigma)(\mathbf{v}_1)_{\sigma(1)}(\mathbf{v}_2)_{\sigma(2)} \vdots (\mathbf{v}_n)_{\sigma(n)}\end{aligned}$$

($\sum_\sigma$ means sum over all $\sigma \in S_n$)

Properties

(i) *Multilinear*

$$\begin{aligned}[\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}, \alpha\mathbf{u} + \beta\mathbf{w}, \mathbf{v}_{p+1}, \ldots, \mathbf{v}_n] =& \alpha[\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}, \mathbf{u}, \mathbf{v}_{p+1}, \ldots, \mathbf{v}_n] \\ &+ \beta[\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}, \mathbf{w}, \mathbf{w}_{p+1}, \ldots, \mathbf{v}_n]\end{aligned}$$

(ii) *Totally antisymmetric*

$$[\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(n)}] = \varepsilon(\sigma)[\mathbf{v}_1, \ldots, \mathbf{v}_n]$$

(iii) $[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = 1$ for $\mathbf{e}_i$ standard basis vectors. Properties (i), (ii), (iii) fix the alternating form, and they also imply

(iv) If $\mathbf{v}_p = \mathbf{v}_q$ for some $p \neq q$ then

$$[\mathbf{v}_1, \ldots, \mathbf{v}_p, \ldots, \mathbf{v}_q, \ldots, \mathbf{v}_n] = 0$$

(from (ii), exchanging $\mathbf{v}_p \leftrightarrow \mathbf{v}_q$ changes sign of alternating form).

(v) If $\mathbf{v}_p = \sum_{i \neq p} \lambda_i \mathbf{v}_i$ then

$$[\mathbf{v}_1, \ldots, \mathbf{v}_p, \ldots, \mathbf{v}_n] = 0$$

(sub in and use (i) and (iv)).

Example. In $\mathbb{C}^4$,

$$\mathbf{v}_1 = \begin{pmatrix} i \\ 0 \\ 0 \\ 2 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \\ 5i \\ 0 \end{pmatrix},$$

$$\mathbf{v}_3 = \begin{pmatrix} 3 \\ 2i \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ -i \\ 1 \end{pmatrix}$$

$$\begin{aligned}
\implies [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4] &= 5i[\mathbf{v}_1, \mathbf{e}_3, \mathbf{v}_3, \mathbf{v}_4] \\
&= 5i[i\mathbf{e}_1 + 2\mathbf{e}_4, \mathbf{e}_3, 3\mathbf{e}_1 + 2i\mathbf{e}_2, \cancel{-i\mathbf{e}_3} + \mathbf{e}_4] \\
&= 5i[i\mathbf{e}_1 + \cancel{2\mathbf{e}_4}, \mathbf{e}_3, 3\mathbf{e}_1 + 2i\mathbf{e}_2, \mathbf{e}_4] \\
&= 5i[i\mathbf{e}_1, \mathbf{e}_3, \cancel{3\mathbf{e}_1} + 2i\mathbf{e}_2, \mathbf{e}_4] \\
&= (5i \cdot i \cdot 2i)[\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_4] \\
&= -10i(-1) \\
&= 10i
\end{aligned}$$

Note. Properties (i) and (iii) immediate from definition.

Proof of property (ii).

$$\begin{aligned}
[\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(n)}] &= \sum_\rho \varepsilon(\rho) \underbrace{[\mathbf{v}_{\sigma(1)}]_{\rho(1)} \cdots [\mathbf{v}_{\sigma(n)}]_{\rho(n)}}_{\text{each term can be rewritten}} \\
&= \sum_\rho \varepsilon(\rho)[\mathbf{v}_1]_{\rho\sigma^{-1}(1)} \cdots [\mathbf{v}_n]_{\rho\sigma^{-1}(n)} \\
&= \sum_\rho \varepsilon(\sigma)\varepsilon(\rho')[\mathbf{v}_1]_{\rho'(1)} \cdots [\mathbf{v}_n]_{\rho'(n)} \\
&= \varepsilon(\sigma) \sum_{\rho'} \sigma(\rho')[\mathbf{v}_1]_{\rho'(1)} \cdots [\mathbf{v}_n]_{\rho'(n)} \\
&= \varepsilon(\sigma)[\mathbf{v}_1, \dots, \mathbf{v}_n]
\end{aligned}$$

as claimed. □

Proposition.

$$[\mathbf{v}_1, \dots, \mathbf{v}_n] \neq 0 \iff \mathbf{v}_1, \dots, \mathbf{v}_n \text{ linearly indendent}$$

Proof. To show "⇒" use property (v). If $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent then $\sum \alpha_i \mathbf{v}_i = \mathbf{0}$ where not all coefficients are zero. Suppose without loss of generality that $\alpha_p \neq 0$, then express $\mathbf{v}_p$ as a linear combination of $\mathbf{v}_i$ $(i \neq p)$ and

$$[\mathbf{v}_1, \dots, \mathbf{v}_n] = 0.$$

To show "⇐" note that $\mathbf{v}_1, \dots, \mathbf{v}_n$ linearly independent means they also span (in $\mathbb{R}^n$ or $\mathbb{C}^n$) so we can write standard basis vectors as

$$\mathbf{e}_i = A_{ai}\mathbf{v}_a$$

for some $A_{ai} \in \mathbb{R}$ or $\mathbb{C}$. But then

$$\begin{aligned}[\mathbf{e}_1, \dots, \mathbf{e}_n] &= [A_{a1}\mathbf{v}_a, A_{b2}\mathbf{v}_b, \dots, A_{cn}\mathbf{v}_c] \\ &= A_{a1}A_{b2}\cdots A_{cn}[\mathbf{v}_a, \mathbf{v}_b, \dots, \mathbf{v}_c] \\ &= A_{a1}A_{b2}\cdots A_{cn}\varepsilon_{ab\cdots c}[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]\end{aligned}$$

and $LHS = 1$, so $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \neq 0$. Example in $\mathbb{C}^4$ above: $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, $\mathbf{v}_4$ linearly independent. □

5.3 Determinants in $\mathbb{R}^n$ and $\mathbb{C}^n$

(a) Definition

For an $n \times n$ matrix M with columns

$$\mathbf{C}_a = M\mathbf{e}_a$$

the *determinant* $\det M$ or $|M| \in \mathbb{R}$ or $\mathbb{C}$ is defined by

$$\begin{aligned}\det M &= [\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_n] \\ &= [M\mathbf{e}_1, M\mathbf{e}_1, \dots, M\mathbf{e}_n] \\ &= \varepsilon_{ij\cdots l}M_{i1}M_{j2}\cdots M_{ln}\end{aligned}$$

$$\sum_\sigma \varepsilon(\sigma)M_{\sigma(1)1}M_{\sigma(2)2}\cdots M_{\sigma(n)n}$$

Proposition (Tanspose Property)**.**

$$\det M = \det M^\top$$

So

$$\begin{aligned}\det(M) &= [\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n] \\ &= \varepsilon_{ij\cdots l}M_{1i}M_{2j}\cdots M_{nl} \\ &= \sum_\sigma \varepsilon(\sigma)M_{1\sigma(1)}M_{2\sigma(2)}\cdots M_{n\sigma(n)}\end{aligned}$$

Example. In $\mathbb{R}^3$ or $\mathbb{C}^3$

$$\begin{aligned}\det M &= \varepsilon_{ijk} M_{i1} M_{j2} M_{k3} \\ &= M_{11} \begin{vmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{vmatrix} - M_{21} \begin{vmatrix} M_{12} & M_{13} \\ M_{32} & M_{33} \end{vmatrix} + M_{31} \begin{vmatrix} M_{12} & M_{13} \\ M_{22} & M_{23} \end{vmatrix}\end{aligned}$$

Properties

$\det M$ is a function of rows or columns of M that is

(i) multilinear

(ii) totally antisymmetric (or alternating)

(iii) $\det I = 1$

Theorem.

$$\begin{aligned}\det M \neq 0 &\iff \text{cols of } M \text{ are linearly independent} \\ &\iff \text{rows of } M \text{ are linearly independent} \\ &\iff \text{rank } M = n \qquad (M \;\; n \times n) \\ &\iff \text{Ker } M = \{\mathbf{0}\} \\ &\iff M^{-1} \text{ exists}\end{aligned}$$

Proof. All equivalences follow immediately from earlier results including discussion in section 5.1. □

Proof of Transpose Property. Suffices to show

$$\sum_{\sigma} \varepsilon(\sigma) M_{\sigma(1)1} \cdots M_{\sigma(n)n} = \sum_{\sigma} \varepsilon(\sigma) M_{1\sigma(1)} \cdots M_{2\sigma(2)}$$

But in a given term on the left hand side,

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

by re-ordering factors, where $\rho = \sigma^{-1}$. Then $\varepsilon(\sigma) = \varepsilon(\rho)$ and $\sum_\sigma$ equivalent to $\sum_\rho$, so result follows. □

(b) Evaluating Determinants: Expanding by Rows or Columns

For M $n \times n$, for each entry M_{ia} define the *minor* M^{ia} to be the determinant of $(n-1) \times (n-1)$ matrix obtained by deleting row i and column a from M.

Proposition.

$$\det M = \sum_i (-1)^{i+a} M_{ia} M^{ia} \qquad a \text{ fixed}$$
$$= \sum_a (-1)^{i+a} M_{ia} M^{ia} \qquad i \text{ fixed}$$

called expanding by (or about) column a or row i respectively.

Proof. See section 5.4. □

Example.

$$M = \begin{pmatrix} i & 0 & 3 & 0 \\ 0 & 0 & 2i & 0 \\ 0 & 5i & 0 & -i \\ 2 & 0 & 0 & 1 \end{pmatrix}$$

Expand by row 3 to find

$$\det M = \sum_a (-1)^{3+a} M_{3a} M^{3a}$$

$$M_{31} = M_{33} = 0;$$

$$M_{32} = 5i, \quad M^{32} = \begin{vmatrix} i & 3 & 0 \\ 0 & 2i & 0 \\ 2 & 0 & 1 \end{vmatrix}$$

$$M_{34} = -i, \quad M^{34} = \begin{vmatrix} i & 0 & 3 \\ 0 & 0 & 2i \\ 2 & 0 & 0 \end{vmatrix}$$

$$M^{32} = i \begin{vmatrix} 2i & 0 \\ 0 & 1 \end{vmatrix} - 3 \begin{vmatrix} 0 & 0 \\ 2 & 1 \end{vmatrix} = i(2i) = -2 \qquad \text{(row 1)}$$

$$M^{34} = i \begin{vmatrix} 0 & 2i \\ 0 & 0 \end{vmatrix} + 3 \begin{vmatrix} 0 & 0 \\ 2 & 0 \end{vmatrix} = 0 \qquad \text{(row 1)}$$

$$\det M = (-1)^{3+2} 5i(-2) = 10i$$

Alternatively we can expand by column 2:

$$\begin{aligned} \det M &= \sum_i (-1)^{2+i} M_{i2} M^{i2} \\ &= (-1)^{2+3} M_{32} M^{32} \\ &= 10i \end{aligned}$$

(Calculated this previously as example of alternating form in $\mathbb{C}^n$)

Lemma. If

$$M = \left(\begin{array}{c|c} A & O \\ \hline O & I \end{array} \right)$$

block form with A an $r \times r$ matrix; I an $(n-r) \times (n-r)$ identity, then $\det M = \det A$.

Proof. For $r = n - 1$, result follows by expanding about column n or row n, and for $r < n - 1$, continue process. □

(c) Simplifying Determinants: Rows and Column Operations

From the definitions of $\det M$ in terms of columns (a) or rows (i) and the properties above (including section 5.2(b)) we note the following

- Row or Column Scalings
  If $\mathbf{R}_i \mapsto \lambda \mathbf{R}_i$ for some (fixed) i or $\mathbf{C}_a \mapsto \lambda \mathbf{C}_i$ for some (fixed) a then $\det M \mapsto \lambda \det M$. If *all* rows or columns are scaled, so $M \mapsto \lambda M$, then $\det M \mapsto \lambda^n \det M$.

- Row or Column Operations
  If $\mathbf{R}_i \mapsto \mathbf{R}_i + \lambda \mathbf{R}_j$ for $i \neq j$ or $\mathbf{C}_a \mapsto \mathbf{C}_a + \lambda \mathbf{C}_b$ for $a \neq b$, then $\det M \mapsto \det M$.

- Row or Column Exchanges
  If $\mathbf{R}_i \leftrightarrow \mathbf{R}_j$ for $i \neq j$ or $\mathbf{C}_a \leftrightarrow \mathbf{C}_b$ for $a \neq b$ then $\det M \mapsto -\det M$.

Example.

$$A = \begin{pmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{pmatrix} \qquad a \in \mathbb{C}$$

Considering $\mathbf{C}_1 \mapsto \mathbf{C}_1 - \mathbf{C}_3$, which keeps the determinant invariant, we get:

$$\begin{aligned} \det A &= \det \begin{pmatrix} 1-a & 1 & a \\ a-1 & 1 & 1 \\ 0 & a & 1 \end{pmatrix} \\ &= (1-a) \det \begin{pmatrix} 1 & 1 & a \\ -1 & 1 & 1 \\ 0 & a & 1 \end{pmatrix} \end{aligned}$$

Now we consider $\mathbf{C}_2 \to \mathbf{C}_2 - \mathbf{C}_3$:

$$\begin{aligned} \det A &= (1-a) \det \begin{pmatrix} 1 & 1-a & a \\ -1 & 0 & 1 \\ 0 & a-1 & 1 \end{pmatrix} \\ &= (1-a)^2 \det \begin{pmatrix} 1 & 1 & a \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \end{aligned}$$

And finally $\mathbf{R}_1 \to \mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3$:

$$\begin{aligned} \det A &= (1-a)^2 \det \begin{pmatrix} 0 & 0 & a+2 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \\ &= (1-a)^2 (a+2) \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} \\ &= (1-a)^2 (a+2) \end{aligned}$$

(d) Multiplicative Property

Theorem. For $n \times n$ matrices M and N,

$$\det(MN) = \det(M) \det(N).$$

This is based on the following lemma.

Lemma.

$$\varepsilon_{i_1 \ldots i_n} M_{i_1 a_1} \cdots M_{i_n a_n} = (\det M) \varepsilon_{a_1 \ldots a_n}$$

Proof of Theorem.

$$\begin{aligned}\det(MN) &= \varepsilon_{i_1\dots i_n}(MN)_{i_11}\cdots(MN)_{i_nn}\\ &= \varepsilon_{i_1\dots i_n}M_{i_1k_1}N_{k_11}\cdots M_{i_nk_n}N_{k_nn}\\ &= \varepsilon_{i_1\dots i_n}M_{i_1k_1}\cdots M_{i_nk_n}N_{k_11}N_{k_nn}\\ &= (\det M)\varepsilon_{k_1\dots k_n}N_{k_11}\cdots N_{k_nn}\\ &= (\det M)(\det N)\end{aligned}$$

as required. □

Proof of Lemma. Use total antisymmetry of left hand side and right hand side and then check by taking $a_1 = 1, \dots, a_n = n$. □

Examples

(i) If

$$M = \left(\begin{array}{c|c} A & O \\ \hline O & B \end{array}\right)$$

(block form) with A an $r \times r$ and B an $(n-r) \times (n-r)$, then

$$\det M = \det A \cdot \det B$$

Since

$$\left(\begin{array}{c|c} A & O \\ \hline O & B \end{array}\right) = \left(\begin{array}{c|c} A & O \\ \hline O & I \end{array}\right)\left(\begin{array}{c|c} I & O \\ \hline O & B \end{array}\right)$$

and we can use Lemma above.

(ii) $M^{-1}M = I \implies \det(M^{-1})\det(M) = \det(I) = 1$ so $\det(M^{-1}) = (\det M)^{-1}$.

(iii) For R real and orthogonal,

$$R^\top R = I \implies \det(R^\top)\det(R) = (\det R^2) = 1$$

$$\implies \det R = \pm 1$$

(iv) For U complex and unitary

$$U^\dagger U = UI \implies \det(U^\dagger)\det(U) = \overline{\det(U)}\det(U) = |\det(U)|^2 = 1$$

$$\implies |\det U| = 1$$

5.4 Minors, Cofactors and Inverses

(a) Cofactors and Determinants

Consider column $\mathbf{C}_a$ of matrix M (a fixed) and write $\mathbf{C}_a = \sum_i M_{ia}\mathbf{e}_i$ in definition of determinant:

$$\begin{aligned}
\det M &= [\mathbf{C}_1, \ldots, \mathbf{C}_{a-1}, \mathbf{C}_a, \mathbf{C}_{a+1}, \ldots, \mathbf{C}_n] \\
&= [\mathbf{C}_1, \ldots, \mathbf{C}_{a-1}, \sum_i M_{ia}\mathbf{e}_i, \mathbf{C}_{a+1}, \ldots, \mathbf{C}_n] \\
&= \sum_i M_ia\Delta_{ia} \qquad \text{no sum over } a
\end{aligned}$$

where the *cofactor* Δ_{ia} is defined by

$$\Delta_{ia} = [\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_{a-1}, \mathbf{e}_i, \mathbf{C}_{a+1}, \ldots \mathbf{C}_n]$$

$$= \cdots$$

$$= \det \begin{pmatrix} A & \vdots & B \\ & 0 & \\ \cdots 0 & 1 & 0 \cdots \\ & 0 & \\ C & \vdots & D \end{pmatrix} \begin{matrix} \\ \\ \leftarrow \text{row } i \\ \\ \\ \end{matrix}$$

$$\qquad\qquad \uparrow \text{col } a$$

$$= (-1)^{n-a}(-1)^{n-i} \det \begin{pmatrix} A & B & \vdots \\ C & D & 0 \\ \cdots & 0 & 1 \end{pmatrix}$$

$$= (-1)^{a+i} M^{ia}$$

$$\text{where } M^{ia} = \det \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

introduced earlier. We have deduced

$$\begin{aligned}
\det M &= \sum_i M_{ia}\Delta_{ia} \\
&= \sum_i M_{ia}(-1)^{i+a}M^{ia}
\end{aligned}$$

proving proposition in section 5.3(b). [Similarly, considering row i, find other expression].

(b) Adjugates and Inverses

Reasoning as in (a) with $\mathbf{C}_b = \sum_i M_{ib}\mathbf{e}_i$

$$\begin{aligned}
[\mathbf{C}_1, \ldots, \mathbf{C}_{a-1}, \mathbf{C}_b, \ldots, \mathbf{C}_{a+1}, \ldots, \mathbf{C}_n] &= \sum_i M_{ib}\Delta_{ia} \\
&= \begin{cases} \det M & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}
\end{aligned}$$

Hence

$$\sum_i M_{ib}\Delta_{ia} = (\det M)\delta_{ab}$$

And similarly

$$\sum_a M_{ja}\Delta_{ia} = (\det M)\delta_{ij}$$

Let Δ be the *matrix* of cofactors with entries Δ_{ia}, and define *adjugate* $\widetilde{M} = \text{adj}(M) = \Delta^\top$. Then relations above because

$$\begin{aligned}
\Delta_{ia}M_{ib} &= (\Delta^\top)_{ai}M_{ib} \\
&= (\Delta^\top M)_{ab} \\
&= (\widetilde{M}M)_{ab} \\
&= (\det M)\delta_{ab}
\end{aligned}$$

and

$$M_{ja}\Delta_{ia} = (M\widetilde{M})_{ji} = (\det M)\delta_{ij}$$

This justifies $(*)$ in section 5.1 with

$$\widetilde{(M)} = \Delta^\top$$

and

$$\Delta_{ia} = (-1)^{i+a}M^{ia}$$

we have

$$\widetilde{M}M = M\widetilde{M} = (\det M)I$$

Hence if $\det M \neq 0$ then it is invertible and

$$M^{-1} = \frac{1}{\det M}\widetilde{M}$$

Example. Consider

$$A = \begin{pmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{pmatrix}$$

previously found $\det A = (a-1)^2(a+2)$. Hence A^{-1} exists if $a \neq 1$, $a \neq -2$. Matrix of cofactors is

$$\Delta = \begin{pmatrix} 1-a & 1-a & a^2-1 \\ a^2-1 & 1-a & 1-a \\ 1-a & a^2-1 & 1-a \end{pmatrix}$$

e.g.

$$A^{12} = \begin{vmatrix} a & 1 \\ 1 & 1 \end{vmatrix} = a - 1$$

$$\Delta_{12} = (-1)^{1+2} A^{12} = 1 - a$$

Adjugate $\widetilde{A} = \Delta^\top$ and

$$\begin{aligned} A^{-1} &= \frac{1}{\det A} \widetilde{A} \\ &= \frac{1}{(1-a)(a+2)} \begin{pmatrix} 1 & -(1+a) & 1 \\ 1 & 1 & -(1+a) \\ -(1+a) & 1 & 1 \end{pmatrix} \end{aligned}$$

if $a \neq 1$, $a \neq -2$.

5.5 Systems of Linear Equations

(a) Introduction and Nature of Solutions

Consider a system of n linear equations in n unknowns x_i written in vector / matrix form

$$A\mathbf{x} = \mathbf{b} \qquad \mathbf{x}, \mathbf{b} \in \mathbb{R}^n$$

and A an $n \times n$ matrix, i.e.

$$\begin{gathered} A_{11}x_1 + \cdots + A_{1n}x_n = b_1 \\ \vdots \\ A_{n1}x_1 + \cdots + A_{nn}x_n = b_n \end{gathered}$$

There are three possibilities:

(i) $\det A \neq 0 \implies A^{-1}$ exists $\implies$ unique solution $\mathbf{x} = A^{-1}\mathbf{b}$

(ii) $\det A = 0$ and $b \notin \text{Im } A \implies$ no solution.

(iii) $\det A = 0$ and $b \in \text{Im } A \implies$ infinitely many solutions.

Elaboration: a solution exists if and only if

$$A\mathbf{x}_0 = \mathbf{b} \text{ for some } \mathbf{x}_0 \iff \mathbf{b} \in \text{Im } A$$

Then $\mathbf{x}$ is also a solution if and only if $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$ satisfies

$$A\mathbf{u} = \mathbf{0}$$

homogeneous problem. Now

$$\begin{aligned} \det A \neq 0 &\iff \text{Im } A = \mathbb{R}^n \\ &\iff \text{Ker } A = \{\mathbf{0}\} \end{aligned}$$

So in (i) there is a unique solution and it can be found using A^{-1}. But

$$\begin{aligned} \det A = 0 &\iff \text{rank}(A) < n \\ &\iff \text{null}(A) > 0 \end{aligned}$$

and then either $\mathbf{b} \notin \text{Im } A$ as in case (ii) or $\mathbf{b} \in \text{Im } A$ as in case (iii). If $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for Ker A then general solution of homogeneous problem is

$$\mathbf{u} = \sum_{i=1}^{k} \lambda_i \mathbf{u}_i$$

Example
$A\mathbf{x} = \mathbf{b}$ with A as in section 5.4 and

$$\mathbf{b} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$$

with $a, c \in \mathbb{R}$.

- $\underline{a \neq 1, -2}$
  Then A^{-1} exists and we have a solution for any c:

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{(1-a)(a+2)} \begin{pmatrix} 2 - c - ca \\ c - a \\ c - a \end{pmatrix}$$

- $\underline{a = 1}$

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

$$\text{Im } A = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \qquad \text{Ker } A = \text{span} \left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$b \in \text{Im } A$ if and only if $c = 1$, particular solution

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

general solution

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{u} = \begin{pmatrix} 1 - \lambda - \mu \\ \lambda \\ \mu \end{pmatrix}$$

case (ii). For $a = 1$ and $c \neq 1$ have no solutions: case (iii).

- $\underline{a = -2}$

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

$$\text{Im } A = \text{span} \left\{ \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\}, \left\{ \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\} \qquad \text{Ker } A = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

$\mathbf{b} \in \text{Im } A$ if and only if $c = -2$, particular solution

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

general solution

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{u} = \begin{pmatrix} 1 + \lambda \\ \lambda \\ \lambda \end{pmatrix}$$

For $c \neq -2$ no solutions.

(b) Geometrical Intepretation in $\mathbb{R}^3$

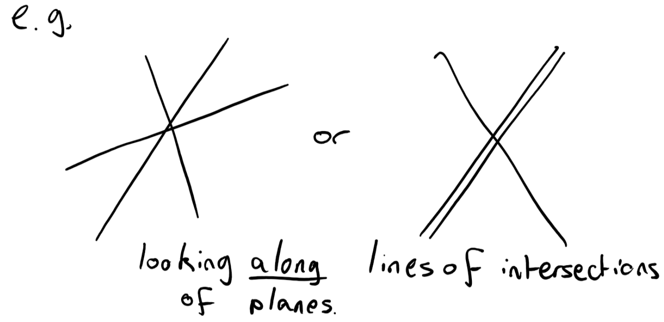
Let $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ be rows of A (3×3).

$$A\mathbf{u} = \mathbf{0} \iff \begin{cases} \mathbf{R}_1 \cdot \mathbf{u} = \mathbf{0} \\ \mathbf{R}_2 \cdot \mathbf{u} = bf0 \\ \mathbf{R}_3 \cdot \mathbf{u} = \mathbf{0} \end{cases}$$

(these are 3 equations of planes through $\mathbf{0}$, normals $\mathbf{R}_i$, assuming $\neq \mathbf{0}$). So solutions of homogeneous problem (finding Ker A) given by intersection of these planes.

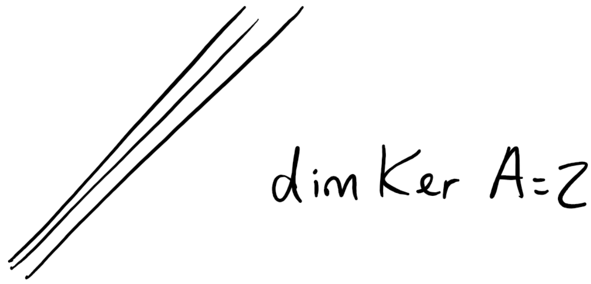
$$\text{rank}(A) = 3 \implies \text{normals linearly independent and planes intersect in } \mathbf{0}$$

$$\text{rank}(A) = 2 \implies \text{normals span a plane and planes intersect in a line}$$



$$\dim \text{Ker } A = 1.$$

$$\text{rank}(A) = 1 \implies \text{normals are parallel and planes coincide}$$



Now consider instead

$$A\mathbf{x} = \mathbf{b} \iff \begin{cases} \mathbf{R}_1 \cdot \mathbf{x} = b_1 \\ \mathbf{R}_2 \cdot \mathbf{x} = b_2 \\ \mathbf{R}_3 \cdot \mathbf{x} = b_3 \end{cases}$$

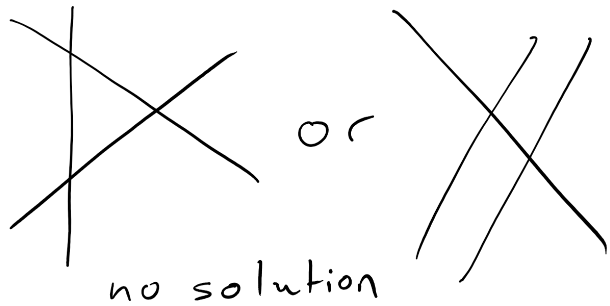
planes with normals $\mathbf{R}_i$ but not passing through $\mathbf{0}$ unless $b_i = 0$.

$$\text{rank}(A) = 3 \iff \det A \neq 0,$$

normals linearly independent; planes intersect in a point and get unique solution for any $\mathbf{b}$.

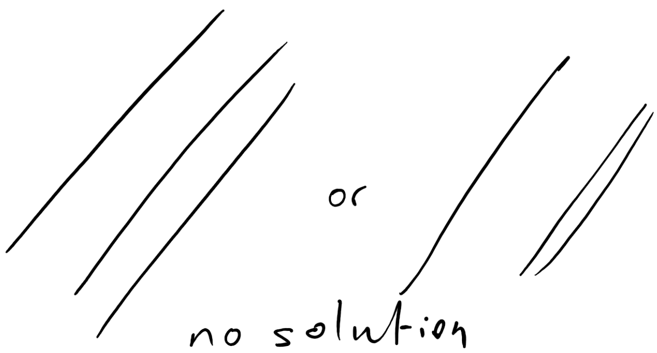
$$\text{rank}(A) = 2 \implies \text{planes may intersect in a line (as in homogeneous case)}$$

but they may not, e.g.



$$\text{rank}(A) = 1 \implies \text{planes may coincide (as in homogeneous case)}$$

but they may not, e.g.



Gaussian Elimination and Echelon Form

Consider $A\mathbf{x} = \mathbf{b}$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ and A an $m \times n$ matrix. Gaussian elimination is a direct approach to solving system of equations:

$$A_{11}x_1 + \cdots + A_{1n}x_n = b_1$$
$$\vdots$$
$$A_{m1}x_1 + \cdots + A_{mn}x_n = b_m$$

Example.

$$3x_1 + 2x_2 + x_3 = b_1 \tag{1}$$

$$6x_1 + 3x_2 + 3x_3 = b_2 \tag{2}$$

$$6x_1 + 2x_2 + 4x_3 = b_3 \tag{3}$$

Step (1): subtract multiples of (1) from (2) and (3) to eliminate x_1:

$$0 - x_2 + x_3 = b_2 - 2b_1 \tag{2'}$$

$$0 - 2x_2 + 2x_3 = b_3 - 2b_1 \tag{3'}$$

Step (2): repeat this using (2′) to eliminate x_2:

$$0 + 0 + 0 = b_3 - 2b_2 + 2b_1 \tag{3''}$$

Now consider new system (1), (2′), (3′)

$$b_3 - 2b_2 + 2b_1 \neq 0 \implies \text{no solution}$$

$$b_3 - 2b_2 + 2b_1 = 0 \text{ then infinitely many solutions}$$

x_3 is arbitrary and then x_2 and x_1 determined from (2′) and (1). In general case we aim to carry out steps as in example until we obtain equivalent system

$$M\mathbf{x} = \mathbf{d} \text{ with } M = \left(\begin{array}{c|c} \hat{M} & \text{numbers} \\ 0 & 0 \end{array}\right)$$

with M an $m \times n$ (block form), with

$$\hat{M} = \begin{pmatrix} M_{11} & & \text{numbers} \\ & \cdots & \\ 0 & & M_{rr} \end{pmatrix}$$

$M_{jj} \neq 0$ for each j. M obtained from A by row operations including row exchanges and column exchanges which relabel variables x_i. Note $x_{r+1}, \ldots, x_n$ undetermined, $d_{r+1}, \ldots, d_m = 0$ else no solution. And if this is satisfied then $x_1, \ldots, x_r$ determined successively.

$$r = \text{rank } M = \text{rank } A$$

If $n = m$ then $\det A = \pm \det M$ and if $r = n = m$ then

$$\det M = M_{11} \cdots M_{rr} \neq 0$$

$$\implies A \text{ and } M \text{ invertible}$$

M as above is an example of *echelon form*.

6 Eigenvalues and Eigenvectors

6.1 Introduction

(a) Definitions

For a linear map $T : V \to V$ (V a real or complex vector space) a vector $\mathbf{v} \in V$ with $\mathbf{v} \neq \mathbf{0}$ is an *eigenvector* of T with *eigenvalue* λ if

$$T(\mathbf{v}) = \lambda \mathbf{v}$$

If $V = \mathbb{R}^n$ or $\mathbb{C}^n$ and T given by an $n \times n$ matrix A, then

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

and for given λ this holds for some $\mathbf{v} \neq \mathbf{0}$ if and only if $\det(A - \lambda I) = 0$ *characteristic equation* i.e. λ is an eigenvalue if and only if it is a root of $\chi_A(t) = \det(A - tI)$ *characteristic polynomial.* $\chi_A(t)$ polynomial of degree n for A $n \times n$. We find eigenvalues as roots of characteristic equation and then find corresponding eigenvectors.

(b) Examples

(i) $V = \mathbb{C}^2$ and

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}$$

then

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & i \\ -i & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = 0$$

if and only if $\lambda = 1$ or 3. To find eigenvectors $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$: $\underline{\lambda = 1}$:

$$(A - I)\mathbf{v} = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0}$$

$$\implies \mathbf{v} = \alpha \begin{pmatrix} 1 \\ i \end{pmatrix} \qquad \text{any } \alpha \neq 0.$$

$\underline{\lambda = 3}$:

$$(A - 3I)\mathbf{v} = \begin{pmatrix} -1 & i \\ -i & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0}$$

$$\implies \mathbf{v} = \beta \begin{pmatrix} 1 \\ -i \end{pmatrix} \qquad \text{any } \beta \neq 0.$$

(ii) $V = \mathbb{R}^2$

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

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 = 0$$
$$\implies \lambda = 1$$

Eigenvector:

$$(A - I)\mathbf{v} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0}$$
$$\implies \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{for any } \alpha \neq 0.$$

(iii) $V = \mathbb{R}^2$ or $\mathbb{C}^2$

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
$$\chi_U(t) = \det(U - tI) = t^2 - 2t\cos\theta + 1$$

Eigenvalues $\lambda = e^{\pm i\theta}$ and eigenvectors

$$\mathbf{v} = \alpha \begin{pmatrix} 1 \\ \mp i \end{pmatrix} \qquad (\alpha \neq 0)$$

(c) Deductions involving $\chi_A(t)$

For A an $n \times n$ matrix, characteristic polynomial has degree n:

$$\chi_A(t) = \det \begin{pmatrix} A_{11} - t & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22-t} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn-t} \end{pmatrix}$$
$$= \sum_{j=0}^{n} c_j t^j$$
$$= (-1)^n (t - \lambda_1) \cdots (t - \lambda_n)$$

(i) There exists at least one eval (one root of χ_A); in fact there exists n roots counted with multiplicity (Fundamental Theorem of Algebra)

(ii) $\mathrm{tr}(A) = A_{ii} = \sum_i \lambda_i$ sum of reals by comparing terms of order $n-1$ in t.

(iii) $\det(A) = \chi_A(0) = \prod_i \lambda_i$ (product of eigenvalues)

(iv) If A is diagonal:

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

with diagonal entries eigenvalues; (ii) and (iii) are then immediate.

(v) If A is real, coefficients c_i are real and $\chi_A(\lambda) = 0 \iff \chi_A(\overline{\lambda}) = 0$: non-real roots occur in conjugate pairs.

6.2 Eigenspaces and Multiplicities

(a) Definitions

For an eigenvalue λ of matrix A, define the *eigenspace*

$$E_\lambda = \{\mathbf{v} : A\mathbf{v} = \lambda\mathbf{v}\} = \mathrm{Ker}(A - \lambda I);$$

the *geometric multiplicity*

$$m_\lambda = \dim E_\lambda = \mathrm{null}(A - \lambda I).$$

(# linearly independent eigenvalues eigenvectors with eval λ);
the *algebraic multiplicity*

$$M_\lambda, \text{ multiplicity of } \lambda \text{ as a root of } \chi_A$$

i.e. $\chi_A(t) = (t - \lambda)^{M_\lambda} f(t)$ with $f(\lambda) \neq 0$.

Proposition. $M_\lambda \geq m_\lambda$

[Further discussion in section 6.3]

(b) Examples

(i) Define:

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

$$\chi_A(t) + \det(A - tI) = (5 - t)(t + 3)^2$$

so we have roots 5 and -3, with $M_5 = 1$ and $M_{-3} = 2$.

- For $\lambda = 5$ we have:

$$(A - 5I)\mathbf{x} = \begin{pmatrix} -7 & 2 & -3 \\ 2 & -4 & -6 \\ -1 & -2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

$$\implies E_5 = \left\{ \alpha \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\}$$

- For $\lambda = -3$ we have

$$(A + 3I)\mathbf{x} = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}$$

Solve to find:

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

or

$$E_{-3} = \left\{ \alpha \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} \right\}$$

So

$$\dim E_5 = m_5 = 1 = M_5$$
$$\dim E_{-3} = m_{-3} = 2 = M_{-3}$$

(ii) Consider

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

Then

$$\chi_A(t) = \det(A - tI) = -(t+2)^3$$

roots are $\lambda = -2$, with $M_{-2} = 3$. To find eigenvectors:

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

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

$$\implies E_{-2} = \left\{ \alpha \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

so $\dim E_{-2} = m_{-2} = 2$ but $M_{-2} = 3$. (So we do have $M_{-2} \geq m_{-2}$.)

(c) Linear Independence of Eigenvectors

Proposition. (i) Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be eigenvectors of matrix A $(n \times n)$ with eigenvalues $\lambda_1, \ldots, \lambda_r$. If the eigenvalues are distinct, $\lambda_i \neq \lambda_j$ for $i \neq j$, then the eigenvectors are linearly independent.

(ii) With conditions as in (i), let $\mathcal{B}_{\lambda_i}$ be a basis for E_{λ_i}, then

$$\mathcal{B}_{\lambda_1} \cup \mathcal{B}_{\lambda_2} \cup \cdots \cup \mathcal{B}_{\lambda_r}$$

is linearly independent.

Proof.

(i) Note

$$\mathbf{w} = \sum_{j=1}^{r} \alpha_j \mathbf{v}_j$$
$$\implies (A - \lambda I)\mathbf{w} = \sum_{j=1}^{r} \alpha_j (\lambda_j - \lambda)\mathbf{v}_j$$

First, suppose eigenvectors are linearly dependent, so there exists linear relations $\mathbf{w} = \mathbf{0}$ with number of non-zero coefficients $p \geq 2$. Pick a $\mathbf{w}$ for which p is least and assume (without loss of generality) that $\alpha_1 \neq 0$. Then

$$(A - \lambda_1 I)\mathbf{w} = \sum_{j>1} \alpha_j (\lambda_j - \lambda_1)\mathbf{v}_j = \mathbf{0},$$

a linear relation with $p - 1$ non-zero coefficients,※ (p was least).
Alternative second proof,

$$\mathbf{w} = \mathbf{0}$$
$$\implies \prod j \neq k(A - \lambda_j I)\mathbf{w} = \alpha_k \left(\prod_{j \neq k} (\lambda_k - \lambda_j) \right) \mathbf{v}_k = \mathbf{0}$$

(for some chosen k).

$$\implies \alpha_k = 0$$

so the eigenvectors are linearly independent.

(ii) It suffices to show that if

$$\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2 + \cdots + \mathbf{w}_r = \mathbf{0}$$

with $\mathbf{w}_i \in E_{\lambda_i}$ then

$$\implies \mathbf{w}_i = \mathbf{0}.$$

This follows by same arguments as in (i).

□

6.3 Diagonalisability and Similarity

(a) Introduction

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

(i) There exists a basis of eigenvectors for V, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ with

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

(no summation convention here!)

(ii) There exists an $n \times n$ invertible matrix P with

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

If either of these conditions holds, A is *diagonalisable*.

Proof. Note that for any matrix P, AP has columns $A\mathbf{C}_i(P)$ and PD has columns $\lambda_i \mathbf{C}_i(P)$ for each i. Then (i) and (ii) are related by

$$\mathbf{v}_i = \mathbf{c}_i(P) : P^{-1}AP = D \iff AP = PD \iff A\mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

□

Example
Refer to section 6.1(b):

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

eigenvalues $e^{\pm i\theta}$ and eigenvectors $\begin{pmatrix} 1 \\ \mp 1 \end{pmatrix}$. Linearly independent over $\mathbb{C}$ so

$$P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \implies P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$$

and

$$P^{-1}UP = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

U diagonalisable over $\mathbb{C}$ but *not* over $\mathbb{R}$.

(b) Criteria for Diagonalisability

Theorem. Let A be an $n \times n$ matrix and $\lambda_1, \ldots, \lambda_r$ all its distinct eigenvalues.

(i) A necessary and sufficient condition: A is diagonalisable if and only if

$$M_{\lambda_i} = m_{\lambda_i} \qquad \text{for } i = 1, \ldots, r$$

(ii) A sufficient condition: A is diagonalisable if there are n distinct eigenvalues, i.e. $r = n$.

Proof. Use Proposition in section 6.2(c)
For (ii) if $r = n$ we have n distinct eigenvalues and hence n linearly independent eigenvalues, which form a basis (for $\mathbb{R}^n$ or $\mathbb{C}^n$).
For (i), choosing bases $\mathcal{B}_{\lambda_i}$ for each eigenspace,

$$\mathcal{B}_{\lambda_i} \cup \mathcal{B}_{\lambda_2} \cup \cdots \cup \mathcal{B}_{\lambda_r}$$

is a linearly independent set of

$$m_{\lambda_1} + m_{\lambda_2} + \cdots + m_{\lambda_r}$$

vectors. It is a basis (for $\mathbb{R}^n$ or $\mathbb{C}^n$) if and only if we have n vectors. But

$$m_{\lambda_i} \leq M_{\lambda_i}$$

and

$$M_{\lambda_1} + M_{\lambda_2} + \cdots + M_{\lambda_r} = n.$$

Hence we have a basis if and only if

$$M_{\lambda_i} = m_{\lambda_i} \qquad \text{for each } i$$

□

Examples
Refer to section 6.2(b)

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

$$\lambda = 5, -3, -3 \qquad M_5 = m_5 = 1 \qquad M_{-3} = m_{-3} = 2$$

hence A diagonalisable.

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

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

as expected.

(ii)

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

$$\lambda = -2, -2, -2 \qquad M_{-2} = 3 > m_{-2} = 2$$

hence A is not diagonalisable. Check: if it was then

$$P^{-1}AP = -2I$$

$$\Longrightarrow A = P(-2I)P^{-1} = -2I \text{ ⨳}.$$

(c) Similarity

Matrices A and B $(n \times n)$ are *similar* if

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

for some invertible P $(n \times n)$. This is an equivalence relation.

Proposition. If A and B are similar, then

(i) $B^r = P^{-1}A^rP$ for $r \geq 0$.

(ii) $B^{-1} = P^{-1}A^{-1}P$ (if either A or B invertible, so is the other).

(iii) $\operatorname{tr}(B) = \operatorname{tr}(A)$.

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

(v) $\chi_B(t) = \chi_A(t)$.

Proof. (i) and (ii) immediate. (iii):

$$\begin{aligned} \operatorname{tr}(B) &= \operatorname{tr}(P^{-1}AP) \\ &= \operatorname{tr}(APP^{-1}) \\ &= \operatorname{tr}(A) \end{aligned}$$

For (iv):

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

For (v):

$$\begin{aligned} \det(B - tI) &= \det(P^{-1}AP - tI) \\ &= \det(P^{-1}(A - tI)P) \\ &= \det(A - tI) \end{aligned}$$

as in (iv). □

6.4 Hermitian and Symmetric Matrices

(a) Real Eigenvalues and Orthogonal Eigenvectors

Recall: matrix A $(n \times n)$ is hermitian if

$$A^\dagger = \overline{A}^\top = A \qquad \text{or} \qquad A_{ij} = \overline{A_{ji}}$$

special case: A is real and symmetric

$$\overline{A} = A \qquad A^\top = A \qquad \text{or} \qquad \begin{cases} A_{ij} = \overline{A_{ij}} \\ A_{ij} = A_{ji} \end{cases}$$

Recall: complex inner-product for $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ IS

$$\mathbf{v}^\dagger \mathbf{w} = \sum_i \overline{v_i} w_i$$

and for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ this reduces to

$$\mathbf{v}^\top \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = \sum_i v_i w_i$$

Observation: if A is hermitian then

$$(A\mathbf{v})^\dagger \mathbf{w} = \mathbf{v}^\dagger (A\mathbf{w}) \;\; \forall \mathbf{v}, \mathbf{w} \in \mathbb{C}^n$$

[since $LHS = (\mathbf{v}^\dagger A^\dagger)\mathbf{w} = \mathbf{v}^\dagger A^\dagger \mathbf{w} = \mathbf{v}^\dagger A \mathbf{w} = RHS$]

Theorem. For a matrix A $(n \times n)$ that is hermitian

(i) Every eigenvalue λ is real

(ii) Eigenvectors $\mathbf{v}$, $\mathbf{w}$ with distinct eigenvalues λ, μ respectively $(\lambda \neq \mu)$ are orthogonal

$$\mathbf{v}^\dagger \mathbf{v} = 0$$

(iii) If A is real and symmetric then for each λ in (i) we can choose a real eigenvector $\mathbf{v}$ and (ii) becomes

$$\mathbf{v}^\top \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$$

Proof.

(i)

$$\begin{aligned} \mathbf{w}^\dagger (A\mathbf{v}) &= (A\mathbf{v})^\dagger \mathbf{v} \\ \implies \mathbf{v}^\dagger (\lambda \mathbf{v}) &= (\lambda \mathbf{v})^\dagger \mathbf{v} \\ \implies \lambda \mathbf{v}^\dagger \mathbf{v} &= \overline{\lambda} \mathbf{v}^\dagger \mathbf{v} \end{aligned}$$

for $\mathbf{v}$ an eigenvector with eigenvalue λ. But $\mathbf{v} \neq \mathbf{0}$ so $\mathbf{v}^\dagger \mathbf{v} \neq 0$ and $\lambda = \overline{\lambda}$.

(ii)

$$
\begin{aligned}
\mathbf{v}^\dagger(A\mathbf{w}) &= (A\mathbf{v})^\dagger\mathbf{w} \\
\implies \mathbf{v}^\dagger(\mu\mathbf{w}) &= (\lambda\mathbf{v})^\dagger)\mathbf{w} \\
\implies \mu\mathbf{v}^\dagger\mathbf{w} &= \overline{\lambda}\mathbf{v}^\dagger\mathbf{w} \\
&= \lambda\mathbf{v}^\dagger\mathbf{w}
\end{aligned}
$$

from (i). But $\lambda \neq \mu$ so $\mathbf{v}^\dagger\mathbf{w} = 0$.

(iii) Given $A\mathbf{v} = \lambda\mathbf{v}$ with $\mathbf{v} \in \mathbb{C}^n$ and A, λ real, let

$$\mathbf{w} = \mathbf{u} + i\mathbf{u}'$$

with $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^n$. Then $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{u}' = \lambda\mathbf{u}'$ but $\mathbf{v} \neq 0$ implies one of $\mathbf{u}$ or $\mathbf{u}'$ is nonzero, so there is at least one real eigenvector.

□

Unitary and Orthogonal Diagonalisation

Theorem. Any $n \times n$ hermitian matrix A is diagonalisable (as in section 6.3(a))

(i) There exists a basis of eigenvectors

$$\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{C}^n$$

with

$$A\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

; or equivalently

(ii) There exists $n \times n$ invertible matrix P with

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

columns of P are eigenvectors $\mathbf{u}_i$.

In addition: the eigenvectors $\mathbf{u}_i$ can be chosen to be orthonormal

$$\mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}.$$

or equivalently the matrix P can be chosen to be unitary

$$P^\dagger = P^{-1} \implies P^\dagger AP = D$$

Special case: for $n \times n$ real symmetric A, can choose eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ with

$$\mathbf{u}_i^\top \mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij}$$

equivalently, the matrix P can be chosen to be orthogonal

$$P^\top = P^{-1} \implies P^\top AP = D$$

Proof of diagonalisability is *not examinable* and remaining statements follow by combining results of section 6.2, 6.3 and choosing *orthonormal* basis for each eigenspace.

Examples

(i) Consider hermitian ($A^\dagger = A$) as in section 6.1(b):

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}$$

then $\lambda_1 = 1$ and $\lambda_2 = 3$ and choose

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

to ensure $\mathbf{u}_1^\dagger \mathbf{u}_1 = \mathbf{u}_2^\dagger \mathbf{u}_2 = 1$ and note

$$\mathbf{u}_1^\dagger \mathbf{u}_2 = \frac{1}{2}(1-i)\begin{pmatrix} 1 \\ -i \end{pmatrix} = 0.$$

Let

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

then $P^\dagger = P^{-1}$ unitary and

$$P^\dagger A P = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

(ii) Consider symmetric matrix

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

then $\lambda_1 = \lambda_2 = -1$ and $\lambda_3 = 2$ and can choose

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \qquad \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Let P be matrix with columns $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ then $P^\top = P^{-1}$ orthogonal

$$P^\top A P = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

6.5 Quadratic Forms

Consider $\mathcal{F} : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\mathcal{F}(\mathbf{x}) = 2x_1^2 - 4x_1x_2 + 5x_2^2$$

This can be expressed

$$\mathbf{F}(\mathbf{x}) = x_1'^2 + 6x_2'^2$$

where

$$x_1' = \frac{1}{\sqrt{5}}(2x_1 + x_2)$$

$$x_2' = \frac{1}{\sqrt{5}}(-x_1 + 2x_2)$$

with $x_1'^2 + x_2'^2 = x_1^2 + x_2^2$. To understand this better, note

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x}$$

where

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

and we can diagonalise A because $\lambda_1 = 1$, $\lambda_2 = 6$, and then we can compute

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Then

$$\begin{aligned} x_1' &= \mathbf{u}_1 \cdot \mathbf{x} \\ x_2' &= \mathbf{u}_2 \cdot \mathbf{x} \end{aligned}$$

give the simplified form for $\mathcal{F}$. In general, a *quadratic form* is a function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^2$ given by

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} = x_i A_{ij} x_j$$

where A is an $n \times n$ real symmetric matrix. From section 6.4,

$$P^\top A P = D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

where λ_i are eigenvalues of A and P orthogonal with columns $\mathbf{u}_i$ orthonormal eigenvectors. Let $\mathbf{x}' = P^\top \mathbf{x}$ or $\mathbf{x} = P\mathbf{x}'$. Then

$$\begin{aligned} \mathcal{F}(\mathbf{x} &= \mathbf{x}^\top A \mathbf{x} \\ &= (P\mathbf{x}')^\top A (P\mathbf{x}') \\ &= (\mathbf{x}')^\top (P^\top A P) \mathbf{x}' \\ &= (\mathbf{x}')^\top D \mathbf{x}' \end{aligned}$$

$\mathcal{F}$ has been *diagonalised.* Now

$$\mathbf{x}' = x_1' \mathbf{e}_1 + \cdots + x_n' \mathbf{e}_n$$

and

$$\begin{aligned} \mathbf{x} &= x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n \\ &= x_1' \mathbf{u} + \cdots + x_n' \mathbf{u}_n \end{aligned}$$

since $x_i' = \mathbf{u}_i \cdot \mathbf{x} \iff \mathbf{x}' P^\top \mathbf{x}$. Thus, x_i' are coordinates with respect to new axes given by orthonormal basis vector $\mathbf{u}_i$ and these called *principal axes* of $\mathcal{F}$. Relation to original axes along standard basis vectors $\mathbf{e}_i$ and coordinates x_i is given by an orthogonal transformation

$$|\mathbf{x}|^2 = x_i x_i = x_i' x_i'$$

(b) Examples in $\mathbb{R}^2$ and $\mathbb{R}^3$

In $\mathbb{R}^2$

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$$

with

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

eigenvalues $\lambda_1 = \alpha + \beta$, $\lambda_2 = \alpha - \beta$. Eigenvectors:

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\begin{aligned} \mathcal{F}(\mathbf{x}) &= \alpha x_1^2 + 2\beta x_1 x_2 + \alpha x_2^2 \\ &= (\alpha+\beta)x_1'^2 + (\alpha-\beta)x_2'^2 \end{aligned}$$

with

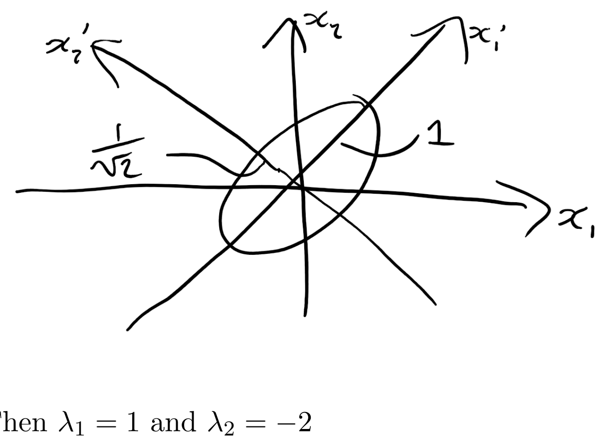
$$x_1' = \frac{1}{\sqrt{2}}(x_1 + x_2$$

$$x_2' = \frac{1}{\sqrt{2}}(-x_1 + x_2$$

(i) $\alpha = \frac{3}{2}$, $\beta = -\frac{1}{2}$. Then $\lambda_1 = 1$, $\lambda_2 = 2$.

$$\mathcal{F}(\mathbf{x}) = x_1'^2 + 2x_2'^2 = 1$$

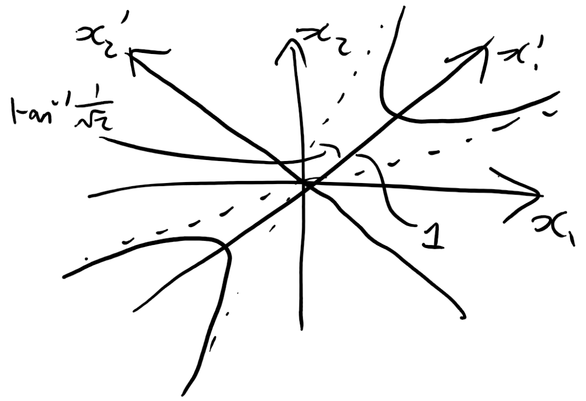
defines an ellipse.



(ii) $\alpha = -\frac{1}{2}$, $\beta = \frac{3}{2}$. Then $\lambda_1 = 1$ and $\lambda_2 = -2$

$$\mathcal{F}(\mathbf{x}) = x_1'^2 - 2x_2'^2 = 1$$

defines a hyperbola.



In $\mathbb{R}^3$

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \lambda_3 x_3'^2$$

after diagonalisation.

(i) If A has eigenvalues $\lambda_1, \lambda_2, \lambda_3 > 0$ then $\mathcal{F} = 1$ defines an ellipsoid.

(ii) From section 6.4,

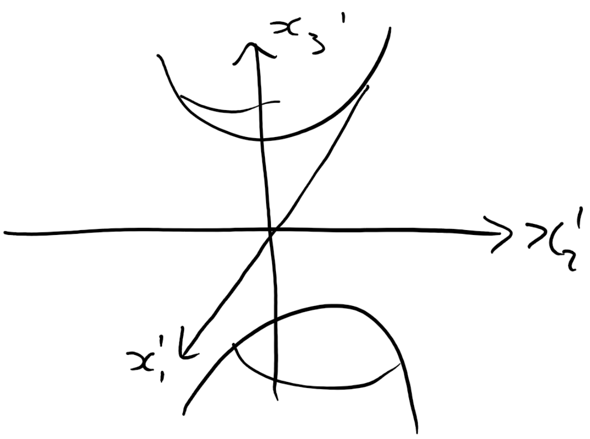
$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

has eigenvalues $\lambda_1 = \lambda_2 = -1$, $\lambda_3 = 2$. Hence

$$\mathcal{F} = 2x_1x_2 + 2x_2x_3 + 2x_3x_1$$
$$= -x_1'^2 - x_2'^2 + 2x_3'^2$$

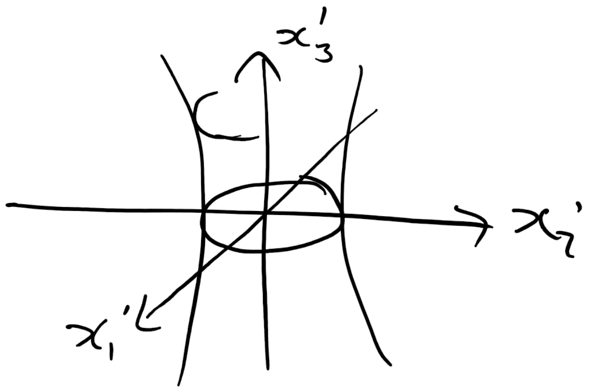
$$\mathcal{F} = 1 \iff 2x_3'^2 = 1 + x_1'^2 + x_2'^2$$

hyperboloid:



$$\mathcal{F} = -1 \iff x_1'^2 + x_2'^2 = 1 + 2x_3'^2$$

2 sheeted hyperboloid:



6.6 Cayley-Hamilton Theorem

If A is an $n \times n$ complex matrix and

$$f(t) = c_0 + c_1 t + \cdots + c_k t^k$$

polynomial of degree k, then

$$f(A) = c_0 I + c_1 A + \cdots + c_k A^k$$

We can also define power series of matrices subject to convergence, for example

$$\exp A = I + A + \cdots + \frac{1}{r!} A^r + \cdots$$

converges for any A. Note

(i) If

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

is some diagonal matrix, then

$$D^r = \begin{pmatrix} \lambda_1^r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^r \end{pmatrix}$$

and

$$f(D) = \begin{pmatrix} f(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f(\lambda_n) \end{pmatrix}$$

(ii) If $B = P^{-1}AP$ for invertible P, i.e. A and B are similar then

$$B^r = P^{-1}AP \qquad \text{and} \qquad f(B) = f(P^{-1}AP) = P^{-1}f(A)P$$

Recall, the characteristic polynomial is

$$\chi_A(t) = \det(A - tI) = c_0 + c_1 t + \cdots c_n t^n$$

where $c_0 = \det A$ and $c_n = (-1)^n$.

Theorem (Cayley-Hamilton)**.**

$$\chi_A(A) = c_0 I + c_1 A + \cdots + c_n A^n = 0$$

"a matrix satisfies its own characteristic equation"

Note. Cayley-Hamilton implies

$$c_0 I = -A(c_1 I + \cdots + c_n A^{n-1})$$

and if $c_0 = \det A \neq 0$ then

$$A^{-1} = -\frac{1}{c_0}(c_1 I + \cdots c_n A^{n-1}).$$

Proof.

(i) General 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \chi_A(t) = t^2 - (a+d)t + (ad - bc)$$

then check by substitution that $\chi_A = 0$ (on example sheet 4).

(ii) Diagonalisable $n \times n$ matrix:
consider A with eigenvalues λ_i and invertible P such that

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

and hence

$$\chi_A(D) = \begin{pmatrix} \chi_A(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \chi_A(\lambda_n) \end{pmatrix} = 0$$

since λ_i are eigenvalues. Then

$$\begin{aligned}\chi_A(A) &= \chi_A(P^{-1}DP) \\ &= P^{-1}\chi_A(D)P \\ &= 0\end{aligned}$$

as required.

(iii) The non diagonalisable case is beyond the scope of this course, but one can use an analytical argument to extend the diagonalisable case.

□

7 Changing Bases, Canonical Forms and Symmetries

7.1 Changing Bases in General

(a) Definitions and Proposition

Recall Section 4.4: given linear map $T : V \to W$ (real or complex vector spaces) and choice of bases

$$\{\mathbf{e}_i\} \quad i = 1, \ldots, n \quad \text{for } V$$

$$\{\mathbf{f}_a\} \quad a = 1, \ldots, m \quad \text{for } W$$

the matrix A $(m \times n)$ with respect to these bases is defined by

$$T(\mathbf{e}_i) = \sum_a \mathbf{f}_a A_{ai}$$

This definition is chosen to ensure

$$\mathbf{y} = T(\mathbf{x}) \iff y_A = \sum_i A_{ai} x_i = A_{ai} x_i$$

where

$$\mathbf{x} = \sum_i x_i \mathbf{e}_i, \qquad \mathbf{y} = \sum_a y_a \mathbf{f}_a,$$

which holds since

$$\begin{aligned} T(\sum_i x_i \mathbf{e}_i) &= \sum_i x_i T(\mathbf{e}_i) \\ &= \sum_i x_i (\sum_a \mathbf{f}_a A_{ai}) \\ &= \sum_a \underbrace{\left(\sum_i A_{ai} x_i \right)}_{= y_a \text{ as required}} \mathbf{f}_a \end{aligned}$$

Same linear map T has matrix A' with respect to bases

$$\{\mathbf{e}'_i\} \qquad i = 1, \ldots, n \quad \text{for } V$$

$$\{\mathbf{f}'_a\} \qquad a = 1, \ldots, m \quad \text{for } W$$

defined by

$$T(\mathbf{e}'_i) = \sum_a \mathbf{f}'_a A'_{ai}$$

To relate A and A' we need to say how bases are related, and *change* of *base* matrices P $(n \times n)$ and Q $(m \times n)$ are defined by

$$\mathbf{e}'_i = \sum_j \mathbf{e}_j P_{ji}, \qquad \mathbf{f}'_a = \sum_b \mathbf{f}_b Q_{ba}$$

Note. P and Q invertible; in relation above we can exchange $\{\mathbf{e}_i\}$ and $\{\mathbf{e}'_i\}$ with $P \to P^{-1}$ and similarly for Q.

Proposition. With definitions as above

$$A' = Q^{-1}AP$$

change of basis formula for matrix of a linear map.

Example. $n = 2$, $m = 3$

$$T(\mathbf{e}_1) = \mathbf{f}_1 + 2\mathbf{f}_2 - \mathbf{f}_3 = \sum_a \mathbf{f}_a A_{a1}$$

$$T(\mathbf{e}_2) = -\mathbf{f}_1 + 2\mathbf{2}_2 + \mathbf{f}_3 = \sum_a \mathbf{f}_a A_{a2}$$

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

New basis for V

$$\mathbf{e}'_1 = \mathbf{e}_1 - \mathbf{e}_2 = \sum_i \mathbf{e}_i P_{i1} \qquad \mathbf{e}'_2 = \mathbf{e}_1 + \mathbf{e}_2 = \sum_i \mathbf{e}_i P_{i2}$$

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

New basis for W

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

$$\implies Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Change of basis formula:

$$A' = Q^{-1}AP = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix}$$

Direct check

$$T(\mathbf{e}'_1) = 2\mathbf{f}'_1 \qquad T(\mathbf{e}'_2) = 4\mathbf{f}'_2$$

which agrees.

(b) Proof of Proposition

$$\begin{aligned}
T(\mathbf{e}'_i) &= T(\sum_j \mathbf{e}_j P_{ji}) && \text{definition of } P \\
&= \sum_j T(\mathbf{e}_j) P_{ji} && T \text{ linear} \\
&= \sum_j \sum_a \mathbf{f}_a A_{aj} P_{ji} && \text{definition of } A
\end{aligned}$$

$$\begin{aligned}
T(\mathbf{e}'_i) &= \sum_b \mathbf{f}'_b A'_{bi} && \text{definition of } A' \\
&= \sum_b \sum_a \mathbf{f}_a Q_{ab} A'_{bi} && \text{definition of } Q
\end{aligned}$$

Comparing coefficients of $\mathbf{f}_a$ (since it's a basis):

$$\sum_j A_{aj} P_{ji} = \sum_b Q_{ab} A'_{bi}$$

or

$$AP = QA'$$

as required.

(c) Approach using vector components

Consider

$$\begin{aligned}
\mathbf{x} &= \sum_j x_j \mathbf{e}_j \\
&= \sum_i x'_i \mathbf{e}'_i \\
&= \sum_j \left(\sum_i P_{ji} x'_i \right) \mathbf{e}_j \\
\implies x_j &= P_{ji} x'_i
\end{aligned}$$

Write

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{and} \qquad X' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

then

$$X = PX' \qquad \text{or} \qquad X' = P^{-1} X$$

Note: some care needed if $V = \mathbb{R}^n$, e.g. $n = 2$ with

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} 5 \\ 1 \end{pmatrix} \in \mathbb{R}^2$$

has $\mathbf{x} = 3\mathbf{e}_1 + 2\mathbf{e}_2$ so

$$X = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Similarly

$$\mathbf{y} = \sum_b y_b \mathbf{f}_b = \sum_a y'_a \mathbf{f}'_a$$

$$\implies y_b = Q_{ba} y'_a$$

Then

$$Y = QY' \qquad \text{or} \qquad Y' = Q^{-1}Y$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \qquad \text{and} \qquad Y' = \begin{pmatrix} y_1 \\ \vdots y'_m \end{pmatrix}$$

Now, marices A, A' are defined to ensure

$$Y = AX \qquad \text{and} \qquad Y' = A'X'$$

But

$$\begin{aligned} Y' &= Q^{-1}Y \\ &= Q^{-1}AX \\ &= (Q^{-1}AP)X' \\ &= A'X' \end{aligned}$$

and true $\forall \mathbf{x}$ so

$$A' = Q^{-1}AP.$$

Comments

(i) Definition of matrix A for $T : V \to W$ with respect to bases $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_a\}$ can be expressed; column i of A consists of components of $T(\mathbf{e}_i)$ with respect to basis $\{\mathbf{f}_a\}$. [For $T : \mathbb{R}^n \to \mathbb{R}^m$ with standard bases, columns of A are images of standard basis vectors.] Similarly, definitions of P and Q say: columns consist of complements of new basis vectors with respect to old.

(ii) With $V = W$ and same bases and $\mathbf{e}_i = \mathbf{f}_i$, $\mathbf{e}'_i = \mathbf{f}'_i$ we have

$$P = Q \qquad \text{and} \qquad A' = P^{-1}AP$$

Matrices representing the same linear map with respect to different bases are similar; conversely if A and A' are similar then we can regard them as representing same linear map with P defining change of basis. In section 6.3, we observed

$$\begin{aligned} \operatorname{tr}(A') &= \operatorname{tr}(A), \\ \det(A') &= \det(A), \\ \chi_{A'}(t) &= \chi_A(t) \end{aligned}$$

so these are properties of linear map.

(iii) $V = W = \mathbb{R}^n$ or $\mathbb{C}^n$, with $\mathbf{e}_i$ standard basis - matrix A is diagonalisable if and only if there exists basis of eigenvectors

$$\mathbf{e}'_i = \mathbf{v}_i$$

with

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \qquad \text{no summation convention!}$$

and then

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

and

$$\mathbf{v}_i = \sum_j \mathbf{e}_j P_{ji}$$

eigenvectors are columns of P. Specialising further $A^\dagger = A$ implies exists basis of orthonormal eigenvectors

$$\mathbf{e}'_i = \mathbf{u}_i \qquad \text{and} \qquad P^\dagger = P^{-1}$$

7.2 Jordan Canonical / Normal Form

This result classifies $n \times n$ complex matrices up to similarity.

Proposition. Any 2×2 complex matrix A is similar to one of the following:

(i) For some $\lambda_1 \neq \lambda_2$

$$A' = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

so

$$\chi_A(t) = (t - \lambda_1)(t - \lambda_2)$$

(ii) For some λ,

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

so

$$\chi_A(t) = (t - \lambda)^2$$

(iii) For some λ,

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

so

$$\chi_A(t) = (t - \lambda)^2$$

Proof. $\chi_A(t)$ has 2 roots over $\mathbb{C}$.

(i) For distinct roots or eigenvalues, λ_1, λ_2, we have $M_1 = m_1 = M_2 = m_2 = 1$ and eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ provide a basis.

(ii) For repeated root / eigenvalue λ, if $M_\lambda = m_\lambda = 2$, then same argument applies.

(iii) For repeated root / eigenvalue λ, with $M_\lambda = 2$ and $m_\lambda = 1$, let $\mathbf{v}$ be eigenvector for λ and $\mathbf{w}$ any linearly independent vector. Then

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

$$A\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{w}$$

say. Matrix of map with respect to basis $\{\mathbf{v}, \mathbf{w}\}$ is

$$\begin{pmatrix} \lambda & \alpha \\ 0 & \beta \end{pmatrix}$$

But $\beta = \lambda$ (otherwise case (i)) and $\alpha \neq 0$ (otherwise case (ii)). Now set $\mathbf{u} = \alpha\mathbf{v}$ and note

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

$$A\mathbf{w} = \mathbf{u} + \lambda\mathbf{w}$$

so with respect to basis $\{\mathbf{u}, \mathbf{w}\}$ matrix is

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

as claimed.

□

Example (using a slightly different approach)**.**

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

$$\implies \chi_A(t) = (t-3)^2$$

and

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

Choose

$$\mathbf{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

not an eigenvector and then

$$\mathbf{u} = (A - 3I)\mathbf{w} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

But $(A - 3I)^2 = 0$, and

$$A\mathbf{u} = 3\mathbf{u}$$

$$A\mathbf{w} = \mathbf{u} + 3\mathbf{w}$$

so basis $\{\mathbf{u}, \mathbf{w}\}$ gives JCF. Check:

$$P = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix} \implies P^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}$$

and

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

Generalisation to larger matrices can be consired, starting with

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$n \times n$. When applied to standard basis vectors get

$$\mathbf{w}_n \mapsto \mathbf{e}_{n-1} \mapsto \cdots \mapsto \mathbf{e}_1 \mapsto \mathbf{0}$$

Note

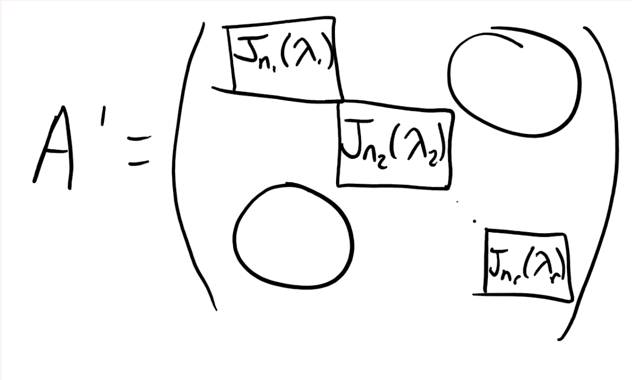
$$J = \lambda I + N$$

then

$$\chi_J(t) = (\lambda - t)^n$$

but $m_\lambda = 1$ $(M_\lambda = n)$.

Theorem. Any $n \times n$ complex matrix A is similar to a matrix A' with block form



where each diagonal block is a *Jordan block* with form

$$J_p(\lambda) = \underbrace{\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}}_{p \times p}$$

with $n_1 + \cdots n_3 = n$ and $\lambda_1, \ldots, \lambda_r$ are eigenvectors of A and A' (same eigenvalue may appear in more than one block). A is diagonalisable if and only if A' consists of 1×1 Jordan blocks only.

Proof. See Linear Algebra and GRM in Part IB. □

7.3 Quadrics and Conics

(a) Quadrics in General

A quadric in $\mathbb{R}^n$ is a hypersurface defined

$$Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} + \mathbf{x}^\top \mathbf{x} + x = 0$$

for some A, $n \times n$ real symmetric, non-zero matrix, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$. So

$$Q(\mathbf{x}) = A_{ij}x_i x_j + b_i x_i + c = 0$$

Consider classifying solutions up to geometrical equivalence: no distinction between solutions related by *isometries* (length preserving maps) in $\mathbb{R}^n$, i.e. related by

(i) translation - change in origin

(ii) orthogonal transformation about origin - change in axes.

If A is invertible (no zero eigenvalues) then by setting $\mathbf{y} = \mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}$ we have

$$\begin{aligned}\mathbf{y}^\top A\mathbf{y} &= (\mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b})^\top A(\mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}) \\ &= \mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + \frac{1}{4}\mathbf{b}^\top A^{-1}\mathbf{b}\end{aligned}$$

[since $(A^{-1}\mathbf{b})^\top = \mathbf{b}^\top(A^{-1})^\top$ and $(A^{-1})^\top = (A^\top)^{-1} = A_{-1}$ in this case.] Then $Q(\mathbf{x}) = 0 \iff \mathcal{F}(\mathbf{y}) = k$ with $\mathcal{F}(\mathbf{y}) = \mathbf{y}^\top A\mathbf{y}$. (quadratic form with respect to new origin $\mathbf{y} = \mathbf{0}$) and $k = \frac{1}{4}\mathbf{b}^\top A^{-1}\mathbf{b} - c$. Diagonalise $\mathcal{F}$ as in section 6.5: orthonormal eigenvectors give principal axes, eigenvalues of A and value of k determine nature of quadric. Example in $\mathbb{R}^3$ given in section 6.5(b)

(i) eigenvalues > 0 and $k > 0$ get ellipsoid

(ii) eigenvalues of different sign and $k \neq 0$ get hyperboloid

If A has one or more zero eigenvalues then analysis changes and simplest standard form may have both linear and quadratic terms.

(b) Conics

Quadrics in $\mathbb{R}^2$ are curves, *conics.*

$\det A \neq 0$.
By completing square and diagonalising we get a standard form

$$\lambda_1 x_1'^2 + \lambda_2 x_2'^2 = k$$

$$\lambda_1, \lambda_2 > 0 \implies \begin{cases} \text{ellipse for } k > 0 \\ \text{point for } k = 0 \\ \text{no solution for } k < 0 \end{cases}$$

$$\lambda_1 > 0, \lambda_2 < 0 \implies \begin{cases} \text{hyperbola for } k > 0 \text{ or } k < 0 \\ \text{pair of lines for } k = 0 \end{cases}$$

e.g.

$$x_1'^2 - x_2'^2 = (x_1 - x_2)(x_1 + x_2) = 0$$

$\det A = 0$ Suppose $\lambda_1 > 0$ and $\lambda_2 = 0$; diagonalise A in original formula to get

$$\lambda_1 x_1'^2 + b_1' x_1' + b_2' x_2' + c = 0$$

$$\iff \lambda_1 x_1''^2 + b_2' x_2' + c' = 0$$

where

$$x_1'' = x_1' + \frac{1}{2\lambda_1} b_1' \qquad \text{and} \qquad c' = c - \frac{b_1'^2}{4\lambda_1^2}$$

If $b_2' = 0$ then we get a pair of lines for $c' < 0$, single line for $c' = 0$ and no solutions for $c' > 0$. If $b_2' \neq 0$ then equation becomes

$$\lambda_1 x_1''^2 + b_2' x_2'' = 0$$

parabola where

$$x_2'' = x_2' + \frac{1}{b_2'} c'$$

7.4 Symmetries and Transformation Groups

(a) Orthogonal Transformation and Rotations in $\mathbb{R}^n$

$$\begin{aligned} R \text{ orthogonal} &\iff R^\top R = RR^\top = I \\ &\iff (R\mathbf{x}) \cdot (R\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \ \forall \mathbf{x}, \mathbf{y} \\ &\iff \text{columns or rows of } R \text{ orthonormal vectors} \end{aligned}$$

The set of such matrices forms the *orthogonal group* $\mathrm{O}(n)$.

$$R \in O(n) \implies \det R = \pm 1$$

$[\det(R^\top) \det(R) = [\det(R)]^2 = 1]$

$$\mathrm{SO}(n) = \{R \in \mathrm{O}(n) : \det R = +1\}$$

is a subgroup, the *special orthogonal group.*

$$R \in \mathrm{O}(n) \implies R \text{ preserves lengths and } |n\text{-dim vol}|$$

$$R \in \mathrm{SO}(n) \implies R \text{ also preserves orientation}$$

$\mathrm{SO}(n)$ consists of all rotations in $\mathbb{R}^n$.

Reflections belong to $\mathrm{O}(n) \setminus \mathrm{SO}(n)$, any element of $\mathrm{O}(n)$ is of the form

$$R \text{ or } RH \text{ with } R \in \mathrm{SO}(n)$$

e.g. if n is odd, we can choose $H = -I$.
Active and Passive Points of View
For a rotation R (matrix), the transformation

$$x_i' = R_{ij} x_j$$

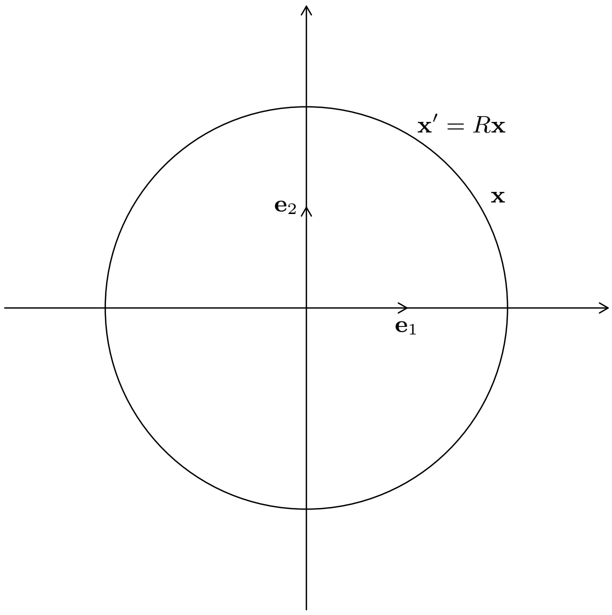
can be viewed in two ways.

Active view point: rotation transforms vectors

$$x_i' \text{ components of new vector}$$

$$\mathbf{x}' = R\mathbf{x} \text{ with respect to standard basis } \{e_i\}$$

e.g. $\mathbb{R}^2$

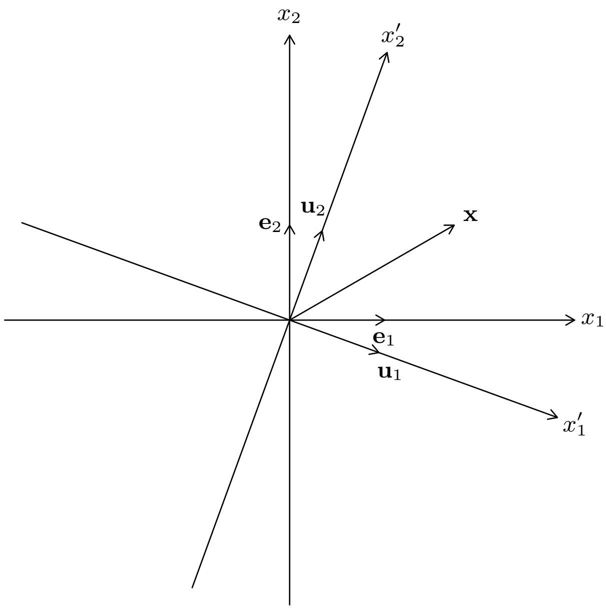


$$|\mathbf{x}'|^2 = |\mathbf{x}|^2$$

Passive view point: rotation changes basis

x'_i components of same vector $\mathbf{x}$ but with respect to new basis $\{u_i\}$

e.g. $\mathbb{R}^2$



$$\begin{aligned} \mathbf{u}_1 &= \sum_j R_{ij}\mathbf{e}_j \\ &= \sum_j \mathbf{e}_j (R^{-1})_{ji} \end{aligned}$$

(compare to section 6.5: $P = R^{-1}$)

(b) 2D Minkowski Space and Lorentz Transformations

Define a new "inner product" on $\mathbb{R}^2$ by

$$(\mathbf{x}, \mathbf{y} = \mathbf{x}^\top J \mathbf{y} = x_0 y_0 - x_1 y_1$$

where

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

where we now label components

$$\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{y} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$$

This is not positive definite, since

$$(\mathbf{x}, \mathbf{x}) = \mathbf{x}^\top J \mathbf{x} = x_0^2 - x_1^2$$

but still bilinear and symmetric. Standard basis vectors are "orthonormal" in generalised sence:

$$\mathbf{e}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{e}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

obey

$$\begin{aligned} (\mathbf{e}_0, \mathbf{e}_0) &= 1 \\ (\mathbf{e}_1, \mathbf{e}_1) &= -1 \\ (\mathbf{e}_0, \mathbf{e}_1) &= 0 \end{aligned}$$

New inner product is called the *Minkowski* metric and $\mathbb{R}^2$ equipped with it is called *Minkowski space*. Consider

$$M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$$

giving a linear map $\mathbb{R}^2 \to \mathbb{R}^2$. This preserves the Minkowski metric if and only if

$$\begin{aligned} (M\mathbf{x}, M\mathbf{y}) &= (\mathbf{x}, \mathbf{y} & \forall \mathbf{x}, athbfy \in \mathbb{R}^2 \\ \iff (M\mathbf{x})^\top J (M\mathbf{y}) &= \mathbf{x}^\top (M^\top J M)\mathbf{y} \\ &= \mathbf{x}^\top J \mathbf{y} & \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^2 \\ \iff M^\top J M &= J \end{aligned}$$

The set of such matrices forms a group. Now

$$\begin{aligned}\det(M^\top JM) &= \det M^\top \det J \det M \\ &= \det J \implies (\det M)^2 \qquad\qquad = 1 \\ \implies \det M &= \pm 1\end{aligned}$$

Furthermore, $|M_{00}| \geq 1$, so

$$M_{00} \geq 1 \qquad \text{or} \qquad M_{00} \leq -1.$$

The subgroup with

$$\det M = +1 \qquad \text{and} \qquad M_{00} \geq 1$$

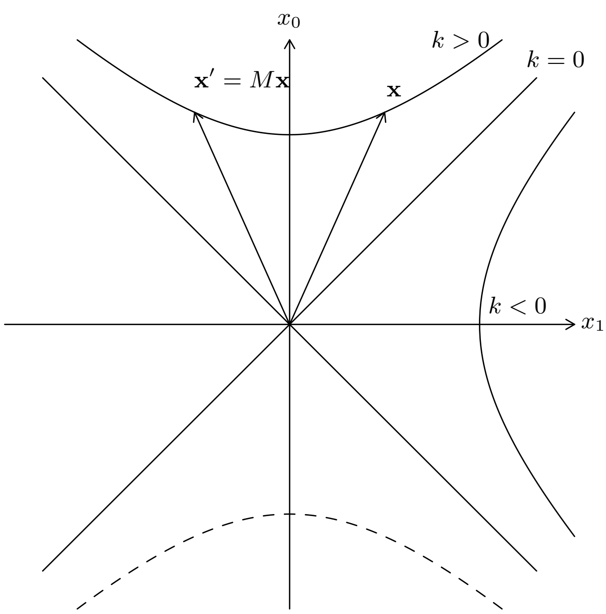
is the *Lorentz group* in 2D.

General form for M: require columns $M\mathbf{e}_0$ and $M\mathbf{e}_1$ to be orthonormal, like $\mathbf{e}_0$, $\mathbf{e}_0$ (with respect to new inner product). This implies

$$M(\theta) = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix}$$

First column fixed by requiring $(M\mathbf{e}_0, M\mathbf{e}_1) = 1$ or $M_{00}^1 - M_{10}^2 = 1$ and $M_{00} \geq 1$. The second column is then fixed by $M\mathbf{e}_0, M\mathbf{e}_1) = 0$, $(M\mathbf{e}_1, M\mathbf{e}_1) = -1$ and $\det M = +1$ (fixes overall sign). For such matrices

$$M(\theta_1)M(\theta_2) = M(\theta_1 + \theta_2)$$



curves with $(\mathbf{x}, \mathbf{x}) = k$, constant, as shown.

Physical application
Set

$$M(\theta) = \gamma(v) \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}$$

$$v = \tanh\theta$$

$$\gamma(v) = (1 - v^2)^{-1/2}, \qquad |v| < 1.$$

Rename $x_0 \to t$ time coordinate and $x_1 \to x$ space coordinate.

$$\mathbf{x}' = M\mathbf{x} \iff \begin{cases} t' = \gamma(t + vx) \\ x' = \gamma(x + vt) \end{cases}$$

Lorentz transformation or *boost* relating observes moving with relative velocity b according to Special Relativity (units with $c = 1$). Factor $\gamma(v) = (1 - v^2)^{-1/2}$ gives rise to effects such as time dilation and length contraction.