Part IA
Vectors and Matrices

Dr Ángela Capel Cuevas
Michaelmas 2025
Version 20251202

These are Zixuan's notes for **Part IA – Vectors and Matrices** at the University of Cambridge in 2025. The notes are not endorsed by the lecturers or the University, and all errors are my own.

The latest version of this document is available at academic.micfong.space. Please direct any comments to my CRSid email or use the contact details listed on the site.

This document is typeset using Typst. All figures are created using Inkscape and Mathematica.

Contents

Syllabus and Overview

Michaelmas Term [24 Lectures]

Complex Numbers [2 Lectures]

Review of complex numbers, including complex conjugate, inverse, modulus, argument and Argand diagram. Informal treatment of complex logarithm, n-th roots and complex powers. de Moivre's theorem.

Vectors [5 Lectures]

Review of elementary algebra of vectors in $\mathbb{R}^3$, including scalar product. Brief discussion of vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$; scalar product and the Cauchy–Schwarz inequality. Concepts of linear span, linear independence, subspaces, basis and dimension.

Matrices [8 Lectures]

Elementary algebra of 3×3 matrices, including determinants. Extension to n×n complex matrices. Trace, determinant, non-singular matrices and inverses. Matrices as linear transformations; examples of geometrical actions including rotations, reflections, dilations, shears; kernel and image, rank–nullity theorem (statement only). [4]

Simultaneous linear equations: matrix formulation; existence and uniqueness of solutions, geometric interpretation; Gaussian elimination. [3]

Symmetric, anti-symmetric, orthogonal, Hermitian and unitary matrices. Decomposition of a general matrix into isotropic, symmetric trace-free and antisymmetric parts. [1]

Eigenvalues and Eigenvectors [9 Lectures]

Eigenvalues and eigenvectors; geometric significance. [2]

Proof that eigenvalues of Hermitian matrix are real, and that distinct eigenvalues give an orthogonal basis of eigenvectors. The effect of a general change of basis (similarity transformations). Diagonalization of general matrices: sufficient conditions; examples of matrices that cannot be diagonalized. Canonical forms for 2×2 matrices. [5]

Discussion of quadratic forms, including change of basis. Classification of conics, cartesian and polar forms. [1]

Rotation matrices and Lorentz transformations as transformation groups. [1]

Lecture 1 · 2025-10-09

1 Complex Numbers

We should be fairly familiar with complex numbers already, but here is a recap of what should be well-covered.

1.1 Definition

We construct $\mathbb{C}$ by adding the element i to $\mathbb{R}$, satisfying $\mathrm{i}^2 = -1$. Then, any complex number $z \in \mathbb{C}$ has the form

$$z = x + \mathrm{i}y$$

where $x, y \in \mathbb{R}$.

Each complex number consists of a **real part** $\mathrm{Re}(z) = x$ and an **imaginary part** $\mathrm{Im}(z) = y$.

1.2 Properties

1. **Addition.** Given $z_1, z_2 \in \mathbb{C}$ where $z_1 = x_1 + \mathrm{i}y_1$, $z_2 = x_2 + \mathrm{i}y_2$, we can add or subtract them by
$$z_1 + z_2 = (x_1 + x_2) + \mathrm{i}(y_1 + y_2).$$
2. **Multiplication.** We can multiply z_1 and z_2 by
$$z_1 z_2 = (x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) = (x_1 x_2 - y_1 y_2) + \mathrm{i}(x_1 y_2 + x_2 y_1).$$

> ***Remark.*** Both addition and multiplication are associative and commutative.

3. **Identity.** The identity element for the addition operation is the element 0. [Thus $(\mathbb{C}, +)$ is an Abelian group with identity element 0.]
4. **Inverse.** For any $z \neq 0$, the inverse of z is given by
$$x^{-1} = \frac{x - \mathrm{i}y}{x^2 + y^2},$$
and it satisfies $z \cdot z^{-1} = 1$. [Thus $(\mathbb{C}^*, \cdot)$ is an Abelian group with identity element 1.]

   Moreover, distributivity is satisfied, *i.e.* if $z_1, z_2, z_3 \in \mathbb{C}$ Then
$$(z_1 + z_2) z_3 = z_1 z_3 + z_2 z_3.$$
5. **Complex conjugate.** For any $z = x + \mathrm{i}y$, the complex conjugate is $\bar{z} = x - \mathrm{i}y$. With this, we can write $\mathrm{Re}(z) = \frac{z+\bar{z}}{2}, \mathrm{Im}(z) = \frac{z-\bar{z}}{2}$.

   Properties of complex conjugates includes:
   1. $\overline{(\bar{z})} = z$.
   2. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$.
   3. $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$.
6. **Modulus.** For any $z = x + \mathrm{i}y$, we define the modulus in $\mathbb{R}_{\geqslant 0}$ to be
$$|z|^2 = x^2 + y^2.$$

We will sometimes denote $|z|$ by r.

7. **Argument.** The argument fo a complex number $z = x + iy \neq 0$ is a rela number, denoted by $\arg(z) = \theta$ such that

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

This is called the **polar form** of z. We can write

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

If θ is an argument of z, then any $\theta + 2\pi n$ where $n \in \mathbb{Z}$ is also an argument of z. Therefore, to make this argument unique, we restrict the range of theta to $(-\pi, \pi]$. We call arguments within this range to be the **principal value**.

We denote the principal value by $\arg(z)$. We also write $\text{Arg}(z) = \{\arg(z) + 2\pi n \mid n \in \mathbb{Z}\}$.

Remark.
1. $\mathbb{R} \subseteq \mathbb{C}$, since for any $\alpha \in \mathbb{R}$, we have $\alpha + i0 \in \mathbb{C}$.
2. Complex numbers of the form $0 + ib$ are called **pure imaginary numbers**.
3. The representation of a complex number in terms of real and imaginary parts is unique.

Once we have the properties above, here are a few more properties we can get to.

1. $(\mathbb{C}, +, \cdot)$ is a field.
2. For the modulus operation, we have
   - $|z_1 + z_2| \leqslant |z_1| + |z_2|$.
   - $|z_1 z_2| = |z_1||z_2|$.
   - $|z_1 - z_2| \geqslant ||z_1| - |z_2||$.

We can also reach the following theorem, though we will not prove it here.

Theorem 1.1 (Fundamental Theorem of Algebra)

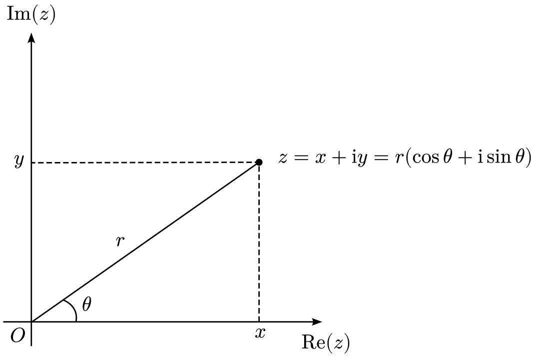
A polynomial of degree n with coefficients in $\mathbb{C}$ can be written as a product of n linear factors:

$$\begin{aligned} p(z) &= c_n z^n + \ldots + c_0 \quad \text{where } c_i \in \mathbb{C} \text{ and } c_n \neq 0 \\ &= c_{n(z-\alpha_1)}(z-\alpha_2)\ldots(z-\alpha_n) \quad \text{where } \alpha_i \in \mathbb{C}. \end{aligned}$$

Hence $p(z) = 0$ has at least one root in $\mathbb{C}$, and n roots α_i connected with multiplicity.

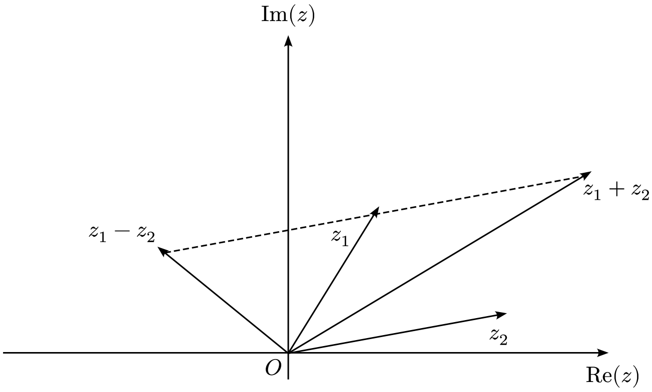
1.3 Argand Diagram

For $z = x + iy \in \mathbb{C}$, we can plot z in a 2-dimensional plot.

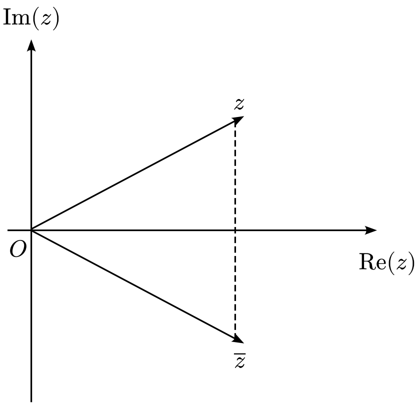


We can therefore demonstrate some operations on the diagram.

1. Addition and subtraction



2. Complex conjugates



This method immediately leads to some properties:

- $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$.
- $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$.
- $|\overline{z}| = |z|$.

1.4 De Moivre's Theorem

Theorem 1.2 (De Moivre's Theorem)

For any $\theta \in \mathbb{R}$, $n \in \mathbb{Z}$, we have

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

Proof. To prove this, we first need a lemma.

Lemma 1.3

Let $z_1 = r_1(\cos\theta_1 + \mathrm{i}\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + \mathrm{i}\sin\theta_2)$. Then

$$z_1 z_2 = r_1 r_2(\cos(\theta_1+\theta_2) + \mathrm{i}\sin(\theta_1+\theta_2)).$$

Proof. Multiplying z_1 and z_2,

$$z_1 z_2 = r_1 r_2\left(\underbrace{\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2}_{\cos(\theta_1+\theta_2)} + \mathrm{i}\underbrace{(\cos\theta_1\sin\theta_2 + \cos\theta_2\sin\theta_1)}_{\sin(\theta_1+\theta_2)}\right).$$

Lecture 2 · 2025-10-11

For $n = 0$, we have $z^0 = 1$, which is true.

For $n \in \mathbb{Z}_{\geqslant 0}$, we shall prove by induction.

- **Base case**. This statement is true for $n = 0$.
- **Inductive step**. Let us assume $(\cos\theta + \mathrm{i}\sin\theta)^n = \cos(n\theta) + \mathrm{i}\sin(n\theta)$ for n. Then consider the case for $n+1$.

$$\begin{aligned}(\cos\theta + \mathrm{i}\sin\theta)^{n+1} &= (\cos\theta + \mathrm{i}\sin\theta)^n(\cos\theta + \mathrm{i}\sin\theta) \\ &= \cos(n\theta) + \mathrm{i}\sin(n\theta)(\cos\theta + \mathrm{i}\sin\theta) \\ \text{by lemma} \quad &= \cos((n+1)\theta) + \mathrm{i}\sin((n+1)\theta).\end{aligned}$$

For $n \in \mathbb{Z}^-$, we write $n = -m$ with $m > 0$. Thus

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

1.5 Exponential and Trigonometric Functions

1.5.1 Exponential Function

Definition 1.4

For $z \in \mathbb{C}$, we define

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

This definition converges $\forall z \in \mathbb{C}$. Some fundamental properties of this function are:

- $\forall z, w \in \mathbb{C}, e^z e^w = e^{z+w}$
- if $z \in \mathbb{R}$, then e^z reduces to the usual exponential for reals
- $e^0 = 1$
- $\forall z \in \mathbb{C}, n \in \mathbb{Z}, (e^z)^n = e^{nz}$

1.5.2 Trigonometric Functions

Definition 1.5 (Complex Trigonometric Functions)

For all $z \in \mathbb{C}$:

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

Analogously,

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

If $z \in \mathbb{R}$, then the definitions produce the analogous results to real numbers.

From these definitions, we can write, for all $z \in \mathbb{C}$,

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

In particular, [Euler's identity]

$$e^{i\pi} = -1.$$

If $x \in \mathbb{R}$,

$$\begin{aligned}
\operatorname{Re}\left(e^{ix}\right) &= \cos x \\
\operatorname{Im}\left(e^{ix}\right) &= \sin x.
\end{aligned}$$

Lemma 1.6

For all $n \in \mathbb{Z}$, $\mathrm{e}^z = 1 \Leftrightarrow z = 2\pi n\mathrm{i}$.

Proof. Write $z = x + \mathrm{i}y$. Then

$$\begin{aligned} \mathrm{e}^z &= \mathrm{e}^x \mathrm{e}^{\mathrm{i}y} \\ &= \mathrm{e}^x(\cos y + \mathrm{i}\sin y) \end{aligned}$$

[$\Rightarrow$] Assume that $\mathrm{e}^x(\cos y + \mathrm{i}\sin y) = 1$. Matching real and imaginary parts gives

$$\begin{cases} \sin y = 0 \Rightarrow y = n\pi & \text{for any } n \in \mathbb{Z} \\ \mathrm{e}^x \cos y = 1 \Rightarrow x = 0, y = 2n\pi & \text{for any } n \in \mathbb{Z} \end{cases}$$

Hence $\mathrm{e}^z = \mathrm{e}^{2n\pi\mathrm{i}}$.

[$\Leftarrow$] Assume that $z = 2\pi n\mathrm{i}$. Then evaluating $\mathrm{e}^{2\pi n\mathrm{i}}$ gives $\cos(2\pi n) + \mathrm{i}\sin(2\pi n) = 1$.

Finally, if $z = r(\cos\theta + \mathrm{i}\sin\theta) = r\mathrm{e}^{\mathrm{i}\theta}$ for $r > 0, \theta \in \mathbb{R}$, then de Moivre's Theorem is immediate from the results above.

1.6 Roots of Unity

Let $r = \mathrm{e}^{\mathrm{i}\theta}$. If for some $N \in \mathbb{N}$ we have $z^N = 1$, then

$$r^N \mathrm{e}^{\mathrm{i}\theta N} = r^N(\cos\theta N + \mathrm{i}\sin\theta N) = 1.$$

Hence $r^N = 1$ leads to $r = 1$.

Also, $\theta N = 2\pi n$ for some $n \in \mathbb{Z}$. Therefore $\theta = \frac{2\pi n}{N}$.

Therefore, we get

$$\begin{aligned} z &= \mathrm{e}^{\frac{2\pi n i}{N}} \quad \text{where } n = 0, 1, \ldots, N-1 \\ &= \omega^n \quad \text{where } \omega = \mathrm{e}^{\frac{2\pi i}{N}}. \end{aligned}$$

We call the roots to be the **roots of unity**.

1.7 Logarithm and Complex Powers

1.7.1 Logarithm

Definition 1.7 (Complex logarithm)

We define, for $z \in \mathbb{C}$ and $z \neq 0$,

$$\omega = \log z$$

such that

$$\mathrm{e}^\omega = z.$$

Hence we have

$$\ln\left(re^{i\theta}\right) = \ln r + i\theta.$$

Note that complex logarithm is multi-valued.

Definition 1.8 (Multivalued complex logarithm)

We write $\mathrm{Log}(z) = \{\log z + 2\pi ni : n \in \mathbb{Z}\}$ to represent the multivalued function.

Remark. To make the result unique, we can restrict the argument to $-\pi < \theta \leqslant \pi$.

1.7.2 Complex Powers

We can define, for $z \in \mathbb{C}$, $z \neq 0$, $\alpha \in \mathbb{C}$ that

$$z^\alpha = e^{\alpha \ln z}.$$

Note that this is multi-valued in general. However

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

gives the same value for $n \in \mathbb{Z}$.

Example 1.9

Consider $i = e^{\frac{i\pi}{2}}$.

Then $\arg(i) = \frac{\pi}{2}$ and $\mathrm{Arg}(i) = \left\{\frac{\pi}{2} + 2\pi n : n \in \mathbb{Z}\right\}$. Hence $\ln i = i\left(\frac{\pi}{2} + 2\pi n\right)$ for all $n \in \mathbb{Z}$.

1.8 Lines and Circles in the Complex Plane

1.8.1 Lines

Taking $z \in \mathbb{C}$ as a point on the line, and $\omega \in \mathbb{C}$ as the direction, then a line can be expressed as

$$z = z_0 + \lambda\omega \quad \lambda \in \mathbb{R}.$$

Taking conjugates, $\overline{z} = \overline{z_0} + \lambda\overline{\omega} \Rightarrow \overline{\omega}z - \omega\overline{z} = \overline{\omega}z_0 - \omega\overline{z_0}$.

1.8.2 Circles

For a center $c \in \mathbb{C}$ and radius $\rho > 0$, we can describe a circle as

$$\begin{aligned} & z = c + \rho e^{i\theta} \quad \theta \in \mathbb{R} \\ \Leftrightarrow \quad & |z - c| = \rho \\ \Leftrightarrow \quad & |z|^2 - c\overline{z} - \overline{c}z = p^2 - |c|^2. \end{aligned}$$

Lecture 3 · 2025-10-14

2 Vectors

A vector can be specified by a (positive) magnitude and a direction in space.

2.1 Introduction on Vectors

We can represent a vector as a line segment between two points A and B, and we write $\boldsymbol{v} = \overrightarrow{AB}$. The vector $\boldsymbol{v}$ has length $|\boldsymbol{v}|$ and direction from A to B.

If we choose O as the origin, then point A has position vector $\boldsymbol{a} = \overrightarrow{OA}$.

Definition 2.1 (Vector space over reals and complex numbers)

A **vector space** V over $\mathbb{C}$ or $\mathbb{R}$ is a set of abstract vectors $\{\boldsymbol{v}\}$ equipped with operations of

- vector addition $\oplus : V \times V \to V$, and
- scalar multiplication $\otimes : \mathbb{R} \times V \to V$

that satisfy the following axioms:

Vector addition axioms
1. **Commutivity**: $\boldsymbol{u} \oplus \boldsymbol{v} = \boldsymbol{v} \oplus \boldsymbol{u}$.
2. **Associtivity**: $(\boldsymbol{u} \oplus \boldsymbol{v}) \oplus \boldsymbol{w} = \boldsymbol{u} \oplus (\boldsymbol{v} \oplus \boldsymbol{w})$.
3. **Additive identity**: $\exists \mathbf{0} \in V$ such that $0 \oplus \boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$.
4. **Additive inverse**: $\forall \mathrm{v} \in V$, $\exists (-\mathrm{v}) \in V$ such that $\boldsymbol{v} \oplus (-\boldsymbol{v}) = \mathbf{0}$.

Scalar multiplication axioms
1. $\lambda \otimes (\boldsymbol{u} \oplus \mathrm{v}) = (\lambda \otimes \boldsymbol{u}) \oplus (\lambda \otimes \mathrm{v})$.
2. $(\lambda \oplus \mu) \otimes \boldsymbol{v} = (\lambda \otimes \mathrm{v}) \oplus (\mu \otimes \mathrm{v})$.
3. $\lambda \otimes (\mu \otimes \mathrm{v}) = (\lambda\mu) \otimes \boldsymbol{v}$.
4. $1 \otimes \boldsymbol{v} = \boldsymbol{v}$.

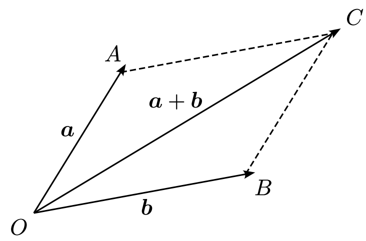
Notation. Usually, we omit the circles of $\oplus$ and $\otimes$, and write then as if they were $+$ and $\times$.

Remark.
- Vectors under $+$ form an Abelian group.

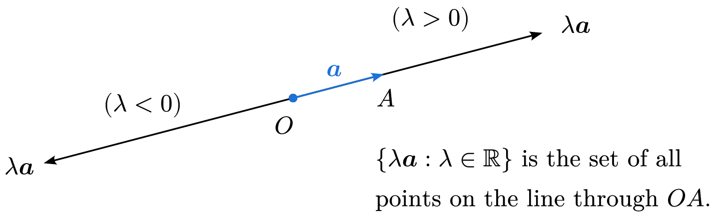
Example 2.2

In $\mathbb{R}^3$, we define the two operations as follows:

- **Vector addition.** Consider $\boldsymbol{a}$ and $\boldsymbol{b}$ the position vectors of two points A, B respectively. We can construct a parallelogram and then do compositions of two vectors, such that $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c}$.



- **Scalar multiplicaton.** Given $\boldsymbol{a}$ the position vector of a point A, and $\lambda \in \mathbb{R}$, $\lambda\boldsymbol{a}$ is a position vector of a point on line OA, with length $|\lambda\boldsymbol{a}| = |\lambda||\boldsymbol{a}|$, in the direction as shown follows.



Definition 2.3 (Unit vector)

A **unit vector** is a vector with length 1. We denote it as $\hat{\boldsymbol{v}}$.

Definition 2.4 (Linear combination)

Consider two vectors $\boldsymbol{a}, \boldsymbol{b}$ and scalars $\alpha, \beta \in \mathbb{R}$. Then

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b}$$

is a **linear combination** of $\boldsymbol{a}$ and $\boldsymbol{b}$.

In general, we denote all possible linear combinations of two given vectors $\boldsymbol{a}, \boldsymbol{b}$ by

$$\{\alpha\boldsymbol{a} + \beta\boldsymbol{b} : \alpha\beta \in \mathbb{R}\} = \operatorname{span}\{\boldsymbol{a}, \boldsymbol{b}\}.$$

This is called that **span** of $\{\boldsymbol{a}, \boldsymbol{b}\}$.

This extends to any number of vectors (possibly more than two).

Definition 2.5 (Parallel)

We say that $\boldsymbol{a}$ and $\boldsymbol{b}$ are **parallel**, denoted $\boldsymbol{a} \parallel \boldsymbol{b}$, if $\boldsymbol{a} = \lambda\boldsymbol{b}$ (or equivalently $\boldsymbol{b} = \lambda\boldsymbol{a}$) for some $\lambda \in \mathbb{R}$. We allow $\lambda = 0$, so $\boldsymbol{0} \parallel \boldsymbol{v}$ for any vector $\boldsymbol{v}$.

Remark. If $\boldsymbol{a} \nparallel \boldsymbol{b}$, then $\operatorname{span}\{\boldsymbol{a}, \boldsymbol{b}\} = \operatorname{span}\{\boldsymbol{a}, \boldsymbol{a} - \boldsymbol{b}\}$ is a plane through O, A, B.

2.2 Scalar Product (Dot Product)

Definition 2.6 (Scalar product in $\mathbb{R}^n$)

For two vectors $\boldsymbol{a}, \boldsymbol{b}$ in $\mathbb{R}^n$, and θ the solid angle between them. Then the **scalar** product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by:

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

Intuitively, this is the product of the parts in $\boldsymbol{a}$ and $\boldsymbol{b}$ which are parallel.

We have some interesting results on scalar products in general.

Proposition 2.7

If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are vectors and $\lambda \in \mathbb{R}$, we have

- $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$
- $\boldsymbol{a} \cdot \boldsymbol{a} = |\boldsymbol{a}|^2 \geqslant 0$, and $|\boldsymbol{a}| = 0$ if and only if $\boldsymbol{a} = \mathbf{0}$
- $(\lambda \boldsymbol{a}) \cdot \boldsymbol{b} = \lambda(\boldsymbol{a} \cdot \boldsymbol{b}) = \boldsymbol{a} \cdot (\lambda \boldsymbol{b})$
- $\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}$

Definition 2.8 (Perpendicularity)

Moreover, we say that $\boldsymbol{a}$ and $\boldsymbol{b}$ are **orthogonal** or **perpendicular** and denote it by $\boldsymbol{a} \perp \boldsymbol{b}$ if

$$\boldsymbol{a} \cdot \boldsymbol{b} = 0.$$

In this case, we allow for $\boldsymbol{a}$ or $\boldsymbol{b}$ to be $\mathbf{0}$.

Using the dot product we can write the projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ as

$$\hat{\boldsymbol{a}}|\boldsymbol{b}| \cos \theta = (\hat{\boldsymbol{a}} \cdot \boldsymbol{b})\hat{\boldsymbol{a}}.$$

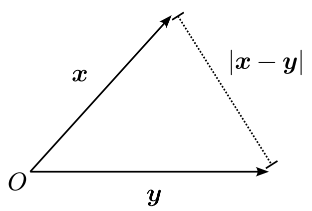
We can actually derive Definition 2.6.

Proposition 2.9

For $x, y \in \mathbb{R}^3$, define $\theta \in [0, \pi]$ to be the solid angle between them. Then

$$\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}|\, |\boldsymbol{y}| \cos \theta.$$

Proof.



For any $x, y \in \mathbb{R}^3$, we have

$$|\boldsymbol{x} - \boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 - 2|\boldsymbol{x}|\, |\boldsymbol{y}| \cos \theta \quad \text{by cosine rule.} \tag*{«1}$$

But from the definition of scalar product,

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

By comparing Equation 1 and Equation 2, we get $\mathbf{x}\cdot\mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\cos\theta$.

Definition 2.10 (Real inner product)

We say, for vector space V, a map $\langle\cdot,\cdot\rangle : V\times V \to \mathbb{R}$ is called an inner product if

1. $\langle\alpha\mathbf{x}+\beta\mathbf{y},\mathbf{z}\rangle = \alpha\langle\mathbf{x},\mathbf{z}\rangle + \beta\langle\mathbf{y},\mathbf{z}\rangle$.
2. $\langle\mathbf{x},\mathbf{y}\rangle = \langle\mathbf{y},\mathbf{x}\rangle$.
3. $\langle\mathbf{x},\mathbf{x}\rangle > 0$ for $\mathbf{x}\neq\mathbf{0}$.

Definition 2.11 (Norm)

Given the inner product on V, we define the norm $|\cdot| : V \to [0,\infty]$ to be $\|\mathbf{x}\| = \sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$.

Lecture 4 · 2025-10-16

We can now form an inequality that we will encounter various times in various forms in later courses, but here we shall see a simplest formation of it.

Theorem 2.12 (Cauchy-Schwarz inequality)

For all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, then

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

Proof. Consider the expression $|\mathbf{x}-\lambda\mathbf{y}|^2$, where $\lambda\in\mathbb{R}$. Then

$$\begin{aligned}|\mathbf{x}-\lambda\mathbf{y}|^2 &\geqslant 0 \\ (\mathbf{x}-\lambda\mathbf{y})\cdot(\mathbf{x}-\lambda\mathbf{y}) &\geqslant 0 \\ |\mathbf{x}|^2+\lambda^2|\mathbf{y}|^2-2\lambda\mathbf{x}\cdot\mathbf{y} &\geqslant 0 \\ \lambda^2|\mathbf{y}|^2-2\lambda\mathbf{x}\cdot\mathbf{y}+|\mathbf{x}|^2 &\geqslant 0 && \text{by rearranging as a quadratic of } \lambda \\ 4(\mathbf{x}\cdot\mathbf{y})^2-4|\mathbf{x}|^2|\mathbf{y}|^2 &\leqslant 0 && \text{by taking discriminant} \\ |\mathbf{x}\cdot\mathbf{y}| &\leqslant |\mathbf{x}||\mathbf{y}|.\end{aligned}$$

Here are some important observations for the Cauchy-Schwarz inequality.

Remark.

- This inequality holds for all scalar product in any real vector space.
- The equality holds if and only if $\mathbf{x}=\lambda\mathbf{y}$ or $\mathbf{y}=\lambda\mathbf{x}$ for some $\lambda\in\mathbb{R}$.
- Definition 2.6 is now well-defined since Cauchy-Schwarz ensures that $-1\leqslant\cos\theta\leqslant 1$.

Corollary 2.13 (Triangle inequality)

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$|\boldsymbol{x} + \boldsymbol{y}| \leqslant |\boldsymbol{x}| + |\boldsymbol{y}|.$$

Proof. We have

$$\begin{aligned} |\boldsymbol{x} + \boldsymbol{y}|^2 &= (\boldsymbol{x} + \boldsymbol{y}) \cdot (\boldsymbol{x} + \boldsymbol{y}) \\ &= |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 + 2\boldsymbol{x} \cdot \boldsymbol{y} \\ &\leqslant |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2 + 2|\boldsymbol{x}||\boldsymbol{y}| \\ &= (|\boldsymbol{x}| + |\boldsymbol{y}|)^2 \end{aligned}$$

The result then follows.

2.3 Orthonormal Bases

Definition 2.14 (Orthonormal)

Vectors are said to be **orthonormal** if they are orthogonal unit vectors.

Consider $\mathbb{R}^3$, and consider vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ that are orthonormal. Then we have

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad \text{for } i, j = 1, 2, 3.$$

This is equivalent to choosing Cartesian axes along these directions. We need a few extra definitions to describe this.

Definition 2.15 (Spanning)

For a vector space V, we say a subset $S = \left\{\boldsymbol{u}_1, ..., \boldsymbol{u}_p\right\}$ is a **spanning set** for V if each of $\boldsymbol{v} \in V$ can be written as a linear combination of the vectors in S.

Definition 2.16 (Linearly independent)

We say the set $T = \left\{\boldsymbol{v}_1, ..., \boldsymbol{v}_q\right\}$ is **linearly independent** if

$$\sum_{i=1}^{q} \lambda_i \boldsymbol{v}_i = 0 \Leftrightarrow \lambda_i = 0 \text{ for } i = 1, ..., q.$$

Definition 2.17 (Basis)

A set of vectors $B = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_2\}$ in V is called a **basis** if it is spanning and linearly independent.

Hence, $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is an orthonormal basis.

We can therefore denote $\boldsymbol{a}$ in the following ways:

$$a = (a_1, a_2, a_3) \quad \text{or} \quad a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Now, for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$, we have

$$\begin{aligned} \boldsymbol{a} \cdot \boldsymbol{b} &= (a_1 \boldsymbol{e}_1 + a_2 \boldsymbol{e}_2 + a_3 \boldsymbol{e}_3) \cdot (b_1 \boldsymbol{e}_1 + b_2 \boldsymbol{e}_2 + b_3 \boldsymbol{e}_3) \\ &= a_1 b_1 + a_2 b_2 + a_3 b_3. \end{aligned}$$

In particular, we can derive the Pythagorean rule, since

$$\boldsymbol{a} \cdot \boldsymbol{a} = |\boldsymbol{a}|^2 = a_1^2 + a_2^2 + a_3^2.$$

For the canonical basis of $\mathbb{R}^3$, the one that we use for the representation in terms of row or column vector,

$$\boldsymbol{e}_1 = (1, 0, 0) \quad \boldsymbol{e}_2 = (0, 1, 0) \quad \boldsymbol{e}_3 = (0, 0, 1),$$

we can represent the vectors by

$$\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}$$

respectively.

2.4 Vector Product (Cross Product) in $\mathbb{R}^3$

Definition 2.18 (Vector Product in $\mathbb{R}^3$)

Consider $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$. Their vector product is defined by

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

where $\hat{\boldsymbol{n}}$ is a unit vector that is perpendicular to both $\boldsymbol{a}$ and $\boldsymbol{b}$, and $(\boldsymbol{a}, \boldsymbol{b}, \hat{\boldsymbol{n}})$ is right-handed.

Remark.

1. If we change θ to $2\pi - \theta$, we obtain $-\hat{\boldsymbol{n}}$ in the definition of $\boldsymbol{a} \times \boldsymbol{b}$ instead.
2. $\hat{\boldsymbol{n}}$ is not defined if $\boldsymbol{a} \parallel \boldsymbol{b}$. However, we immediately have $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0}$.
3. θ is not defined if $|\boldsymbol{a}| = 0$ or $|\boldsymbol{b}| = 0$.

Notation. $\boldsymbol{a} \wedge \boldsymbol{b} \equiv \boldsymbol{a} \times \boldsymbol{b}$ for vector product.

Proposition 2.19 (Properties of vector product)

If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are vectors in $\mathbb{R}^3$, then we have

1. $\boldsymbol{a} \times \boldsymbol{b} = -\boldsymbol{b} \times \boldsymbol{a}$.
2. $\boldsymbol{a} \times \boldsymbol{a} = \boldsymbol{0}$.
3. $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0} \Leftrightarrow \boldsymbol{a} = \lambda \boldsymbol{b}$ for some $\lambda \in \mathbb{R}$, or either vector is the zero vector.
4. $(\lambda \boldsymbol{a}) \times \boldsymbol{b} = \lambda(\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{a} \times (\lambda \boldsymbol{b})$.

5. $\boldsymbol{a} \times (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \times \boldsymbol{b} + \boldsymbol{a} \times \boldsymbol{c}$.
6. $\boldsymbol{a} \cdot (\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{b} \cdot (\boldsymbol{a} \times \boldsymbol{b}) = 0$.

Proposition 2.20 (Geometric interpretations of vector product)

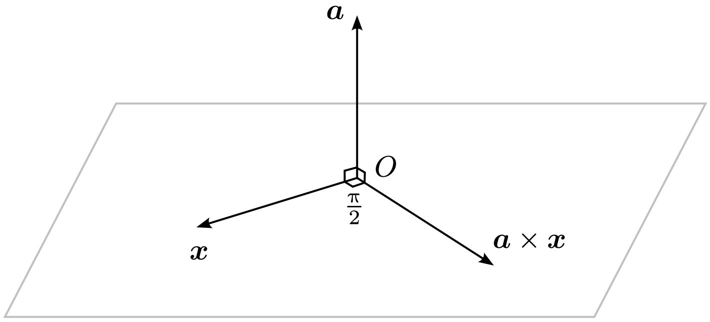
For two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$, then $|\boldsymbol{a} \times \boldsymbol{b}|$ is the area of the parallelogram formed by $\boldsymbol{a}$ and $\boldsymbol{b}$.

If $\boldsymbol{a} = \overrightarrow{OA}$ and $\boldsymbol{b} = \overrightarrow{OB}$, then the area of the triangle OAB is given by $\frac{1}{2}|\boldsymbol{a} \times \boldsymbol{b}|$.

Lecture 5 · 2025-10-18

Proposition 2.21 (Alternative geometric interpretations of vector product)

Fix a vector $\boldsymbol{a}$ and consider $\boldsymbol{x} \perp \boldsymbol{a}$. Then, computing $\boldsymbol{a} \times \boldsymbol{x}$ gives a vector that scales $\boldsymbol{x}$ by $|\boldsymbol{a}|$ and rotates it by $\frac{\pi}{2}$ in a plane that is orthogonal to $\boldsymbol{a}$.



2.4.1 Component Expressions

Let $\boldsymbol{e}_1 = (1, 0, 0), \boldsymbol{e}_2 = (0, 1, 0), \boldsymbol{e}_3 = (0, 0, 1)$. Then

$$\begin{aligned} \boldsymbol{e}_1 \times \boldsymbol{e}_2 &= \boldsymbol{e}_3 = -\boldsymbol{e}_2 \times \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \times \boldsymbol{e}_3 &= \boldsymbol{e}_1 = -\boldsymbol{e}_3 \times \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \times \boldsymbol{e}_1 &= \boldsymbol{e}_2 = -\boldsymbol{e}_1 \times \boldsymbol{e}_3 \end{aligned}$$

Consider $\boldsymbol{a} = (a_1, a_2, a_3)$ and $\boldsymbol{b} = (b_1, b_2, b_3)$. We have

$$\begin{aligned} \boldsymbol{a} \times \boldsymbol{b} = (&a_2 b_3 - a_3 b_2)\boldsymbol{e}_1 \\ &+ (a_3 b_1 - a_2 b_3)\boldsymbol{e}_2 \\ &+ (a_1 b_2 - a_2 b_1)\boldsymbol{e}_3. \end{aligned}$$

This is also equivalent to

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

2.5 Triple Products

2.5.1 Scalar Triple Product

Definition 2.22 (Scalar triple product)

Consider $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$. We write

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

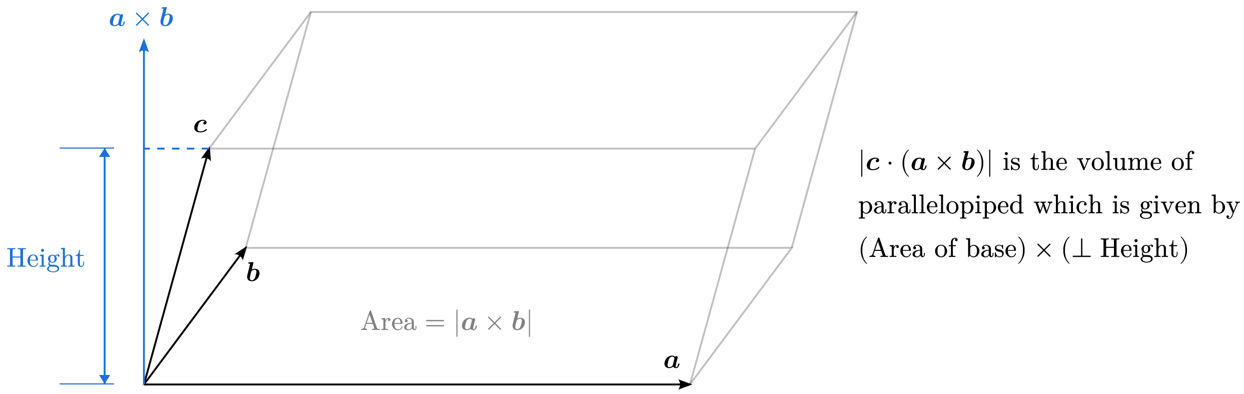
to be the **scalar triple product** between $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$.

Proposition 2.23

For $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$, we have

$$\begin{aligned} \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b} \cdot (\boldsymbol{c} \times \boldsymbol{a}) = \boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b}) \\ = -\boldsymbol{a} \cdot (\boldsymbol{c} \times \boldsymbol{b}) = -\boldsymbol{b} \cdot (\boldsymbol{a} \times \boldsymbol{c}) = -\boldsymbol{c} \cdot (\boldsymbol{b} \times \boldsymbol{a}). \end{aligned}$$

We can interpret Proposition 2.23 using a parallelpiped.



Note that $\boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b}) = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})$ is a signed volume:

- If $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) > 0$, then $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ constitute a right-handed set.
- $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = 0$ iff $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are coplanar. *i.e.* one of the them is a linear combination of the other two.

2.5.2 Vector Triple Product

Definition 2.24 (Vector triple product)

Consider $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$. We call

$$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c})$$

to be the **scalar triple product** between $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$.

Proposition 2.25

For $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$, we have

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

Note that the vector triple product is not associative. This is because

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

but

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

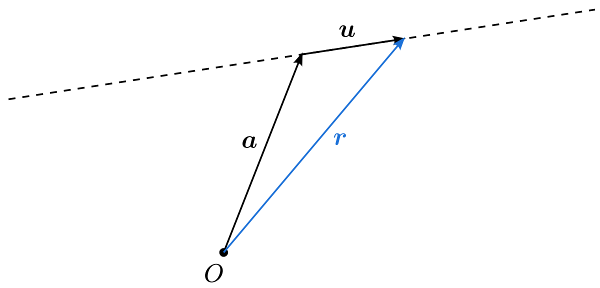
2.6 Lines, Planes and Vector Equations

2.6.1 Lines

Proposition 2.26 (Parametric form of a line)

Any point on a line through $\boldsymbol{a}$ with direction $\boldsymbol{u} \neq 0$ has position vector $\boldsymbol{r}$ given by

$$\boldsymbol{r} = \boldsymbol{a} + \lambda \boldsymbol{u}. \quad (\lambda \in \mathbb{R})$$



This form is equivalent to

$$\boldsymbol{u} \times (\boldsymbol{c} - \boldsymbol{a}) = \boldsymbol{0} \Leftrightarrow \boldsymbol{u} \times \boldsymbol{r} = \boldsymbol{b}$$

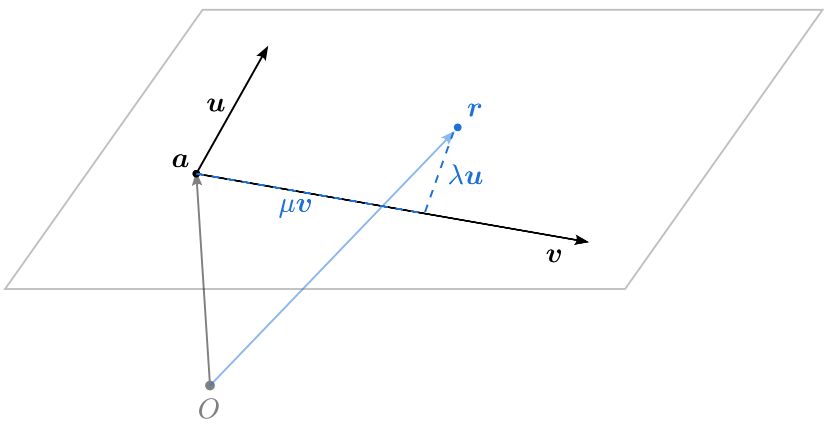
where $\boldsymbol{b}$ is a constant vector.

2.6.2 Planes

Proposition 2.27 (Parametric form of a plane)

Any point on a plane through $\boldsymbol{a}$ can be described using directions $\boldsymbol{u}, \boldsymbol{v}$ where $\boldsymbol{u} \nparallel \boldsymbol{v}$, with the position vector

$$\boldsymbol{r} = \boldsymbol{a} + \lambda \boldsymbol{u} + \mu \boldsymbol{v}. \quad (\lambda, \mu \in \mathbb{R})$$



The normal vector to the plane $\boldsymbol{r} = \boldsymbol{a} + \lambda \boldsymbol{u} + \mu \boldsymbol{v}$ is

$$\boldsymbol{n} = \boldsymbol{u} \times \boldsymbol{v}.$$

This normal vector is not a unit vector in general.

Then, we can write

$$\boldsymbol{r} \cdot \boldsymbol{n} = \underbrace{\boldsymbol{a} \cdot \boldsymbol{n}}_{k = \text{ constant}} \Leftrightarrow (\boldsymbol{r} - \boldsymbol{a}) \cdot \boldsymbol{n} = 0$$

The component of $\boldsymbol{r}$ along $\boldsymbol{n}$ is

$$\hat{\boldsymbol{n}} \cdot \boldsymbol{r} = \frac{\boldsymbol{n} \cdot \boldsymbol{r}}{|\boldsymbol{n}|} = \frac{k}{|\boldsymbol{n}|}$$

and $\frac{k}{|\boldsymbol{n}|}$ is the perpendicular distance from the origin to the plane.

Remark. If $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ lie in the plane, then we can write the equation of the plane by

$$(\boldsymbol{r} - \boldsymbol{a}) \cdot [(\boldsymbol{b} - \boldsymbol{a}) \times (\boldsymbol{c} - \boldsymbol{a})] = 0$$

Example 2.28 (Intersection of a line and a plane)

Consider the point of intersection between

$$\begin{aligned} \text{Line:} \quad & \boldsymbol{u} \times \boldsymbol{r} = \boldsymbol{u} \times \boldsymbol{a} \quad \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3 \\ \text{Plane:} \quad & \boldsymbol{n} \cdot \boldsymbol{r} = \boldsymbol{n} \cdot \boldsymbol{b}. \end{aligned}$$

The line equation can be re-written as $\boldsymbol{r} \times \boldsymbol{u} = \boldsymbol{a} \times \boldsymbol{u}$. Taking vector product of this with $\boldsymbol{n}$ gives

$$(\boldsymbol{r} \times \boldsymbol{u}) \times \boldsymbol{n} = (\boldsymbol{a} \times \boldsymbol{u}) \times \boldsymbol{n}$$

Applying vector triple product property in Proposition 2.25 gives

$$\begin{aligned} (\boldsymbol{r} \times \boldsymbol{u}) \times \boldsymbol{n} &= (\boldsymbol{r} \cdot \boldsymbol{n})\boldsymbol{u} - (\boldsymbol{u} \cdot \boldsymbol{b})\boldsymbol{r} \\ &= (\boldsymbol{b} \cdot \boldsymbol{n})\boldsymbol{u} - (\boldsymbol{u} \cdot \boldsymbol{n})\boldsymbol{r}. \end{aligned}$$

Hence

$$(\boldsymbol{u} \cdot \boldsymbol{n})\boldsymbol{r} = (\boldsymbol{b} \cdot \boldsymbol{n})\boldsymbol{u} - (\boldsymbol{a} \times \boldsymbol{u}) \times \boldsymbol{n}.$$

If $\boldsymbol{u} \cdot \boldsymbol{n} \neq 0$, then we can compute

$$\boldsymbol{r} = \frac{(\boldsymbol{b} \cdot \boldsymbol{n})\boldsymbol{u} - (\boldsymbol{a} \times \boldsymbol{u}) \times \boldsymbol{n}}{\boldsymbol{u} \cdot \boldsymbol{n}}$$

as the position vector of the point of intersection.

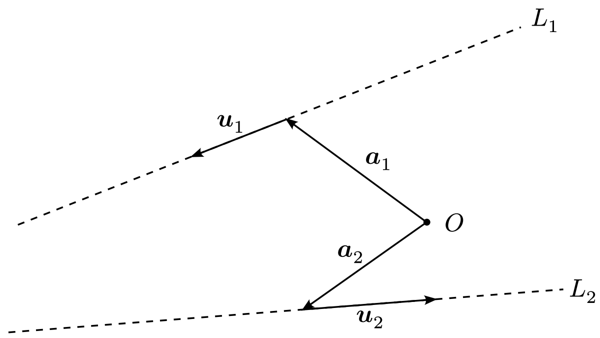
Otherwise, if $\boldsymbol{u} \cdot \boldsymbol{n} = 0$, $\boldsymbol{u}$ is orthogonal to $\boldsymbol{n}$. So either
- the line is parallel to the plane and never intersects the plane, or
- the line is contained within the plane.

Lecture 6 · 2025-10-21

Example 2.29 (Shortest distance between two lines)

Consider two lines

$$L_1 : \boldsymbol{u}_1 \times (\boldsymbol{r} - \boldsymbol{a}_1) = \boldsymbol{0}$$
$$L_2 : \boldsymbol{u}_2 \times (\boldsymbol{r} - \boldsymbol{a}_2) = \boldsymbol{0}.$$



Then, the shortest distance between L_1 and L_2 is attained at a line perpendicular to both lines, with direction $\boldsymbol{u}_1 \times \boldsymbol{u}_2$.

The shortest distance s is then computed by projecting the vector $\boldsymbol{a}_2 - \boldsymbol{a}_1$ onto the unit vector in the direction of $\boldsymbol{u}_1 \times \boldsymbol{u}_2$, giving

$$s = \left| (\boldsymbol{a}_1 - \boldsymbol{a}_2) \cdot \frac{\boldsymbol{u}_1 \times \boldsymbol{u}_2}{|\boldsymbol{u}_1 \times \boldsymbol{u}_2|} \right|.$$

2.6.3 Spheres

A sphere in $\mathbb{R}^3$ with centre $\boldsymbol{0}$ and radius $r \in \mathbb{R}$ is given by

$$\Sigma = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| = r, r > 0 \right\}.$$

In general, in $\mathbb{R}^n$, a hypersphere with center $\boldsymbol{a} \in \mathbb{R}^n$ and radius $r \in \mathbb{R}$ is given by

$$\Sigma = \{ \boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{x} - \boldsymbol{a}| = r, r > 0 \}.$$

2.6.4 Vector Equations

Our goal is to solve equations of the form

$$r + a \times (b \times r) = c \qquad «3$$

for r, where a, b, c are known vectors.

Using the vector triple product identity in Proposition 2.25, we have

$$a \times (b \times r) = (a \cdot r)b - (a \cdot b)r,$$

so that Equation 3 becomes

$$r + (a \cdot r)b - (a \cdot b)r = c \qquad «4$$

Taking the dot product of both sides of Equation 4 with a gives

$$a \cdot r + (a \cdot r)(a \cdot b) - (a \cdot b)(a \cdot r) = a \cdot c$$

so we obtain

$$a \cdot r = a \cdot c.$$

Hence, substituting back into Equation 4 gives

$$\begin{aligned} r + (a \cdot c)b - (a \cdot b)r &= c \\ r(1 - (a \cdot b)) &= c - (a \cdot c)b. \end{aligned} \qquad «5$$

- If $a \cdot b \neq 1$, the there is a unique solution given by

$$r = \frac{c - (a \cdot c)b}{1 - (a \cdot b)}.$$

- If $a \cdot b = 1$, then by Equation 5, either
  - there is no solution if $c - (a \cdot c)b \neq 0$, or
  - there are infinitely many solutions if $c - (a \cdot c)b = 0$. The set of solutions is given by our derived condition

$$a \cdot r = a \cdot c,$$

  which represents a plane.

2.7 Index Notation & Summation Conventions

Consider an orthonormal right-handed basis $\{e_1, e_2, e_3\}$. We write vectors a, b, *etc.* in terms of coordinates in this basis.

From now on, we will use indices i, j, k that take values $1, 2, 3$.

Definition 2.30 (Kronecker delta)

The **Kronecker delta** δ_{ij} is defined as

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

Proposition 2.31 (Properties of Kronecker delta)

- It is symmetric: $\delta_{ij} = \delta_{ji}$. Note that we can write

$$e_i \cdot e_j = \delta_{ij}.$$

- For vectors $\boldsymbol{a}, \boldsymbol{b}$, we can write

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{3} a_i b_i = \sum_{i,j=1}^{3} \delta_{ij} a_i b_j.$$

Definition 2.32 (Levi-Civita epsilon)

The **Levi-Civita epsilon** ε_{ijk} is defined as

$$\varepsilon_{ijk} = \begin{cases} 1 & (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & (i,j,k) \text{ is an odd permutation of } (1,2,3) \\ 0 & \text{if any two indices are equal.} \end{cases}.$$

This is to say, that

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$$
$$\varepsilon_{321} = \varepsilon_{132} = \varepsilon_{213} = -1$$

and all other combinations are zero.

Proposition 2.33 (Properties of Levi-Civita epsilon)

- It is antisymmetric. We can write

$$\boldsymbol{e}_i \times \boldsymbol{e}_j = \sum_{k=1}^{3} \varepsilon_{ijk} \boldsymbol{e}_k.$$

- For vectors $\boldsymbol{a}, \boldsymbol{b}$, we can write

$$\boldsymbol{a} \times \boldsymbol{b} = \sum_{i,j,k=1}^{3} \varepsilon_{ijk} a_i b_j \boldsymbol{e}_k.$$

2.7.1 Einstein Summation Convention

Now, we can use a more efficient notation.

Definition 2.34 (Einstein summation convention)

In index notation, an index variable that appears twice in an expression are normally summed. To simplify notation, we omit the summation sign for repeated indices and sum over them. This is called the **Einstein summation convention**.

This notation follows the following rules:

- If an index appears only once in an expression, it is a free index, so it must appear in every term of the equation, and can take any value. [We are not summing over it.]

- If an index appears twice in a term, it is a contracted index, and we sum over all its possible values. [We are summing over it.]
- No index can appear more than twice in a term.

Example 2.35

Using Einstein summation convention, we can write

- $a_i \delta_{ij} = a_j$ (which means $\sum_{i=1}^{3} a_i \delta_{ij} = a_j$)
- $\boldsymbol{a} \cdot \boldsymbol{b} = \delta_{ij} a_i b_j = a_i b_i$
- $(\boldsymbol{a} \times \boldsymbol{b})_i = \varepsilon_{ijk} a_j b_k$
- $\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \varepsilon_{ijk} a_i b_j c_k$
- $\delta_{ii} = 3$

Proposition 2.36 (Important identities involving delta and epsilon)

For indices i, j, k, l taking values $1, 2, 3$, we have

1. $\varepsilon_{ijk}\varepsilon_{pqr} = \delta_{ip}\delta_{jq}\delta_{kr} - \delta_{jp}\delta_{iq}\delta_{kr} + \delta_{ji}\delta_{kq}\delta_{ir} - \delta_{kp}\delta_{jq}\delta_{ir} + \delta_{kp}\delta_{iq}\delta_{jr} - \delta_{ip}\delta_{kq}\delta_{jk}$.
2. $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$
3. $\varepsilon_{ijk}\varepsilon_{pjk} = 2\delta_{ip}$
4. $\varepsilon_{ijk}\varepsilon_{ijk} = 6$

2.7.2 Proofs Using Index Notation

We can now use index notation to prove the vector triple product identity.

Example 2.37 (Proof of vector triple product identity)

We want to show that for $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$,

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

Proof. Using index notation, the ith component of the left-hand side is

$$\begin{aligned}
(\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}))_i &= \varepsilon_{ijk} a_j (\boldsymbol{b} \times \boldsymbol{c})_k \\
&= \varepsilon_{ijk} a_j \varepsilon_{kpq} b_p c_q \\
&= \left(\varepsilon_{ijk}\varepsilon_{kpq}\right) a_j b_p c_q \\
&= \left(\varepsilon_{ijk}\varepsilon_{pqk}\right) a_j b_p c_q \\
&= \left(\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}\right) a_j b_p c_q \quad \text{so } i = p \text{ in the first term, and } j = q \text{ in the second} \\
&= a_j c_j b_i - a_j b_j c_i \\
&= (\boldsymbol{a} \cdot \boldsymbol{c}) b_i - (\boldsymbol{a} \cdot \boldsymbol{b}) c_i.
\end{aligned}$$

> This is precisely the ith component of the right-hand side.

Lecture 7 · 2025-10-23

2.7.3 Spherical Trigonometry

With index notation, we can also consider **spherical trigonometry**.

Proposition 2.38

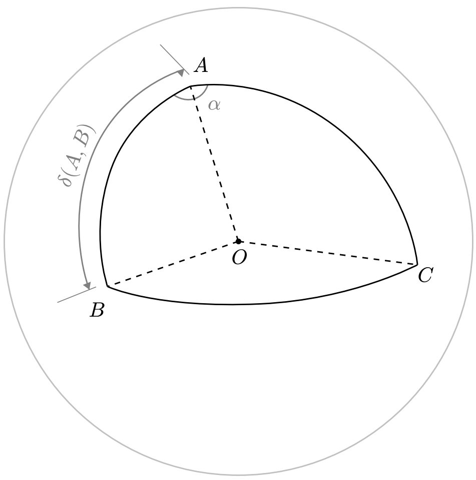
For $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$, then

$$(\boldsymbol{a} \times \boldsymbol{b}) \cdot (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{b})(\boldsymbol{b} \cdot \boldsymbol{c}) - (\boldsymbol{a} \cdot \boldsymbol{c})|\boldsymbol{b}|^2.$$

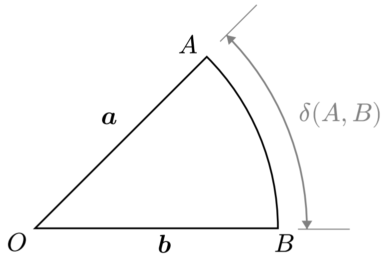
Proof.

$$\begin{aligned}
\text{LHS} &= (\boldsymbol{a} \times \boldsymbol{b})_i \cdot (\boldsymbol{b} \times \boldsymbol{c})_i \\
&= \varepsilon_{ijk} a_j b_k \varepsilon_{ipq} b_p c_q \\
&= \left(\varepsilon_{ijk} \varepsilon_{ipq}\right) a_j b_k b_p c_q \\
&= \left(\varepsilon_{ijk} \varepsilon_{pqi}\right) a_j b_k b_p c_q \\
&= \left(\delta_{jp} \delta_{kq} - \delta_{jq} \delta_{kp}\right) a_j b_k b_p c_q \\
&= a_j b_j b_k c_k - a_j b_k b_k c_j \\
&= (\boldsymbol{a} \cdot \boldsymbol{b})(\boldsymbol{b} \cdot \boldsymbol{c}) - (\boldsymbol{a} \cdot \boldsymbol{c})|\boldsymbol{b}|^2.
\end{aligned}$$

Now consider a unit sphere in $\mathbb{R}^3$ with centre $\mathbf{O}$, and points A, B, C on the surface of the sphere with position vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ respectively.



The distance from A to B, $\delta(A, B)$, is an arc length on the sphere.



$$\begin{aligned}\angle AOB &= \delta(A,B)\\ \boldsymbol{a}\cdot\boldsymbol{b} &= \cos\angle AOB\\ &= \cos\delta(A,B)\end{aligned}$$

In the same way, $|\boldsymbol{a}\times\boldsymbol{b}| = \sin\delta(A,B)$.

Hence, we have

$$\begin{aligned}\cos\alpha &= \frac{(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{a}\times\boldsymbol{c})}{|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{a}\times\boldsymbol{c}|}\\ &= -\frac{(\boldsymbol{b}\times\boldsymbol{a})\cdot(\boldsymbol{a}\times\boldsymbol{c})}{|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{a}\times\boldsymbol{c}|}\\ &= \frac{(\boldsymbol{b}\cdot\boldsymbol{c})|\boldsymbol{a}|^2-(\boldsymbol{b}\cdot\boldsymbol{a})(\boldsymbol{a}\cdot\boldsymbol{c})}{|\boldsymbol{a}\times\boldsymbol{b}||\boldsymbol{a}\times\boldsymbol{c}|}.\end{aligned}$$

$$\cos\alpha\sin\delta(A,B)\sin\delta(A,C) = \cos\delta(B,A) - \cos\delta(B,A)\cos\delta(A,C).$$

Which is the cosine rule for spherical triangles.

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

We define the following operations for vectors in $\mathbb{R}^n$.

Definition 2.39 (Addition and Scalar Multiplication in $\mathbb{R}^n$)

Addition. For $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n$, we define

$$\boldsymbol{a}+\boldsymbol{b} = (a_1+b_1, a_2+b_2, \dots, a_n+b_n).$$

Scalar Multiplication. For $\boldsymbol{a}\in\mathbb{R}^n$ and $\lambda\in\mathbb{R}$, we define

$$\lambda\boldsymbol{a} = (\lambda a_1, \lambda a_2, \dots, \lambda a_n).$$

Any $\boldsymbol{x}\in\mathbb{R}^n$ can be written as

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i\boldsymbol{e}_i$$

where $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ is the standard basis for $\mathbb{R}^n$ with 1 in the ith position and 0 elsewhere for $\boldsymbol{e}_i$.

Definition 2.40 (Dot Product in $\mathbb{R}^n$)

For $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}^n$, we define their dot product to be

$$\boldsymbol{a}\cdot\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i.$$

Proposition 2.41

$$e_i \cdot e_j = \delta_{ij}.$$

Hence, the components of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ can be determined by

$$x_i = \boldsymbol{x} \cdot \boldsymbol{e}_i.$$

Notation. If we write vectors in $\mathbb{R}^n$ as columns, then for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{x}^\top$ and $\boldsymbol{y}^\top$ denote their transposes, and that their inner product can be written as

$$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y}.$$

2.8.1 Summation Convention

We have

$$\boldsymbol{x} \cdot \boldsymbol{y} = \delta_{ij} x_i y_j = x_i y_i.$$

We define $\varepsilon_{\underbrace{i,j,\ldots,l}_{n \text{ indices}}}$ to be the extension of the Levi-Civita epsilon (Definition 2.32) to n dimensions.

In $\mathbb{R}^2$, it can be used to define an additional scalar product:

$$[\boldsymbol{a}, \boldsymbol{b}] = \varepsilon_{ij} a_i b_j = a_1 b_2 - a_2 b_1$$

Geometrically, this represents the signed area of the parallelogram formed by $\boldsymbol{a}$ and $\boldsymbol{b}$.

Remark. One can compare this to $[\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}]$, which represents the signed volume of the parallelepiped formed by $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ in $\mathbb{R}^3$.

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

We define the following operations for vectors in $\mathbb{C}^n$.

Definition 2.42 (Addition and Scalar Multiplication in $\mathbb{C}^n$)

Addition. For $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$, we define

$$\boldsymbol{z} + \boldsymbol{w} = (z_1 + w_1, z_2 + w_2, \ldots, z_n + w_n).$$

Scalar Multiplication. For $\boldsymbol{z} \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$, we define

$$\lambda \boldsymbol{z} = (\lambda z_1, \lambda z_2, \ldots, \lambda z_n).$$

- If $\lambda \in \mathbb{R}$, then $\mathbb{C}_n$ is a real vector space.
- If $\lambda \in \mathbb{C}$, then $\mathbb{C}_n$ is a complex vector space.

For any $\boldsymbol{z} \in \mathbb{C}^n$, we have

$$z_j = x_j + \mathrm{i} y_j.$$

If we are only allowing real scalars, then we can write

$$z = \sum_j \left(x_j + \mathrm{i}y_j\right)\boldsymbol{e}_j = \sum_j x_j\boldsymbol{e}_j + \sum_j y_j\boldsymbol{f}_j = \boldsymbol{x} + \mathrm{i}\boldsymbol{y}.$$

where $\boldsymbol{f}_j$ is defined to be the vector with i in the jth position of the imaginary part and 0 elsewhere.

Note that $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n, \boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_n\}$ forms a basis for $\mathbb{C}^n$ as a real vector space, with dimension $2n$.

If we allow complex scalars, then we can define

$$\boldsymbol{f}_j = \mathrm{i}\boldsymbol{e}_j,$$

and thus $\boldsymbol{z} = \sum_j z_j\boldsymbol{e}_j$. Hence $\mathbb{C}^n$ is a complex vector space with dimension n. Note that $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ forms a basis for $\mathbb{C}^n$ as a complex vector space, with dimension n.

2.9.1 Inner Product in $\mathbb{C}^n$

Definition 2.43 (Inner Product in $\mathbb{C}^n$)

For $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$, we define their inner product to be

$$(\boldsymbol{z}, \boldsymbol{w}) = \sum_{i=1}^{n} \overline{z_i} w_i.$$

Lecture 8 · 2025-10-25

Remark. This definition, including a complex conjugate, allows us to proceed with a definition for the norm.

2.9.1.1 Properties of the Inner Product

Proposition 2.44 (Properties of the inner product in $\mathbb{C}^n$)

1. **Hermitianity.** $(\boldsymbol{z}, \boldsymbol{w}) = \overline{(\boldsymbol{w}, \boldsymbol{z})}$.
2. **Linearity and anti-linearity.** $\forall \boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n, \forall \mu, \mu', \lambda, \lambda' \in \mathbb{C}$,
   - $(\boldsymbol{z}, \lambda\boldsymbol{w}' + \lambda'\boldsymbol{w}) = \lambda(\boldsymbol{z}, \boldsymbol{w}) + \lambda'(\boldsymbol{z}, \boldsymbol{w}')$
   - $(\mu\boldsymbol{z} + \mu'\boldsymbol{z}', \boldsymbol{w}) = \overline{\mu}(\boldsymbol{z}, \boldsymbol{w}) + \overline{\mu'}(\boldsymbol{z}', \boldsymbol{w})$
3. **Positive definite.** $(\boldsymbol{z}, \boldsymbol{z}) = \sum_j \left|z_j\right|^2 \geqslant 0$. Equality holds iff $\boldsymbol{z} = \boldsymbol{0}$.

Definition 2.45 (Norm in $\mathbb{C}^n$)

For $\boldsymbol{z} \in \mathbb{C}^n$, we define its norm to be

$$|\boldsymbol{z}|^2 = (\boldsymbol{z}, \boldsymbol{z}) = \sum_{i=1}^{n} |z_i|^2.$$

Definition 2.46 (Orthogonality in $\mathbb{C}^n$)

We say that $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$ are **orthogonal** if

$$(\boldsymbol{z}, \boldsymbol{w}) = 0.$$

Remark. The standard basis for $\mathbb{C}^n$ is orthonormal, and

$$\left(\boldsymbol{e}_i, \boldsymbol{e}_j\right) = \delta_{ij}.$$

2.9.1.2 From Complex to Real Inner Products

For $n = 1$, take $z, w \in \mathbb{C}$, then

$$(z, w) = \bar{z}w.$$

Now, write $z = a_1 + \mathrm{i}a_2$ and $w = b_1 + \mathrm{i}b_2$ where $a_1, a_2, b_1, b_2 \in \mathbb{R}$. Then we can identify z and w as vectors in $\mathbb{R}^2$, with $\boldsymbol{a} = (a_1, a_2)$ and $\boldsymbol{b} = (b_1, b_2)$ respectively.

Then,

$$\bar{z}w = \boldsymbol{a} \cdot \boldsymbol{b} + \mathrm{i}[\boldsymbol{a}, \boldsymbol{b}]$$

where $[\boldsymbol{a}, \boldsymbol{b}] = a_1 b_2 - a_2 b_1$ is product defined in Section 2.8.1, recovers both scalar products in $\mathbb{R}^2$.

2.10 General Vector Spaces

Definition 2.47 (Vector space)

A vector space V is a collection of vectors with two operations defined on them: vector addition and scalar multiplication, which satisfies the axioms in Definition 2.1.

- If the scalar field is $\mathbb{R}$, then V is a **real vector space**.
- If the scalar field is $\mathbb{C}$, then V is a **complex vector space**.

Consider a real vector space V, and consider $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r \in V$, we can write a linear combination:

$$\lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + \ldots + \lambda_r \boldsymbol{v}_r \in V$$

for any $\lambda_1, \lambda_2, \ldots, \lambda_r \in \mathbb{R}$.

Definition 2.48 (Span)

The **span** of $\{\boldsymbol{v}_1, \ldots \boldsymbol{v}_r\}$ is defined as

$$\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} = \left\{\sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i : \lambda_i \in \mathbb{R}\right\}.$$

Definition 2.49 (Subspace)

A **subspace** U of a vector space V is a subset of V that is also a vector space under the same operations of addition and scalar multiplication defined on V.

Equivalently, a non-empty subset $U \subseteq V$ is a subspace if it satisfies that for every $\boldsymbol{u}, \boldsymbol{v} \in U$ and $\lambda \in \mathbb{R}$, we have $\lambda \boldsymbol{v} + \mu \boldsymbol{u} \in U$.

In particular, for any $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in V$,

$$\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$$

is a subspace of V.

> ***Remark.*** The two trivial subspaces of any vector space V are $\{\mathbf{0}\}$ and V itself.

2.10.1 Linear Independence and Dependence

Definition 2.50 (Linear independence and dependence)

Consider a vector space V, and vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in V$. Consider a linear combination of these vectors:

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \dots + \lambda_r \mathbf{v}_r \quad \lambda_1, \dots, \lambda_r \in \mathbb{R} \text{ or } \mathbb{C}.$$

If $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \dots + \lambda_r \mathbf{v}_r = \mathbf{0}$ implies $\lambda_1 = \lambda_2 = \dots = \lambda_r = 0$, then the vectors are **linearly independent**.

If there exists $\lambda_1, \lambda_2, \dots, \lambda_r$, not all zero, such that $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \dots + \lambda_r \mathbf{v}_r = \mathbf{0}$, then the vectors are **linearly dependent**.

Remark.

- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ is linearly dependent iff one of the vectors can be expressed as a linear combination of the others.
- In $\mathbb{R}^3$, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are linearly independent iff

  $$\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) \neq 0.$$

  This can be geomtrically interpreted as the vectors not being coplanar [the LHS represents the volume of the parallelepiped spanned by the vectors].

Example 2.51

1. $\left\{\binom{1}{0}, \binom{0}{1}, \binom{0}{2}\right\}$ in $\mathbb{R}^2$ is linearly dependent, noting that $2\binom{0}{1} = \binom{0}{2}$.
2. $\left\{\binom{1}{0}, \binom{0}{1}\right\}$ in $\mathbb{R}^2$ is linearly independent.
3. Any set containing $\mathbf{0}$ is linearly dependent.

2.10.2 Inner Products

Definition 2.52 (Inner product)

An **inner product** on a vector space V is a function that assigns to each pair of vectors $\boldsymbol{v}, \boldsymbol{w} \in V$ a scalar $(\boldsymbol{v}, \boldsymbol{w}) \in \mathbb{R}$ or $\mathbb{C}$, satisfying

1. **Hermitianity**: $(\mathbf{v},\mathbf{w}) = \overline{(\mathbf{w},\mathbf{v})}$.
2. **Linearity and anti-linearity**: $\forall \mathbf{u},\mathbf{v},\mathbf{w} \in V, \forall \mu,\mu',\lambda,\lambda' \in \mathbb{C}$ or $\mathbb{R}$,
   - $(\mathbf{v}, \lambda\mathbf{w} + \lambda'\mathbf{w}') = \lambda(\mathbf{v},\mathbf{w}) + \lambda'(\mathbf{v},\mathbf{w}')$
   - $(\mu\mathbf{v} + \mu'\mathbf{v}', \mathbf{w}) = \overline{\mu}(\mathbf{v},\mathbf{w}) + \overline{\mu'}(\mathbf{v}',\mathbf{w})$
3. **Positive definiteness**: $(\mathbf{v},\mathbf{v}) \geqslant 0$ with equality iff $\mathbf{v} = \mathbf{0}$.

Definition 2.53 (Orthogonality)

We say that $\mathbf{v},\mathbf{w} \in V$ are **orthogonal** if

$$(\mathbf{v},\mathbf{w}) = 0.$$

Proposition 2.54

If vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ are non-zero and orthogonal, then they are linearly independent.

Proof. Suppose for contradiction that the vectors are linearly dependent. Then there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ or $\mathbb{C}$, not all zero, such that

$$\sum_i \alpha_i \mathbf{v}_i = \mathbf{0}.$$

Then

$$\begin{aligned} 0 = \left(\mathbf{v}_j, \sum_i \alpha_i \mathbf{v}_i\right) &= \sum_i \alpha_i \left(\mathbf{v}_j, \mathbf{v}_i\right) && \text{by linearity} \\ &= \alpha_j \left(\mathbf{v}_j, \mathbf{v}_j\right) && \text{by orthogonality} \\ &= \alpha_j \left|\mathbf{v}_j\right|^2. \end{aligned}$$

By positive definiteness, $\left|\mathbf{v}_j\right|^2 > 0$, so we must have $\alpha_j = 0$. This holds for all j, contradicting our assumption that not all α_i are zero.

2.10.3 Basis and Dimension

Definition 2.55 (Basis)

A **basis** of a vector space V is a set of vectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ in V that

1. B spans V,
2. the vectors in B are linearly independent.

Remark. This implies that the coefficients in the linear combination $\lambda_1 \mathbf{v_1} + \lambda_2 \mathbf{v_2} + \ldots + \lambda_r \mathbf{v_r}$ are unique for any vector in V. The set of coefficients are called the **components** of the vector with respect to the basis B.

Theorem 2.56

If $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ and $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_m\}$ are bases for the same vector space V, then $n = m$. The number n is called the **dimension** of V.

Proposition 2.57

If V is a vector space of dimension n. Then,

1. if $Y = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$ spans V, and that $m > n$, we can remove vectors from Y to get a basis.
2. If $Z = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is a linearly independent set in V with $k < n$, we can add vectors to Z to get a basis.

Lecture 9 · 2025-10-28

3 Matrices

3.1 Linear Maps

Definition 3.1 (Linear map)

For two vector spaces V and W, a **linear map** is a function

$$T : V \to W$$

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 all scalars λ, μ.

Definition 3.2

Let $T : V \to W$ be a linear map.

- The **image** of $\mathbf{x} \in V$ under T is the vector $\mathbf{x}' = T(\mathbf{x}) \in W$.

  The **image** of T is the set

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

  It forms a subspace of W.

- If $\mathbf{x} \in V$ such that $T(\mathbf{x}) = \mathbf{0}$, then $\mathbf{x}$ is in the **kernel** of T.

  The **kernel** of T is the set

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

  It forms a subspace of V.

- For $T : V \to W$, V is called the **domain** of T and W the **codomain** of T.
- The dimension of the image of T, $\dim(\operatorname{Im} T)$, is called the **rank** of T, denoted $\operatorname{rank}(T)$.
- The dimension of the kernel of T, $\dim(\ker T)$, is called the **nullity** of T, denoted $\operatorname{null}(T)$.

Remark. For $T : V \to W$, we have

$$\dim(\ker T) \leqslant \dim V, \quad \dim(\operatorname{Im} T) \leqslant \dim W.$$

Example 3.3

1. The **zero linear map** $T : V \to W$ is defined by $T(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in V$.

   It has $\operatorname{Im} T = \{\mathbf{0}\}$ and $\ker T = V$.
2. The **identity map** $T : V \to V$ is defined by $T(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in V$.

It has $\operatorname{Im} T = V$ and $\ker T = \{\mathbf{0}\}$.

3. Consider $V = W = \mathbb{R}^2$ and $T(\mathbf{x}) = \mathbf{x}'$, with

$$x_1' = 2x_1 + x_2$$
$$x_2' = x_1 - 4x_2.$$

This is a linear map. In this case, $\operatorname{Im} T = \left\{\lambda\binom{2}{1} + \mu\binom{1}{-3} : \lambda, \mu \in \mathbb{R}\right\} = \mathbb{R}^2$ and $\ker T = \{\mathbf{0}\}$.

We can carry out several operations on linear maps.

- **Linear combination**

  Let $T, S : V \to W$ be linear maps. Then,
  $$\alpha T + \beta S : V \to W$$
  is still a linear map, defined by
  $$(\alpha T + \beta S)(\mathbf{x}) = \alpha T(\mathbf{x}) + \beta S(\mathbf{x})$$
  for all $\mathbf{x} \in V$ and all scalars α, β.

- **Composition**

  Let $T : V \to W$, $S : U \to V$ be linear maps. Then,
  $$T \circ S : U \to W$$
  is still a linear map, defined by
  $$(T \circ S)(\mathbf{x}) = T(S(\mathbf{x}))$$
  for all $\mathbf{x} \in U$.

Theorem 3.4 (Rank-Nullity Theorem)

Let $T : V \to W$ be a linear map, where V is finite-dimensional. Then,

$$\dim V = \operatorname{rank} T + \operatorname{null} T.$$

Proof. Let us call $n = \dim V$ and $\mathbf{M} = \operatorname{null} T$. Since $\dim(\ker T) \leqslant \dim V$, we have $\mathbf{M} \leqslant n$. We have two cases:

1. $|\mathbf{M}| = n$. Then, $\ker T = V$, so T is the zero map. Thus, $\operatorname{Im} T = \{\mathbf{0}\}$ and $\operatorname{rank} T = 0$. Therefore, $\dim V = n = 0 + n = \operatorname{rank} T + \operatorname{null} T$.
2. $|\mathbf{M}| < n$. Then let $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} \subseteq V$ be a basis of $\ker T$. Then, $T(\mathbf{e}_i) = \mathbf{0}$ for all i.

   We can extend $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ to the basis of the whole V:
   $$\{\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{e}_{\mathrm{M}+1}, \ldots, \mathbf{e}_n\}.$$
   We need to show that $\{T(\mathbf{e}_{\mathrm{M}+1}), \ldots, T(\mathbf{e}_n)\}$ is a basis of $\operatorname{Im} T$.

   - **Spanning**. To show that $\{T(\mathbf{e}_{\mathrm{M}+1}), \ldots, T(\mathbf{e}_n)\}$ spans $\operatorname{Im} T$, take $y \in \operatorname{Im} T$. Then $\exists \mathbf{x} \in V$ such that

$$T(\boldsymbol{x}) = y.$$

Since $\boldsymbol{x} \in V$, we can write

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i$$

for some scalars $\alpha_1, ..., \alpha_n$. Thus,

$$y = T(\boldsymbol{x}) = T\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i\right) = \sum_{i=1}^{n} \alpha_i T(\boldsymbol{e}_i) = \sum_{i=\mathrm{M}+1}^{n} \alpha_i T(\boldsymbol{e}_i).$$

Therefore, y is in the span of $\{T(\boldsymbol{e}_{\mathrm{M}+1}), ..., T(\boldsymbol{e}_n)\}$.

- **Linear independence**. To show that $\{T(\boldsymbol{e}_{\mathrm{M}+1}), ..., T(\boldsymbol{e}_n)\}$ is linearly independent, suppose that

$$\sum_{i=\mathrm{M}+1}^{n} \alpha_i T(\boldsymbol{e}_i) = \boldsymbol{0}$$

for some scalars $\alpha_{\mathrm{M}+1}, ..., \alpha_n$. Then, by linearity of T, we can write

$$T\left(\underbrace{\sum_{i=\mathrm{M}+1}^{n} \alpha_i \boldsymbol{e}_i}_{x}\right) = \boldsymbol{0}.$$

Thus, $\boldsymbol{x} \in \ker T$. Therefore, since we supposed that $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_m\}$ is a basis of $\ker T$, we write

$$\boldsymbol{x} = \sum_{i=1}^{\mathrm{M}} \beta_i \boldsymbol{e}_i$$

for some scalars $\beta_1, ..., \beta_m$. But since $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_n\}$ is a basis of V, the representation of $\boldsymbol{x}$ is unique. Thus, $\alpha_{\mathrm{M}+1} = ... = \alpha_n = 0$.

Example 3.5

- **Zero linear map.** We have $\operatorname{null} T = \dim V$ and $\operatorname{rank} T = 0$. Then $\dim V = \dim V + 0$.
- **Identity map.** We have $\operatorname{null} T = 0$ and $\operatorname{rank} T = \dim V$. Then $\dim V = 0 + \dim V$.

3.2 Matrices as Linear Maps

Let $\mathbf{M}$ be a matrix with entries $M_{ij} \in \mathbb{R}$. define

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

such that

$$T(\boldsymbol{x}) = \mathbf{M}\boldsymbol{x} = \boldsymbol{x}' \quad \text{for } \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^n$$

where

$$x_i' = M_{ij} x_j.$$

Given $n = 2$, with

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

we have

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} M_{11}x_1 + M_{12}x_2 \\ M_{21}x_1 + M_{22}x_2 \end{pmatrix}.$$

Consider $\boldsymbol{R}_i \in \mathbb{R}^n$ the rows, and $\boldsymbol{C}_i \in \mathbb{R}^n$ the columns of $\mathbf{M}$.

Lecture 10 · 2025-10-30

Proposition 3.6

The image and kernel of the linear map T defined by the matrix $\mathbf{M}$ are given by

$$\operatorname{Im} T = \operatorname{Im} \mathbf{M} = \operatorname{span}\{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_n\}$$

and

$$\ker T = \ker \mathbf{M} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{R}_i \cdot \boldsymbol{x} = 0 \text{ for all } i = 1, \ldots, n\}.$$

Proof. Let us consider the image and kernel of $\mathbf{M}$. The components are related in the following form:

$$M_{ij} = \left(\boldsymbol{C}_j\right)_i = \left(\boldsymbol{R}_i\right)_j.$$

If $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is the standard basis of $\mathbb{R}^n$, then, under T,

$$\boldsymbol{e}_i \mapsto T(\boldsymbol{e}_i) = \mathbf{M}\boldsymbol{e}_i = \boldsymbol{C}_i.$$

Since T is a linear map, we can write

$$\boldsymbol{x} = \sum_i x_i \boldsymbol{e}_i \mapsto T(\boldsymbol{x}) = \sum_i x_i T(\boldsymbol{e}_i) = \sum_i x_i \boldsymbol{C}_i = x_i \boldsymbol{C}_i.$$

Thus, $\operatorname{Im} T = \operatorname{Im} \mathbf{M} = \operatorname{span}\{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_n\}$, which is the span of the columns of $\mathbf{M}$.

Now, for the kernel, consider $x_i' = M_{ij}x_j = (\boldsymbol{R}_i)_j x_j = \boldsymbol{R}_i \cdot \boldsymbol{x}$.

If $\boldsymbol{x}' = \mathbf{0}$, then $\boldsymbol{R}_i \cdot \boldsymbol{x} = 0$ for all i. Thus, $\ker T = \ker \mathbf{M}$ is the set of vectors orthogonal to all the rows of $\mathbf{M}$.

Example 3.7 (Examples of matrices as linear maps)

1. **Zero map**. The zero map is defined by taking $\mathbf{M} = \mathbf{0}$.
2. **Identity map**. The identity map is defined by taking $\mathbf{M} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.
3. Consider the map $T : V \to W$ where $\boldsymbol{x}' = T(\boldsymbol{x}) = \mathbf{M}\boldsymbol{x}$. Let T be defined by

$$\begin{cases} x_1' = 3x_1 + x_2 + 5x_3 \\ x_2' = -x_1 - 2x_3 \\ x_3' = 2x_1 + x_2 + 3x_3 \end{cases}$$

then, the matrix associated to T is

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

with columns

$$\boldsymbol{C}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \boldsymbol{C}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{C}_3 = \begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix}.$$

and rows

$$\boldsymbol{R}_1 = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}, \boldsymbol{R}_2 = \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}, \boldsymbol{R}_3 = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Hence, the image and kernel of the linear map T are given by

$$\operatorname{Im} T = \operatorname{Im} \mathbf{M} = \operatorname{span}\{\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3\} = \operatorname{span}\{\boldsymbol{C}_1, \boldsymbol{C}_2\} \Rightarrow \operatorname{rank}(T) = 2,$$

because we have that $\begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$.

Then, for the kernel, we need

$$\boldsymbol{R}_2 \times \boldsymbol{R}_3 = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -1 & 0 & -2 \\ 2 & 1 & 3 \end{vmatrix} = \begin{pmatrix} 0 - (-2) \\ 3 + (-4) \\ -1 - 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}.$$

Hence

$$\ker T = \ker \mathbf{M} = \operatorname{span}\left\{\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}\right\}.$$

3.3 Geometric examples in $\mathbb{R}^2$ and $\mathbb{R}^3$

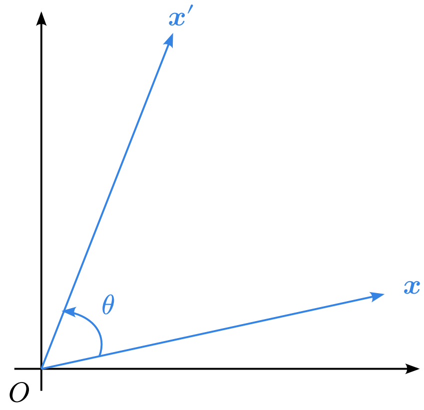
3.3.1 In $\mathbb{R}^2$

1. **Rotations.**

   Consider θ such that $-\pi < \theta \leqslant \pi$. Then, a rotation by an angle θ about the origin in $\mathbb{R}^2$ is given by the matrix

$$\mathbf{Rot}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

   Note that $\det \mathbf{Rot}(\theta) = \cos^2(\theta) + \sin^2(\theta) = 1$.

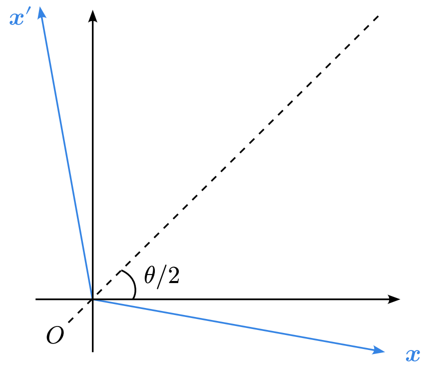


2. **Reflections.**

   Consider θ with $-\pi < \theta \leqslant \pi$. Then, a reflection of angle $\frac{\theta}{2}$ in $\mathbb{R}^2$is given by the matrix

$$\mathbf{Ref}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

   Note that $\det \mathbf{Ref}(\theta) = -\left(\cos^2(\theta) + \sin^2(\theta)\right) = -1.$



Proposition 3.8 (Properties of rotations and reflections)

- $\mathbf{Rot}(\theta)\,\mathbf{Rot}(\varphi) = \mathbf{Rot}(\theta + \varphi)$
- $\mathbf{Ref}(\theta)\,\mathbf{Ref}(\varphi) = \mathbf{Rot}(\theta - \varphi)$
- $\mathbf{Rot}(\theta)\,\mathbf{Ref}(2\varphi) = \mathbf{Ref}(2\varphi + \theta)$
- $\mathbf{Ref}(2\theta)\,\mathbf{Rot}(\varphi) = \mathbf{Ref}(2\varphi - \theta)$

3.3.2 In $\mathbb{R}^3$

1. **Rotations.**

   - Consider a rotation by an angle θ about axis $\boldsymbol{e}_3$. This is given by the matrix

$$\mathbf{Rot}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

   - Consider a rotation by an angle θ about the unit vector $\boldsymbol{n}$. In this case, we have

$$\mathbf{x}' = \mathbf{R}\mathbf{x}$$

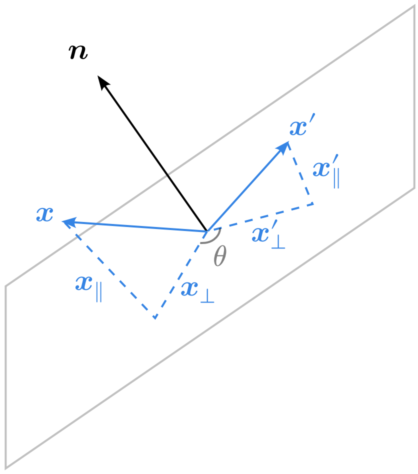
where $\boldsymbol{x} \in \mathbb{R}^3$ and $x_i' = R_{ij}x_j$.

Then,

$$\boldsymbol{x}' = (\cos\theta)\boldsymbol{x} + (1-\cos\theta)(\boldsymbol{n}\cdot\boldsymbol{x})\boldsymbol{n} + (\sin\theta)\boldsymbol{n}\times\boldsymbol{x}$$

or equivalently,

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



This can be derived by decomposing $\boldsymbol{x}$ into components parallel and perpendicular to $\boldsymbol{n}$, and then rotating the perpendicular component in the plane orthogonal to $\boldsymbol{n}$.

$$\boldsymbol{x} = \boldsymbol{x}_{\parallel} + \boldsymbol{x}_{\perp}$$

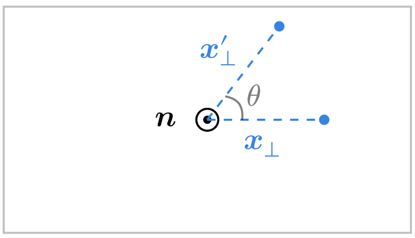
with

$$\begin{aligned}\boldsymbol{x}_{\parallel} &= (\boldsymbol{n}\cdot\boldsymbol{x})\boldsymbol{n},\\ \boldsymbol{x}_{\perp} &= \boldsymbol{x} - \boldsymbol{x}_{\parallel}.\end{aligned}$$

After applying $\mathbf{R}$, we have

$$\begin{aligned}\boldsymbol{x}'_{\parallel} &= \boldsymbol{x}_{\parallel},\\ \boldsymbol{x}'_{\perp} &= (\cos\theta)\boldsymbol{x}_{\perp} + (\sin\theta)\boldsymbol{n}\times\boldsymbol{x}_{\perp}.\end{aligned}$$

[looking down $\boldsymbol{n}$]



2. **Reflections.**

   Reflections in a plane through the origin with normal unit vector $\boldsymbol{n}$ are given by

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

   Thus we have

$$x'_i = H_{ij} x_j,$$

where

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

Lecture 11 · 2025-11-01

3. **Dilations.**

   Dilations from the origin with scale factor λ are given by

   $$\boldsymbol{x}' = \mathbf{D}(\lambda)\boldsymbol{x} = \lambda \boldsymbol{x}.$$

   Thus, we have

   $$\boldsymbol{x}'_i = D_{ij}(\lambda) x_j,$$

   where

   $$D_{ij}(\lambda) = \lambda \delta_{ij}.$$

4. **Shears.**

   Given a, b with $|a| = |b| = 1$ and such that $\boldsymbol{a} \cdot \boldsymbol{b} = 0$, a shear with parameter λ is defined by

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

   Thus, we have

   $$\boldsymbol{x}'_i = S_{ij}(\lambda) x_j,$$

   where

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

3.4 Matrices in General

3.4.1 Definitions

Definition 3.9 (Matrix)

Consider a linear map $T : V \to W$, with $\dim V = n$ and $\dim W = m$, and take two bases $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_n\}$ of V and $\{\boldsymbol{f}_1, ..., \boldsymbol{f}_m\}$ of W.

Then, T can be represented by $\mathbf{M}$, which is an $m \times n$ array with entries $M_{ij} \in \mathbb{R}$ or $\mathbb{C}$ for $i = 1, ..., m$ as the rows and $j = 1, ..., n$ as the columns, such that

$$T\left(\boldsymbol{e}_j\right) = \sum_{i=1}^{m} M_{ij} \boldsymbol{f}_i$$

for $j = 1, ..., n$. This automatically ensures that for any $\boldsymbol{x} \in V$, $\boldsymbol{x}' = T(\boldsymbol{x})$, we can always write $\boldsymbol{x}'$ and $\boldsymbol{x}$ in terms of the bases:

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j, \quad \mathbf{x}' = \sum_{i=1}^{m} x_i' \mathbf{f}_i.$$

This means that any coefficient from the image can be written as

$$x_i' = \sum_{j=1}^{n} M_{ij} x_j.$$

To summarise, given V and W which are real or complex vector spaces with $\dim V = n$ and $\dim W = m$, and given bases $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of V and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ of W, then

- V is identified with $\mathbb{R}^n$ or $\mathbb{C}^n$.
- W is identified with $\mathbb{R}^m$ or $\mathbb{C}^m$.
- We identify the linear map $T : V \to W$ with the matrix $\mathbf{M}$ such that $\mathbf{x}' = \mathbf{M}\mathbf{x}$.

Remark. Consider another linear map $S : V \to W$ with matrix representation $\mathbf{N}$ with respect to the same bases. Then, for scalars α, β,

$$\alpha T + \beta S$$

is represented by matrix

$$\alpha \mathbf{M} + \beta \mathbf{N}$$

with coefficients

$$(\alpha \mathbf{M} + \beta \mathbf{N})_{ij} = \alpha M_{ij} + \beta N_{ij}.$$

This is because addition and scalar multiplication in matrices takes place entry-wise.

Example 3.10

Consider $V = M_{2\times 2}(\mathbb{R})$ and $W = \mathbb{R}^3$. Hence $\dim V = 4$ and $\dim W = 3$. Consider the map

$$T : V \to W \quad \text{with} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a+b \\ c \\ d \end{pmatrix}.$$

The map is linear. We want to find the matrix representation of T with respect to the bases

$$\left\{ \mathbf{e}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \mathbf{e}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

of V and

$$\left\{ \mathbf{f}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{f}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

of W.

To determine $\mathbf{M}$, we need to compute $T(\mathbf{e}_i)$ for $i = 1, \dots, 4$:

$$T(\boldsymbol{e}_1) = T\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$T(\boldsymbol{e}_2) = T\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$T(\boldsymbol{e}_3) = T\left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$T(\boldsymbol{e}_4) = T\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Therefore, for $a, b, c, d \in \mathbb{R}$,

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = aT(\boldsymbol{e}_1) + bT(\boldsymbol{e}_2) + cT(\boldsymbol{e}_3) + dT(\boldsymbol{e}_4) = a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + d\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a + b \\ c \\ d \end{pmatrix}.$$

Thus, the matrix representation of T with respect to the given bases is

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

3.4.2 Matrix Multiplication

Consider linear maps T and S such that

$$T : V \to W, \quad S : U \to V.$$

We wish to compose them. The composition is given by

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

such that

$$(T \circ S)(\boldsymbol{x}) = T(S(\boldsymbol{x}))$$

for all $\boldsymbol{x} \in U$.

If T is represented by the matrix $\mathbf{M}$ and S is represented by the matrix $\mathbf{N}$, then $T \circ S$ is represented by the matrix $\mathbf{L} = \mathbf{MN}$.

Lecture 12 · 2025-11-04

Let

- $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be a basis of V $(\dim V = n)$,
- $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_m\}$ be a basis of W $(\dim W = m)$,
- $\{\boldsymbol{g}_1, \dots, \boldsymbol{g}_l\}$ be a basis of U $(\dim U = l)$.

If we consider $T \circ S$ so that $T \circ S$ is represented by the matrix $\mathbf{L} = \mathbf{MN}$, with coefficients given by

$$L_{ik} = M_{ij} N_{jk}.$$

Note that
- $\mathbf{M}$ is an $m \times n$ matrix,
- $\mathbf{N}$ is an $n \times l$ matrix,
- $\mathbf{L}$ is an $m \times l$ matrix.

Remark.
1. The number of columns of **M** must equal the number of rows of **N** for the product **MN** to be defined.
2. **L** has the same number of rows as **M** and the same number of columns as **N**.

We can also write

$$\begin{aligned} L_{ik} &= (\mathbf{MN})_{ik} \\ &= \left[\boldsymbol{R}_i(\mathbf{M})\right]_j \left[\boldsymbol{C}_k(\mathbf{N})\right]_j \\ &= \boldsymbol{R}_i(\mathbf{M}) \cdot \boldsymbol{C}_k(\mathbf{N}). \end{aligned}$$

If we apply **MN** to a $\boldsymbol{x} \in U$, we obtain

$$(\mathbf{MN})\mathbf{x} = \mathbf{M}(\mathbf{N}\mathbf{x})$$

with

$$[\mathbf{M}(\mathbf{N}\mathbf{x})]_i = M_{ij}[\mathbf{N}\mathbf{x}]_j$$

and Thus

$$(\mathbf{MN})_{ik} x_k = M_{ij} N_{jk} x_k.$$

Proposition 3.11 (Matrix properties)

For any three matrices $\mathbf{L}, \mathbf{M}, \mathbf{N}$ such that the products below are defined, and for any scalars λ, μ,

- $(\lambda\mathbf{M} + \mu\mathbf{N})\mathbf{L} = \lambda(\mathbf{ML}) + \mu(\mathbf{NL})$
- $\mathbf{L}(\lambda\mathbf{M} + \mu\mathbf{N}) = \lambda(\mathbf{LM}) + \mu(\mathbf{LN})$
- $\mathbf{M}(\mathbf{NL}) = (\mathbf{MN})\mathbf{L}$.

3.4.3 Matrix Inverses

Consider three matrices $\boldsymbol{M}, \boldsymbol{N}, \boldsymbol{L}$, satisfiying

- The size of $\mathbf{N}$ is $m \times n$,
- The size of $\mathbf{M}$ is $n \times m$,
- The size of $\mathbf{L}$ is $n \times m$.

We say that **L** is a **left inverse** of **N** if

$$\mathbf{LN} = \mathbf{I}_n$$

where $\mathbf{I}$ is the identity matrix of size $n \times n$.

We say that $\mathbf{M}$ is a **right inverse** of $\mathbf{N}$ if

$$\mathbf{N}\mathbf{M} = \mathbf{I}_m.$$

If $\mathbf{N}$ is a square matrix (i.e., $m = n$), then

$$\mathbf{L} = \mathbf{L}(\mathbf{N}\mathbf{M}) = (\mathbf{L}\mathbf{N})\mathbf{M} = \mathbf{M},$$

so the left and right inverses coincide. In this case, we say that $\mathbf{N}$ is **invertible** (or **non-singular**) and we denote its inverse by $\mathbf{N}^{-1}$.

> ***Remark.*** If $\mathbf{N}$ has an inverse, then $\mathbf{N}$ is a square matrix.
>
> Not all square matrices are invertible. For example, the zero matrix is not invertible.

Proposition 3.12

For two invertible matrices $\mathbf{M}$ and $\mathbf{N}$ of the same size,

$$(\mathbf{M}\mathbf{N})^{-1} = \mathbf{N}^{-1}\mathbf{M}^{-1}.$$

Proof.

$$\left(\mathbf{N}^{-1}\mathbf{M}^{-1}\right)(\mathbf{M}\mathbf{N}) = \mathbf{N}^{-1}\left(\mathbf{M}^{-1}\mathbf{M}\right)\mathbf{N} = \mathbf{N}^{-1}\mathbf{I}\mathbf{N} = \mathbf{N}^{-1}\mathbf{N} = \mathbf{I}.$$

Example 3.13

1. **Rotation.** For $\mathbf{Rot}(\theta, \boldsymbol{n})$, we have
$$\mathbf{Rot}(\theta, \boldsymbol{n})^{-1} = \mathbf{Rot}(-\theta, \boldsymbol{n}).$$
2. **Shear.** Fix $\boldsymbol{a}, \boldsymbol{b}$. Then, for $\mathbf{S}(\lambda)$, we have
$$\mathbf{S}(\lambda)^{-1} = \mathbf{S}(-\lambda).$$
3. **Reflection.** If $\mathbf{H}$ is a reflection in a plane with normal $\boldsymbol{n}$, then
$$\mathbf{H}^{-1} = \mathbf{H}.$$

3.4.4 Transpose and Hermitian Conjugate

Definition 3.14 (Transpose)

Consider a matrix $\mathbf{M}$ of size $m \times n$. Then, the **transpose** of $\mathbf{M}$ is the matrix $\mathbf{M}^\top$ of size $n \times m$ with entries

$$\left(\mathbf{M}^\top\right)_{ij} = M_{ji}.$$

Proposition 3.15 (Properties of the transpose)

1. $\left(\mathbf{M}^\top\right)^\top = \mathbf{M}$

2. If $\boldsymbol{x}$ is a column vector $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, then $\mathbf{x}^\top$ is the row vector $(x_1 \ \vdots \ x_n)$.
3. $(\mathbf{MN})^\top = \mathbf{N}^\top \mathbf{M}^\top$
4. $(\alpha \mathbf{M} + \beta \mathbf{N})^\top = \alpha \mathbf{M}^\top + \beta \mathbf{N}^\top$

Definition 3.16 (Symmetric and antisymmetric matrices)

If $\mathbf{M}$ is a square matrix, then $\mathbf{M}$ is

- **symmetric** if $\mathbf{M}^\top = \mathbf{M}$,
- **antisymmetric** if $\mathbf{M}^\top = -\mathbf{M}$

Definition 3.17 (Hermitian conjugate)

Consider a matrix $\mathbf{M}$ of size $m \times n$ with complex entries. Then, the **Hermitian conjugate** of $\mathbf{M}$ is the matrix $\mathbf{M}^\dagger$ of size $n \times m$ is the matrix

$$\mathbf{M}^\dagger = \overline{\mathbf{M}^\top}$$

with entries

$$\left(\mathbf{M}^\dagger\right)_{ij} = \overline{M_{ji}}$$

where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$.

Proposition 3.18 (Properties of the Hermitian conjugate)

1. $(\alpha \mathbf{M} + \beta \mathbf{N})^\dagger = \overline{\alpha} \mathbf{M}^\dagger + \overline{\beta} \mathbf{N}^\dagger$
2. $(\mathbf{MN})^\dagger = \mathbf{N}^\dagger \mathbf{M}^\dagger$

Definition 3.19 (Hermitian and anti-Hermitian matrices)

If $\mathbf{M}$ is a square, then $\mathbf{M}$ is

- **Hermitian** if $\mathbf{M}^\dagger = \mathbf{M}$, *i.e.* $M_{ij} = \overline{M_{ji}}$ for all i, j,
- **anti-Hermitian** (or **skew-Hermitian**) if $\mathbf{M}^\dagger = -\mathbf{M}$, *i.e.* $M_{ij} = -\overline{M_{ji}}$ for all i, j.

3.4.5 Trace

Definition 3.20 (Trace)

Consider any $n \times n$ matrix $\mathbf{M}$, the trace is defined by

$$\operatorname{tr}(\mathbf{M}) = M_{ii}.$$

i.e. the sum of the diagonal entries.

Proposition 3.21 (Properties of the trace)

1. $\mathrm{tr}(\alpha\mathbf{M} + \beta\mathbf{N}) = \alpha\,\mathrm{tr}(\mathbf{M}) + \beta\,\mathrm{tr}(\mathbf{N})$
2. $\mathrm{tr}(\mathbf{MN}) = \mathrm{tr}(\mathbf{NM})$
3. $\mathrm{tr}(\mathbf{M}^\top) = \mathrm{tr}(\mathbf{M})$
4. $\mathrm{tr}(\mathbf{I}) = n$ for the identity matrix of size $n \times n$.

3.4.6 Decomposition of $n \times n$ Matrices

Any $n \times n$ matrix is a sum of symmetric and antisymmetric parts. For a matrix $\mathbf{M}$ that is square with real entries, we can write $\mathbf{M}$ as $\mathbf{S} + \mathbf{A}$, where

$$\mathbf{S} = \frac{1}{2}(\mathbf{M} + \mathbf{M}^\top)$$

is the symmetric part and

$$\mathbf{A} = \frac{1}{2}(\mathbf{M} - \mathbf{M}^\top)$$

is the antisymmetric part.

The symmetric part can be further decomposed:

$$\mathbf{T} = \mathbf{S} - \frac{1}{n}\,\mathrm{tr}(\mathbf{S})\mathbf{I}$$

Note that $\mathrm{tr}(\mathbf{T}) = 0$, and we call $\mathbf{T}$ to be traceless. Noting that $\mathrm{tr}(\mathbf{S}) = \mathrm{tr}(\mathbf{M})$ and $\mathrm{tr}(\mathbf{A}) = 0$, we can write

$$\mathbf{M} = \underbrace{\mathbf{T}}_{\substack{\text{symmetric}\\ \text{traceless}}} + \underbrace{\frac{1}{n}\,\mathrm{tr}(\mathbf{M})\mathbf{I}}_{\text{isotropic part}} + \underbrace{\mathbf{A}}_{\text{antisymmetric part}}.$$

Lecture 13 · 2025-11-06

3.4.7 Orthogonal and Unitary Matrices

Definition 3.22 (Orthogonal matrix)

A real $n \times n$ matrix $\mathbf{U}$ is **orthogonal** if and only if

$$\mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}$$

or equivalently,

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

This means that columns and rows of $\mathbf{U}$ are orthonormal vectors. Equivalently, $\mathbf{U}$ is orthogonal if and only if $\mathbf{U}$ preserves the dot product, *i.e.* for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$(\mathbf{U}\boldsymbol{x}) \cdot (\mathbf{U}\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y},$$

and in this cases, $\mathbf{U}$ preserves lengths and angles.

Definition 3.23 (Unitary matrix)

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

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$$

or equivalently,

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

Equivalently, $\mathbf{U}$ is unitary iff it preserves the complex inner product, *i.e.* for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$,

$$(\mathbf{U}\boldsymbol{x})^\dagger(\mathbf{U}\boldsymbol{y}) = \boldsymbol{x}^\dagger\boldsymbol{y},$$

and in this cases, $\mathbf{U}$ preserves lengths and angles.

Example 3.24

In 2×2, consider $\mathbf{U}$ as an orthogonal matrix. Consider the basis

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

- $\mathbf{U}$ preserves norms

$$\mathbf{U}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \quad \text{for } \theta \in \mathbb{R}$$

- $\mathbf{U}$ preserves angles, in particular, orthogonality

$$\mathbf{U}\begin{pmatrix}0\\1\end{pmatrix} = \pm\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix} \quad \text{for } \theta \in \mathbb{R}$$

Thus, we have either

$$\mathbf{U} = \mathbf{Rot}(\theta) = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$$

or

$$\mathbf{U} = \mathbf{Ref}(\theta) = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix}.$$

3.5 Determinant

Consider a map $\mathbb{R}^n \to \mathbb{R}^n$ given by a real $n \times n$ matrix $\mathbf{M}$, where

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

for all $\boldsymbol{x} \in \mathbb{R}^n$.

Assume that $\mathbf{M}^{-1}$ exists, then

$$\boldsymbol{x} = \mathbf{M}^{-1}\boldsymbol{x}'.$$

3.5.1 In $\mathbb{R}^2$

Consider $\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$, and let $\tilde{\mathbf{M}} = \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}$. Then,

$$\boldsymbol{x}' = \mathbf{M}\boldsymbol{x} \quad \Rightarrow \quad \tilde{\mathbf{M}}\boldsymbol{x}' = \mathbf{M}\tilde{\mathbf{M}}\boldsymbol{x} = \det(\mathbf{M})\boldsymbol{x}.$$

with $\det \mathbf{M} = M_{11}M_{22} - M_{12}M_{21}$.

Note that $\det \mathbf{M} = [\mathbf{M}\boldsymbol{e}_1, \mathbf{M}\boldsymbol{e}_2]$.

Therefore, if $\det \mathbf{M} \neq 0$, then $\mathbf{M}^{-1} = \frac{1}{\det \mathbf{M}}\tilde{\mathbf{M}}$.

3.5.2 In $\mathbb{R}^3$

We shall attempt to generalise our construction of the $\det \mathbf{M}$ to $\mathbb{R}^3$. Take $\mathbf{x} \mapsto \mathbf{x}' = \mathbf{M}\boldsymbol{x}$ where $\mathbf{M}$ is a 3×3 matrix with real entries. We seek a matrix $\tilde{\mathbf{M}}$ and a scalar $\det \mathbf{M}$ such that

$$\tilde{\mathbf{M}}\mathbf{M} = (\det \mathbf{M})\mathbf{I}.$$

We call this scalar $\det \mathbf{M}$ the **determinant** of $\mathbf{M}$.

Recall that the scalar triple product of three vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$ is defined by

$$[\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}] = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = \varepsilon_{ijk} a_i b_j c_k$$

which describes the volume of the parallelepiped formed by the three vectors.

Under the action of a 3×3 matrix $\mathbf{M}$, volumes are scaled by a factor $\det \mathbf{M}$, where

$$\begin{aligned} [\mathbf{M}\boldsymbol{e}_1, \mathbf{M}\boldsymbol{e}_2, \mathbf{M}\boldsymbol{e}_3] &= [\mathbf{C}_1(\mathbf{M}), \mathbf{C}_2(\mathbf{M}), \mathbf{C}_3(\mathbf{M})] \\ &= \left[M_{i1}\mathbf{e}_1, M_{j2}\mathbf{e}_2, M_{k3}\mathbf{e}_3\right] \\ &= M_{i1}M_{j2}M_{k3}[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3] \\ &= \varepsilon_{ijk}M_{i1}M_{j2}M_{k3} \\ &=: \det \mathbf{M}. \end{aligned}$$

Thus, in $\mathbb{R}^3$, the determinant of a matrix $\mathbf{M}$ is given by

$$\det \mathbf{M} = \varepsilon_{ijk}M_{i1}M_{j2}M_{k3}.$$

To construct $\tilde{\mathbf{M}}$, note

$$\begin{aligned} \boldsymbol{R}_1\left(\tilde{\mathbf{M}}\right) &= \mathbf{C}_2(\mathbf{M}) \times \mathbf{C}_3(\mathbf{M}) \\ \boldsymbol{R}_2\left(\tilde{\mathbf{M}}\right) &= \mathbf{C}_3(\mathbf{M}) \times \mathbf{C}_1(\mathbf{M}) \\ \boldsymbol{R}_3\left(\tilde{\mathbf{M}}\right) &= \mathbf{C}_1(\mathbf{M}) \times \mathbf{C}_2(\mathbf{M}) \end{aligned}$$

so that

$$\boldsymbol{R}_i\left(\tilde{\mathbf{M}}\right) \cdot \boldsymbol{C}_j(\mathbf{M}) = \mathbf{C}_1(\mathbf{M}) \cdot (\mathbf{C}_2(\mathbf{M}) \times \mathbf{C}_3(\mathbf{M}))\delta_{ij}.$$

Thus,

$$\left(\tilde{\mathbf{M}}\mathbf{M}\right)_{ij} = (\det \mathbf{M})\delta_{ij}.$$

And hence $\det \mathbf{M} \neq 0$ iff $\{\mathbf{M}e_1, \mathbf{M}e_2, \mathbf{M}e_3\}$ is linearly independent. This is equivalent to saying $\operatorname{Im}(\mathbf{M}) = \mathbb{R}^3$, or that $\operatorname{rank}(\mathbf{M}) = 3$.

Remark. General 3×3 determinants can be expanded in terms of 2×2 determinants. For example,

$$\begin{vmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{vmatrix} = M_{11} \begin{vmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{vmatrix} - M_{12} \begin{vmatrix} M_{21} & M_{23} \\ M_{31} & M_{33} \end{vmatrix} + M_{13} \begin{vmatrix} M_{21} & M_{22} \\ M_{31} & M_{32} \end{vmatrix}.$$

3.5.3 Permutations

Our goal is to generalise the Levi-Civita symbol to n dimensions to define the determinant of an $n \times n$ matrix.

Definition 3.25 (Permuation)

A **permutation** of a set S is a bijection $\varepsilon : S \to S$.

Notation. We write S_n to be the set of all permutations of the set $\{1, 2, \dots, n\}$. Note that $|S_n| = n!$.

Consider $\rho \in S_n$ with

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \rho(1) & \rho(2) & \dots & \rho(n) \end{pmatrix}$$

Definition 3.26 (Fixed point)

A **fixed point** of a permutation $\rho \in S_n$ is an element $i \in \{1, \dots, n\}$ such that $\rho(i) = i$. We normally omit fixed points when writing permutations.

Definition 3.27 (Disjoint permutations)

Two permutations are **disjoint** if members moved by one permutation are not moved by the other, *i.e.* they have no common elements that are not fixed points.

Example 3.28

We can write

$$\begin{aligned} \rho &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 4 & 5 & 6 \\ 5 & 6 & 1 & 4 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 & 5 \\ 5 & 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 6 \\ 6 & 2 \end{pmatrix} \\ &= (5\ \ 1\ \ 4)(6\ \ 2) \end{aligned}$$

where $(5\ 4\ 1)$ and $(6\ 2)$ are called cycles.

Note that disjoint permutations commute, but in general permutations do not commute.

Definition 3.29 (Transposition)

A **transposition** is a 2-cycle.

Proposition 3.30

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

Proof. This is because we can write

$$(1\ 2\ \dots\ n) = (1\ 2)(2\ 3)\dots(n-1\ n).$$

Definition 3.31 (Sign of permutation)

The **sign** of a permutation ρ is defined by

$$\varepsilon(\rho) = (-1)^r$$

where r is the number of 2-cycles of p when written as a product of 2-cycles.

In particular, if $\varepsilon(\rho) = 1$, then ρ is an **even** permutation, and if $\varepsilon(\rho) = -1$, then ρ is an **odd** permutation.

Remark. $\varepsilon(\rho\sigma) = \varepsilon(\rho)\varepsilon(\sigma)$ and $\varepsilon\left(\rho^{-1}\right) = \varepsilon(\rho)$.

Definition 3.32 (Levi-Civita symbol)

The **Levi-Civita symbol** in n dimensions is defined by

$$\varepsilon_{i_1 i_2 \dots i_n} = \begin{cases} +1 & \text{if } \left(i_1, i_2, \dots, i_n\right) \text{ is an even permutation of } (1, 2, \dots, n) \\ -1 & \text{if } \left(i_1, i_2, \dots, i_n\right) \text{ is an odd permutation of } (1, 2, \dots, n) \\ 0 & \text{if any two indices are equal} \end{cases}$$

It is totally antisymmetric.

Lecture 14 · 2025-11-08

3.5.4 Alternating Forms

Definition 3.33 (Alternating form)

For vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, the rank n **alternating form** is defined by

$$
\begin{aligned}
[\mathbf{v}_1, \ldots, \mathbf{v}_n] &= \varepsilon_{j_1,\ldots,j_n} (v_1)_{j1} (v_2)_{j2} \cdots (v_n)_{jn} \\
&= \sum_{\rho} \varepsilon(\rho) (v_1)_{\rho(1)} (v_2)_{\rho(2)} \cdots (v_n)_{\rho(n)}
\end{aligned}
$$

Proposition 3.34 (Properties of alternating forms)

1. $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is multilinear in its arguments. *i.e.*

$$[\mathbf{v}_1, \ldots, \alpha \mathbf{v}_i + \beta \mathbf{u}_i, \ldots, \mathbf{v}_n] = \alpha[\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n] + \beta[\mathbf{v}_1, \ldots, \mathbf{u}_i, \ldots, \mathbf{v}_n]$$

2. It is totally antisymmetric: $\left[\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n\right] = -\left[\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n\right]$ for all $i \neq j$.

   Alternatively, $\left[\mathbf{v}_{\rho(1)}, \ldots, \mathbf{v}_{\rho(n)}\right] = \varepsilon(\rho)[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ for any permutation ρ.

3. $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = 1$.

> ***Remark.*** Properties (1) (2) (3) uniquely define the alternating forms. Note that exchanging two vectors changes the sign of the alternating form, so if any two vectors are equal, the alternating form is zero.

4. If $\mathbf{v}_p = \mathbf{v}_q$ for some $p \neq q$, then $[\mathbf{v}_1, \ldots, \mathbf{v}_n] = 0$. [Follows from (2).]

5. If $\mathbf{v}_p = \sum_{i \neq p} \lambda_i \mathbf{v}_i$ for some scalars λ_i, then $[\mathbf{v}_1, \ldots, \mathbf{v}_n] = 0$. [Follows from (1) and (4).]

Proposition 3.35

$$[\mathbf{v}_1, \ldots, \mathbf{v}_n] \neq 0 \Leftrightarrow \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \text{ is linearly independent.}$$

Proof.

[$\Rightarrow$] If the vectors are linearly dependent, then one of them can be written as a linear combination of the others. By property (5), the alternating form is zero.

[$\Leftarrow$] If the vectors are linearly independent, then they span $\mathbb{R}^n$ or $\mathbb{C}^n$. In particular, for some matrix $\mathbf{U}$, we can write

$$e_j = U_{ij} \mathbf{v}_i.$$

Hence,

$$
\begin{aligned}
[\mathbf{e}_1, \ldots, \mathbf{e}_n] &= U_{i_1 1} U_{i_2 2} \ldots U_{i_n n} \left[\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_n}\right] \\
&= U_{i_1 1} U_{i_2 2} \ldots U_{i_n n} \varepsilon_{i_1 i_2 \ldots i_n} [\mathbf{v}_1, \ldots, \mathbf{v}_n].
\end{aligned}
$$

Since $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = 1$, we have

$$[\mathbf{v}_1, \ldots, \mathbf{v}_n] \neq 0.$$

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

Definition 3.36 (Determinant)

Consider an $n \times n$ matrix $\mathbf{M}$ with columns given by

$$\mathbf{C}_i = \mathbf{M}\boldsymbol{e}_i.$$

The **determinant** of $\mathbf{M}$ is defined by

$$\begin{aligned}\det \mathbf{M} &= [\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_n] \\ &= [\mathbf{M}\boldsymbol{e}_1, \mathbf{M}\boldsymbol{e}_2, \ldots, \mathbf{M}\boldsymbol{e}_n] \\ &= \varepsilon_{i_1 i_2 \ldots i_n} M_{i_1 1} M_{i_2 2} \ldots M_{i_n n} \\ &= \sum_{\rho} \varepsilon(\rho) M_{\rho(1)1} M_{\rho(2)2} \ldots M_{\rho(n)n}\end{aligned}$$

where $\varepsilon(\rho)$ is the sign of the permutation $\rho \in S_n$. We can also write

$$\det \mathbf{M} = \begin{vmatrix} M_{11} & M_{12} & \ldots & M_{1n} \\ M_{21} & M_{22} & \ldots & M_{2n} \\ \vdots & & & \vdots \\ M_{n1} & M_{n2} & \ldots & M_{nn} \end{vmatrix}$$

Proposition 3.37 (Properties of the determinant)

1. The determinant is multilinear in the columns of the matrix. In particular,

   $$\det(\lambda \mathbf{M}) = \lambda^n \det(\mathbf{M})$$

   for any scalar λ and any $n \times n$ matrix $\mathbf{M}$.
2. The determinant is totally antisymmetric in the columns of the matrix. In particular, if we exchange two columns of $\mathbf{M}$, then the determinant changes sign.
3. $\det(\mathbf{I}) = 1$ for the identity matrix of any size $n \times n$.
4. If two rows or two columns of $\mathbf{M}$ are equal, then $\det(\mathbf{M}) = 0$.
5. If two rows or two columns of $\mathbf{M}$ are linearly dependent, then $\det(\mathbf{M}) = 0$.
6. $\det(\mathbf{M}) \neq 0$ if and only if the columns of $\mathbf{M}$ are linearly independent.

   As a consequence, under a column operation $\mathbf{C}_i \mapsto \mathbf{C}_i + \lambda \mathbf{C}_j$ for some $j \neq i$, the determinant is unchanged.
7. $\det(\mathbf{M}) = \det(\mathbf{M}^\top)$.

   Hence, all properties above also hold for rows.
8. For any two $n \times n$ matrices $\mathbf{M}$ and $\mathbf{N}$,

   $$\det(\mathbf{MN}) = \det(\mathbf{M}) \det(\mathbf{N}).$$

   In particular, if $\mathbf{M}$ is invertible, then

$$\det\left(\mathbf{M}^{-1}\right) = (\det(\mathbf{M}))^{-1}.$$

9. If $\mathbf{M}$ is orthogonal, then $\det(\mathbf{M}) = \pm 1$.
10. If $\mathbf{M}$ is unitary, then $|\det(\mathbf{M})| = 1$.

Proof. For (5), Suppose $\mathbf{C}_i(\mathbf{M}) + \lambda \mathbf{C}_j(\mathbf{M}) = 0$ for some $i \neq j$ and scalar λ. Define $\mathbf{N}$ given by

$$N_{is} = \begin{cases} M_{is} & \text{if } s \neq i \\ M_{is} + \lambda M_{js} & \text{if } s = i \end{cases}.$$

Then

$$\det(\mathbf{N}) = \det(\mathbf{M}) + \underbrace{\lambda \det(\text{matrix with column } i = \text{column } j)}_{=0} = \det(\mathbf{M}).$$

Then, the ith column of $\mathbf{N}$ is all zeros. And thus $\det(\mathbf{N}) = 0 = \det(\mathbf{M})$.

Lecture 15 · 2025-11-11

For (7), take a single term $M_{\rho(1)1}, \ldots, M_{\rho(n)n}$, and a σ in S_n. We have

$$M_{\rho(1)1} M_{\rho(2)2} \ldots M_{\rho(n)n} = M_{\rho(\sigma(1))\sigma(1)} M_{\rho(\sigma(2))\sigma(2)} \ldots M_{\rho(\sigma(n))\sigma(n)}).$$

Take $\rho = \sigma^{-1}$. Since $\varepsilon(\rho) = \varepsilon(\sigma)$, we have

$$\begin{aligned} \det \mathbf{M} &= \sum_{\sigma \in S_n} \varepsilon(\sigma) M_{1\sigma(1)}, \ldots, M_{n\sigma(n)} \\ &= \det(\mathbf{M}^\top). \end{aligned}$$

For (8), note that swapping columns an even/odd number of times introduces a factor of ± 1. Hence,

$$\begin{aligned} \det(\mathbf{MN}) &= \sum_{\sigma} \varepsilon(\sigma) (\mathbf{MN})_{\sigma(1)1} (\mathbf{MN})_{\sigma(2)2} \ldots (\mathbf{MN})_{\sigma(n)n} \\ &= \sum_{\sigma} \varepsilon(\sigma) \sum_{k_1 \cdots k_n = 1}^{n} M_{\sigma(1)k_1} N_{k_1 1} \ldots M_{\sigma(n)k_n} N_{k_n n} \\ &= \sum_{k_1 \cdots k_n = 1}^{n} N_{k_1 1} \ldots N_{k_n n} \underbrace{\left[\sum_{\sigma} \varepsilon(\sigma) M_{\sigma(1)k_1} \ldots M_{\sigma(n)k_n} \right]}_{S}. \end{aligned}$$

If in S two indices $k_p = k_q$ for some $p \neq q$, then $S = 0$ by property (4). Hence, we only need to consider the case where $k_1, \ldots, k_n$ are all distinct. This means that there exists a permutation ρ such that $k_i = \rho(i)$ for all i. Thus,

$$\begin{aligned} S &= \sum_{\sigma} \varepsilon(\sigma) M_{\sigma(1)\rho(1)} \ldots M_{\sigma(n)\rho(n)} \\ &= \varepsilon(\rho) \det(\mathbf{M}). \end{aligned}$$

Therefore,

$$\det(\mathbf{MN}) = \sum_{\rho} N_{\rho(1)1} \dots N_{\rho(n)n} \varepsilon(\rho) \det(\mathbf{M})$$
$$= \det(\mathbf{M}) \det(\mathbf{N}).$$

For (9), if $\mathbf{M}$ is orthogonal, then $\mathbf{M}^\top \mathbf{M} = \mathbf{I}$, and thus

$$\det(\mathbf{M}^\top) \det(\mathbf{M}) = \det(\mathbf{I}) = 1.$$

Hence, $\det(\mathbf{M}) = \pm 1$.

For (10), if $\mathbf{M}$ is unitary, then $\mathbf{M}^\dagger \mathbf{M} = \mathbf{I}$, and thus

$$\det\left(\mathbf{M}^\dagger\right) \det(\mathbf{M}) = \det(\mathbf{I}) = 1.$$

Hence, $|\det(\mathbf{M})| = 1$.

3.5.6 Minors and Cofactors

We want to find a way to compute determinants of $n \times n$ matrices in an efficient way. We do this by defining minors and cofactors.

Definition 3.38 (Minor)

For an $n \times n$ matrix $\mathbf{M}$, consider the $(n-1) \times (n-1)$ matrix in row i and column j obtained by deleting row i and column j from $\mathbf{M}$. The determinant of this $(n-1) \times (n-1)$ matrix is called the **minor** of entry M_{ij} and is denoted by M^{ij}.

Definition 3.39 (Cofactor)

For an $n \times n$ matrix $\mathbf{M}$, the **cofactor** of entry M_{ij} is defined by

$$\Delta_{ij} = (-1)^{i+j} M^{ij}.$$

Consider the columns and rows of $\mathbf{M}$ given by

$$\boldsymbol{C}_j = \sum_i M_{ij} \boldsymbol{e}_i, \quad \boldsymbol{R}_i = \sum_j M_{ij} \boldsymbol{e}_j.$$

Then, the determinant of $\mathbf{M}$ can be written as (see proof in Theorem 3.40):

$$\det \mathbf{M} = [\boldsymbol{C}_1, \dots, \boldsymbol{C}_n] = \sum_i M_{ij} \Delta_{ij} = \sum_j M_{ij} \Delta_{ij}.$$

We have

$$
\begin{aligned}
\Delta_{ij} &= \left[\mathbf{C}_1, \ldots, \mathbf{C}_{j-1}, \mathbf{e}_i, \mathbf{C}_{j+1}, \ldots, \mathbf{C}_n\right] \\
&= \left[\mathbf{R}_1, \ldots, \mathbf{R}_{i-1}, \mathbf{e}_j, \mathbf{R}_{i+1}, \ldots, \mathbf{R}_n\right] \\
&= \begin{vmatrix}
M_{11} & \cdots & M_{1(j-1)} & 0 & M_{1(j+1)} & \cdots & M_{1n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M_{(i-1)1} & \cdots & M_{(i-1)(j-1)} & 0 & M_{(i-1)(j+1)} & \cdots & M_{(i-1)n} \\
0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
M_{(i+1)1} & \cdots & M_{(i+1)(j-1)} & 0 & M_{(i+1)(j+1)} & \cdots & M_{(i+1)n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
M_{n1} & \cdots & M_{n(j-1)} & 0 & M_{n(j+1)} & \cdots & M_{nn}
\end{vmatrix}.
\end{aligned}
$$

i.e. the cofactor Δ_{ij} is the determinant of the matrix obtained from $\mathbf{M}$ by replacing the entry M_{ij} with 1 and all other entries in row i and column j with 0.

Hence, we can write the determinant of $\mathbf{M}$ as

$$\det \mathbf{M} = \sum_i M_{ij}\Delta_{ij} = \sum_i M_{ij}(-1)^{i+j}M^{ij}$$

for any fixed column j. Alternatively, we can write

$$\det \mathbf{M} = \sum_j M_{ij}\Delta_{ij} = \sum_j M_{ij}(-1)^{i+j}M^{ij}$$

for any fixed row i.

Theorem 3.40 (Laplace expansion formula)

Consider $\mathbf{M}$ an $n \times n$ matrix. Then, for any fixed j,

$$\det \mathbf{M} = \sum_{i=1}^{n} M_{ij}\Delta_{ij}.$$

Proof.

Notation. For an object j, we write $\left\{1, \ldots, \bar{j}, \ldots, n\right\}$ to be the set of indices $\{1, 2, \ldots, n\} \setminus \{j\}$.

We have

$$
\begin{aligned}
\det \mathbf{M} &= \sum_{i_1 \ldots i_n = 1}^{n} \varepsilon_{i_1 \ldots i_n} M_{i_1 1} \ldots M_{i_n n} \\
&= \sum_{i_j = 1}^{n} M_{i_j j} \sum_{i_1 \ldots \bar{i_j} \ldots i_n = 1}^{n} \varepsilon_{i_1 \ldots i_n} M_{i_1 1} \ldots \overline{M_{i_j j}} \ldots M_{i_n n}.
\end{aligned}
$$

Consider $\sigma \in S_n$ the permutation that moves i_j to the jth position, and leaves everything else in its natural order:

$$\sigma = \begin{pmatrix} 1 & \dots & j & j+1 & j+2 & \dots & i_j - 1 & i_j & i_j + 1 & \dots & n \\ 1 & \dots & i_j & j & j+1 & \dots & i_j - 2 & i_j - 1 & i_j + 1 & \dots & n \end{pmatrix}$$

Assume $i_j > j$ (we can do a similar argument for $i_j < j$). Since we have to perform $\left|j - i_j\right|$ transpositions for σ, $\varepsilon(\sigma) = (-1)^{j-i_j}$. Now consider the permutation $\rho \in S_{n-1}$,

$$\rho = \begin{pmatrix} 1 & \dots & \overline{i_j} & \dots & n \\ i_1 & \dots & \overline{i_j} & \dots & i_n \end{pmatrix}$$

Note that $\rho\sigma$ reorders $(1, \dots, n)$ to $(i_1, \dots, i_n)$. Thus,

$$\begin{aligned} \sigma(\rho\sigma) = \varepsilon_{i_1 \dots i_n} &= \varepsilon(\rho)\varepsilon(\sigma) \\ &= (-1)^{j-i_j} \varepsilon_{i_1 \dots \overline{i_j} \dots i_n}. \end{aligned}$$

Hence, we can rewrite

$$\begin{aligned} \det \mathbf{M} &= \sum_{i_j=1}^{n} M_{i_j j} \sum_{i_1 \dots \overline{i_j} \dots i_n = 1}^{n} (-1)^{j-i_j} \varepsilon_{i_1 \dots \overline{i_j} \dots i_n} M_{i_1 1} \dots \overline{M_{i_j j}} \dots M_{i_n n} \\ &= \sum_{i_j=1}^{n} M_{i_j j} (-1)^{j-i_j} M^{i_j j} \\ &= \sum_{i_j=1}^{n} M_{i_j j} \Delta_{i_j j} \\ &= \sum_{i=1}^{n} M_{ij} \Delta_{ij}. \end{aligned}$$

Definition 3.41 (Adjugate matrix)

Reasoning as above, if $\boldsymbol{C}_k = \sum_i M_{ik} \boldsymbol{e}_i$ then

$$\left[\boldsymbol{C}_1, \dots, \boldsymbol{C}_{j-1}, \boldsymbol{C}_k, \boldsymbol{C}_{j+1}, \dots, \boldsymbol{C}_n\right] = \sum_i M_{ik} \Delta_{ij} = \begin{cases} \det \mathbf{M} & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}.$$

Hence

$$\begin{aligned} \sum_i M_{ik} \Delta_{ij} &= (\det \mathbf{M}) \delta_{jk} \\ \sum_j M_{lj} \Delta_{ij} &= (\det \mathbf{M}) \delta_{li}. \end{aligned}$$

Lecture 16 · 2025-11-13

The **adjugate** of a matrix is defined to be

$$\tilde{\mathbf{M}} = \operatorname{adj}(\mathbf{M}) = \boldsymbol{\Delta}^{\mathsf{T}}.$$

where $\boldsymbol{\Delta}$ is the matrix with entries of cofactors Δ_{ij}.

Remark. From the expression above, note that

$$\tilde{\mathbf{M}}\mathbf{M} = \mathbf{M}\tilde{\mathbf{M}} = (\det \mathbf{M})\mathbf{I}.$$

and if $\det \mathbf{M} \neq 0$, then

$$\mathbf{M}^{-1} = \frac{\tilde{\mathbf{M}}}{\det \mathbf{M}}.$$

This suggests a way to compute the inverse of a matrix using only determinants of smaller matrices.

Example 3.42

Consider the matrix

$$\mathbf{M} = \begin{pmatrix} 1 & x & 1 \\ 1 & 1 & x \\ x & 1 & 1 \end{pmatrix}$$

for some arbitrary scalar $x \in \mathbb{R}$. We want to compute $\det \mathbf{M}$.

By the fact that determinants are conserved under operations of the form $\boldsymbol{C}_i \to \boldsymbol{C}_i + \lambda \boldsymbol{C}_j$,

$$\begin{aligned}
\det \mathbf{M} &= \begin{vmatrix} 1 & x & 1 \\ 1 & 1 & x \\ x & 1 & 1 \end{vmatrix} \\
&= \begin{vmatrix} 0 & x & 1 \\ 1-x & 1 & x \\ x-1 & 1 & 1 \end{vmatrix} && \text{by } \boldsymbol{C}_1 \to \boldsymbol{C}_1 - \boldsymbol{C}_3 \\
&= \begin{vmatrix} 0 & x & 1 \\ 2-2x & 0 & x-1 \\ x-1 & 1 & 1 \end{vmatrix} && \text{by } \boldsymbol{R}_2 \to \boldsymbol{R}_2 - \boldsymbol{R}_3 \\
&= \begin{vmatrix} 0 & x & 1 \\ 2-2x & 0 & x-1 \\ x-1 & 1-x & 0 \end{vmatrix} && \text{by } \boldsymbol{R}_3 \to \boldsymbol{R}_3 - \boldsymbol{R}_1 \\
&= (x-1)^2 \begin{vmatrix} 0 & x & 1 \\ -2 & 0 & 1 \\ 1 & -1 & 0 \end{vmatrix} && \text{by scaling in } \boldsymbol{R}_2 \text{ and } \boldsymbol{R}_3 \\
&= (x-1)^2 \begin{vmatrix} 0 & x+2 & 0 \\ -2 & 0 & 1 \\ 1 & -1 & 0 \end{vmatrix} && \text{by } \boldsymbol{R}_1 \to \boldsymbol{R}_1 - 2\boldsymbol{R}_3 - \boldsymbol{R}_2 \\
&= (x-1)^2(x+2) \begin{vmatrix} 0 & 1 & 0 \\ -2 & 0 & 1 \\ 1 & -1 & 0 \end{vmatrix} && \text{by scaling in } \boldsymbol{R}_1 \\
&= (x-1)^2(x+2)(-1) && \text{by direct computation} \\
&= -(x-1)^2(x+2).
\end{aligned}$$

3.6 Systems of Linear Equations

3.6.1 2×2 Case

Consider the system of equations given by

$$\begin{cases} A_{11}x_1 + A_{12}x_2 = b_1 & (1) \\ A_{21}x_1 + A_{22}x_2 = b_2 & (2) \end{cases}$$

We can write this system in matrix form as

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

where

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Consider $(1) \times A_{22} - (2) - A_{12}$, we have

$$(A_{11}A_{22} - A_{21}A_{12})x_1 = b_1A_{22} - b_2A_{12}.$$

Similarly, consider $(2) \times A_{11} - (1) \times A_{21}$, we have

$$(A_{11}A_{22} - A_{21}A_{12})x_2 = b_2A_{11} - b_1A_{21}.$$

Note that $\det \mathbf{A}$ is $A_{11}A_{22} - A_{21}A_{12}$. Thus, we can write

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Equivalently, given $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, if $\mathbf{A}^{-1}$ exists, we can write

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$$

3.6.2 General Case

Consider a system of n linear equations in n unknown x_o written is matrix form as

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

where $\mathbf{A}$ is an $n \times n$ matrix, $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$.

We shall consider three possible scenarios.

1. If $\det \mathbf{A} \neq 0$, then $\mathbf{A}^{-1}$ exists, and therefore there is a unique solution given by

$$\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}.$$

2. If $\det \mathbf{A} = 0$ and $\mathbf{b} \notin \operatorname{Im} \mathbf{A}$, then there is no solution.
3. If $\det \mathbf{A} = 0$ and $\mathbf{b} \in \operatorname{Im} \mathbf{A}$, then there are infinitely many solutions. We can find these solutions by considering

$$\boldsymbol{x} + \boldsymbol{x}_0 + \boldsymbol{u}$$

   where $\boldsymbol{x}_0$ is a particular solution to the system, and $\boldsymbol{u} \in \ker \mathbf{A}$.

In more detail, a solution exists for

$$\mathbf{A}\boldsymbol{x}_0 = \boldsymbol{b}$$

if and only if we can find $\mathbf{A}\boldsymbol{x}_0 = \boldsymbol{b}$ for some $\boldsymbol{x}_0 \in \mathbb{R}^n$. This is equivalent to saying that $\boldsymbol{b} \in \operatorname{Im}\mathbf{A}$. Then, $\boldsymbol{x}$ is also a solution if and only if

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}_0$$

satisfies

$$\mathbf{A}\boldsymbol{u} = \mathbf{0}.$$

Thus, the general solution is given by

$$\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{u}$$

for any $\boldsymbol{u} \in \ker \mathbf{A}$.

Remark. In the first case, note that

$$\det \mathbf{A} \neq 0 \Leftrightarrow \operatorname{Im}\mathbf{A} = \mathbb{R}^n \Leftrightarrow \ker \mathbf{A} = \{\mathbf{0}\}.$$

In this case, if $\mathbf{A}\boldsymbol{u} = \mathbf{0}$ then we must have $\boldsymbol{u} = \mathbf{0}$. Hence there is a unique solution.

For the other cases,

$$\det \mathbf{A} = \mathbf{0} \Leftrightarrow \operatorname{Im}\mathbf{A} \neq \mathbb{R}^n \Leftrightarrow \ker \mathbf{A} \neq \{\mathbf{0}\}.$$

and thus either

$$\begin{cases} \boldsymbol{b} \notin \operatorname{Im}\mathbf{A} & \text{as in (2)} \\ \boldsymbol{b} \in \operatorname{Im}\mathbf{A} & \text{as in (3)} \end{cases}$$

If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is a basis for $\ker \mathbf{A}$, then the general solution for $\mathbf{A}\boldsymbol{u} = \mathbf{0}$ is

$$\boldsymbol{u} = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i$$

for any scalars $\lambda_1, \ldots, \lambda_k$, where $k = \operatorname{null}\mathbf{A}$.

Example 3.43

Consider the equation

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

with

$$\mathbf{A} = \begin{pmatrix} 1 & x & 1 \\ 1 & 1 & x \\ x & 1 & 1 \end{pmatrix}$$

and

$$\boldsymbol{b} = \begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix}.$$

where $x, y \in \mathbb{R}$ are some scalars. We saw before that

$$\det \mathbf{A} = -(x-1)^2(x+2).$$

1. Assume $\det \mathbf{A} \neq 0 \Leftrightarrow x \neq 1, -2$. Then $\mathbf{A}^{-1}$ exists, and we can construct it from the matrix of cofactors.

$$\mathbf{A}^{-1} = \frac{\boldsymbol{\Delta}^{\mathsf{T}}}{\det \mathbf{A}}.$$

[It can be computed that $\boldsymbol{\Delta} = \begin{pmatrix} 1-x & x^2-1 & 1-x \\ 1-x & 1-x & x^2-1 \\ x^2-1 & 1-x & 1-x \end{pmatrix}$.]

We have

$$\boldsymbol{\Delta}^{\mathsf{T}} = \begin{pmatrix} 1-x & 1-x & x^2-1 \\ x^2-1 & 1-x & 1-x \\ 1-x & x^2-1 & 1-x \end{pmatrix}.$$

Note that $x^2 - 1 = (x+1)(x-1)$. This indicates that we can simplify our matrix. Hence, the solution to the equation is

$$\begin{aligned} \boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b} &= \frac{1}{(1-x)(x+2)} \begin{pmatrix} 1 & 1 & -x-1 \\ -x-1 & 1 & 1 \\ 1 & -x-1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix} \\ &= \frac{1}{(1-x)(x+2)} \begin{pmatrix} y-x \\ -x+y \\ 2-xy-y \end{pmatrix}. \end{aligned}$$

The solution is a point in $\mathbb{R}^3$.

2. Assume that $x = 1$, then

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

and then $\mathrm{Im}(\mathbf{A}) = \mathrm{span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}$ with $\ker \mathbf{A} = \mathrm{span}\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right\}$.

The image suggests that we must have $y = 1$ to have a solution. In this case, one particular solution is given by $\boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Hence, the general solution is given by

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

for any scalars $\lambda, \mu \in \mathbb{R}$, *i.e.*

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

If $y \neq 1$, then there is no solution.

3. The case $x = -2$ is similar to case 2.

Lecture 17 · 2025-11-15

3.6.3 The Homogeneous Case – Geometrical Interpretation

Consider the equation

$$\mathbf{A}\mathbf{u} = \mathbf{0}.$$

Then, if $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ are the rows of $\mathbf{A}$, then

$$\mathbf{A}\mathbf{u} = \mathbf{0} \Leftrightarrow \begin{cases} \mathbf{R}_1 \cdot \mathbf{u} = 0 \\ \mathbf{R}_2 \cdot \mathbf{u} = 0 \\ \mathbf{R}_3 \cdot \mathbf{u} = 0 \end{cases}$$

Each equation represents a plane in $\mathbb{R}^3$ that passes through the origin with normal $\mathbf{R}_i$. The solution to the system, which is $\ker \mathbf{A}$, is the intersection of these planes.

The possible scenarios are as follows:

1. $\operatorname{rank} \mathbf{A} = 3 \Leftrightarrow \operatorname{null} \mathbf{A} = 0$, so $\ker \mathbf{A} = \{\mathbf{0}\}$. This means that all the normals of the three planes are linearly independent, and thus the only intersection point is the origin.
2. $\operatorname{rank} \mathbf{A} = 2 \Leftrightarrow \operatorname{null} \mathbf{A} = 1$. The intersection of the three planes is a line through the origin, and the three normals span a plane.
3. $\operatorname{rank} \mathbf{A} = 1 \Leftrightarrow \operatorname{null} \mathbf{A} = 2$. The intersection of the three planes is a plane through the origin, so all three planes coincide. In this case, all normals are parallel.

3.6.4 The General Case – Geometrical Interpretation

Consider the equation

$$\mathbf{A}\mathbf{u} = \mathbf{b}.$$

Then,

$$\mathbf{A}\mathbf{u} = \mathbf{b} \Leftrightarrow \begin{cases} \mathbf{R}_1 \cdot \mathbf{u} = b_1 \\ \mathbf{R}_2 \cdot \mathbf{u} = b_2 \\ \mathbf{R}_3 \cdot \mathbf{u} = b_3 \end{cases}$$

These are three planes in $\mathbb{R}^3$ with normals $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$, and in general do not pass through the origin.

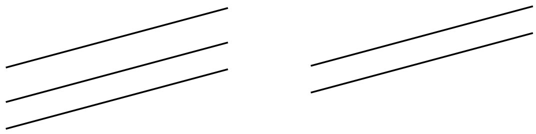
The possible scenarios are as follows:

1. $\operatorname{rank} \mathbf{A} = 3 \Leftrightarrow \det \mathbf{A} \neq 0$. All the normals are linearly independent, and thus the three planes intersect at a single point. There is a unique solution for any $\mathbf{b} \in \mathbb{R}^3$.
2. $\operatorname{rank} \mathbf{A} < 3 \Leftrightarrow \det \mathbf{A} = 0$.

The existence of solutions depends on $\boldsymbol{b}$. More specifically, whether $\boldsymbol{b}$ is in the image of $\mathbf{A}$.

- if $\operatorname{rank}\mathbf{A} = 2$, then the planes may intersect in a line as in the homogeneous case, or there is no solution.

- if $\operatorname{rank}\mathbf{A} = 1$, then either all three planes coincide as in the homogeneous case, or there is no solution.



3.6.5 Gaussian Elimination

Consider a system of m equations in n unknowns:

$$\begin{cases} A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \dots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + A_{m2}x_2 + \dots + A_{mn}x_n &= b_m \end{cases}$$

WLOG, we can assume that $A_{11} \neq 0$ (since we can always swap rows).

> ***Notation.*** We will use a superscript (i) to denote the value in the ith step of the algorithm.

Step 1. We subtract multiples of the first equation from all other equations to make the coefficients of x_1 zero in all equations except the first one.

$$\begin{cases} A_{11}^{(1)}x_1 + A_{12}^{(1)}x_2 + \dots + A_{1n}^{(1)}x_n &= b_1^{(1)} \\ \qquad\qquad A_{22}^{(1)}x_2 + \dots + A_{2n}^{(1)}x_n &= b_2^{(1)} \\ &\vdots \\ \qquad\qquad A_{m2}^{(1)}x_2 + \dots + A_{mn}^{(1)}x_n &= b_m^{(1)} \end{cases}$$

Step 2. Repeat (1) for A_{22}^1 and coefficients of x_2 in all equations except the second one, and so on.

$$\begin{cases} A_{11}^{(1)}x_1 + A_{12}^{(1)}x_2 + A_{13}^{(1)}x_3 + \dots + A_{1n}^{(1)}x_n &= b_1^{(1)} \\ \qquad\qquad A_{22}^{(2)}x_2 + A_{23}^{(2)}x_3 + \dots + A_{2n}^{(2)}x_n &= b_2^{(2)} \\ \qquad\qquad\qquad\qquad A_{33}^{(3)}x_3 + \dots + A_{3n}^{(3)}x_n &= b_3^{(3)} \\ &\vdots \\ \qquad\qquad\qquad\qquad A_{rr}^{(r)}x_r + \dots + A_{rn}^{(r)}x_n &= b_r^{(r)} \\ 0 &= b_{r+1}^{(r)} \\ &\vdots \\ 0 &= b_m^{(r)} \end{cases}$$

In all these equations, $A_{ii}^i \neq 0$.

The possible cases are as follows.

1. $r = n \leqslant m$ and $b_i^r = 0$ for all $i = r+1, \ldots, m$. Then, there is a unique solution. To obtain it, we can first find x_n from the nth equation, then substitute it into the $(n-1)$th equation to find x_{n-1}, and so on.
2. $r < n$ and $b_i^r \neq 0$ for some $i = r+1, \ldots, m$. Then, there is no solution.
3. $r = m$ (and not necessarily $n = m$. Then $x_{r+1} \ldots x_n$ are undetermined. So, given any values of $b_1, \ldots, b_r$, we can solve $x_1, \ldots, x_r$. Then, there are infinitely many solutions given by varying $x_{r+1}, \ldots, x_n$.

Note that this algorithm can also be written in matrix form by

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

where $\mathbf{A}$ is an $m \times n$ matrix. This algorithm can be reexpressed to obtain

$$\mathbf{M}\boldsymbol{x} = \boldsymbol{d}$$

with

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & \ldots & M_{1r} & \ldots & M_{1n} \\ 0 & M_{22} & \ldots & M_{2r} & \ldots & M_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & M_{rr} & \ldots & M_{rn} \\ 0 & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & \ldots & 0 \end{pmatrix}$$

which is called the **row echelon form** of $\mathbf{A}$. Note that

- the first $r \times r$ block is upper triangular with non-zero entries on the diagonal.
- $r = \operatorname{rank} \mathbf{M} = \operatorname{rank} \mathbf{A}$
- if $n = m$, $\det \mathbf{A} \pm \det \mathbf{M}$, and if $n = m = r$, then $\det \mathbf{A} = \det \mathbf{M} = M_{11} M_{22} \ldots M_{nn} \neq 0$. Then, both $\mathbf{A}$ and $\mathbf{M}$ are invertible.

4 Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors can be used to analyse and simplify matrices.

4.1 Introduction

Theorem 4.1 (Fundamental Theorem of Algebra)

Let $p(z)$ be a polynomial of degree $m \geqslant 1$. Then

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

where $c_j \in \mathbb{C}$ and $c_m \neq 0$.

Then $p(z) = 0$ has precisely m roots in $\mathbb{C}$ (counting with multiplicities).

Definition 4.2 (Multiplicity of a Root)

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

Definition 4.3 (Eigenvector and Eigenvalue)

Let $T : V \to V$ (for a real or complex vector space V) be a linear map. Then, a vector $\boldsymbol{v} \in V$ with $\boldsymbol{v} \neq \mathbf{0}$ is an **eigenvector** of T if there exists a scalar $\lambda \in \mathbb{R}$ (or $\mathbb{C}$) such that

$$T(\mathbf{v}) = \lambda \mathbf{v}.$$

The scalar λ is called the **eigenvalue** corresponding to the eigenvector $\boldsymbol{v}$.

If $V \in \mathbb{R}^n$ or $\mathbb{C}^n$, and T is given in terms of a $n \times n$ matrix $\mathbf{A}$, then

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

and for a given λ, this holds for some vector $\mathbf{v} \neq 0$ if and only if $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$. This is called the **characteristic equation** of the matrix $\mathbf{A}$.

Furthermore, the polynomial $\chi_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$ is called the **characteristic polynomial** of degree n of the matrix $\mathbf{A}$.

Lecture 18 · 2025-11-18

Remark. From the definition of the determinant,

$$\begin{aligned}
\chi_{\mathbf{A}}(t) &= \det(\mathbf{A} - t\mathbf{I}) \\
&= \varepsilon_{j_1,\dots,j_n}\left(A_{j_1 1} - t\delta_{j_1 1}\right)\dots\left(A_{j_n n} - t\delta_{j_n n}\right) \\
&= c_0 + c_1 t + c_2 t^2 + \dots + c_n t^n
\end{aligned}$$

for some coefficients $c_0, c_1, \dots, c_n$. From here we can conclude

1. $\chi_{\mathbf{A}}(t)$ has degree n, and thus n roots by the fundamental theorem of algebra. Hence, an $n \times n$ matrix has n eigenvalues (counting multiplicities).
2. If $\mathbf{A}, \mathbf{B}$ are real, then the coefficients $c_0, c_1, \ldots, c_n$ are real, and thus the eigenvalues are either real or come in complex conjugate pairs.
3. $c_n = (-1)^n$ and $c_{n-1} = (-1)^{n-1}(\operatorname{tr} \mathbf{A})$. By Vieta's formulas, the sum of the eigenvalues is equal to the trace of the matrix:
$$\sum_{i=1}^{n} \lambda_i = \operatorname{tr} \mathbf{A}.$$
4. Finally,
$$c_0 = \chi_{\mathbf{A}}(0) = \det \mathbf{A}.$$
By Vieta's formulas, the product of the eigenvalues is equal to the determinant of the matrix:
$$\prod_{i=1}^{n} \lambda_i = \det \mathbf{A}.$$

Example 4.4

1. Consider $V = \mathbb{C}^2$ and $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ [representing a 90° rotation]. Then
$$\chi_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I}) = \det\begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} = t^2 + 1.$$
Hence the eigenvalues are $\lambda_1 = \mathrm{i}$ and $\lambda_2 = -\mathrm{i}$. To find eigenvectors, for $\lambda = \mathrm{i}$, we have
$$\begin{pmatrix} -\mathrm{i} & -1 \\ 1 & -\mathrm{i} \end{pmatrix}\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \alpha\begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix}.$$
For $\lambda = -\mathrm{i}$, we have
$$\begin{pmatrix} \mathrm{i} & -1 \\ 1 & \mathrm{i} \end{pmatrix}\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \beta\begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix}.$$
2. Consider $V = \mathbb{R}^2$ with $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then
$$\chi_{\mathbf{A}}(t) = (t-1)^2.$$
Hence the only eigenvalue is $\lambda = 1$ with multiplicity 2. To find eigenvectors, we have
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \alpha\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
for any $\alpha \in \mathbb{R}$.

4.2 Eigenspaces and Multiplicity

Definition 4.5 (Eigenspace)

For an eigenvalue λ of a matrix $\mathbf{A}$, we define its **eigenspace** as

$$E_\lambda = \{\mathbf{v} : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\} = \ker(\mathbf{A} - \lambda\mathbf{I}).$$

Definition 4.6 (Algebraic multiplicity)

The **algebraic multiplicity** of an eigenvalue λ, $M(\lambda)$ or M_λ, is its multiplicity as a root of the characteristic polynomial $\chi_{\mathbf{A}}(\lambda)$.

By the fundamental theorem of algebra, the sum of the algebraic multiplicities of all eigenvalues of an $n \times n$ matrix is n.

Definition 4.7 (Geometric multiplicity)

The **geometric multiplicity** of an eigenvalue λ, $m(\lambda)$ or m_λ, is the dimension of its eigenspace E_λ, *i.e.* the maximum number of linearly independent eigenvectors corresponding to λ.

Proposition 4.8

Consider $\mathbf{A}$ an $n \times n$ matrix, and λ an eigenvalue of $\mathbf{A}$. Then

$$m_\lambda \leqslant M_\lambda \leqslant n.$$

Definition 4.9 (Defect)

The **defect** of an eigenvalue λ is defined as

$$\Delta_\lambda = M_\lambda - m_\lambda.$$

By Proposition 4.8, $\Delta_\lambda \geqslant 0$.

Example 4.10

1. Consider

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix}, \quad \chi_{\mathbf{A}}(t) = (4-t)^3.$$

The eigenvalue is $\lambda = 4$ with algebraic multiplicity $M_4 = 3$. To find the eigenspace, we solve

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} = \alpha\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore, eigenvector is $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ with geometric multiplicity $m_4 = 1$.

Eigenspace is $E_4 = \operatorname{span}\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$ with $\dim E_4 = 1$.

2. Consider a reflection matrix in $\mathbb{R}^3$ in plane through $\mathbf{0}$ with normal $\boldsymbol{n}$. Then we have

$$\mathbf{H}\boldsymbol{n} = -\boldsymbol{n}, \quad \mathbf{H}\boldsymbol{v} = 1\boldsymbol{v} \quad \forall \boldsymbol{v} \perp \boldsymbol{n}.$$

Hence the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 1$. We have

$$\begin{aligned} E_{-1} &= \{\alpha \boldsymbol{n}\} & M_{-1} &= m_{-1} = 1. \\ E_1 &= \{\boldsymbol{x} : \boldsymbol{x} \cdot \boldsymbol{n} = 0\} & M_1 &= m_1 = 2. \end{aligned}$$

3. Consider a rotation in $\mathbb{R}^2$

$$\mathbf{Rot}(\theta) = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}.$$

We have

$$\begin{aligned} \chi_{\mathbf{Rot}(\theta)}(t) &= t^2 - 2(\cos\theta)t + 1 \\ \lambda_{1,2} &= \mathrm{e}^{\pm \mathrm{i}\theta} \\ \boldsymbol{v} &= \alpha\begin{pmatrix}1\\ \mp\mathrm{i}\end{pmatrix}. \end{aligned}$$

4. Consider a rotation by angle θ about $\boldsymbol{n}$. Then

$$\mathbf{Rot}(\theta, \boldsymbol{n})\boldsymbol{n} = \boldsymbol{n},$$

and we have an eigenvalue $\lambda = 1$ with eigenspace

$$E_1 = \operatorname{span}\{\boldsymbol{n}\}.$$

There are no other real eigenvalues unless $\theta = k\pi$ for some integer k. A rotation restricted to the plane that is perpendicular to $\boldsymbol{n}$ has eigenvalues $\mathrm{e}^{\pm\mathrm{i}\theta}$.

5. Consider the matrix

$$\mathbf{A} = \begin{pmatrix}-3 & -1 & 1\\ -1 & -3 & 1\\ -2 & -2 & 0\end{pmatrix}, \quad \chi_{\mathbf{A}}(t) = -(t+2)^3.$$

Then the only eigenvalue is $\lambda = -2$ with algebraic multiplicity $M_{-2} = 3$. To find the eigenspace, we solve

$$\begin{pmatrix}-1 & -1 & 1\\ -1 & -1 & 1\\ -2 & -2 & 2\end{pmatrix}\boldsymbol{v} = \mathbf{0}$$

and we have a general solution $\begin{pmatrix}x\\y\\x+y\end{pmatrix}$. Therefore, the eigenspace is

$$E_{-2} = \operatorname{span}\left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right\}$$

with geometric multiplicity $m_{-2} = 2$.

The defect is $\Delta_{-2} = 3 - 2 = 1$. The eigenvectors do not form a basis for $\mathbb{C}^3$.

4.3 Diagonolisation and Similarity

Proposition 4.11

For an $n \times n$ matirx $\mathbf{A}$ acting on $V = \mathbb{R}^n$ or $\mathbb{C}^n$, the following are equivalent:

- There exists a basis of V consisting of eigenvectors of $\mathbf{A}$. *i.e.* we have $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ where
$$\mathbf{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$
for some eigenvalue λ_i.
- $\mathbf{A}$ is diagonalisable, *i.e.* there exists an $n \times n$ invertible matrix $\mathbf{P}$ such that
$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$
where $\mathbf{D}$ is a diagonal matrix, with the eigenvalues of $\mathbf{A}$ on the diagonal:
$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Definition 4.12 (Diagonalisable Matrix)

An $n \times n$ matrix $\mathbf{A}$ is called **diagonalisable** if it satisfies the conditions of Proposition 4.11.

We will prove Proposition 4.11 in the following section.

Lecture 19 · 2025-11-20

4.3.1 Linearly Independent Eigenvectors

Theorem 4.13

Suppose that an $n \times n$ matrix $\mathbf{M}$ has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$. Then the corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$ are linearly independent.

Remark. Let B_λ be a basis for eigenspace E_λ associated to λ. If $\lambda_1, \ldots, \lambda_r$ are distinct, then $B_{\lambda_1} \cup B_{\lambda_2} \cup \ldots \cup B_{\lambda_r}$ is linearly independent.

Proof. We shall prove this by contradiction. Suppose that $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$ are linearly dependent, such that

$$\sum_{j=1}^{r} \alpha_j \mathbf{v}_j = \mathbf{0}$$

for some scalars α_j, not all zero.

Take the minimal p for which $\exists \alpha_1, ..., \alpha_p \neq 0$ with [reordering if necessary]

$$\sum_{j=1}^{p} \alpha_j \mathbf{v}_j = \mathbf{0}.$$

Then, applying $\mathbf{A} - \lambda_1 \mathbf{I}$ gives

$$(\mathbf{A} - \lambda_1 \mathbf{I}) \sum_{j=1}^{p} \alpha_j \mathbf{v}_j = \sum_{j>1} \alpha_j \left(\lambda_j - \lambda_1\right) \mathbf{v}_j = \mathbf{0}.$$

which is a linear combination of eigenvectors with $p - 1$ non-zero coefficients. This contradicts the minimality of p. ※

Now, we can prove Proposition 4.11.

Proof. [of Proposition 4.11]

For any matrix $\mathbf{P}$,

- $\mathbf{AP}$ has columns $\mathbf{A}\boldsymbol{C}_i(\mathbf{P})$
- $\mathbf{PD}$ has columns $\lambda_i \boldsymbol{C}_i(\mathbf{P})$

where $\boldsymbol{C}_i(\mathbf{P})$ is the ith column of $\mathbf{P}$.

This means that

$$\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D} \Leftrightarrow \mathbf{AP} = \mathbf{PD} \Leftrightarrow \mathbf{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad \forall i = 1, ..., n$$

where $\boldsymbol{v}_i = \boldsymbol{C}_i(\mathbf{P})$.

[$\Rightarrow$] Given a basis of eigenvectors, we can construct $\mathbf{P}$ with these eigenvectors as columns, and the above holds.

[$\Leftarrow$] Given $\mathbf{P}$ such that the above holds, the columns of $\mathbf{P}$ are eigenvectors of $\mathbf{A}$. Since $\mathbf{P}$ is invertible, its columns form a basis of V.

4.3.2 Criteria for Diagonalisability

1. [*Sufficient* but *not necessary*] An $n \times n$ matrix $\mathbf{A}$ with n distinct eigenvalues is diagonalisable.

   This implies the existence of n eigenvectors which are linearly independent, and then this provides a basis for $\mathbb{R}^n$ or $\mathbb{C}^n$.

2. [*Sufficient* and *necessary*] For any eigenvalue λ,

   $$M_\lambda = m_\lambda.$$

   If λ_i with $i = 1, ..., r$ are all the different eigenvalues of a matrix, then $B_{\lambda_1} \cup ... \cup B_{\lambda_r}$ is a linearly independent set, and its number of elements is

$$\sum_{i=1}^{r} m_{\lambda_i} = \sum_{i=1}^{r} M_{\lambda_i} = n.$$

Hence, it forms a basis of $\mathbb{R}^n$ or $\mathbb{C}^n$.

4.3.3 Similarity

Definition 4.14

We say that two matrices $\mathbf{A}$ and $\mathbf{B}$ of size $n \times n$ are similar if

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

for some invertible matrix $\mathbf{P}$.

Proposition 4.15

If $\mathbf{A}$ and $\mathbf{B}$ are similar, then,

1. $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{B})$
2. $\det(\mathbf{A}) = \det(\mathbf{B})$
3. $\chi_{\mathbf{A}}(t) = \chi_{\mathbf{B}}(t)$

So similar matrices represent the same linear map with respect to different bases.

Remark. For the articular case of

$$\mathbf{B} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

this means that $\mathbf{A}$ is diagonalisable, and then

1. $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$
2. $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
3. $\chi_{\mathbf{A}}(t) = \prod_{i=1}^{n}(\lambda_i - t)$

4.3.4 Hermitian and Symmetric Matrices

Recall that a matrix $\mathbf{A}$ is called Hermitian if $\mathbf{A} = \mathbf{A}^{\dagger}$, and symmetric if $\mathbf{A} = \mathbf{A}^{\mathsf{T}}$.

Recall that the complex inner product is defined as $\mathbf{v}^{\dagger}\mathbf{w}$. For $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}$, this reduces to the dot product $\boldsymbol{v} \cdot \boldsymbol{w}$.

Remark. If $\mathbf{A}$ is Hermitian,

$$(\mathbf{A}\mathbf{v})^{\dagger}\mathbf{w} = \mathbf{v}^{\dagger}(\mathbf{A}\mathbf{w})$$

Theorem 4.16

For a Hermitian matrix $\mathbf{A}$ of size $n \times n$,

1. Every eigenvalue of $\mathbf{A}$ is real.
2. Eigenvectors $\boldsymbol{v}, \boldsymbol{w}$ corresponding to distinct eigenvalues λ, μ are orthogonal.
3. If $\mathbf{A}$ is symmetric, then for each eigenvalue λ, we can choose a real eigenvector $\mathbf{v}$ so that (2) becomes

$$\mathbf{v}^\top \boldsymbol{w} = 0.$$

Proof.

1. Consider an eigenvector $\mathbf{v}$ with eigenvalue λ. We have

$$\begin{aligned} \mathbf{v}^\dagger(\mathbf{A}\mathbf{v}) &= (\mathbf{A}\mathbf{v})^\dagger \mathbf{v} \\ \Leftrightarrow \lambda \mathbf{v}^\dagger \mathbf{v} &= \left(\overline{\lambda} \mathbf{v}\right)^\dagger \mathbf{v}. \end{aligned}$$

   Since $\boldsymbol{v} \neq 0$, we have $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

2. Let $\boldsymbol{v}, \boldsymbol{w}$ be eigenvectors with eigenvalues λ, μ. Then

$$\begin{aligned} \mathbf{v}^\dagger(\mathbf{A}\boldsymbol{w}) &= (\mathbf{A}\boldsymbol{v})^\dagger \boldsymbol{w} \\ \Leftrightarrow \mu \mathbf{v}^\dagger \boldsymbol{w} &= \overline{\mu} \mathbf{v}^\dagger \mathbf{w} &= \mu \mathbf{v}^\dagger \boldsymbol{w} \quad (\lambda \text{ real}). \end{aligned}$$

   Since $\lambda \neq \mu$, we have $\mathbf{v}^\dagger \boldsymbol{w} = 0$.

3. We have $\mathbf{A}\boldsymbol{v} = \lambda \boldsymbol{v}$ with $\boldsymbol{v} \in \mathbb{C}^n$ and $\mathbf{A}, \lambda$ are real. Let $\boldsymbol{v} = \boldsymbol{u} + \mathrm{i}\boldsymbol{u}'$, with $\boldsymbol{u}, \boldsymbol{u}' \in \mathbb{R}^n$. Then we have

$$\begin{cases} \mathbf{A}\boldsymbol{u} = \lambda \boldsymbol{u} \\ \mathbf{A}\boldsymbol{u}' = \lambda \boldsymbol{u}' \end{cases}$$

   but $\mathbf{v} \neq \mathbf{0}$ since it is an eigenvector, so at least one of $\boldsymbol{u}, \boldsymbol{u}'$ is non-zero, and we can choose this as a real eigenvector.

Lecture 20 · 2025-11-22

4.3.5 Gram-Schmidt Orthogonalisation

Given a linearly independent set of vectors in $\mathbb{C}^n$, say $\{\boldsymbol{w}_1, ..., \boldsymbol{w}_r\}$. We can construct a sequence of sets of the form:

- $\{\boldsymbol{u}_1, \boldsymbol{w}_2', ..., \boldsymbol{w}_r'\}$
- $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{w}_3'', ..., \boldsymbol{w}_r''\}$
- ...
- $\{\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_r\}$

so that each set has the same span, each is linearly independent, and $\boldsymbol{u}_i$ are orthonormal to each other, and orthogonal to the $\boldsymbol{w}$-vectors.

We construct this as follows:

- **First step.** Let $\boldsymbol{u}_1 = \frac{w_1}{|w_1|}$ and $\boldsymbol{w}_j' = \boldsymbol{w}_j - \left(\boldsymbol{u}_1^\dagger \boldsymbol{w}_j\right)\boldsymbol{u}_1$.

  This guarantees that $|\boldsymbol{u}_1| = 1$ and $\boldsymbol{u}_1^\dagger \boldsymbol{w}_j' = 0$ for all $j \geqslant 2$.

- **Next step.** Let $\boldsymbol{u}_2 = \frac{\boldsymbol{w}_2'}{|\boldsymbol{w}_2'|}$ and $\boldsymbol{w}_j'' = \boldsymbol{w}_j' - \left(\boldsymbol{u}_2^\dagger \boldsymbol{w}_j'\right)\boldsymbol{u}_2$.

  This guarantees that $|\boldsymbol{u}_2| = 1$ and

$$\begin{cases} \boldsymbol{u}_2^\dagger \boldsymbol{u}_1 = 0 \\ \boldsymbol{u}_1^\dagger \boldsymbol{w}_j'' = 0 \\ \boldsymbol{u}_2^\dagger \boldsymbol{w}_j'' = 0 \end{cases}$$

  for all $j \geqslant 3$.

- **Continue similarly** until we reach $\boldsymbol{u}_r$.

We then find an orthonormal basis B_λ for each eigenspace E_λ of a Hermitian matrix **A**.

Then, if $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of **A**, we have that

$$B = B_{\lambda_1} \cup B_{\lambda_2} \cup \ldots \cup B_{\lambda_r}$$

is an orthonormal set of $\mathbb{C}^n$ consisting of eigenvectors of **A**.

4.3.6 Unitary and Orthogonal Diagonolisation

Theorem 4.17

Let **A** be a hermitian matrix of size $n \times n$. Then, **A** is diagonalisable.

More specifically,

1. There exists a basis of eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in \mathbb{C}^n$ with

$$\mathbf{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$$

   for eigenvalues λ_i;

and equivalently,

2. There exists an $n \times n$ invertible matrix **P** such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

   with the columns of **P** representing the eigenvectors $\boldsymbol{v}_i$.

Remark. In addition, the eigenvectors $\mathbf{u}_\mathrm{i}$ can be chosen to be orthonormal, so that $\boldsymbol{u}_i^\dagger \boldsymbol{u}_j = \delta_{ij}$.

Equivalently, the matrix **P** can be chosen to be unitary, so that $\mathbf{P}^\dagger = \mathbf{P}^{-1}$, and that

$$\mathbf{P}^\dagger \mathbf{A}\mathbf{P} = \mathbf{D}.$$

Remark. For an $n \times n$ real symmetric matrix **A**, the eigenvectors can be taken to be $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n \in \mathbb{R}^n$, and can be chosen such that

$$\boldsymbol{u}_i^\top \boldsymbol{u}_j = \delta_{ij}.$$

Equivalently, $\mathbf{P}$ can be chosen to be orthogonal, so that $\mathbf{P}^\top = \mathbf{P}^{-1}$, and that

$$\mathbf{P}^\top \mathbf{A}\mathbf{P} = \mathbf{D}.$$

4.4 Change of Basis

Consider V, W to be real or complex vector spaces, with

$$\dim V = n, \quad \dim W = m, \quad T : V \to W \quad \text{(linear map)}.$$

and

- $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ to be a basis of V;
- $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_m\}$ to be a basis of W,

such that T is represented by the $m \times n$ matrix $\mathbf{A}$ with respect to these bases. This means that

$$T(\boldsymbol{e}_i) = \sum_{j=1}^{m} A_{ji}\boldsymbol{f}_j.$$

Now consider

- $\{\boldsymbol{e}'_1, \dots, \boldsymbol{e}'_n\}$ to be another basis of V;
- $\{\boldsymbol{f}'_1, \dots, \boldsymbol{f}'_m\}$ to be another basis of W.

In this case, T is represented by another $m \times n$ matrix $\mathbf{B}$ with respect to these new bases, such that

$$T(\boldsymbol{e}'_i) = \sum_{j=1}^{m} B_{ji}\boldsymbol{f}'_j.$$

Suppose that the bases are related by

$$\boldsymbol{e}'_i = \sum_k P_{ki}\boldsymbol{e}_k, \quad \boldsymbol{f}'_j = \sum_l Q_{lj}\boldsymbol{f}_l$$

where $\mathbf{P}$ of size $n \times n$ and $\mathbf{Q}$ of size $m \times m$ are invertible matrices.

Proposition 4.18

With $\mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q}$ as above, we have

$$\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{P}.$$

This defines the change of basis formula for matrices representing linear maps.

$\mathbf{P}$ and $\mathbf{Q}$ are called the **change of basis matrices**.

Proof. We have

$$
\begin{aligned}
T(\boldsymbol{e}_i') &= T\left(\sum_k P_{ki}\boldsymbol{e}_k\right) \\
&= \sum_k P_{ki} T(\boldsymbol{e}_k) \\
&= \sum_{k,j} \boldsymbol{f}_j A_{jk} P_{ki}
\end{aligned}
$$

and also

$$
\begin{aligned}
T(\boldsymbol{e}_i') &= \sum_j \boldsymbol{f}_j' B_{ji} \\
&= \sum_{j,l} \boldsymbol{f}_l Q_{lj} B_{ji} \\
&= \sum_{k,j} \boldsymbol{f}_j Q_{jk} B_{ki}. \quad \text{(exchanging indices)}
\end{aligned}
$$

Comparing coefficients of $\boldsymbol{f}_j$, we have, in summation notation,

$$
A_{jk} P_{ki} = Q_{jk} B_{ki}.
$$

Therefore,

$$
\mathbf{AP} = \mathbf{QB} \Leftrightarrow \mathbf{B} = \mathbf{Q}^{-1}\mathbf{AP}.
$$

Remark.

- The definition of $\mathbf{A}$ which represents T with respect to $\{\boldsymbol{e}_i\}$ and $\{\boldsymbol{f}_j\}$ implies that the column i of $\mathbf{A}$ consists of the components of $T(\boldsymbol{e}_i)$ in the basis $\{\boldsymbol{f}_j\}$.
- Similarly, the column i of $\mathbf{P}$ consists of the components of $\boldsymbol{e}_i'$ in the basis $\{\boldsymbol{e}_j\}$.
- If we instead change in the other direction, *i.e.* from $\{\boldsymbol{e}_i'\}$ to $\{\boldsymbol{e}_i\}$ and from $\{\boldsymbol{f}_j'\}$ to $\{\boldsymbol{f}_j\}$, then $\mathbf{P}' = \mathbf{P}^{-1}$ and $\mathbf{Q}' = \mathbf{Q}^{-1}$, such that

$$
\boldsymbol{e}_i = \sum_k P_{ki}' \boldsymbol{e}_k', \quad \boldsymbol{f}_j = \sum_l Q_{lj}' \boldsymbol{f}_l'.
$$

Example 4.19

Consider $\dim V = n = 2$ and $\dim W = m = 3$, with

$$
\begin{aligned}
T(\boldsymbol{e}_1) &= \boldsymbol{f}_1 + 2\boldsymbol{f}_2 - \boldsymbol{f}_3 \\
T(\boldsymbol{e}_2) &= -\boldsymbol{f}_1 + 2\boldsymbol{f}_2 + \boldsymbol{f}_3.
\end{aligned}
$$

Thus, $\mathbf{A}$ is represented by

$$
\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 2 & 2 \\ -1 & 1 \end{pmatrix}.
$$

Now consider a basis for V formed by $\{\boldsymbol{e}_1', \boldsymbol{e}_2'\}$ that relates to $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ by d

$$e_1' = e_1 - e_2, \quad e_2' = e_1 + e_2.$$

Hence we have

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

For W, consider a basis formed by $\{f_1', f_2', f_3'\}$ that relates to $\{f_1, f_2, f_3\}$ by

$$f_1' = f_1 - f_3, \quad f_2' = f_2, \quad f_3' = f_1 + f_3.$$

Hence we have

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

Therefore, the matrix **B** representing T with respect to the new bases is given by

$$\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix} \Rightarrow \begin{cases} T(e_1') = 2f_1' \\ T(e_2') = 4f_2' \end{cases}$$

Remark. [Special cases]

1. If $V = W$ with the same basis change, *i.e.* $e_i' = e_i$ and $f_j' = f_j$, then $\mathbf{P} = \mathbf{Q}$ and

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

   Therefore, matrices represent the same linear map $T : V \to V$ iff they are similar.

2. If $V = W = \mathbb{R}^n$ or $\mathbb{C}^n$, consider for both the standard basis $\{e_i\}$, then if there exists a basis of eigenvectors of T denoted by $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, denote $\{e_i' = \mathbf{v}_i\}$, and define **B** to be the matrix representing T with respect to this basis. Then,

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

   where **P** has columns given by the eigenvectors $\mathbf{v}_i$. By Proposition 4.11, **B** is diagonal, with the eigenvalues of T on the diagonal. So

$$\mathbf{B} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

   where $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for each i, and thus $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ such that

   Since $\mathbf{v}_i = \sum_j e_j P_{ji}$ is the ith eigenvector of T, the columns of **P** are the eigenvectors of T expressed in the standard basis. Therefore, **P** is the change of basis matrix, and is also the matrix that diagonalises **A**.

Lecture 21 · 2025-11-25

4.4.1 Changes in Vector Components Under Change of Basis

Consider V a vector space and $\boldsymbol{x} \in V$. Assume that $\{\boldsymbol{e}_i\}$ and $\{\boldsymbol{e}'_i\}$ are two different bases of V, related by $\mathbf{P}$ and

$$\boldsymbol{x} = x_i \boldsymbol{e}_i = x'_i \boldsymbol{e}'_i.$$

Then, taking into account that $\boldsymbol{e}'_j = \boldsymbol{e}_i P_{ij}$, we have

$$\boldsymbol{x} = \underbrace{x_i \boldsymbol{e}_i} = x'_j \boldsymbol{e}_i P_{ij} = \underbrace{\left(P_{ij} x'_j\right)} \boldsymbol{e}_i$$

and hence

$$x_i = P_{ij} x'_j.$$

and this is the relation between vector components with respect to bases related by $\mathbf{P}$. We can write

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} P_{11} & \dots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \dots & P_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \mathbf{P} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

and thus

$$\boldsymbol{x} = \mathbf{P}\boldsymbol{x}'.$$

Similarly, consider vector space W and $\boldsymbol{y} \in W$. Assume that $\{\boldsymbol{f}_j\}$ and $\{\boldsymbol{f}'_j\}$ are two different bases of W such that

$$\boldsymbol{y} = y_j \boldsymbol{f}_j = y'_j \boldsymbol{f}'_j$$

and with bases related by $\mathbf{Q}$. Then, we have

$$\boldsymbol{y} = \mathbf{Q}\boldsymbol{y}'.$$

Now, if we consider the definition of a linear map $T : V \to W$ in terms of a matrix $\mathbf{A}$, we have

$$\boldsymbol{y} = T(\boldsymbol{x}) \Leftrightarrow \boldsymbol{y} = \mathbf{A}\boldsymbol{x} \quad \text{and} \quad \boldsymbol{y}' = \mathbf{B}\boldsymbol{x}'.$$

Therefore,

$$\boldsymbol{y}' = \mathbf{Q}^{-1}\boldsymbol{y} = \mathbf{Q}^{-1}\mathbf{A}\boldsymbol{x} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{P}\boldsymbol{x}' = \mathbf{B}\boldsymbol{x}'.$$

This recovers the change of basis formula for matrices representing linear maps:

$$\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{P}.$$

4.5 Cayley-Hamilton Theorem

Theorem 4.20 (Cayley-Hamilton Theroem)

Let $\mathbf{A}$ be an $n \times n$ matrix with

$$\chi_{\mathbf{A}}(t) = \det(\mathbf{A} - t\mathbf{I}) = \sum_{r=0}^{n} c_r t^r.$$

Then,

$$\chi_{\mathbf{A}}(\mathbf{A}) = \sum_{r=0}^{n} c_r \mathbf{A}^r = \mathbf{0}.$$

Proof.

1. Consider a general matrix of size 2×2. Then

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \Rightarrow \chi_{\mathbf{A}}(t) = t^2 - (A_{11} + A_{22})t + (A_{11}A_{22} - A_{12}A_{21}).$$

   Then, checking by direct substitution gives the result.

2. For a diagonolisable $n \times n$ matrix, consider $\mathbf{A}$ with eigenvalues λ_i, along with an invertible matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

   Note that we can compute powers of $\mathbf{D}$ easily:

$$\mathbf{D}^r = \begin{pmatrix} \lambda_1^r & 0 & \dots & 0 \\ 0 & \lambda_2^r & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^r \end{pmatrix}.$$

   Thus

$$\chi_{\mathbf{A}}(\mathbf{D}) = \sum_{r=0}^{n} c_r \mathbf{D}^r = \begin{pmatrix} \chi_{\mathbf{A}}(\lambda_1) & 0 & \dots & 0 \\ 0 & \chi_{\mathbf{A}}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \chi_{\mathbf{A}}(\lambda_n) \end{pmatrix} = \mathbf{0}.$$

   But $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, and so $\mathbf{A}^r = \left(\mathbf{P}\mathbf{D}\mathbf{P}^{-1}\right)^r = \mathbf{P}\mathbf{D}^r\mathbf{P}^{-1}$. Therefore,

$$\begin{aligned} \chi_{\mathbf{A}}(\mathbf{A}) &= \sum_{r=0}^{n} c_r \mathbf{A}^r \\ &= \sum_{r=0}^{n} c_r \left(\mathbf{P}\mathbf{D}^r\mathbf{P}^{-1}\right) \\ &= \mathbf{P}\left(\sum_{r=0}^{n} c_r \mathbf{D}^r\right)\mathbf{P}^{-1} \\ &= \mathbf{P}\mathbf{0}\mathbf{P}^{-1} = \mathbf{0}. \end{aligned}$$

3. [Non-examinable.] In the general case, let $\mathbf{M} := \mathbf{A} - t\mathbf{I}$, and $\chi_{\mathbf{A}}(t) = \det \mathbf{M}$.

   Recall that the adjugate matrix $\widetilde{\mathbf{M}}$ is defined such that

$$\tilde{\mathbf{M}}\mathbf{M} = \det(\mathbf{M})\mathbf{I}.$$

We will use

$$\tilde{\mathbf{M}} = \sum_{r=0}^{n-1} \mathbf{B_r} t^r.$$

Comparing coefficients of t^n on both sides,

$$\begin{aligned}\tilde{\mathbf{M}}\mathbf{M} &= \sum_{r=0}^{n-1} \mathbf{B_r} t^r (\mathbf{A} - t\mathbf{I}) \\ &= \mathbf{B_0}\mathbf{A} + (\mathbf{B_1}\mathbf{A} - \mathbf{B_0})t + \dots + (\mathbf{B_{n-1}}\mathbf{A} - \mathbf{B_{n-2}})t^{n-1} - \mathbf{B_{n-1}}t^n\end{aligned}$$

and hence

$$\begin{aligned} c_0\mathbf{I} &= \mathbf{B_0}\mathbf{A} \\ c_1\mathbf{I} &= \mathbf{B_1}\mathbf{A} - \mathbf{B_0} \\ \vdots &= \vdots \\ c_{n-1}\mathbf{I} &= \mathbf{B_{n-1}}\mathbf{A} - \mathbf{B_{n-2}} \\ c_n\mathbf{I} &= -\mathbf{B_{n-1}}. \end{aligned}$$

evaluating in $\mathbf{A}$ gives

$$\begin{aligned} c_0\mathbf{I} &= \mathbf{B_0}\mathbf{A} \\ c_1\mathbf{A} &= \mathbf{B_1}\mathbf{A}^2 - \mathbf{B_0}\mathbf{A} \\ \vdots &= \vdots \\ c_{n-1}\mathbf{A}^{n-1} &= \mathbf{B_{n-1}}\mathbf{A}^n - \mathbf{B_{n-2}}\mathbf{A}^{n-1} \\ c_n\mathbf{A}^n &= -\mathbf{B_{n-1}}\mathbf{A}^n. \end{aligned}$$

Adding these equations gives

$$\sum_{r=0}^{n} c_r \mathbf{A}^r = \mathbf{0}.$$

This completes the proof.

4.6 Quadratic Forms

We wish to study functions of the form $x_1^2 + x_2^2$ or $2x_1^2 + 2x_1x_2 + 5x_2^2$ in $\mathbb{R}^2$, or more generally, a quadratic homogeneous polynomial of degree 2 in n variables $x_1, \dots, x_n$. It turns out that these can be written in matrix form as $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ for some symmetric matrix $\mathbf{A}$.

Definition 4.21

A **quadratic form** is a function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top\mathbf{A}\mathbf{x} = x_i A_{ij} x_j$$

where $\mathbf{A}$ is a real symmetric matrix of size $n \times n$.

We can hence write

$$\mathbf{P}^\top \mathbf{A}\mathbf{P} = \mathbf{D}$$

where $\mathbf{D}$ is diagonal with eigenvalues $\lambda_1, \ldots, \lambda_n$ on the diagonal, and $\mathbf{P}$ is a real orthogonal matrix of size $n \times n$ with columns given by orthonormal eigenvectors of $\mathbf{A}$.

Setting $\boldsymbol{x}' = \mathbf{P}^\top \boldsymbol{x} \Leftrightarrow \boldsymbol{x} = \mathbf{P}\boldsymbol{x}'$, we can diagonalise the quadratic form:

$$\begin{aligned}\mathcal{F}(\mathbf{x}) &= \mathbf{x}^\top \mathbf{A}\mathbf{x} \\ &= (\mathbf{P}\boldsymbol{x}')^\top \mathbf{A}(\mathbf{P}\boldsymbol{x}') \\ &= \boldsymbol{x}'^\top(\mathbf{P}^\top \mathbf{A}\mathbf{P})\boldsymbol{x}' \\ &= \boldsymbol{x}'^\top \mathbf{D}\boldsymbol{x}'\end{aligned}$$

Therefore,

$$\mathcal{F}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i (x_i')^2.$$

Note that $\boldsymbol{x}'$ is the representation of $\boldsymbol{x}$ in the orthonormal basis of eigenvectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ of $\mathbf{A}$, where $\boldsymbol{u}_i$ is the eigenvector corresponding to eigenvalue λ_i. Indeed, since the columns of $\mathbf{P}$ are the eigenvectors $\boldsymbol{u}_i$, we have

$$\begin{aligned}\boldsymbol{x}' &= x_1'\boldsymbol{e}_1 + \ldots + x_n'\boldsymbol{e}_n \\ \boldsymbol{x} &= x_1\boldsymbol{e}_1 + \ldots + x_n\boldsymbol{e}_n \\ &= x_1'\boldsymbol{u}_1 + \ldots + x_n'\boldsymbol{u}_n \quad (\text{since } \boldsymbol{x} = \mathbf{P}\boldsymbol{x}')\end{aligned}$$

and

$$x_i' = \boldsymbol{u}_i \cdot \mathbf{x}$$

are the components of $\boldsymbol{x}$ in the orthonormal basis of eigenvectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$, with the new axes along these direction called the **principal axes** of the quadratic form.

Since these are related to the standard axes by orthogonal $\mathbf{P}$, we have

$$|\mathbf{x}|^2 = x_i x_i = x_i' x_i'.$$

Example 4.22

In $\mathbb{R}^2$, consider $\mathcal{F}(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ with $\mathbf{A} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$.

The eigenvalues are $\lambda_1 = \alpha + \beta, \lambda_2 = \alpha - \beta$ and the eigenvectors are

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Then

$$\mathcal{F}(\mathbf{x}) = \alpha x_1^2 + 2\beta x_1 x_2 + \alpha x_2^2 = (\alpha + \beta)(x_1')^2 + (\alpha - \beta)(x_2')^2$$

with

$$x'_1 = \frac{1}{\sqrt{2}}(x_1 + x_2), \quad x'_2 = \frac{1}{\sqrt{2}}(-x_1 + x_2)$$

e.g. take $\alpha = \frac{3}{2}$, $\beta = -\frac{1}{2}$, then $\lambda_1 = 1$, $\lambda_2 = 2$. Then if we set $\mathcal{F} = 1$,

$$\mathcal{F}(\mathbf{x}) = \frac{3}{2}x_1^2 - x_1x_2 + \frac{3}{2}x_2^2 = (x'_1)^2 + 2(x'_2)^2 = 1.$$

defines an ellipse.

e.g. take $\alpha = -\frac{1}{2}$, $\beta = \frac{3}{2}$, then $\lambda_1 = 1, \lambda_2 = -2$. Then if we set $\mathcal{F} = 1$,

$$\mathcal{F}(\mathbf{x}) = -\frac{1}{2}x_1^2 + 3x_1x_2 - \frac{1}{2}x_2^2 = (x'_1)^2 - 2(x'_2)^2 = 1.$$

defines a hyperbola.

Example 4.23

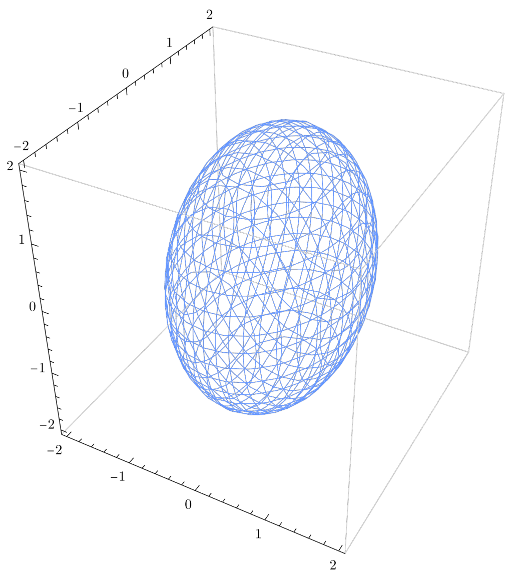
Consider

$$\mathcal{F}(\mathbf{x}) = \mathbf{x}^{\mathrm{tr}}\mathbf{A}\mathbf{x} = \lambda_1(x'_1)^2 + \lambda_2(x'_2)^2 + \lambda_3(x'_3)^2.$$

1. If $\lambda_1, \lambda_2, \lambda_3 > 0$, then

$$\mathcal{F}(\mathbf{x}) = 1$$

defines an ellipsoid.



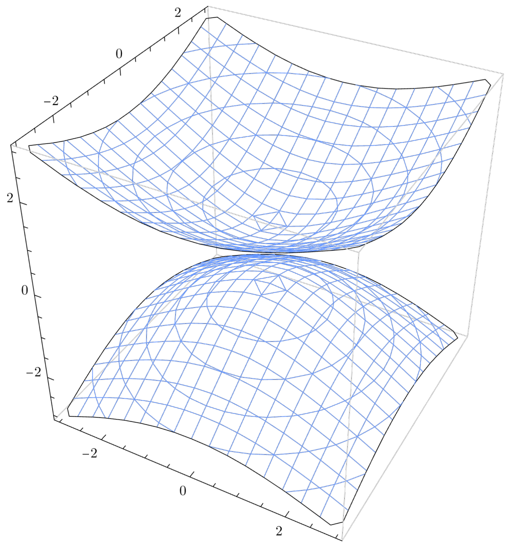
2. If $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, then the eigenvalues are $\lambda_1 = \lambda_2 = -1, \lambda_3 = 2$, and the eigenvectors are

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \quad \boldsymbol{u}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

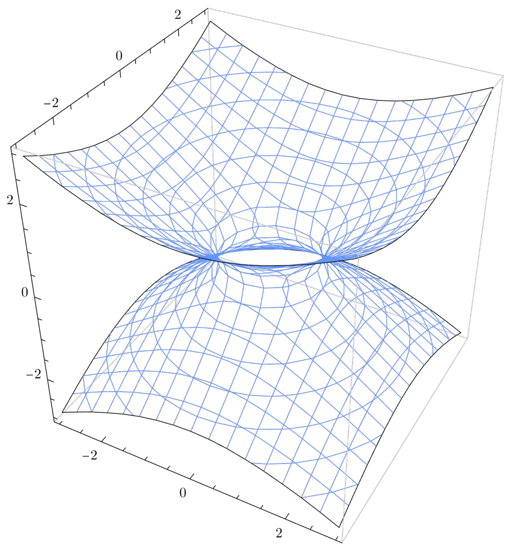
Then,

$$\mathcal{F}(\boldsymbol{x}) = 2(x_3')^2 - (x_1')^2 - (x_2')^2 = 2x_1x_2 + 2x_2x_3 + 2x_3x_1.$$

If we set $\mathcal{F}(\boldsymbol{x}) = 1$, it defines a two-sheeted hyperboloid.



If we set $\mathcal{F}(\boldsymbol{x}) = -1$, it defines a one-sheeted hyperboloid.



Lecture 22 · 2025-11-27

Remark. Given a matrix $\mathbf{M}$, $\mathbf{M}$ can be decomposed as

$$\mathbf{M} = \mathbf{S} + \mathbf{A}$$

where $\mathbf{S}$ is symmetric and $\mathbf{A}$ is antisymmetric. Note that since $\mathbf{A}$ is antisymmetric, $\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} = 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Therefore,

$$\boldsymbol{x}^\top \mathbf{M}\boldsymbol{x} = \boldsymbol{x}^\top \mathbf{S}\boldsymbol{x}.$$

This is why we only consider symmetric matrices in the definition of quadratic forms.

4.7 Quadrics and Conics

4.7.1 Quadrics

Definition 4.24 (Quadric)

A **quadric** in $\mathbb{R}^n$ is a hypersurface defined by

$$Q(\boldsymbol{x}) = \boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} + 2\boldsymbol{b}^\top \boldsymbol{x} + c = 0$$

for some $n \times n$ real symmetric matrix $\mathbf{A}$, $\boldsymbol{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Hence,

$$Q(\boldsymbol{x}) = A_{ij}x_i x_j + 2b_i x_i + c = 0.$$

The purpose of this section is to classify the solutions of this kind of equations up to geomtrical equivalence. *i.e.* there is no distinction between solutions related by isometries of $\mathbb{R}^n$, including

- translations,
- orthogonal transformations about the origin.

If $\mathbf{A}$ to be invertible, we can complete the square by setting

$$\boldsymbol{y} = \boldsymbol{x} + \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{b},$$

then

$$\begin{aligned}
\mathcal{F}(\boldsymbol{y}) = \boldsymbol{y}^\top \mathbf{A}\boldsymbol{y} &= \boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b} \\
&= \left(\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c\right) + \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b} - c \\
&= Q(\boldsymbol{x}) + \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b} - c \\
&= Q(\boldsymbol{x}) + k
\end{aligned}$$

where $k = \frac{1}{4}\boldsymbol{b}^\top \mathbf{A}^{-1}\boldsymbol{b} - c$ is a constant.

Hence we have

$$\mathcal{F}(\boldsymbol{y}) = k \Leftrightarrow Q(\boldsymbol{x}) = 0.$$

Now we diagonolise $\mathcal{F}$ as for the quadratic forms before. The orthonormal eigenvectors of $\mathbf{A}$ define principal axes (the new coordinate axes), and the eigenvalues of $\mathbf{A}$ along with k determine the shape of the quadric.

1. If all eigenvalues > 0 and $k > 0$, then we have an ellipsoid.
2. If eigenvalues are of both signs and $k \neq 0$, then we have a hyperboloid.
3. If $\mathbf{A}$ has one or more zero eigenvalues, then our analysis changes. It is simplest in the standard form, where we have linear and quadratic terms to analyse.

4.7.2 Conics

Definition 4.25 (Conic)

Quadrics in $\mathbb{R}^2$ are curves called **conics**.

- If $\det \mathbf{A} \neq 0$, we get the form
$$\lambda_1(x_1')^2 + \lambda_2(x_2')^2 = k$$
which represents
  1. if $\lambda_1, \lambda_2 > 0$, then
     - if $k > 0$, an ellipse;
     - if $k = 0$, a point;
     - if $k < 0$, no solutions.
  2. if λ_1 and λ_2 have opposite signs, then
     - if $k \neq 0$, a hyperbola;
     - if $k = 0$, a pair of lines.
- If $\det \mathbf{A} = 0$, consider $\lambda_1 > 0$ and $\lambda_2 = 0$. We can diagnoalise $\mathbf{A}$ into the original formula for quadrics to get
$$\begin{aligned} \lambda_1(x_1')^2 + b_1'x_1' + b_2'x_2' + c = 0. \\ \lambda_1(x_1')^2 + b_1'x_1' + \frac{1}{4\lambda_1}(b_1')^2 - \frac{1}{4\lambda_1^2}(b_1')^2 + b_2'x_2' + c = 0 \\ \lambda_1 \underbrace{\left(x_1' + \frac{b_1'}{2\lambda_1}\right)^2}_{x_1''} + b_2'x_2' + \underbrace{\left(c - \frac{(b_1')^2}{4\lambda_1}\right)}_{c'} = 0 \\ \lambda_1 \mathbf{x}_1''^2 + b_2'x_2' + c' = 0. \end{aligned}$$
  1. If $b_2' = 0$, then the equation reduces to $\lambda_1 \mathbf{x}_1''^2 + c' = 0$. This represents
     - if $c' < 0$, a pair of lines;
     - if $c' = 0$, a single line;
     - if $c' > 0$, no solutions.
  2. If $b_2' \neq 0$, we can write
  $$\lambda_1(x_1'')^2 + b_2' + x_2'' = 0$$
  for $x_2'' = x_2' + \frac{c'}{b_2'}$. This represents a parabola.

  Note that all changes of coordinates used here are isometries of $\mathbb{R}^2$.

4.7.3 Standard Forms for Conics in Cartesian Coordinates

Example 4.26

Consider

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

- If eccentricity $e < 1$, this is an ellipse, where the semi-major axis is $\max(a, b)$ and the semi-minor axis is $\min(a, b)$. We can write
$$b^2 = a^2\left(1 - e^2\right)$$
and the foci are at $x = \pm ae$.
- If eccentricity $e = 1$, this is a parabola, with focus at $\left(\frac{a}{2}, 0\right)$ and
$$y^2 = 4ax.$$
- If eccentricity $e > 1$, this is a hyperbola, with semi-major axis a and semi-minor axis b related by
$$b^2 = a^2\left(e^2 - 1\right).$$
The foci are at $x = \pm ae$.

Lecture 23 · 2025-11-29

4.7.4 Focus-Directrix Property of Conics

The four types of conics (ellipse, parabola, hyperbola, circle) are essentially four different types of cross-sections of a cone. They can also be defined in terms of a focus point and a directrix line.

Consider the expression for the conic:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = k.$$

Conic sections can be defined in terms of the followings.

Definition 4.27 (Eccentricity)

The **eccentricity** e is a non-negative parameter. The eccentricity and scale properties of a conic section satisfy

- the foci of a conic are $(\pm ae, 0)$;
- the directrices are the vertical lines $x = \pm\frac{a}{e}$.

A conic is the set of points whose distance from the focus is

$$e \times \text{distance from the closest directrix,}$$

unless $e = 1$, in which we will take the other directrix.

We have the following cases.

1. $e < 1$, the conic is an ellipse.

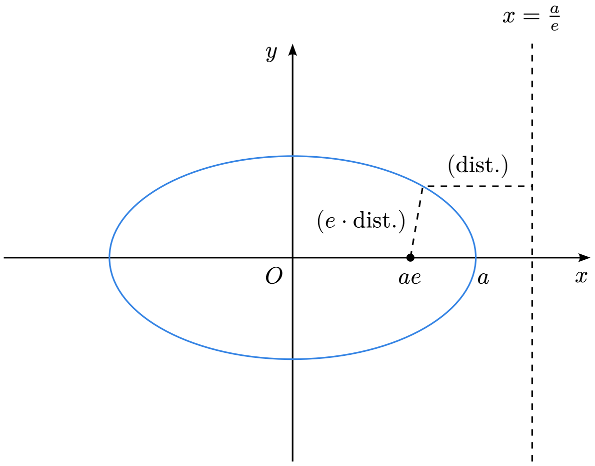
The equation of the ellipse is

$$\sqrt{(x-ae)^2+y^2} = e \cdot \left(\frac{a}{e} - x\right)$$

or equivalently,

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1$$

where $b = a\sqrt{1-e^2}$.



In this case, the semi-major axis is a and the semi-minor axis is b. Additionally, if $e = 0$, the ellipse is a circle of radius a.

2. If $e > 1$, the conic is a hyperbola.

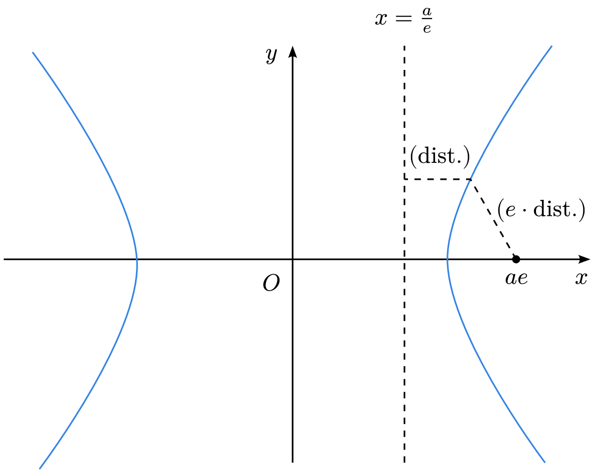
The equation of the hyperbola is

$$\sqrt{(x-ae)^2+y^2} = e \cdot \left(x - \frac{a}{e}\right)$$

or equivalently,

$$\frac{x^2}{a^2} - \frac{y^2}{a^2(e^2-1)} = 1$$

where $b = a\sqrt{e^2-1}$.



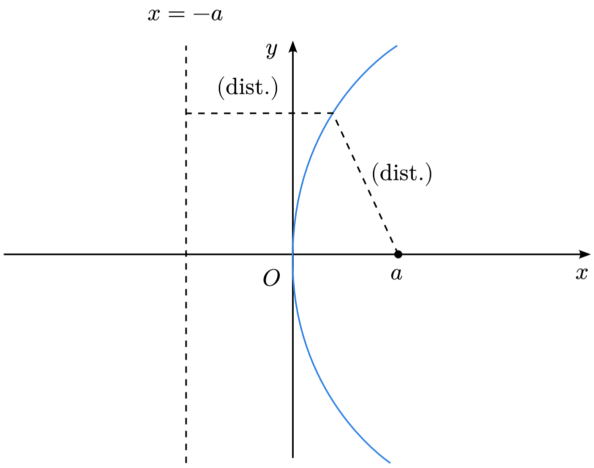
3. If $e = 1$, the conic is a parabola.

   The equation of the parabola is

$$\sqrt{(x-a)^2 + y^2} = (x+a)$$

   or equivalently,

$$y^2 = 4ax.$$



4.7.5 Polar Coordinates

We introduce a new parameter l such that $\frac{l}{e}$ is the distance from the focus to the directrix. Then,

$$l = a\left|1 - e^2\right|.$$

We can use polar coordinates (r, θ) centered on a focus, such that the focus-directrix property is:

$$r = e\left(\frac{l}{e} - r\cos\theta\right) \Leftrightarrow r = \frac{l}{1 + e\cos\theta}.$$

Example 4.28

1. For an ellipse with $e < 1$, we have

$$r = \frac{a\left(1 - e^2\right)}{1 + e\cos\theta} \quad \text{where} \quad l = a\left(1 - e^2\right).$$

2. For a hyperbola with $e > 1$, we have

$$r = \frac{a\left(e^2 - 1\right)}{1 + e\cos\theta} \quad \text{where} \quad l = a\left(e^2 - 1\right).$$

3. For a parabola with $e = 1$, we have

$$r = \frac{2a}{1 + \cos\theta} \quad \text{where} \quad l = 2a.$$

4.8 Jordan Normal Forms

This gives us a classification for $n \times n$ complex matrices up to similarity.

Consider a matrix $\mathbf{A}$ of size $n \times n$ corresponding to a linear map $T : \mathbb{C}^n \to \mathbb{C}^n$ and is similar to a matrix $\mathbf{B}$ after a change of basis.

Proposition 4.29

Any 2×2 complex matrix $\mathbf{A}$ is similar to one of the followings:

1. $\mathbf{B} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ with $\lambda_1 \neq \lambda_2$, with $\chi_{\mathbf{A}}(t) = (t - \lambda_1)(t - \lambda_2)$.
2. $\mathbf{B} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, with $\chi_{\mathbf{A}}(t) = (t - \lambda)^2$.
3. $\mathbf{B} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, with $\chi_{\mathbf{A}}(t) = (t - \lambda)^2$.

Proof. $\chi_{\mathbf{A}}(t)$ has 2 roots, counting multiplicities, in $\mathbb{C}$. We have the following cases.

1. For distinct roots (eigenvalues) λ_1, λ_2, we have $M_{\lambda_1} = m_{\lambda_1} = 1 = m_{\lambda_2} = M_{\lambda_2}$. And thus eigenvectors $\boldsymbol{v_1}, \boldsymbol{v_2}$ form a basis of $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, diagonolised with the eigenvectors as columns of $\mathbf{P}$.
2. For repeated root $\lambda_1 = \lambda_2 = \lambda$, with $M_\lambda = m_\lambda = 2$, then the same argument as above applies, and $\mathbf{B}$ is diagonolised.
3. For repeated root $\lambda_1 = \lambda_2 = \lambda$, with $M_\lambda = 2$ and $m_\lambda = 1$. Let $\boldsymbol{v}$ to be an eigenvector for λ and extend it to a basis $\{\boldsymbol{v}, \boldsymbol{w}\}$, where $\boldsymbol{w}$ is any vector linearly independent of $\boldsymbol{v}$. Hence,

$$\mathbf{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \quad \mathbf{A}\boldsymbol{w} = \alpha \boldsymbol{v} + \beta \boldsymbol{w}.$$

   Then, the matrix of the linear map w.r.t. the basis $\{\boldsymbol{v}, \boldsymbol{w}\}$ is

$$\mathbf{B} = \begin{pmatrix} \lambda & \alpha \\ 0 & \beta \end{pmatrix}.$$

   Note that we will only consider $\beta = \lambda$, otherwise we will return to case (1). Also, $\alpha \neq 0$, otherwise we will return to case (2).

   Now, defining $\boldsymbol{u} = \alpha \boldsymbol{v}$. Then we have that, with respect to the basis $\{\boldsymbol{u}, \boldsymbol{w}\}$, the matrix of the linear map is

$$\mathbf{B} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix},$$

   with $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, and the columns of $\mathbf{P}$ given by $\boldsymbol{u}, \boldsymbol{w}$.

Theorem 4.30 (General Jordan Normal Form)

Any $n \times n$ complex matrix $\mathbf{A}$ is similar to a matrix $\mathbf{B}$ with block form given by

$$\mathbf{B} = \begin{pmatrix} \begin{bmatrix} J_{n_1}(\lambda_1) \end{bmatrix} & & & \\ & \begin{bmatrix} J_{n_2}(\lambda_2) \end{bmatrix} & & \\ & & \ddots & \\ & & & \begin{bmatrix} J_{n_r}(\lambda_r) \end{bmatrix} \end{pmatrix}$$

where each Jordan block is a matrix of the form

$$J_p(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix} \quad \text{of size } p \times p$$

with $n_1 + n_2 + \dots + n_r = n$, and $\lambda_1, \lambda_2, \dots, \lambda_r$ are the eigenvalues of **A** and **B** (because they are similar).

Note that the same eigenvalue may appear in multiple Jordan blocks.

A is diagonalisable iff all Jordan blocks are of size 1×1.

4.9 Symmetries and Transformation Groups

4.9.1 Orthogonal Transformations and Rotations in $\mathbb{R}^n$

[This topic is discussed in more detail in IA Groups.]

Recall that **R** is an orthogonal is equivalent to

- $\mathbf{R}^\top\mathbf{R} = \mathbf{R}\mathbf{R}^\top = \mathbf{I}$,
- $(\mathbf{R}\boldsymbol{x}) \cdot (\mathbf{R}\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,
- The columns or rows of **R** form an orthonormal basis of $\mathbb{R}^n$.

Definition 4.31 (Orthogonal group)

The set of orthogonal matrices $\mathbb{R}$ of size $n \times n$ is a group, denoted $O(n)$, is called the **orthogonal group**.

Recall that $\det(\mathbf{R}) = \pm 1$ for any orthogonal matrix **R**.

Definition 4.32 (Special orthogonal group)

A subgroup of $O(n)$ formed by orthogonal matrices with determinant 1 is called the **special orthogonal group**, denoted $\mathrm{SO}(n)$.

- $\mathbf{R} \in O(n)$ preserves lengths and n-dimensional (absolute) volumes.

- $\mathbf{R} \in \mathrm{SO}(n)$ preserves orientations, given by the signs of the volumes.

Geometrically, $\mathrm{SO}(n)$ consists of all rotations in $\mathbb{R}^n$, and reflections belong to $O(n)$ but not to $\mathrm{SO}(n)$.

Any element of $O(n)$ is of the form:

- $\mathbf{R} \in \mathrm{SO}(n)$,
- $\mathbf{RH}$ where $\mathbf{R} \in \mathrm{SO}(n)$ and $\mathbf{H} \in O(n) \setminus \mathrm{SO}(n)$.

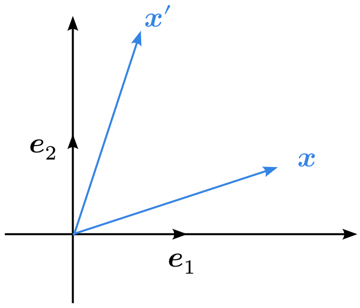
Lecture 24 · 2025-12-02

For a rotation matrix $\mathbf{R}$, consider

$$x_i' = R_{ij} x_j.$$

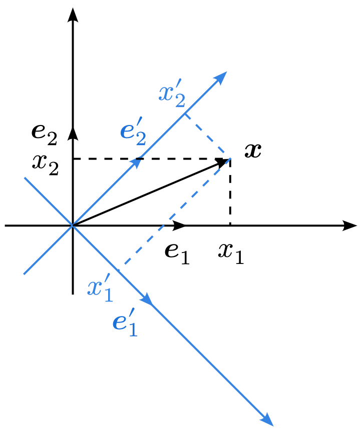
We can view this in two ways:

- transformation of vectors (active point of view)



  We have $|\boldsymbol{x}'| = |\boldsymbol{x}|$ where x_i' are component sof the new vector $\boldsymbol{x}'$ after $\boldsymbol{x}' = \mathbf{R}\boldsymbol{x}$ with respect to the standard basis.

- change of basis (passive point of view)



  Now x_i' are components of the same vector $\boldsymbol{x}$ but with respect to a new orthonormal basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ where

$$\boldsymbol{u}_i = \sum_j R_{ij} \boldsymbol{e}_j = \sum_j e_j \left(\boldsymbol{R}^{-1}\right)_{ji}$$

> ***Remark.*** Compare this to the standard notation for the matrix of change of basis, with $\mathbf{P} = \mathbf{R}^{-1}$.

4.9.2 2D Minkowski Space and Lorentz Transformations

Consider the *inner product* on $\mathbb{R}^2$ given by $(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top \mathbf{J}\boldsymbol{y}$ where $\mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

If $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then

$$(\boldsymbol{x}, \boldsymbol{y}) = x_1 y_1 - x_2 y_2.$$

This *inner product* is not positive definite, since

$$(\boldsymbol{x}, \boldsymbol{x}) = x_1^2 - x_2^2$$

which is not always positive. Nonetheless, it is still bilinear and symmetric.

Now let us consider how the standard basis vectors behave under this inner product. Consider $\boldsymbol{e}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ amd $\boldsymbol{e}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. They are *orthonormal* with respect to this inner product, in the sense that

$$(\boldsymbol{e}_0, \boldsymbol{e}_0) = 1, \quad (\boldsymbol{e}_1, \boldsymbol{e}_1) = -1, \quad (\boldsymbol{e}_0, \boldsymbol{e}_1) = 0.$$

Definition 4.33 (Minkowski metric and Minkowski space)

The inner product defined $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ by

$$(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top \mathbf{J}\boldsymbol{y}$$

where $\mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is called the **Minkowski metric**.

$\mathbb{R}^2$ equipped with the Minkowski metric is called a **Minkowski space**.

Consider $\mathbf{M} = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$ associated to a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$. This preserves the Minkowski metric iff

$$\begin{aligned}
&(\mathbf{M}\boldsymbol{x}, \mathbf{M}\boldsymbol{y}) &&= (\boldsymbol{x}, \boldsymbol{y}) && \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2 \\
\Leftrightarrow\ &(\mathbf{M}\boldsymbol{x})^\top \mathbf{J}(\mathbf{M}\boldsymbol{y}) &&= \boldsymbol{x}^\top \mathbf{J}\boldsymbol{y} && \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2 \\
\Leftrightarrow\ &\boldsymbol{x}^\top (\mathbf{M}^\top \mathbf{J}\mathbf{M})\boldsymbol{y} &&= \boldsymbol{x}^\top \mathbf{J}\boldsymbol{y} && \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2 \\
\Leftrightarrow\ &\mathbf{M}^\top \mathbf{J}\mathbf{M} &&= \mathbf{J}.
\end{aligned}$$

The matrices $\mathbf{M}$ satisfying this condition form a group, with

$$(\det \mathbf{M}^\top)(\det \mathbf{J})(\det \mathbf{M}) = \det \mathbf{J} \Rightarrow (\det \mathbf{M})^2 = 1 \Rightarrow \det \mathbf{M} = \pm 1.$$

Definition 4.34 (Lorentz group)

The Lorentz group is the subgroup of the group above that satisfies $\det \mathbf{M} = 1$ and $M_{00} > 0$.

4.9.2.1 General Form of Lorentz Transformations

We shall determine a general form for matrices in the Lorentz group given the conditions above.

A similar argument as for orthogonal matrices will be followed.

Using

- $(\boldsymbol{e}_0, \boldsymbol{e}_0) = 1$, which gives $(\mathbf{M}\boldsymbol{e}_0, \mathbf{M}\boldsymbol{e}_0) = 1$.

We have

$$(1\ \ 0)\begin{pmatrix} M_{00} & M_{10} \\ M_{01} & M_{11} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = M_{00}^2 - M_{10}^2 = 1.$$

- $(\boldsymbol{e}_1, \boldsymbol{e}_1) = -1$, which gives $(\mathbf{M}\boldsymbol{e}_1, \mathbf{M}\boldsymbol{e}_1) = -1$.

  We have

$$(0\ \ 1)\begin{pmatrix} M_{00} & M_{10} \\ M_{01} & M_{11} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = M_{01}^2 - M_{11}^2 = -1.$$

- $(\boldsymbol{e}_0, \boldsymbol{e}_1) = 0$, which gives $(\mathbf{M}\boldsymbol{e}_0, \mathbf{M}\boldsymbol{e}_1) = 0$.

  This similarly gives $M_{00}M_{01} - M_{10}M_{11} = 0$.
- $M_{00} > 0$

Combining these equations, we can derive a general form

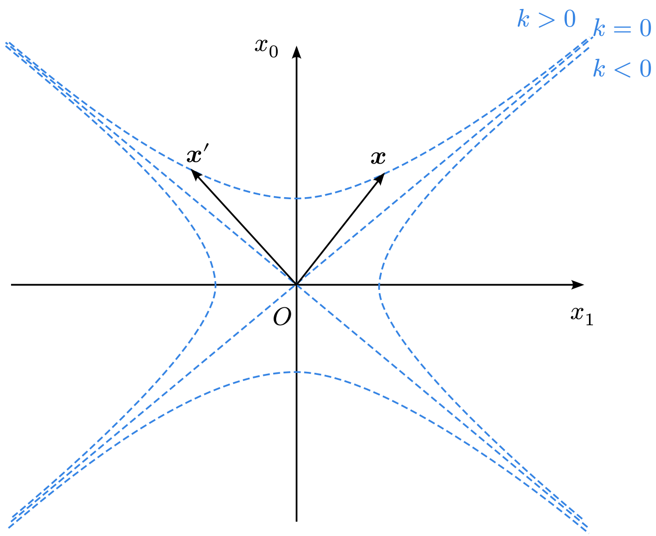
$$\mathbf{M} = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \quad \text{for some } \theta \in \mathbb{R}.$$

For elements of the Lorentz group, we have

$$\mathbf{M}(\theta_1)\mathbf{M}(\theta_2) = \mathbf{M}(\theta_1 + \theta_2)$$

using hyperbolic trigonometric identities.

4.9.2.2 Physical Interpretation of Lorentz Transformations



Fix $(\boldsymbol{x}, \boldsymbol{x}) = k$ to be constant. Any Lorentz transformation over $\boldsymbol{x}$ maps it to other vector $\boldsymbol{x}'$ on a same curve.

Note that $\boldsymbol{x}$ and $\boldsymbol{x}'$ must lie on the same branch of the curve, since $M_{00} > 0$.

We have

$$\mathbf{M}(\theta) = \frac{1}{\sqrt{1 - (\tanh\theta)^2}}\begin{pmatrix} 1 & \tanh\theta \\ \tanh\theta & 1 \end{pmatrix}$$

since $\cosh^2\theta - \sinh^2\theta = 1$ and $\tanh\theta = \frac{\sinh\theta}{\cosh\theta}$.

Now, for a physical interpretation, define $v \coloneqq \tanh\theta$ with $-1 < v < 1$. [With the speed of light $c = 1$.]

Rename $x_0 \to t$ [time coordinate] and $x_1 \to x$ [space coordinate]. Then, we can interpret

$$\mathbf{x}' = \mathbf{M}\mathbf{x} \text{ with } \begin{cases} t' = \frac{1}{\sqrt{1-v^2}}(t + vx) \\ x' = \frac{1}{\sqrt{1-v^2}}(x + vt) \end{cases}$$

so Lorentz transformations boost the time and space coordinates for an observer moving at speed v relative to another observer. More details on this topic will be covered in IA Dynamics and Relativity.

The factor $\frac{1}{\sqrt{1-v^2}}$ in Lorentz transformations gives rise to time dilation and length contraction.

For composition of velocities,

$$\mathbf{M}(\theta_1)\mathbf{M}(\theta_2) = \mathbf{M}\underbrace{(\theta_1 + \theta_2)}_{\theta_3}$$

where $v_i \coloneqq \tanh\theta_i$, we get

$$v_3 = \tanh(\theta_3) = \tanh(\theta_1 + \theta_2) = \frac{\tanh\theta_1 + \tanh\theta_2}{1 + \tanh\theta_1 \tanh\theta_2} = \frac{v_1 + v_2}{1 + v_1 v_2}.$$

This is the relativistic velocity addition formula.

END OF DOCUMENT ■