Mathematical Tripos: IA Algebra & Geometry (Part I)

Contents

0 Introduction

0.1 Schedule

This is a copy from the booklet of schedules.[1] Schedules are minimal for lecturing and maximal for examining; that is to say, all the material in the schedules will be lectured and only material in the schedules will be examined. The numbers in square brackets at the end of paragraphs of the schedules indicate roughly the number of lectures that will be devoted to the material in the paragraph.

ALGEBRA AND GEOMETRY **48 lectures, Michaelmas term**

Review of complex numbers, modulus, argument and de Moivre's theorem. Informal treatment of complex logarithm, *n-th* roots and complex powers. Equations of circles and straight lines. Examples of Möbius transformations. [3]

Vectors in $\mathbb{R}^3$. Elementary algebra of scalars and vectors. Scalar and vector products, including triple products. Geometrical interpretation. Cartesian coordinates; plane, spherical and cylindrical polar coordinates. Suffix notation: including summation convention and δ_{ij}, ϵ_{ijk}. [5]

Vector equations. Lines, planes, spheres, cones and conic sections. Maps: isometries and inversions. [2]

Introduction to $\mathbb{R}^n$, scalar product, Cauchy–Schwarz inequality and distance. Subspaces, brief introduction to spanning sets and dimension. [4]

Linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$ with emphasis on $m, n \leqslant 3$. Examples of geometrical actions (reflections, dilations, shears, rotations). Composition of linear maps. Bases, representation of linear maps by matrices, the algebra of matrices. [5]

Determinants, non-singular matrices and inverses. Solution and geometric interpretation of simultaneous linear equations (3 equations in 3 unknowns). Gaussian Elimination. [3]

Discussion of $\mathbb{C}^n$, linear maps and matrices. Eigenvalues, the fundamental theorem of algebra (statement only), and its implication for the existence of eigenvalues. Eigenvectors, geometric significance as invariant lines. [3]

Discussion of diagonalization, examples of matrices that cannot be diagonalized. A real 3×3 orthogonal matrix has a real eigenvalue. Real symmetric matrices, proof that eigenvalues are real, and that distinct eigenvalues give orthogonal basis of eigenvectors. Brief discussion of quadratic forms, conics and their classification. Canonical forms for 2×2 matrices; discussion of relation between eigenvalues of a matrix and fixed points of the corresponding Möbius map. [5]

Axioms for groups; subgroups and group actions. Orbits, stabilizers, cosets and conjugate subgroups. Orbit-stabilizer theorem. Lagrange's theorem. Examples from geometry, including the Euclidean groups, symmetry groups of regular polygons, cube and tetrahedron. The Möbius group; cross-ratios, preservation of circles, informal treatment of the point at infinity. [11]

Isomorphisms and homomorphisms of abstract groups, the kernel of a homomorphism. Examples. Introduction to normal subgroups, quotient groups and the isomorphism theorem. Permutations, cycles and transpositions. The sign of a permutation. [5]

Examples (only) of matrix groups; for example, the general and special linear groups, the orthogonal and special orthogonal groups, unitary groups, the Lorentz groups, quaternions and Pauli spin matrices. [2]

Appropriate books

M.A. Armstrong *Groups and Symmetry.* Springer–Verlag 1988 (£33.00 hardback).
† Alan F Beardon *Algebra and Geometry.* CUP 2005 (£21.99 paperback, £48 hardback).

[1] See `http://www.maths.cam.ac.uk/undergrad/schedules/`.

D.M. Bloom *Linear Algebra and Geometry.* Cambridge University Press 1979 (out of print).
D.E. Bourne and P.C. Kendall *Vector Analysis and Cartesian Tensors.* Nelson Thornes 1992 (£30.75 paperback).
R.P. Burn *Groups, a Path to Geometry.* Cambridge University Press 1987 (£20.95 paperback).
J.A. Green *Sets and Groups: a first course in Algebra.* Chapman and Hall/CRC 1988 (£38.99 paperback).
E. Sernesi *Linear Algebra: A Geometric Approach.* CRC Press 1993 (£38.99 paperback).
D. Smart *Linear Algebra and Geometry.* Cambridge University Press 1988 (out of print).

0.2 Lectures

- Lectures will start at 11:05 promptly with a summary of the last lecture. Please be on time since it is distracting to have people walking in late.
- I will endeavour to have a 2 minute break in the middle of the lecture for a rest and/or jokes and/or politics and/or paper aeroplanes[2]; students seem to find that the break makes it easier to concentrate throughout the lecture.[3]
- I will aim to finish by 11:55, but am not going to stop dead in the middle of a long proof/explanation.
- I will stay around for a few minutes at the front after lectures in order to answer questions.
- By all means chat to each other *quietly* if I am unclear, but please do not discuss, say, last night's football results, or who did (or did not) get drunk and/or laid. Such chatting is a distraction.
- I want you to learn. I will do my best to be clear but you *must* read through and understand your notes before the next lecture . . . otherwise you will get hopelessly lost (especially at 6 lectures a week). An understanding of your notes will not diffuse into you just because you have carried your notes around for a week . . . or put them under your pillow.
- I welcome *constructive* heckling. If I am inaudible, illegible, unclear or just plain wrong then please shout out.
- I aim to avoid the words *trivial*, *easy*, *obvious* and *yes*[4]. Let me know if I fail. I will occasionally use *straightforward* or *similarly to last time*; if it is not, email me (`S.J.Cowley@damtp.cam.ac.uk`) or catch me at the end of the next lecture.
- Sometimes I may confuse both you and *myself* (I am not infallible), and may not be able to extract myself in the middle of a lecture. Under such circumstances I will have to plough on as a result of time constraints; however I will clear up any problems at the beginning of the next lecture.
- The course is somewhere between *applied* and *pure* mathematics. Hence do not always expect pure mathematical levels of rigour; having said that all the outline/sketch 'proofs' could in principle be tightened up given sufficient time.
- If anyone is colour blind please come and tell me which colour pens you cannot read.
- Finally, I was in your position 32 years ago and nearly gave up the Tripos. If you feel that the course is going over your head, or you are spending more than 20-24 hours a week on it, come and chat.

0.3 Printed Notes

- I hope that printed notes will be handed out for the course . . . so that you can listen to me rather than having to scribble things down (but that depends on my typing keeping ahead of my lecturing). If it is not in the notes or on the example sheets it should not be in the exam.

[2] If you throw paper aeroplanes please pick them up. I will pick up the first one to stay in the air for 5 seconds.
[3] Having said that, research suggests that within the first 20 minutes I will, at some point, have lost the attention of all of you.
[4] But I will fail miserably in the case of *yes*.

- Any notes will only be available in lectures and only *once* for each set of notes.
- I do not keep back-copies (otherwise my office would be an even worse mess) ... from which you may conclude that I will not have copies of last time's notes (so do not ask).
- There will only be approximately as many copies of the notes as there were students at the lecture on the previous Saturday.[5] We are going to fell a forest as it is, and I have no desire to be even more environmentally unsound.
- Please do not take copies for your absent friends unless they are *ill*, but if they are ill then please take copies.[6]
- The notes are deliberately *not* available on the WWW; they are an adjunct to lectures and are not meant to be used independently.
- If you do not want to attend lectures then there are a number of excellent textbooks that you can use in place of my notes.
- With one or two exceptions figures/diagrams are deliberately omitted from the notes. I was taught to do this at my teaching course on *How To Lecture* ... the aim being that it might help you to stay awake if you have to write something down from time to time.
- There are a number of unlectured worked examples in the notes. In the past I have been tempted to not include these because I was worried that students would be unhappy with material in the notes that was not lectured. However, a vote in one of my previous lecture courses was overwhelming in favour of including unlectured worked examples.
- Please email me corrections to the notes and example sheets (`S.J.Cowley@damtp.cam.ac.uk`).

0.4 Example Sheets

- There will be four main example sheets. They will be available on the WWW at about the same time as I hand them out (see `http://damtp.cam.ac.uk/user/examples/`). There will also be two supplementary 'study' sheets.
- You should be able to do example sheets 1/2/3/4 after lectures 6/12/18/23 respectively, or thereabouts. Please bear this in mind when arranging supervisions. Personally I suggest that you do not have your first supervision before the middle of week 2 of lectures.
- There is some repetition on the sheets by design; pianists do scales, athletes do press-ups, mathematicians do algebra/manipulation.
- Your *supervisors* might like to know that the example sheets will be the same as last year (but that they will not be the same next year).

0.5 Acknowledgements.

The following notes were adapted (i.e. stolen) from those of Peter Haynes, my esteemed Head of Department.

[5] With the exception of the first three lectures for the pedants.
[6] If you really have been ill and cannot find a copy of the notes, then come and see me, but bring your sick-note.

0.6 Revision.

You should check that you recall the following.

The Greek alphabet.

A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ	delta	Π	π	pi
E	ϵ	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ	sigma
H	η	eta	T	τ	tau
Θ	θ	theta	Υ	υ	upsilon
I	ι	iota	Φ	ϕ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

There are also typographic variations on epsilon (i.e. ε), phi (i.e. φ), and rho (i.e. ϱ).

The exponential function. The exponential function, $\exp(x)$, is defined by the series

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}\,. \tag{0.1a}$$

It has the following properties

$$\begin{aligned} \exp(0) &= 1\,, && (0.1b)\\ \exp(1) &= e \approx 2.71828183\,, && (0.1c)\\ \exp(x) &= e^x\,, && (0.1d)\\ e^{x+y} &= e^x e^y\,. && (0.1e) \end{aligned}$$

The logarithm. The logarithm of $x > 0$, i.e. $\log x$, is defined as the unique solution y of the equation

$$\exp(y) = x\,. \tag{0.2a}$$

It has the following properties

$$\begin{aligned} \log(1) &= 0\,, && (0.2b)\\ \log(e) &= 1\,, && (0.2c)\\ \log(\exp(x)) &= x\,, && (0.2d)\\ \log(xy) &= \log x + \log y\,. && (0.2e) \end{aligned}$$

The sine and cosine functions. The sine and cosine functions are defined by the series

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-)^n x^{2n+1}}{(2n+1)!}\,, \tag{0.3a}$$

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-)^n x^{2n}}{2n!}\,. \tag{0.3b}$$

Certain trigonometric identities. You should recall the following

$$\begin{aligned}
\sin(x \pm y) &= \sin(x)\cos(y) \pm \cos(x)\sin(y)\,, && (0.4a)\\
\cos(x \pm y) &= \cos(x)\cos(y) \mp \sin(x)\sin(y)\,, && (0.4b)\\
\tan(x \pm y) &= \frac{\tan(x) \pm \tan(y)}{1 \mp \tan(x)\tan(y)}\,, && (0.4c)\\
\cos(x) + \cos(y) &= 2\cos\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)\,, && (0.4d)\\
\sin(x) + \sin(y) &= 2\sin\left(\frac{x+y}{2}\right)\cos\left(\frac{x-y}{2}\right)\,, && (0.4e)\\
\cos(x) - \cos(y) &= -2\sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)\,, && (0.4f)\\
\sin(x) - \sin(y) &= 2\cos\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)\,. && (0.4g)
\end{aligned}$$

The cosine rule.

Let ABC be a triangle. Let the lengths of the sides opposite vertices A, B and C be a, b and c respectively. Further suppose that the angles subtended at A, B and C are α, β and γ respectively. Then the cosine rule (also known as the cosine formula or law of cosines) states that

$$\begin{aligned}
a^2 &= b^2 + c^2 - 2bc\cos\alpha\,, && (0.5a)\\
b^2 &= a^2 + c^2 - 2ac\cos\beta\,, && (0.5b)\\
c^2 &= a^2 + b^2 - 2ab\cos\gamma\,. && (0.5c)
\end{aligned}$$

The equation of a line. In 2D Cartesian co-ordinates, (x, y), the equation of a line with slope m which passes through (x_0, y_0) is given by

$$y - y_0 = m(x - x_0)\,. \tag{0.6a}$$

In parametric form the equation of this line is given by

$$x = x_0 + a\,t\,, \quad y = y_0 + a\,m\,t\,, \tag{0.6b}$$

where t is the parametric variable and a is an arbitrary real number.

The equation of a circle. In 2D Cartesian co-ordinates, (x, y), the equation of a circle of radius r and centre (p, q) is given by

$$(x - p)^2 + (y - q)^2 = r^2\,. \tag{0.7}$$

Suggestions.

Examples.

1. Introduce a sheet 0 covering revision, e.g. $\log(xy) = \log(x) + \log(y)$.
2. Include a complex numbers study sheet, using some of the Imperial examples?
3. Use the following alternative proof of Schwarz's inequality as an example:

$$\begin{aligned} \|\mathbf{x}\|^2\|\mathbf{y}\|^2 - |\mathbf{x}\cdot\mathbf{y}|^2 &= x_i x_i y_j y_j - x_i y_i x_j y_j \\ &= \tfrac{1}{2}x_i x_i y_j y_j + \tfrac{1}{2}x_j x_j y_i y_i - x_i y_i x_j y_j \\ &= \tfrac{1}{2}(x_i y_j - x_j y_i)(x_i y_j - x_j y_i) \\ &\geqslant 0\,. \end{aligned}$$

Additions/Subtractions?

1. Add more grey summation signs when introducing the summation convention.
2. Add a section on diagonal matrices, e.g. multiplication (using suffix notation).
3. Change of basis: add pictures and an example of diagonalisation (e.g. rotated reflection matrix, or reflection and dilatation).
4. Add proof of rank-nullity theorem, and simplify notes on Gaussian elimination.
5. Add construction of the inverse matrix when doing Gaussian elimination.
6. $n \times n$ determinants, inverses, and Gaussian elimination.
7. Do (Shear Matrix)^n, and geometric interpretation.
8. Deliberate mistake a day (with mars bar).

1 Complex Numbers

1.0 Why Study This?

For the same reason as we study real numbers: because they are useful and occur throughout mathematics. For many of you this section is revision, for a few of you who have not done, say, OCR FP1 and FP3 this will be new. For those for which it is revision, please do not fall asleep; instead note the speed that new concepts are being introduced.

1.1 Introduction

1.1.1 Real numbers

The real numbers are denoted by $\mathbb{R}$ and consist of:

integers,	denoted by $\mathbb{Z}$,	$\ldots -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ \ldots$
rationals,	denoted by $\mathbb{Q}$,	p/q where $p,\ q$ are integers ($q \neq 0$)
irrationals,		the rest of the reals, e.g. $\sqrt{2},\ e,\ \pi,\ \pi^2$.

We sometimes visualise real numbers as lying on a line (e.g. between any two points on a line there is another point, and between any two real numbers there is always another real number).

1.1.2 The general solution of a quadratic equation

Consider the quadratic equation

$$\alpha z^2 + \beta z + \gamma = 0 \quad : \quad \alpha, \beta, \gamma \in \mathbb{R} \quad , \alpha \neq 0 \,,$$

where $\in$ means 'belongs to'. This has two roots

$$z_1 = -\frac{\beta + \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \quad \text{and} \quad z_2 = -\frac{\beta - \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \,. \tag{1.1}$$

If $\beta^2 \geqslant 4\alpha\gamma$ then the roots are real (there is a repeated root if $\beta^2 = 4\alpha\gamma$). If $\beta^2 < 4\alpha\gamma$ then the square root is not equal to any real number. In order that we can *always* solve a quadratic equation, we introduce

$$i = \sqrt{-1}\,. \tag{1.2}$$

Remark. Note that i is sometimes denoted by j by engineers.

If $\beta^2 < 4\alpha\gamma$, (1.1) can now be rewritten

$$z_1 = -\frac{\beta}{2\alpha} + i\frac{\sqrt{4\alpha\gamma - \beta^2}}{2\alpha} \quad \text{and} \quad z_2 = -\frac{\beta}{2\alpha} - i\frac{\sqrt{4\alpha\gamma - \beta^2}}{2\alpha} \,, \tag{1.3}$$

where the square roots are now real [numbers]. Subject to us being happy with the introduction and existence of i, we can now *always* solve a quadratic equation.

1.1.3 Complex numbers

Complex numbers are denoted by $\mathbb{C}$. We define a complex number, say z, to be a number with the form

$$z = a + ib, \quad \text{where} \quad a, b \in \mathbb{R}, \tag{1.4}$$

and i is as defined in (1.2). We say that $z \in \mathbb{C}$. Note that z_1 and z_2 in (1.3) are of this form.

For $z = a + ib$, we sometimes write

$$\begin{aligned} a = \operatorname{Re}(z) \quad &: \quad \text{the real part of } z, \\ b = \operatorname{Im}(z) \quad &: \quad \text{the imaginary part of } z. \end{aligned}$$

Remarks.

1. Extending the number system from real ($\mathbb{R}$) to complex ($\mathbb{C}$) allows a number of important generalisations in addition to always being able to solve a quadratic equation, e.g. it makes solving certain differential equations much easier.
2. $\mathbb{C}$ contains all real numbers since if $a \in \mathbb{R}$ then $a + i.0 \in \mathbb{C}$.
3. A complex number $0 + i.b$ is said to be *pure imaginary*.
4. Complex numbers were first used by Tartaglia (1500-1557) and Cardano (1501-1576). The terms *real* and *imaginary* were first introduced by Descartes (1596-1650).

Theorem 1.1. *The representation of a complex number z in terms of its real and imaginary parts is unique.*

Proof. Assume $\exists\ a, b, c, d \in \mathbb{R}$ such that

$$z = a + ib = c + id.$$

Then $a - c = i\,(d - b)$, and so $(a - c)^2 = -\,(d - b)^2$. But the only number greater than or equal to zero that is equal to a number that is less than or equal to zero, is zero. Hence $a = c$ and $b = d$. $\square$

Corollary 1.2. *If $z_1 = z_2$ where $z_1, z_2 \in \mathbb{C}$, then* $\mathrm{Re}\,(z_1) = \mathrm{Re}\,(z_2)$ *and* $\mathrm{Im}\,(z_1) = \mathrm{Im}\,(z_2)$.

1.2 Algebraic Manipulation of Complex Numbers

In order to manipulate complex numbers simply follow the rules for reals, but adding the rule $i^2 = -1$. Hence for $z_1 = a + ib$ and $z_2 = c + id$, where $a, b, c, d \in \mathbb{R}$, we have that

$$\text{addition/subtraction}: \quad z_1 + z_2 = (a + ib) \pm (c + id) = (a \pm c) + i\,(b \pm d)\,; \tag{1.5}$$

$$\begin{aligned} \text{multiplication}: \quad z_1\, z_2 &= (a + ib)\,(c + id) = ac + ibc + ida + (ib)\,(id) \\ &= (ac - bd) + i\,(bc + ad)\,; \end{aligned} \tag{1.6}$$

$$\text{inverse}: \quad z_1^{-1} = \frac{1}{z} = \frac{1}{a + ib}\,\frac{a - ib}{a - ib} = \frac{a}{a^2 + b^2} - \frac{ib}{a^2 + b^2}. \tag{1.7}$$

Exercise: For z_1^{-1} as defined in (1.7), check that $z_1\, z_1^{-1} = 1 + i.0$.

Remark. All the above operations on elements of $\mathbb{C}$ result in new elements of $\mathbb{C}$. This is described as *closure*: $\mathbb{C}$ is closed under addition and multiplication.

1.3 Functions of Complex Numbers

We may extend the idea of functions to complex numbers. A complex-valued function f is one that takes a complex number as 'input' and defines a new complex number $f(z)$ as 'output'.

1.3.1 Complex conjugate

The *complex conjugate* of $z = a + ib$, which is usually written as $\overline{z}$ but sometimes as z^*, is defined as $a - ib$, i.e.

$$\text{if} \quad z = a + ib \quad \text{then} \quad \overline{z} = a - ib. \tag{1.8}$$

Exercises. Show that

1. $$\overline{\overline{z}} = z\,. \tag{1.9a}$$

2. $$\overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}. \tag{1.9b}$$

3. $$\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}. \tag{1.9c}$$

4. $$\overline{(z^{-1})} = (\overline{z})^{-1}. \tag{1.9d}$$

Definition. Given a complex-valued function f, the complex conjugate function $\overline{f}$ is defined by

$$\overline{f}\,(\overline{z}) = \overline{f\,(z)}, \quad \text{and hence from (1.9a)} \quad \overline{f}\,(z) = \overline{f\,(\overline{z})}. \tag{1.10}$$

Example. Let $f\,(z) = pz^2 + qz + r$ with $p, q, r \in \mathbb{C}$ then by using (1.9b) and (1.9c)

$$\overline{f}\,(\overline{z}) \equiv \overline{f\,(z)} = \overline{pz^2 + qz + r} = \overline{p}\,\overline{z}^2 + \overline{q}\,\overline{z} + \overline{r}.$$

Hence $\overline{f}\,(z) = \overline{p}\,z^2 + \overline{q}\,z + \overline{r}$.

1.3.2 Modulus

The *modulus* of $z = a + ib$, which is written as $|z|$, is defined as

$$|z| = \left(a^2 + b^2\right)^{1/2}\,. \tag{1.11}$$

Exercises. Show that

1. $$|z|^2 = z\,\overline{z}. \tag{1.12a}$$

2. $$z^{-1} = \frac{\overline{z}}{|z|^2}\,. \tag{1.12b}$$

1/03

1.4 The Argand Diagram

Consider the set of points in two dimensional (2D) space referred to Cartesian axes. Then we can represent each $z = x + iy \in \mathbb{C}$ by the point (x, y), i.e. the real and imaginary parts of z are viewed as coordinates in an xy plot. We label the 2D vector between the origin and (x, y), say $\overrightarrow{OP}$, by the complex number z. Such a plot is called an *Argand diagram.*

Remarks.

1. The xy plane is referred to as the *complex plane*. We refer to the x-axis as the *real axis*, and the y-axis as the *imaginary axis.*
2. The Argand diagram was invented by Caspar Wessel (1797), and re-invented by Jean-Robert Argand (1806).

Modulus. The modulus of z corresponds to the magnitude of the vector $\vec{OP}$ since

$$|z| = \left(x^2 + y^2\right)^{1/2} .$$

Complex conjugate. If $\vec{OP}$ represents z, then $\vec{OP'}$ represents $\overline{z}$, where P' is the point $(x, -y)$; i.e. P' is P reflected in the x-axis.

Addition. Let $z_1 = x_1 + iy_1$ be associated with P_1, and $z_2 = x_2 + iy_2$ be associated with P_2. Then

$$z_3 = z_1 + z_2 = (x_1 + x_2) + i\,(y_1 + y_2) \ ,$$

is associated with the point P_3 that is obtained by completing the parallelogram $P_1\,O\,P_2P_3$. In terms of vector addition

$$\vec{OP_3} = \vec{OP_1} + \vec{OP_2} \,,$$

which is sometimes called the *triangle law.*

1/02

Theorem 1.3. *If $z_1, z_2 \in \mathbb{C}$ then*

$$\begin{aligned} |z_1 + z_2| &\leqslant |z_1| + |z_2| \,, && (1.13a)\\ |z_1 - z_2| &\geqslant \big|\,|z_1| - |z_2|\,\big| \,. && (1.13b) \end{aligned}$$

Remark. Result (1.13a) is known as the *triangle inequality* (and is in fact one of many).

Proof. By the cosine rule

$$\begin{aligned} |z_1 + z_2|^2 &= |z_1|^2 + |z_2|^2 - 2|z_1|\,|z_2|\cos\psi \\ &\leqslant |z_1|^2 + |z_2|^2 + 2|z_1|\,|z_2| \\ &= \left(|z_1| + |z_2|\right)^2 \,. \end{aligned}$$

(1.13b) follows from (1.13a). Let $z_1' = z_1 + z_2$ and $z_2' = z_2$, so that $z_1 = z_1' - z_2'$. Then (1.13a) implies that

$$|z_1'| \leqslant |z_1' - z_2'| + |z_2'| \,,$$

and hence that

$$|z_1' - z_2'| \geqslant |z_1'| - |z_2'| \,.$$

Interchanging z_1' and z_2' we also have that

$$|z_2' - z_1'| = |z_1' - z_2'| \geqslant |z_2'| - |z_1'| \,.$$

(1.13b) follows. $\square$

1.5 Polar (Modulus/Argument) Representation

Another helpful representation of complex numbers is obtained by using plane polar co-ordinates to represent position in Argand diagram. Let $x = r\cos\theta$ and $y = r\sin\theta$, then

$$\begin{aligned} z = x + iy &= r\cos\theta + ir\sin\theta \\ &= r(\cos\theta + i\sin\theta)\ . \end{aligned} \qquad (1.14)$$

Note that

$$|z| = \left(x^2 + y^2\right)^{1/2} = r\,. \qquad (1.15)$$

- Hence r is the modulus of z (mod(z) for short).
- θ is called the *argument* of z ($\arg(z)$ for short).
- The expression for z in terms of r and θ is
1/06 called the *modulus/argument form*.

The pair (r, θ) specifies z uniquely. However, z does *not* specify (r, θ) uniquely, since adding $2n\pi$ to θ ($n \in \mathbb{Z}$, i.e. the integers) does not change z. For each z there is a unique value of the argument θ such that $-\pi < \theta \leqslant \pi$, sometimes called the *principal value* of the argument.

Remark. In order to get a unique value of the argument it is sometimes more convenient to restrict θ to $0 \leqslant \theta < 2\pi$

1.5.1 Geometric interpretation of multiplication

Consider z_1, z_2 written in modulus argument form:

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

Then

$$\begin{aligned} z_1 z_2 &= r_1 r_2 (\cos\theta_1 . \cos\theta_2 - \sin\theta_1 . \sin\theta_2 \\ &\qquad + i(\sin\theta_1 . \cos\theta_2 + \sin\theta_2 . \cos\theta_1)) \\ &= r_1 r_2 (\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2))\ . \end{aligned} \qquad (1.16)$$

Hence

$$|z_1 z_2| = |z_1|\,|z_2|\,, \qquad (1.17a)$$

$$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) \quad (+2n\pi \text{ with } n \text{ an arbitrary integer}). \qquad (1.17b)$$

In words: multiplication of z_1 by z_2 scales z_1 by $|\, z_2 \,|$ and rotates z_1 by $\arg(z_2)$.

Exercise. Find equivalent results for z_1/z_2.

1.6 The Exponential Function

1.6.1 The real exponential function

The real exponential function, $\exp(x)$, is defined by the power series

$$\exp(x) = \exp x = 1 + x + \frac{x^2}{2!} \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} . \tag{1.18}$$

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

Worked exercise. Show for $x, y \in \mathbb{R}$ that

$$(\exp x)\,(\exp y) = \exp\,(x+y) . \tag{1.19}$$

Solution.

$$\begin{aligned}
\exp x\ \exp y &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{\infty} \frac{y^m}{m!} \\
&= \sum_{r=0}^{\infty} \sum_{m=0}^{r} \frac{x^{r-m}}{(r-m)!} \frac{y^m}{m!} \quad \text{for } n = r - m \\
&= \sum_{r=0}^{\infty} \frac{1}{r!} \sum_{m=0}^{r} \frac{r!}{(r-m)!m!}\ x^{r-m} y^m \\
&= \sum_{r=0}^{\infty} \frac{(x+y)^r}{r!} \quad \text{by the binomial theorem} \\
&= \exp(x+y) .
\end{aligned}$$

Definition. We write

$$\exp(1) = e .$$

Worked exercise. Show for $n, p, q \in \mathbb{Z}$, where without loss of generality (wlog) $q > 0$, that:

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

Solution. For $n = 1$ there is nothing to prove. For $n \geqslant 2$, and using (1.19),

$$\exp(n) = \exp(1)\exp(n-1) = e\ \exp(n-1) , \quad \text{and thence by induction} \quad \exp(n) = e^n .$$

From the power series definition (1.18) with $n = 0$:

$$\exp(0) = 1 = e^0 .$$

Also from (1.19) we have that

$$\exp(-1)\exp(1) = \exp(0) , \quad \text{and thence} \quad \exp(-1) = \frac{1}{e} = e^{-1} .$$

For $n \leqslant -2$ proceed by induction as above.

Next note from applying (1.19) q times that

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

Thence on taking the positive qth root

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

Definition. For irrational x, define

$$e^x = \exp(x)\,.$$

From the above it follows that if $y \in \mathbb{R}$, then it is consistent to write $\exp(y) = e^y$.

1.6.2 The complex exponential function

Definition. For $z \in \mathbb{C}$, the *complex exponential* is defined by

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}\,. \tag{1.20}$$

This series converges for all finite $|z|$ (again see the *Analysis I* course).

Remarks. When $z \in \mathbb{R}$ this definition is consistent with (1.18). For $z_1, z_2 \in \mathbb{C}$,

$$(\exp z_1)\,(\exp z_2) = \exp\,(z_1 + z_2)\ ,$$

with the proof essentially as for (1.19).

Definition. For $z \in \mathbb{C}$ and $z \notin \mathbb{R}$ we define

$$e^z = \exp(z)\,,$$

Remark. This definition is consistent with the earlier results and definitions for $z \in \mathbb{R}$.

1.6.3 The complex trigonometric functions

Theorem 1.4. *For $w \in \mathbb{C}$*

$$\exp\,(iw) \equiv e^{iw} = \cos\, w + i \sin\, w\,. \tag{1.21}$$

Proof. First consider w real. Then from using the power series definitions for cosine and sine when their arguments are real, we have that

$$\begin{aligned}
\exp\,(iw) &= \sum_{n=0}^{\infty} \frac{(iw)^n}{n!} = 1 + iw - \frac{w^2}{2} - i\frac{w^3}{3!}\ldots \\
&= \left(1 - \frac{w^2}{2!} + \frac{w^4}{4!}\ldots\right) + i\left(w - \frac{w^3}{3!} + \frac{w^5}{5!}\ldots\right) \\
&= \sum_{n=0}^{\infty} (-1)^n \frac{w^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{w^{2n+1}}{(2n+1)!} \\
&= \cos\, w + i \sin\, w\,,
\end{aligned}$$

which is as required (as long as we do not mind living dangerously and re-ordering infinite series). Next, for $w \in \mathbb{C}$ *define* the complex trigonometric functions by[7]

$$\cos\, w = \sum_{n=0}^{\infty} (-1)^n \frac{w^{2n}}{(2n)!} \quad \text{and} \quad \sin\, w = \sum_{n=0}^{\infty} (-1)^n \frac{w^{2n+1}}{(2n+1)!}\,. \tag{1.22}$$

The result (1.21) then follows for w complex. $\square$

[7] Again note that these definitions are consistent with the definitions when the arguments are real.

Remarks.

1. From taking the complex conjugate of (1.21), or otherwise,
$$\exp(-iw) \equiv e^{-iw} = \cos w - i \sin w\,. \tag{1.23}$$

2. From (1.21) and (1.23) it follows that
$$\cos w = \tfrac{1}{2}\left(e^{iw} + e^{-iw}\right)\,, \quad \text{and} \quad \sin w = \frac{1}{2i}\left(e^{iw} - e^{-iw}\right)\,. \tag{1.24}$$

2/02

1.6.4 Relation to modulus/argument form

Let $w = \theta$ where $\theta \in \mathbb{R}$. Then from (1.21)
$$e^{i\theta} = \cos\theta + i\sin\theta\,. \tag{1.25}$$

It follows from the polar representation (1.14) that
$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}\,, \tag{1.26}$$
with (again) $r = |z|$ and $\theta = \arg z$. In this representation the multiplication of two complex numbers is rather elegant:
$$z_1 z_2 = \left(r_1 e^{i\theta_1}\right)\left(r_2 e^{i\theta_2}\right) = r_1 r_2\, e^{i(\theta_1+\theta_2)}\,,$$
confirming (1.17a) and (1.17b).

1.6.5 Modulus/argument expression for 1

Consider solutions of
$$z = re^{i\theta} = 1\,.$$
Since by definition $r, \theta \in \mathbb{R}$, it follows that $r = 1$,
$$e^{i\theta} = \cos\theta + i\sin\theta = 1\,,$$
and thence that $\cos\theta = 1$ and $\sin\theta = 0$. We deduce that
$$\theta = 2k\pi\,, \quad \text{for} \quad k \in \mathbb{Z}.$$

2/03

1.7 Roots of Unity

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

Theorem 1.5. *There are n solutions of $z^n = 1$ (i.e. there are n 'nth roots of unity')*

Proof. One solution is $z = 1$. Seek more general solutions of the form $z = r\,e^{i\theta}$ with the restriction $0 \leqslant \theta < 2\pi$ so that θ is not multi-valued. Then
$$\left(r\,e^{i\theta}\right)^n = r^n e^{in\theta} = 1\,, \tag{1.27}$$
and hence from §1.6.5, $r^n = 1$ and $n\,\theta = 2k\pi$ with $k \in \mathbb{Z}$. We conclude that within the requirement that $0 \leqslant \theta < 2\pi$, there are n distinct roots given by
$$\theta = \frac{2k\pi}{n} \quad \text{with} \quad k = 0, 1, \ldots, n-1. \tag{1.28}$$
$\square$

Remark. If we write $\omega = e^{2\pi i/n}$, then the roots of $z^n = 1$ are $1, \omega, \omega^2, \dots, \omega^{n-1}$. Further, for $n \geqslant 2$

$$1 + \omega + \cdots + \omega^{n-1} = \sum_{k=0}^{n-1} \omega^k = \frac{1-\omega^n}{1-\omega} = 0\,, \tag{1.29}$$

because $\omega^n = 1$.

Example. Solve $z^5 = 1$.

Solution. Put $z = e^{i\theta}$, then we require that

$$e^{5i\theta} = e^{2\pi k i} \quad \text{for} \quad k \in \mathbb{Z}.$$

There are thus five distinct roots given by

$$\theta = 2\pi k/5 \quad \text{with} \quad k = 0, 1, 2, 3, 4.$$

Larger (or smaller) values of k yield no new roots. If we write $\omega = e^{2\pi i/5}$, then the roots are $1, \omega, \omega^2, \omega^3, \omega^4$, and

$$1 + \omega + \omega^2 + \omega^3 + \omega^4 = 0\,.$$

Each root corresponds to a vertex of a pentagon.

2/06

1.8 De Moivre's Theorem

Theorem 1.6. *De Moivre's theorem states that for $\theta \in \mathbb{R}$ and $n \in \mathbb{Z}$*

$$\cos n\theta + i \sin n\theta = (\cos\theta + i\sin\theta)^n\,. \tag{1.30}$$

Proof. From (1.25)

$$\begin{aligned} \cos n\theta + i\sin n\theta &= e^{i\,(n\theta)} \\ &= \left(e^{i\,\theta}\right)^n \\ &= (\cos\,\theta + i\sin\,\theta)^n \end{aligned}$$

□

Remark. Although De Moivre's theorem requires $\theta \in \mathbb{R}$ and $n \in \mathbb{Z}$, (1.30) holds for $\theta, n \in \mathbb{C}$ in the sense that $(\cos n\theta + i\sin n\theta)$ (single valued) is one value of $(\cos\theta + i\sin\theta)^n$ (multivalued).

Alternative proof (unlectured). (1.30) is true for $n = 0$. Now argue by induction.

Assume true for $n = p > 0$, i.e. assume that $(\cos\theta + i\sin\theta)^p = \cos p\theta + i\sin p\theta$. Then

$$\begin{aligned} (\cos\theta + i\sin\theta)^{p+1} &= (\cos\theta + i\sin\theta)\,(\cos\theta + i\sin\theta)^p \\ &= (\cos\theta + i\sin\theta)\,(\cos p\theta + i\sin p\theta) \\ &= \cos\theta.\cos p\theta - \sin\theta.\sin p\theta + i\,(\sin\theta.\cos p\theta + \cos\theta.\sin p\theta) \\ &= \cos\,(p+1)\,\theta + i\sin\,(p+1)\,\theta\,. \end{aligned}$$

Hence the result is true for $n = p+1$, and so holds for all $n \geqslant 0$. Now consider $n < 0$, say $n = -p$. Then, using the proved result for $p > 0$,

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

Hence De Moivre's theorem is true $\forall\ n \in \mathbb{Z}$.

1.9 Logarithms and Complex Powers

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

Definition. We define $\log z$ for $z \in \mathbb{C}$ as 'the' solution w of

$$e^w = z\,. \tag{1.31}$$

To understand the nature of the complex logarithm let $w = u + iv$ with $u, v \in \mathbb{R}$. Then $e^{u+iv} = z = re^{i\theta}$, and hence

$$\begin{aligned} e^u &= |z| = r\,. \\ v &= \arg z = \theta + 2k\pi \quad \text{for any } k \in \mathbb{Z}\,. \end{aligned}$$

Thus

$$w = \log z = \log |z| + i \arg z\,. \tag{1.32a}$$

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

Definition. The principal value of $\log z$ is such that

$$-\pi < \arg z = \operatorname{Im}(\log z) \leqslant \pi. \tag{1.32b}$$

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

$$\begin{aligned} \log z &= \log |-x| + i \arg(-x) \\ &= \log |x| + (2k+1)i\pi \quad \text{for any } k \in \mathbb{Z}\,. \end{aligned}$$

The principal value of $\log(-x)$ is $\log|x| + i\pi$.

1.9.1 Complex powers

Recall the definition of x^a, for $x, a \in \mathbb{R}$, $x > 0$ and a irrational, namely

$$x^a = e^{a \log x} = \exp(a \log x)\,.$$

Definition. For $z \neq 0$, $z, w \in \mathbb{C}$, define z^w by

$$z^w = e^{w \log z}. \tag{1.33}$$

Remark. Since $\log z$ is multi-valued so is z^w, i.e. z^w is only defined upto an arbitrary multiple of $e^{2ki\pi w}$, for any $k \in \mathbb{Z}$.

Example. The value of i^i is given by

$$\begin{aligned} i^i &= e^{i \log i} \\ &= e^{i(\log|i| + i \arg i)} \\ &= e^{i(\log 1 + 2ki\,\pi + i\pi/2)} \\ &= e^{-\left(2k+\frac{1}{2}\right)\pi} \quad \text{for any } k \in \mathbb{Z} \quad \text{(which is real).} \end{aligned}$$

1.10 Lines and Circles in the Complex Plane

1.10.1 Lines

For fixed $z_0, w \in \mathbb{C}$ with $w \neq 0$, and varying $\lambda \in \mathbb{R}$, the equation

$$z = z_0 + \lambda w \tag{1.34a}$$

represents in the Argand diagram (complex plane) points on straight line through z_0 and parallel to w.

Remark. Since $\lambda = (z - z_0)/w \in \mathbb{R}$, it follows that $\lambda = \bar{\lambda}$, and hence that

$$\frac{z - z_0}{w} = \frac{\bar{z} - \bar{z}_0}{\bar{w}} .$$

Thus

$$z\bar{w} - \bar{z}w = z_0\bar{w} - \bar{z}_0 w \tag{1.34b}$$

is an alternative representation of the line.

Worked exercise. Show that $z_0\bar{w} - \bar{z}_0 w = 0$ if and only if (iff) the line (1.34a) passes through the origin.

Solution. If the line passes through the origin then put $z = 0$ in (1.34b), and the result follows. If $z_0\bar{w} - \bar{z}_0 w = 0$, then the equation of the line is $z\bar{w} - \bar{z}w = 0$. This is satisfied by $z = 0$, and hence the line passes through the origin.

Exercise. Show that if $z\bar{w} - \bar{z}w = 0$, then $z = \gamma w$ for some $\gamma \in \mathbb{R}$.

1.10.2 Circles

In the Argand diagram, a circle of radius $r \neq 0$ and centre v ($r \in \mathbb{R}$, $v \in \mathbb{C}$) is given by

$$S = \{z \in \mathbb{C} \ : \mid z - v \mid = r\} , \tag{1.35a}$$

i.e. the set of complex numbers z such that $|z - v| = r$.

Remarks.

- If $z = x + iy$ and $v = p + iq$ then
$$|z - v|^2 = (x - p)^2 + (y - q)^2 = r^2 ,$$
which is the equation for a circle with centre (p, q) and radius r in Cartesian coordinates.
- Since $|z - v|^2 = (\bar{z} - \bar{v})(z - v)$, an alternative equation for the circle is
$$|z|^2 - \bar{v}z - v\bar{z} + |v|^2 = r^2 . \tag{1.35b}$$

3/02
3/03

1.11 Möbius Transformations

Consider a *map* of $\mathbb{C} \to \mathbb{C}$ ('$\mathbb{C}$ into $\mathbb{C}$') defined by

$$z \mapsto z' = f(z) = \frac{az + b}{cz + d} , \tag{1.36}$$

where $a, b, c, d \in \mathbb{C}$ are constants. We require that

(i) c and d are not both zero, so that the map is finite (except at $z = -d/c$);

(ii) different points map to different points, i.e. if $z_1 \neq z_2$ then $z'_1 \neq z'_2$, i.e. we require that

$$\frac{az_1 + b}{cz_1 + d} \neq \frac{az_2 + b}{cz_2 + d}, \quad \text{or equivalently} \quad (ad - bc)(z_1 - z_2) \neq 0, \quad \text{i.e.} \quad (ad - bc) \neq 0.$$

Remarks.

- Condition (i) is a subset of condition (ii), hence we need only require that $(ad - bc) \neq 0$.
- $f(z)$ maps every point of the complex plane, except $z = -d/c$, into another ($z = -d/c$ is mapped to infinity).
- Adding the 'point at infinity' makes f complete.

1.11.1 Composition

Consider a second Möbius transformation

$$z' \mapsto z'' = g(z') = \frac{\alpha z' + \beta}{\gamma z' + \delta} \quad \text{where} \quad \alpha, \beta, \gamma, \delta \in \mathbb{C}, \quad \text{and} \quad \alpha\delta - \beta\gamma \neq 0.$$

Then the combined map $z \mapsto z''$ is also a Möbius transformation:

$$\begin{aligned} z'' = g(z') &= g(f(z)) \\ &= \frac{\alpha\left(\frac{az+b}{cz+d}\right) + \beta}{\gamma\left(\frac{az+b}{cz+d}\right) + \delta} \\ &= \frac{\alpha(az+b) + \beta(cz+d)}{\gamma(az+b) + \delta(cz+d)} \\ &= \frac{(\alpha a + \beta c)z + (\alpha b + \beta d)}{(\gamma a + \delta c)z + (\gamma b + \delta d)}, \end{aligned} \tag{1.37}$$

where we note that $(\alpha a + \beta c)(\gamma b + \delta d) - (\alpha b + \beta d)(\gamma a + \delta c) = (ad - bc)(\alpha\delta - \beta\gamma) \neq 0$. Hence the set of all Möbius maps is therefore *closed* under composition.

1.11.2 Inverse

For the $a, b, c, d \in \mathbb{R}$ as in (1.36), consider the Möbius map

$$z'' = \frac{-dz' + b}{cz' - a}, \tag{1.38}$$

i.e. $\alpha = -d$, $\beta = b$, $\gamma = c$ and $\delta = -a$. Then from (1.37), $z'' = z$. We conclude that (1.38) is the *inverse* to (1.36), and vice versa.

Remarks.

1. (1.36) maps $\mathbb{C} \setminus \{-d/c\}$ to $\mathbb{C} \setminus \{a/c\}$, while (1.38) maps $\mathbb{C} \setminus \{a/c\}$ to $\mathbb{C} \setminus \{-d/c\}$.
2. The inverse (1.38) can be deduced from (1.36) by formal manipulation.

Exercise. For Möbius maps f, g and h demonstrate that the maps are *associative*, i.e. $(fg)h = f(gh)$. Those who have taken Further Mathematics A-level should then conclude something.

1.11.3 Basic Maps

Translation. Put $a = 1$, $c = 0$ and $d = 1$ to obtain

$$z' = z + b\,. \tag{1.39a}$$

This map represents a translation; e.g lines map to parallel lines, while circles map to circles of the same radius but with a centre offset by b.

3/06

Dilatation and rotation. Next put $b = 0$, $c = 0$ and $d = 1$ so that

$$z' = az = |a||z|\, e^{i(\arg a + \arg z)}\,. \tag{1.39b}$$

This map scales z by $|a|$ and rotates z by $\arg a$ about the origin O.

The line $z = z_0 + \lambda w$, where $\lambda \in \mathbb{R}$ and $w \in \mathbb{C}$, becomes

$$z' = az_0 + \lambda aw = z_0' + \lambda w'\,,$$

where $z_0' = az_0$ and $w' = aw$, which is just another line.

The circle $|z - v| = r$ becomes

$$\left|\frac{z'}{a} - v\right| = r \quad \text{or equivalently} \quad |z' - v'| = r'\,,$$

where $v' = av$ and $r' = |a|r$, which is just another circle.

Inversion and reflection. Now put $a = 0$, $b = 1$, $c = 1$ and $d = 0$, so that $z' = \frac{1}{z}$. Thus if $z = re^{i\theta}$ then

$$z' = \frac{1}{r}e^{-i\theta}\,, \quad \text{i.e. } |z'| = |z|^{-1} \text{ and } \arg z' = -\arg z.$$

Hence this map represents *inversion in the unit circle* centred on the origin O, and *reflection in the real axis*.

The line $z = z_0 + \lambda w$, or equivalently (see (1.34b))

$$z\bar{w} - \bar{z}w = z_0\bar{w} - \bar{z}_0 w\,,$$

becomes

$$\frac{\bar{w}}{z'} - \frac{w}{\bar{z}'} = z_0\bar{w} - \bar{z}_0 w\,.$$

By multiplying by $|z'|^2$, etc., this equation can be rewritten successively as

$$\begin{aligned}
\bar{z}'\bar{w} - z'w &= (z_0\bar{w} - \bar{z}_0 w)\, z'\bar{z}' \\
z'\bar{z}' - \frac{\bar{z}'\bar{w}}{z_0\bar{w} - \bar{z}_0 w} - \frac{z'w}{\bar{z}_0 w - z_0\bar{w}} &= 0 \\
\left|z' - \frac{\bar{w}}{z_0\bar{w} - \bar{z}_0 w}\right|^2 &= \left|\frac{\bar{w}}{z_0\bar{w} - \bar{z}_0 w}\right|^2
\end{aligned}$$

From (1.35a) this is a circle (which passes through the origin) with centre $\frac{\bar{w}}{z_0\bar{w} - \bar{z}_0 w}$ and radius $\left|\frac{\bar{w}}{z_0\bar{w} - \bar{z}_0 w}\right|$. The exception is when $z_0\bar{w} - \bar{z}_0 w = 0$, in which case the original line passes through the origin, and the mapped curve, $\bar{z}'\bar{w} - z'w = 0$, is also a straight line through the origin.

Further, under the map $z' = \frac{1}{z}$ the circle $| z - v |= r$ becomes

$$\left|\frac{1}{z'} - v\right| = r\,, \quad \text{i.e.} \quad |1 - vz'| = r|z'|\,.$$

Hence

$$\left(1 - vz'\right)\left(1 - \bar{v}\bar{z}'\right) \;=\; r^2\bar{z}'z'\,,$$

or equivalently

$$z'\bar{z}'\left(|v|^2 - r^2\right) - vz' - \bar{v}\bar{z}' + 1 \;=\; 0\,.$$

or equivalently

$$z'\bar{z}' - \frac{v}{(|v|^2 - r^2)}z' - \frac{\bar{v}}{(|v|^2 - r^2)}\bar{z}' + \frac{|v|^2}{(|v|^2 - r^2)^2} \;=\; \frac{r^2}{(|v|^2 - r^2)^2}\,.$$

From (1.35b) this is the equation for a circle with centre $\bar{v}/(|v|^2 - r^2)$ and radius $r/(|v|^2 - r^2)$. The exception is if $|v|^2 = r^2$, in which case the original circle passed through the origin, and the map reduces to

$$vz' + \bar{v}\bar{z}' = 1\,,$$

which is the equation of a straight line.

Summary. Under inversion and reflection, circles and straight lines which do not pass through the origin map to circles, while circles and straight lines that do pass through origin map to straight lines.

1.11.4 The general Möbius map

The reason for introducing the basic maps above is that the general Möbius map can be generated by composition of translation, dilatation and rotation, and inversion and reflection. To see this consider the sequence:

dilatation and rotation	$z \mapsto z_1 = cz$	$(c \neq 0)$
translation	$z_1 \mapsto z_2 = z_1 + d$	
inversion and reflection	$z_2 \mapsto z_3 = 1/z_2$	
dilatation and rotation	$z_3 \mapsto z_4 = \left(\dfrac{bc - ad}{c}\right) z_3$	$(bc \neq ad)$
translation	$z_4 \mapsto z_5 = z_4 + a/c$	$(c \neq 0)$

Exercises.

(i) Show that

$$z_5 = \frac{az + b}{cz + d}\,.$$

(ii) Construct a similar sequence if $c = 0$ and $d \neq 0$.

The above implies that a general Möbius map transforms circles and straight lines to circles and straight lines (since each constituent transformation does so).

2 Vector Algebra

2.0 Why Study This?

Many scientific quantities just have a magnitude, e.g. time, temperature, density, concentration. Such quantities can be completely specified by a single number. We refer to such numbers as *scalars*. You have learnt how to manipulate such scalars (e.g. by addition, subtraction, multiplication, differentiation) since your first day in school (or possibly before that). A scalar, e.g. temperature T, that is a function of position (x, y, z) is referred to as a *scalar field*; in the case of our example we write $T \equiv T(x, y, z)$.

However other quantities have both a magnitude and a direction, e.g. the position of a particle, the velocity of a particle, the direction of propagation of a wave, a force, an electric field, a magnetic field. You need to know how to manipulate these quantities (e.g. by addition, subtraction, multiplication and, next term, differentiation) if you are to be able to describe them mathematically.

2.1 Vectors

Definition. A quantity that is specified by a [positive] magnitude and a direction in space is called a *vector*.

Remarks.

- For the purpose of this course the notes will represent vectors in bold, e.g. $\mathbf{v}$. On the overhead/blackboard I will put a squiggle under the $\underset{\sim}{v}$.[8]
- The magnitude of a vector $\mathbf{v}$ is written $|\mathbf{v}|$.
- Two vectors $\mathbf{u}$ and $\mathbf{v}$ are equal if they have the same magnitude, i.e. $|\mathbf{u}| = |\mathbf{v}|$, and they are in the same direction, i.e. $\mathbf{u}$ is parallel to $\mathbf{v}$ and in both vectors are in the same direction/sense.
- A vector, e.g. force $\mathbf{F}$, that is a function of position (x, y, z) is referred to as a *vector field*; in the case of our example we write $\mathbf{F} \equiv \mathbf{F}(x, y, z)$.

2.1.1 Geometric representation

We represent a vector $\mathbf{v}$ as a line segment, say $\vec{AB}$, with length $|\mathbf{v}|$ and direction that of $\mathbf{v}$ (the direction/sense of $\mathbf{v}$ is from A to B).

Examples.

(i) Every point P in 3D (or 2D) space has a *position vector*, $\mathbf{r}$, from some chosen origin O, with $\mathbf{r} = \vec{OP}$ and $r = OP = |\mathbf{r}|$.

Remarks.

- Often the position vector is represented by $\mathbf{x}$ rather than $\mathbf{r}$, but even then the length (i.e. 4/03 magnitude) is usually represented by r.
- The position vector is an example of a vector field.

(ii) Every complex number corresponds to a unique point in the complex plane, and hence to the position vector of that point.

[8] Sophisticated mathematicians use neither **bold** nor *squiggles*.

2.2 Properties of Vectors

2.2.1 Addition

Vectors add according to the *parallelogram rule*:

$$\mathbf{a}+\mathbf{b} \quad = \quad \mathbf{c}\,, \tag{2.1a}$$

or equivalently

$$\overrightarrow{OA}+\overrightarrow{OB} \quad = \quad \overrightarrow{OC}\,, \tag{2.1b}$$

where $OACB$ is a parallelogram.

Remarks.

(i) Since $\overrightarrow{OB}=\overrightarrow{AC}$, it is also true that

$$\overrightarrow{OA}+\overrightarrow{AC}=\overrightarrow{OC}\ . \tag{2.2}$$

(ii) Addition is *commutative*, i.e.

$$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}\,. \tag{2.3}$$

(iii) Addition is *associative*, i.e.

$$\mathbf{a}+(\mathbf{b}+\mathbf{c}) \quad = \quad (\mathbf{a}+\mathbf{b})+\mathbf{c}\,. \tag{2.4}$$

(iv) There is a triangle inequality analogous to (1.13a), i.e.

$$|\mathbf{a}+\mathbf{b}| \leqslant |\mathbf{a}|+|\mathbf{b}|\,. \tag{2.5}$$

This is proved in an analogous manner to (1.13a) using the cosine rule.

(v) If $|\mathbf{a}|=0$, write $\mathbf{a}=\mathbf{0}$, where $\mathbf{0}$ is the null vector or zero vector.[9] For all vectors $\mathbf{b}$

$$\mathbf{b}+\mathbf{0}=\mathbf{b}\,, \quad \text{and from (2.3)} \quad \mathbf{0}+\mathbf{b}=\mathbf{b}\,. \tag{2.6}$$

(vi) Define the vector $-\mathbf{a}$ to be parallel to $\mathbf{a}$, to have the same magnitude as $\mathbf{a}$, but to have the opposite direction/sense (so that it is *anti-parallel*). This is called the *negative* of $\mathbf{a}$ and is such

$$(-\mathbf{a})+\mathbf{a}=\mathbf{0}\,. \tag{2.7a}$$

Define subtraction of vectors by

$$\mathbf{b}-\mathbf{a}\equiv\mathbf{b}+(-\mathbf{a})\,. \tag{2.7b}$$

4/02
2.2.2 Multiplication by a scalar

If $\lambda\in\mathbb{R}$ then $\lambda\mathbf{a}$ has magnitude $|\lambda||\mathbf{a}|$, is parallel to $\mathbf{a}$, and it has the same direction/sense as $\mathbf{a}$ if $\lambda>0$, but the opposite direction/sense as $\mathbf{a}$ if $\lambda<0$ (see below for $\lambda=0$).

A number of properties follow from the above definition. In what follows $\lambda,\mu\in\mathbb{R}$.

Distributive law:

$$(\lambda+\mu)\mathbf{a} \quad = \quad \lambda\mathbf{a}+\mu\mathbf{a}\,, \tag{2.8a}$$
$$\lambda(\mathbf{a}+\mathbf{b}) \quad = \quad \lambda\mathbf{a}+\lambda\mathbf{b}\,. \tag{2.8b}$$

[9] I have attempted to always write $\mathbf{0}$ in bold in the notes; if I have got it wrong somewhere then please let me know. However, on the overhead/blackboard you will need to be more 'sophisticated'. I will try and get it right, but I will not always. Depending on context 0 will sometimes mean 0 and sometimes $\mathbf{0}$.

Associative law:

$$\lambda(\mu\mathbf{a}) \quad = \quad (\lambda\mu)\mathbf{a}\,. \tag{2.9}$$

Multiplication by 0, 1 and -1:

$$\begin{aligned}
0\,\mathbf{a} &= \mathbf{0} \quad \text{since } 0\,|\mathbf{a}| = 0, & (2.10\text{a})\\
1\,\mathbf{a} &= \mathbf{a}\,, & (2.10\text{b})\\
(-1)\,\mathbf{a} &= -\mathbf{a} \quad \text{since } |-1|\,|\mathbf{a}| = |\mathbf{a}| \text{ and } -1 < 0. & (2.10\text{c})
\end{aligned}$$

We can also recover the final result without appealing to geometry by using (2.4), (2.6), (2.7a), (2.7b), (2.8a), (2.10a) and (2.10b) since

$$\begin{aligned}
(-1)\mathbf{a} &= (-1)\mathbf{a} + \mathbf{0}\\
&= (-1)\mathbf{a} + (\mathbf{a} - \mathbf{a})\\
&= ((-1)\mathbf{a} + 1\,\mathbf{a}) - \mathbf{a}\\
&= (-1+1)\mathbf{a} - \mathbf{a}\\
&= \mathbf{0} - \mathbf{a}\\
&= -\mathbf{a}\,.
\end{aligned}$$

Definition. The vector $\mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b}$ is described as a *linear combination* of $\mathbf{a}$ and $\mathbf{b}$.

Unit vectors. Suppose $\mathbf{a} \neq \mathbf{0}$, then define

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|}\,. \tag{2.11}$$

$\hat{\mathbf{a}}$ is termed a *unit vector* since

$$|\hat{\mathbf{a}}| = \left|\frac{1}{|\mathbf{a}|}\right| |\mathbf{a}| = \frac{1}{|\mathbf{a}|}|\mathbf{a}| = 1\,.$$

A ˆ is often used to indicate a unit vector, but note that this is a convention that is often broken (e.g. see §2.7.1).

2.2.3 Example: the midpoints of the sides of any quadrilateral form a parallelogram

This an example of the fact that the rules of vector manipulation and their geometric interpretation can be used to prove geometric theorems.

Denote the vertices of the quadrilateral by A, B, C and D. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ represent the sides $\overrightarrow{DA}$, $\overrightarrow{AB}$, $\overrightarrow{BC}$ and $\overrightarrow{CD}$, and let P, Q, R and S denote the respective midpoints. Then since the quadrilateral is closed

$$\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d} = 0\,. \tag{2.12}$$

Further

$$\overrightarrow{PQ} = \overrightarrow{PA} + \overrightarrow{AQ} = \tfrac{1}{2}\mathbf{a} + \tfrac{1}{2}\mathbf{b}\,.$$

Similarly, from using (2.12),

$$\begin{aligned}
\overrightarrow{RS} &= \tfrac{1}{2}(\mathbf{c} + \mathbf{d})\\
&= -\tfrac{1}{2}(\mathbf{a} + \mathbf{b})\\
&= -\overrightarrow{PQ}\,,
\end{aligned}$$

and thus $\overrightarrow{SR} = \overrightarrow{PQ}$. Since PQ and SR have equal magnitude and are parallel, $PQSR$ is a parallelogram. 04/06

2.3 Scalar Product

Definition. The *scalar product* of two vectors **a** and **b** is defined to be the real (scalar) number

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta\,, \tag{2.13}$$

where $0 \leqslant \theta \leqslant \pi$ is the angle between **a** and **b** measured in the positive (anti-clockwise) sense once they have been placed 'tail to tail' or 'head to head'.

Remark. The *scalar product* is also referred to as the *dot product.*

2.3.1 Properties of the scalar product

(i) The scalar product of a vector with itself is the square of its modulus:

$$\mathbf{a}\cdot\mathbf{a} = |\mathbf{a}|^2 = \mathrm{a}^2\,. \tag{2.14}$$

(ii) The scalar product is *commutative*:

$$\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}\,. \tag{2.15}$$

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

(iv) If $\mathbf{a} \neq 0$ and $\mathbf{b} \neq 0$ and $\mathbf{a}\cdot\mathbf{b} = 0$, then **a** and **b** must be orthogonal (i.e. $\theta = \frac{1}{2}\pi$).

(v) Suppose $\lambda \in \mathbb{R}$. If $\lambda > 0$ then

$$\begin{aligned} \mathbf{a}\cdot(\lambda\mathbf{b}) &= |\mathbf{a}||\lambda\mathbf{b}|\cos\theta \\ &= |\lambda||\mathbf{a}||\mathbf{b}|\cos\theta \\ &= |\lambda|\,\mathbf{a}\cdot\mathbf{b} \\ &= \lambda\,\mathbf{a}\cdot\mathbf{b}\,. \end{aligned}$$

If instead $\lambda < 0$ then

$$\begin{aligned} \mathbf{a}\cdot(\lambda\mathbf{b}) &= |\mathbf{a}||\lambda\mathbf{b}|\cos(\pi-\theta) \\ &= -|\lambda||\mathbf{a}||\mathbf{b}|\cos\theta \\ &= -|\lambda|\,\mathbf{a}\cdot\mathbf{b} \\ &= \lambda\,\mathbf{a}\cdot\mathbf{b}\,. \end{aligned}$$

Similarly, or by using (2.15), $(\lambda\mathbf{a})\cdot\mathbf{b} = \lambda\,\mathbf{a}\cdot\mathbf{b}$. In summary

$$\mathbf{a}\cdot(\lambda\mathbf{b}) = (\lambda\mathbf{a})\cdot\mathbf{b} = \lambda\,\mathbf{a}\cdot\mathbf{b}\,. \tag{2.16}$$

2.3.2 Projections

Before deducing one more property of the scalar product, we need to discuss projections. The *projection* of **a** onto **b** is that part of **a** that is parallel to **b** (which here we will denote by $\mathbf{a}'$).

From geometry, $|\mathbf{a}'| = |\mathbf{a}|\,\cos\theta$ (assume for the time being that $\cos\theta \geqslant 0$). Thus since $\mathbf{a}'$ is parallel to **b**,

$$\mathbf{a}' = |\mathbf{a}'|\frac{\mathbf{b}}{|\mathbf{b}|} = |\mathbf{a}|\,\cos\theta\,\frac{\mathbf{b}}{|\mathbf{b}|}\,. \tag{2.17a}$$

Exercise. Show that (2.17a) remains true if $\cos\theta < 0$.

Hence from (2.13)

$$\mathbf{a}' = |\mathbf{a}|\,\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}\,\frac{\mathbf{b}}{|\mathbf{b}|} = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\,\mathbf{b} = (\mathbf{a}\cdot\hat{\mathbf{b}})\,\hat{\mathbf{b}}\,, \tag{2.17b}$$

where $\hat{\mathbf{b}}$ is the unit vector in the direction of $\mathbf{b}$.

2.3.3 Another property of the scalar product

We wish to show that

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}\,. \tag{2.18}$$

The result is [clearly] true if $\mathbf{a} = 0$, so henceforth assume $\mathbf{a} \neq 0$. Then from (2.17b) (after exchanging $\mathbf{a}$ for $\mathbf{b}$, and $\mathbf{b}$ or $\mathbf{c}$ or $(\mathbf{b}+\mathbf{c})$ for $\mathbf{a}$, etc.)

$$\begin{aligned}
\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|^2}\,\mathbf{a} + \frac{\mathbf{a}\cdot\mathbf{c}}{|\mathbf{a}|^2}\,\mathbf{a} &= \{\text{projection of } \mathbf{b} \text{ onto } \mathbf{a}\} + \{\text{projection of } \mathbf{c} \text{ onto } \mathbf{a}\}\\
&= \{\text{projection of } (\mathbf{b}+\mathbf{c}) \text{ onto } \mathbf{a}\} \qquad\qquad \text{(by geometry)}\\
&= \frac{\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})}{|\mathbf{a}|^2}\,\mathbf{a}\,.
\end{aligned}$$

Now use (2.8a) on the LHS before 'dotting' both sides with $\mathbf{a}$ to obtain the result (i.e. if $\lambda\mathbf{a} + \mu\mathbf{a} = \gamma\mathbf{a}$, then $(\lambda+\mu)\mathbf{a} = \gamma\mathbf{a}$, hence $(\lambda+\mu)|\mathbf{a}|^2 = \gamma|\mathbf{a}|^2$, and so $\lambda+\mu = \gamma$).

2.3.4 Example: the cosine rule

$$\begin{aligned}
BC^2 \;\equiv\; |\,\overrightarrow{BC}\,|^2 \;&=\; |\,\overrightarrow{BA} + \overrightarrow{AC}\,|^2\\
&= \left(\overrightarrow{BA} + \overrightarrow{AC}\right)\cdot\left(\overrightarrow{BA} + \overrightarrow{AC}\right)\\
&= \overrightarrow{BA}\cdot\overrightarrow{BA} + \overrightarrow{BA}\cdot\overrightarrow{AC} + \overrightarrow{AC}\cdot\overrightarrow{BA} + \overrightarrow{AC}\cdot\overrightarrow{AC}\\
&= BA^2 + 2\,\overrightarrow{BA}\cdot\overrightarrow{AC} + AC^2\\
&= BA^2 + 2\,BA\;AC\;\cos\theta + AC^2\\
&= BA^2 - 2\,BA\;AC\;\cos\alpha + AC^2\,.
\end{aligned}$$

5/03

2.4 Vector Product

Definition. The *vector product* $\mathbf{a}\times\mathbf{b}$ of an ordered pair $\mathbf{a}$, $\mathbf{b}$ is a vector such that

(i)
$$|\mathbf{a}\times\mathbf{b}| = |\mathbf{a}||\mathbf{b}|\sin\theta\,, \tag{2.19}$$
with $0 \leqslant \theta \leqslant \pi$ defined as before;

(ii) $\mathbf{a}\times\mathbf{b}$ is perpendicular/orthogonal to both $\mathbf{a}$ and $\mathbf{b}$ (if $\mathbf{a}\times\mathbf{b} \neq \mathbf{0}$);

(iii) $\mathbf{a}\times\mathbf{b}$ has the sense/direction defined by the 'right-hand rule', i.e. take a right hand, point the index finger in the direction of $\mathbf{a}$, the second finger in the direction of $\mathbf{b}$, and then $\mathbf{a}\times\mathbf{b}$ is in the direction of the thumb.

Remarks.

(i) The vector product is also referred to as the *cross product*.

5/02 (ii) An alternative notation (that is falling out of favour except on my overhead/blackboard) is $\mathbf{a} \wedge \mathbf{b}$.

2.4.1 Properties of the vector product

(i) The vector product is *not* commutative (from the right-hand rule):

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}\,. \tag{2.20a}$$

(ii) The vector product of a vector with itself is zero:

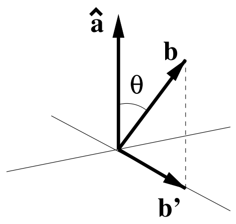
$$\mathbf{a} \times \mathbf{a} = \mathbf{0}\,. \tag{2.20b}$$

(iii) If $\mathbf{a} \neq 0$ and $\mathbf{b} \neq 0$ and $\mathbf{a} \times \mathbf{b} = \mathbf{0}$, then $\theta = 0$ or $\theta = \pi$, i.e. $\mathbf{a}$ and $\mathbf{b}$ are parallel (or equivalently there exists $\lambda \in \mathbb{R}$ such that $\mathbf{a} = \lambda \mathbf{b}$).

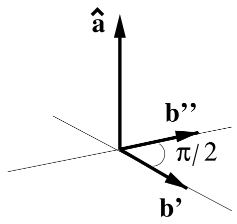
(iv) It follows from the definition of the vector product

$$\mathbf{a} \times (\lambda \mathbf{b}) = \lambda\, (\mathbf{a} \times \mathbf{b})\ . \tag{2.20c}$$

(v) Consider $\mathbf{b}'' = \hat{\mathbf{a}} \times \mathbf{b}$. This vector can be constructed by two operations. First project $\mathbf{b}$ onto a plane orthogonal to $\hat{\mathbf{a}}$ to generate the vector $\mathbf{b}'$, then rotate $\mathbf{b}'$ about $\hat{\mathbf{a}}$ by $\frac{\pi}{2}$ in an 'anti-clockwise' direction ('anti-clockwise' when looking in the opposite direction to $\hat{\mathbf{a}}$).



b' is the projection of **b** onto the plane perpendicular to **â**



b'' is the result of rotating the vector **b'** through an angle π/2 anti−clockwise about **â** (looking in the opposite direction to **â**)

By geometry $|\mathbf{b}'| = |\mathbf{b}|\,\sin\theta = |\hat{\mathbf{a}} \times \mathbf{b}|$, and $\mathbf{b}''$ has the same magnitude as $\mathbf{b}'$. Further, by construction $\mathbf{b}''$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$, and thus has the correct magnitude and sense/direction to be equated to $\hat{\mathbf{a}} \times \mathbf{b}$.

(vi) We can use this geometric interpretation to show that

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}\,, \tag{2.20d}$$

by first noting by geometry that

$$\begin{aligned} \{\text{projection of } \mathbf{b} \text{ onto plane} \perp \text{to } \mathbf{a}\} \quad &+ \quad \{\text{projection of } \mathbf{c} \text{ onto plane} \perp \text{to } \mathbf{a}\} \\ &= \quad \{\text{projection of } (\mathbf{b}+\mathbf{c}) \text{ onto plane} \perp \text{to } \mathbf{a}\}\,, \end{aligned}$$

i.e.

$$\mathbf{b}' + \mathbf{c}' \quad = \quad (\mathbf{b}+\mathbf{c})'\ ,$$

and by then rotating $\mathbf{b}'$, $\mathbf{c}'$ and $(\mathbf{b}+\mathbf{c})'$ by $\frac{\pi}{2}$ 'anti-clockwise' about $\mathbf{a}$ to show that

$$\mathbf{b}'' + \mathbf{c}'' = (\mathbf{b}+\mathbf{c})'' \ .$$

2.4.2 Vector area of a triangle/parallelogram

Let O, A and B denote the vertices of a triangle, and let NB be the *altitude* through B. Denote $\overrightarrow{OA}$ and $\overrightarrow{OB}$ by $\mathbf{a}$ and $\mathbf{b}$ respectively. Then

$$\text{Area of triangle} = \tfrac{1}{2}\, OA \,.\, NB = \tfrac{1}{2}\, OA \,.\, OB \,\sin\theta = \tfrac{1}{2}|\mathbf{a}\times\mathbf{b}| \,.$$

$\frac{1}{2}\mathbf{a}\times\mathbf{b}$ is referred to as the *vector area* of the triangle. It has the same magnitude as the area of the triangle, and is normal to OAB, i.e. normal to the plane containing $\mathbf{a}$ and $\mathbf{b}$.

Let O, A, C and B denote the vertices of a parallelogram, with $\overrightarrow{OA}$ and $\overrightarrow{OB}$ as before. Then

$$\text{Area of parallelogram} = |\mathbf{a}\times\mathbf{b}| \,,$$

and the *vector area* is $\mathbf{a}\times\mathbf{b}$.

2.5 Triple Products

Given the scalar ('dot') product and the vector ('cross') product, we can form two triple products.

Scalar triple product:

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = -\,(\mathbf{b}\times\mathbf{a})\cdot\mathbf{c}\,, \tag{2.21}$$

from using (2.15) and (2.20a).

Vector triple product:

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = -\mathbf{c}\times(\mathbf{a}\times\mathbf{b}) = -\,(\mathbf{b}\times\mathbf{a})\times\mathbf{c} = \mathbf{c}\times(\mathbf{b}\times\mathbf{a}) \ , \tag{2.22}$$

from using (2.20a).

5/06

2.5.1 Properties of the scalar triple product

Volume of parallelepipeds. The volume of a parallelepiped (or parallelipiped or parallelopiped or parallelopipede or parallelopipedon) with edges $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is

$$\begin{aligned} \text{Volume} &= \text{Base Area} \ \times \ \text{Height} \\ &= |\mathbf{a}\times\mathbf{b}|\,|\mathbf{c}|\,\cos\phi \\ &= |\,(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}| && (2.23\text{a}) \\ &= (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} \geqslant 0\,, && (2.23\text{b}) \end{aligned}$$

if $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ have the sense of the right-hand rule.

Identities. Assume that the ordered triple $(\mathbf{a},\mathbf{b},\mathbf{c})$ has the sense of the right-hand rule. Then so do the ordered triples $(\mathbf{b},\mathbf{c},\mathbf{a})$, and $(\mathbf{c},\mathbf{a},\mathbf{b})$. Since the ordered scalar triple products will all equal the volume of the same parallelepiped it follows that

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = (\mathbf{b}\times\mathbf{c})\cdot\mathbf{a} = (\mathbf{c}\times\mathbf{a})\cdot\mathbf{b}\,. \tag{2.24a}$$

Further the ordered triples $(\mathbf{a},\mathbf{c},\mathbf{b})$, $(\mathbf{b},\mathbf{a},\mathbf{c})$ and $(\mathbf{c},\mathbf{b},\mathbf{a})$ all have the sense of the left-hand rule, and so their scalar triple products must all equal the 'negative volume' of the parallelepiped; thus

$$(\mathbf{a}\times\mathbf{c})\cdot\mathbf{b} = (\mathbf{b}\times\mathbf{a})\cdot\mathbf{c} = (\mathbf{c}\times\mathbf{b})\cdot\mathbf{a} = -\,(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\,. \tag{2.24b}$$

It also follows from (2.15) and (2.24a) that

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\ , \tag{2.24c}$$

and hence the order of the 'cross' and 'dot' is inconsequential.[10] For this reason we sometimes use the notation

$$[\mathbf{a},\mathbf{b},\mathbf{c}] = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\,. \tag{2.24d}$$

Coplanar vectors. If $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are coplanar then

$$[\mathbf{a},\mathbf{b},\mathbf{c}] = 0\,,$$

since the volume of the parallelepiped is zero. Conversely if non-zero $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are such that $[\mathbf{a},\mathbf{b},\mathbf{c}] = 0$, then $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are coplanar.

6/03

2.6 Bases and Components

2.6.1 Two dimensional space

First consider 2D space, an origin O, and two non-zero and non-parallel vectors $\mathbf{a}$ and $\mathbf{b}$. Then the position vector $\mathbf{r}$ of any point P in the plane can be expressed as

$$\mathbf{r} = \overrightarrow{OP} = \lambda\mathbf{a} + \mu\mathbf{b}\,, \tag{2.25}$$

for suitable and unique real scalars λ and μ.

Geometric construction. Draw a line through P parallel to $\overrightarrow{OA} = \mathbf{a}$ to intersect $\overrightarrow{OB} = \mathbf{b}$ (or its extension) at N (all non-parallel lines intersect). Then there exist $\lambda,\mu\in\mathbb{R}$ such that

$$\overrightarrow{ON} = \mu\mathbf{b} \quad\text{and}\quad \overrightarrow{NP} = \lambda\mathbf{a}\,,$$

and hence

$$\mathbf{r} = \overrightarrow{OP} = \lambda\mathbf{a} + \mu\mathbf{b}\,.$$

Definition. We say that the set $\{\mathbf{a},\mathbf{b}\}$ *spans* the set of vectors lying in the plane.

Uniqueness. Suppose that λ and μ are not-unique, and that there exists $\lambda,\lambda',\mu,\mu'\in\mathbb{R}$ such that

$$\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} = \lambda'\mathbf{a} + \mu'\mathbf{b}\,.$$

Hence

$$(\lambda-\lambda')\mathbf{a} = (\mu'-\mu)\mathbf{b}\,,$$

and so $\lambda-\lambda' = \mu-\mu' = 0$, since $\mathbf{a}$ and $\mathbf{b}$ are not parallel (or 'cross' first with $\mathbf{a}$ and then $\mathbf{b}$).

[10] What is important is the order of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$.

Definition. If for two vectors $\mathbf{a}$ and $\mathbf{b}$ and $\alpha, \beta \in \mathbb{R}$,

$$\alpha\mathbf{a} + \beta\mathbf{b} = \mathbf{0} \quad \Rightarrow \quad \alpha = \beta = 0\,, \tag{2.26}$$

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

Definition. We say that the set $\{\mathbf{a}, \mathbf{b}\}$ is a *basis* for the set of vectors lying the in plane if it is a spanning set and $\mathbf{a}$ and $\mathbf{b}$ are linearly independent.

Remark. $\{\mathbf{a}, \mathbf{b}\}$ do not have to be orthogonal to be a basis.

2.6.2 Three dimensional space

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

$$\mathbf{r} = \overrightarrow{OP} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}\,, \tag{2.27}$$

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

Geometric construction. Let $\Pi_{\mathbf{ab}}$ be the plane containing $\mathbf{a}$ and $\mathbf{b}$. Draw a line through P parallel to $\overrightarrow{OC} = \mathbf{c}$. This line cannot be parallel to $\Pi_{\mathbf{ab}}$ because $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are not coplanar. Hence it will intersect $\Pi_{\mathbf{ab}}$, say at N, and there will exist $\nu \in \mathbb{R}$ such that $\overrightarrow{NP} = \nu\mathbf{c}$. Further, since $\overrightarrow{ON}$ lies in the plane $\Pi_{\mathbf{ab}}$, from §2.6.1 there exists $\lambda, \mu \in \mathbb{R}$ such that $\overrightarrow{ON} = \lambda\mathbf{a} + \mu\mathbf{b}$. It follows that

$$\begin{aligned} \mathbf{r} = \overrightarrow{OP} &= \overrightarrow{ON} + \overrightarrow{NP} \\ &= \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}\,, \end{aligned}$$

for some $\lambda, \mu, \nu \in \mathbb{R}$.

We conclude that if $[\mathbf{a}, \mathbf{b}, \mathbf{c}] \neq 0$ then the set $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ *spans* 3D space. 6/02

Uniqueness. We can show that λ, μ and ν are unique by construction. Suppose that $\mathbf{r}$ is given by (2.27) and consider

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

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

$$\lambda = \frac{[\mathbf{r}, \mathbf{b}, \mathbf{c}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]}\,, \quad \mu = \frac{[\mathbf{r}, \mathbf{c}, \mathbf{a}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]}\,, \quad \nu = \frac{[\mathbf{r}, \mathbf{a}, \mathbf{b}]}{[\mathbf{a}, \mathbf{b}, \mathbf{c}]}\,. \tag{2.28}$$

The uniqueness of λ, μ and ν implies that the set $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is *linearly independent*, and thus since the set also spans 3D space, we conclude that $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a *basis* for 3D space.

Remark. $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ do not have to be mutually orthogonal to be a basis.

2.6.3 Higher dimensional spaces

We can 'boot-strap' to higher dimensional spaces; in n-dimensional space we would find that the basis had n vectors. However, this is not necessarily the best way of looking at higher dimensional spaces.

Definition. In fact we define the *dimension* of a space as the numbers of [different] vectors in the basis.

2.7 Components

Definition. If $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is a *basis* for 3D space, and if for a vector $\mathbf{r}$,

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

then we call (λ, μ, ν) the *components* of $\mathbf{r}$ with respect to $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$.

2.7.1 The Cartesian or standard basis in 3D

We have noted that $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ do not have to be mutually orthogonal (or right-handed) to be a basis. However, matters are simplified if the basis vectors are mutually orthogonal and have unit magnitude, in which case they are said to define a *orthonormal basis*. It is also conventional to order them so that they are right-handed.

Let OX, OY, OZ be a right-handed set of Cartesian axes. Let

$\mathbf{i}$ be the unit vector along OX,
$\mathbf{j}$ be the unit vector along OY,
$\mathbf{k}$ be the unit vector along OZ,

where it is not conventional to add a $\hat{}$. Then $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ form a basis for 3D space satisfying

$$\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1\,, \tag{2.29a}$$
$$\mathbf{i}\cdot\mathbf{j} = \mathbf{j}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{i} = 0\,, \tag{2.29b}$$
$$\mathbf{i}\times\mathbf{j} = \mathbf{k}\,, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}\,, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}\,, \tag{2.29c}$$
$$[\mathbf{i}, \mathbf{j}, \mathbf{k}] = 1\,. \tag{2.29d}$$

Definition. If for a vector $\mathbf{v}$ and a Cartesian basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$,

$$\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k}\,, \tag{2.30}$$

where $v_x, v_y, v_z \in \mathbb{R}$, we define (v_x, v_y, v_z) to be the *Cartesian components* of $\mathbf{v}$ with respect to $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$.

By 'dotting' (2.30) with $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ respectively, we deduce from (2.29a) and (2.29b) that

$$v_x = \mathbf{v}\cdot\mathbf{i}\,, \quad v_y = \mathbf{v}\cdot\mathbf{j}\,, \quad v_z = \mathbf{v}\cdot\mathbf{k}\,. \tag{2.31}$$

Hence for all 3D vectors $\mathbf{v}$

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{i})\,\mathbf{i} + (\mathbf{v}\cdot\mathbf{j})\,\mathbf{j} + (\mathbf{v}\cdot\mathbf{k})\,\mathbf{k}\,. \tag{2.32}$$

Remarks.

(i) Assuming that we know the basis vectors (and remember that there are an uncountably infinite number of Cartesian axes), we often write

$$\mathbf{v} = (v_x, v_y, v_z)\;. \tag{2.33}$$

In terms of this notation

$$\mathbf{i} = (1, 0, 0)\,, \quad \mathbf{j} = (0, 1, 0)\,, \quad \mathbf{k} = (0, 0, 1)\,. \tag{2.34}$$

(ii) If the point P has the Cartesian co-ordinates (x, y, z), then the position vector

$$\vec{OP} = \mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j} + z\,\mathbf{k}\,, \quad \text{i.e.} \quad \mathbf{r} = (x, y, z)\,. \tag{2.35}$$

(iii) Every vector in 2D/3D space may be uniquely represented by two/three real numbers, so we often write $\mathbb{R}^2/\mathbb{R}^3$ for 2D/3D space.

2.7.2 Direction cosines

If $\mathbf{t}$ is a unit vector with components (t_x, t_y, t_z) with respect to the basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, then

$$t_x = \mathbf{t}\cdot\mathbf{i} = |\mathbf{t}|\,|\mathbf{i}|\cos\alpha\,, \tag{2.36a}$$

6/06

where α is the angle between $\mathbf{t}$ and $\mathbf{i}$. Hence if β and γ are the angles between $\mathbf{t}$ and $\mathbf{j}$, and $\mathbf{t}$ and $\mathbf{k}$, respectively, the *direction cosines* of $\mathbf{t}$ are defined by

$$\mathbf{t} = (\cos\alpha, \cos\beta, \cos\gamma)\ . \tag{2.36b}$$

2.8 Vector Component Identities

Suppose that

$$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}\,, \quad \mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k} \quad \text{and} \quad \mathbf{c} = c_x\mathbf{i} + c_y\mathbf{j} + c_z\mathbf{k}\,. \tag{2.37}$$

Then we can deduce a number of vector identities for components (and one true vector identity).

Addition. From repeated application of (2.8a), (2.8b) and (2.9)

$$\lambda\mathbf{a} + \mu\mathbf{b} = (\lambda a_x + \mu b_x)\mathbf{i} + (\lambda a_y + \mu b_y)\mathbf{j} + (\lambda a_z + \mu b_z)\mathbf{k}\,. \tag{2.38}$$

Scalar product. From repeated application of (2.9), (2.15), (2.16), (2.18), (2.29a) and (2.29b)

$$\begin{aligned}
\mathbf{a}\cdot\mathbf{b} &= (a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k})\cdot(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) \\
&= a_x\mathbf{i}\cdot b_x\mathbf{i} + a_x\mathbf{i}\cdot b_y\mathbf{j} + a_x\mathbf{i}\cdot b_z\mathbf{k} + \ldots \\
&= a_x\,b_x\,\mathbf{i}\cdot\mathbf{i} + a_x\,b_y\,\mathbf{i}\cdot\mathbf{j} + a_x\,b_z\,\mathbf{i}\cdot\mathbf{k} + \ldots \\
&= a_x\,b_x + a_y\,b_y + a_z\,b_z\,.
\end{aligned} \tag{2.39}$$

Vector product. From repeated application of (2.9), (2.20a), (2.20b), (2.20c), (2.20d), (2.29c)

$$\begin{aligned}
\mathbf{a}\times\mathbf{b} &= (a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k})\times(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) \\
&= a_x\mathbf{i}\times b_x\mathbf{i} + a_x\mathbf{i}\times b_y\mathbf{j} + a_x\mathbf{i}\times b_z\mathbf{k} + \ldots \\
&= a_x\,b_x\,\mathbf{i}\times\mathbf{i} + a_x\,b_y\,\mathbf{i}\times\mathbf{j} + a_x\,b_z\,\mathbf{i}\times\mathbf{k} + \ldots \\
&= (a_y\,b_z - a_z\,b_y)\mathbf{i} + (a_z\,b_x - a_x\,b_z)\mathbf{j} + (a_x\,b_y - a_y\,b_x)\mathbf{k}\,.
\end{aligned} \tag{2.40}$$

Scalar triple product. From (2.39) and (2.40)

$$\begin{aligned}
(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} &= ((a_y\,b_z - a_z\,b_y)\mathbf{i} + (a_z\,b_x - a_x\,b_z)\mathbf{j} + (a_x\,b_y - a_y\,b_x)\mathbf{k})\cdot(c_x\mathbf{i} + c_y\mathbf{j} + c_z\mathbf{k}) \\
&= a_x\,b_y\,c_z + a_y\,b_z\,c_x + a_z\,b_x\,c_y - a_x\,b_z\,c_y - a_y\,b_x\,c_z - a_z\,b_y\,c_x\,.
\end{aligned} \tag{2.41}$$

Vector triple product. We wish to show that

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\,. \tag{2.42}$$

Remark. Identity (2.42) has no component in the direction $\mathbf{a}$, i.e. no component in the direction of the vector outside the parentheses.

To prove this, begin with the x-component of the left-hand side of (2.42). Then from (2.40)

$$\begin{aligned}
\big(\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\big)_x \equiv \big(\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\big)\cdot\mathbf{i} &= a_y\,(\mathbf{b}\times\mathbf{c})_z - a_z\,(\mathbf{b}\times\mathbf{c})_y \\
&= a_y\,(b_x\,c_y - b_y\,c_x) - a_z\,(b_z\,c_x - b_x\,c_z) \\
&= (a_y c_y + a_z c_z) b_x + a_x b_x c_x - a_x b_x c_x - (a_y b_y + a_z b_z) c_x \\
&= (\mathbf{a}\cdot\mathbf{c}) b_x - (\mathbf{a}\cdot\mathbf{b}) c_x \\
&= \big((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\big)\cdot\mathbf{i} \equiv \big((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\big)_x\,.
\end{aligned}$$

Now proceed similarly for the y and z components, or note that if its true for one component it must be true for all components because of the arbitrary choice of axes. 7/03

2.9 Polar Co-ordinates

Polar co-ordinates are alternatives to Cartesian co-ordinates systems for describing positions in space. They naturally lead to alternative sets of orthonormal basis vectors.

2.9.1 2D plane polar co-ordinates

Define 2D plane polar co-ordinates (r, θ) in terms of 2D Cartesian co-ordinates (x, y) so that

$$x = r\cos\theta\,, \quad y = r\sin\theta\,, \tag{2.43a}$$

where $0 \leqslant r < \infty$ and $0 \leqslant \theta < 2\pi$. From inverting (2.43a) it follows that

$$r = \left(x^2 + y^2\right)^{\frac{1}{2}}\,, \quad \theta = \arctan\left(\frac{y}{x}\right)\,, \tag{2.43b}$$

where the choice of arctan should be such that $0 < \theta < \pi$ if $y > 0$, $\pi < \theta < 2\pi$ if $y < 0$, $\theta = 0$ if $x > 0$ and $y = 0$, and $\theta = \pi$ if $x < 0$ and $y = 0$.

Remark. The curves of constant r and curves of constant θ intersect at right angles, i.e. are orthogonal.

We use $\mathbf{i}$ and $\mathbf{j}$, orthogonal unit vectors in the directions of increasing x and y respectively, as a basis in 2D Cartesian co-ordinates. Similarly we can use the unit vectors in the directions of increasing r and θ respectively as 'a' basis in 2D plane polar co-ordinates, but a key difference in the case of polar co-ordinates is that the unit vectors are position dependent.

Define $\mathbf{e}_r$ as unit vector*s* orthogonal to lines of constant r in the direction of increasing r. Similarly, define $\mathbf{e}_\theta$ as unit vector*s* orthogonal to lines of constant θ in the direction of increasing θ. Then

$$\begin{aligned}
\mathbf{e}_r &= \cos\theta\,\mathbf{i} + \sin\theta\,\mathbf{j}\,, &(2.44a)\\
\mathbf{e}_\theta &= -\sin\theta\,\mathbf{i} + \cos\theta\,\mathbf{j}\,. &(2.44b)
\end{aligned}$$

Equivalently

$$\begin{aligned}
\mathbf{i} &= \cos\theta\,\mathbf{e}_r - \sin\theta\,\mathbf{e}_\theta\,, &(2.45a)\\
\mathbf{j} &= \sin\theta\,\mathbf{e}_r + \cos\theta\,\mathbf{e}_\theta\,. &(2.45b)
\end{aligned}$$

Exercise. Confirm that

$$\mathbf{e}_r\cdot\mathbf{e}_r = 1\,, \quad \mathbf{e}_\theta\cdot\mathbf{e}_\theta = 1\,, \quad \mathbf{e}_r\cdot\mathbf{e}_\theta = 0\,. \tag{2.46}$$

Given the components of a vector $\mathbf{v}$ with respect to a basis $\{\mathbf{i},\mathbf{j}\}$ we can deduce the components with respect to the basis $\{\mathbf{e}_r,\mathbf{e}_\theta\}$:

$$\begin{aligned}\mathbf{v} &= v_x\mathbf{i}+v_y\mathbf{j}\\ &= v_x(\cos\theta\,\mathbf{e}_r-\sin\theta\,\mathbf{e}_\theta)+v_y(\sin\theta\,\mathbf{e}_r+\cos\theta\,\mathbf{e}_\theta)\\ &= (v_x\cos\theta+v_y\sin\theta)\,\mathbf{e}_r+(-v_x\sin\theta+v_y\cos\theta)\,\mathbf{e}_\theta\,.\end{aligned}\tag{2.47}$$

Example. For the case of the position vector $\mathbf{r}=x\mathbf{i}+y\mathbf{j}$ it follows from using (2.43a) and (2.47) that

$$\mathbf{r}=(x\cos\theta+y\sin\theta)\,\mathbf{e}_r+(-x\sin\theta+y\cos\theta)\,\mathbf{e}_\theta=r\mathbf{e}_r\,.\tag{2.48}$$

Remarks.

(i) It is crucial to note that $\{\mathbf{e}_r,\mathbf{e}_\theta\}$ vary with θ. This means that even constant vectors have components that vary with position, e.g. from (2.45a) the components of $\mathbf{i}$ with respect to the basis $\{\mathbf{e}_r,\mathbf{e}_\theta\}$ are $(\cos\theta,\sin\theta)$.

(ii) The polar co-ordinates of a point P are (r,θ), while the components of $\mathbf{r}=\overrightarrow{OP}$ with respect to the basis $\{\mathbf{e}_r,\mathbf{e}_\theta\}$ are $(r,0)$ (in the case of components the θ dependence is 'hiding' in the basis).

(iii) An alternative notation to (r,θ) is (ρ,ϕ), which as shall become clear has its advantages.

2.9.2 Cylindrical polar co-ordinates

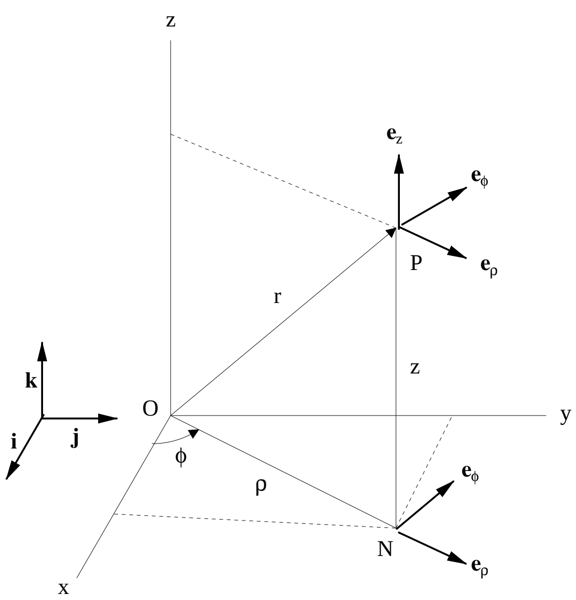
Define 3D cylindrical polar co-ordinates (ρ,ϕ,z)[11] in terms of 3D Cartesian co-ordinates (x,y,z) so that

$$x=\rho\cos\phi\,,\quad y=\rho\sin\phi\,,\quad z=z\,,\tag{2.49a}$$

where $0\leqslant\rho<\infty$, $0\leqslant\phi<2\pi$ and $-\infty<z<\infty$. From inverting (2.49a) it follows that

$$\rho=\left(x^2+y^2\right)^{\frac{1}{2}}\,,\quad \phi=\arctan\left(\frac{y}{x}\right),\tag{2.49b}$$

where the choice of arctan should be such that $0<\phi<\pi$ if $y>0$, $\pi<\phi<2\pi$ if $y<0$, $\phi=0$ if $x>0$ and $y=0$, and $\phi=\pi$ if $x<0$ and $y=0$.



[11] Note that ρ and ϕ are effectively aliases for the r and θ of plane polar co-ordinates. However, in 3D there is a good reason for not using r and θ as we shall see once we get to spherical polar co-ordinates. IMHO there is also a good reason for not using r and θ in plane polar co-ordinates.

Remarks.

(i) Cylindrical polar co-ordinates are helpful if you, say, want to describe fluid flow down a long [straight] cylindrical pipe (e.g. like the new gas pipe between Norway and the UK).

(ii) At any point P on the surface, the vectors orthogonal to the surfaces $\rho =$ constant, $\phi =$ constant and $z =$ constant, i.e. the normals, are mutually orthogonal. 7/02

Define $\mathbf{e}_\rho$, $\mathbf{e}_\phi$ and $\mathbf{e}_z$ as unit vector*s* orthogonal to lines of constant ρ, ϕ and z in the direction of increasing ρ, ϕ and z respectively. Then

$$\begin{aligned} \mathbf{e}_\rho &= \cos\phi\,\mathbf{i} + \sin\phi\,\mathbf{j}\,, && (2.50a)\\ \mathbf{e}_\phi &= -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}\,, && (2.50b)\\ \mathbf{e}_z &= \mathbf{k}\,. && (2.50c) \end{aligned}$$

Equivalently

$$\begin{aligned} \mathbf{i} &= \cos\phi\,\mathbf{e}_\rho - \sin\phi\,\mathbf{e}_\phi\,, && (2.51a)\\ \mathbf{j} &= \sin\phi\,\mathbf{e}_\rho + \cos\phi\,\mathbf{e}_\phi\,, && (2.51b)\\ \mathbf{k} &= \mathbf{e}_z\,. && (2.51c) \end{aligned}$$

Exercise. Show that $\{\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z\}$ are a right-handed triad of mutually orthogonal unit vectors, i.e. show that (cf. (2.29a), (2.29b), (2.29c) and (2.29d))

$$\mathbf{e}_\rho\cdot\mathbf{e}_\rho = 1\,,\quad \mathbf{e}_\phi\cdot\mathbf{e}_\phi = 1\,,\quad \mathbf{e}_z\cdot\mathbf{e}_z = 1\,, \tag{2.52a}$$

$$\mathbf{e}_\rho\cdot\mathbf{e}_\phi = 0\,,\quad \mathbf{e}_\rho\cdot\mathbf{e}_z = 0\,,\quad \mathbf{e}_\phi\cdot\mathbf{e}_z = 0\,, \tag{2.52b}$$

$$\mathbf{e}_\rho\times\mathbf{e}_\phi = \mathbf{e}_z\,,\quad \mathbf{e}_\phi\times\mathbf{e}_z = \mathbf{e}_\rho\,,\quad \mathbf{e}_z\times\mathbf{e}_\rho = \mathbf{e}_\phi\,, \tag{2.52c}$$

$$[\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z] = 1\,. \tag{2.52d}$$

Remark. Since $\{\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z\}$ satisfy analogous relations to those satisfied by $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, i.e. (2.29a), (2.29b), (2.29c) and (2.29d), we can show that analogous vector component identities to (2.38), (2.39), (2.40) and (2.41) also hold.

Component form. Since $\{\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z\}$ form a basis we can write any vector $\mathbf{v}$ in component form as

$$\mathbf{v} = v_\rho\mathbf{e}_\rho + v_\phi\mathbf{e}_\phi + v_z\mathbf{e}_z\,. \tag{2.53}$$

Example. With respect to $\{\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z\}$ the position vector $\mathbf{r}$ can be expressed as

$$\begin{aligned} \mathbf{r} = \overrightarrow{ON} + \overrightarrow{NP} &= \rho\,\mathbf{e}_\rho + z\,\mathbf{e}_z && (2.54a)\\ &= (\rho, 0, z)\,. && (2.54b) \end{aligned}$$

7/06

2.9.3 Spherical polar co-ordinates (cf. height, latitude and longitude)

Define 3D spherical polar co-ordinates (r, θ, ϕ)[12] in terms of 3D Cartesian co-ordinates (x, y, z) so that

$$x = r\sin\theta\cos\phi\,,\quad y = r\sin\theta\sin\phi\,,\quad z = r\cos\theta\,, \tag{2.55a}$$

where $0 \leqslant r < \infty$, $0 \leqslant \theta \leqslant \pi$ and $0 \leqslant \phi < 2\pi$. From inverting (2.55a) it follows that

$$r = \left(x^2+y^2+z^2\right)^{\frac{1}{2}}\,,\quad \theta = \arctan\left(\frac{\left(x^2+y^2\right)^{\frac{1}{2}}}{z}\right)\,,\quad \phi = \arctan\left(\frac{y}{x}\right)\,, \tag{2.55b}$$

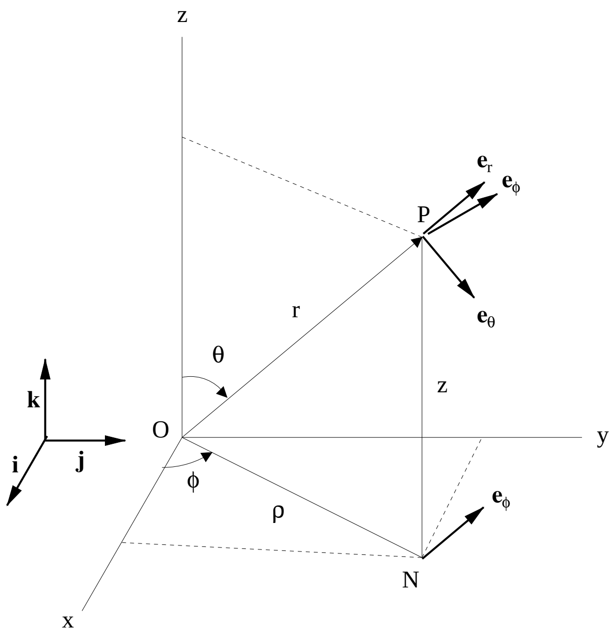
where again some care is needed in the choice of arctan.

[12] Note that r and θ here are *not* the r and θ of plane polar co-ordinates. However, ϕ is the ϕ of cylindrical polar co-ordinates.

Remarks.

(i) Spherical polar co-ordinates are helpful if you, say, want to describe atmospheric (or oceanic, or both) motion on the earth, e.g. in order to understand global warming.

(ii) The normals to the surfaces r = constant, θ = constant and ϕ = constant at any point P are mutually orthogonal.



Define $\mathbf{e}_r$, $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$ as unit vector*s* orthogonal to lines of constant r, θ and ϕ in the direction of increasing r, θ and ϕ respectively. Then

$$\begin{aligned} \mathbf{e}_r &= \sin\theta\cos\phi\,\mathbf{i} + \sin\theta\sin\phi\,\mathbf{j} + \cos\theta\,\mathbf{k}\,, & (2.56a)\\ \mathbf{e}_\theta &= \cos\theta\cos\phi\,\mathbf{i} + \cos\theta\sin\phi\,\mathbf{j} - \sin\theta\,\mathbf{k}\,, & (2.56b)\\ \mathbf{e}_\phi &= -\sin\phi\,\mathbf{i} + \cos\phi\,\mathbf{j}\,. & (2.56c) \end{aligned}$$

Equivalently

$$\begin{aligned} \mathbf{i} &= \cos\phi\,(\sin\theta\,\mathbf{e}_r + \cos\theta\,\mathbf{e}_\theta) - \sin\phi\,\mathbf{e}_\phi\,, & (2.57a)\\ \mathbf{j} &= \sin\phi\,(\sin\theta\,\mathbf{e}_r + \cos\theta\,\mathbf{e}_\theta) + \cos\phi\,\mathbf{e}_\phi\,, & (2.57b)\\ \mathbf{k} &= \cos\theta\,\mathbf{e}_r - \sin\theta\,\mathbf{e}_\theta\,. & (2.57c) \end{aligned}$$

Exercise. Show that $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi\}$ are a right-handed triad of mutually orthogonal unit vectors, i.e. show that (cf. (2.29a), (2.29b), (2.29c) and (2.29d))

$$\mathbf{e}_r\cdot\mathbf{e}_r = 1\,,\quad \mathbf{e}_\theta\cdot\mathbf{e}_\theta = 1\,,\quad \mathbf{e}_\phi\cdot\mathbf{e}_\phi = 1\,, \tag{2.58a}$$

$$\mathbf{e}_r\cdot\mathbf{e}_\theta = 0\,,\quad \mathbf{e}_r\cdot\mathbf{e}_\phi = 0\,,\quad \mathbf{e}_\phi\cdot\mathbf{e}_\theta = 0\,, \tag{2.58b}$$

$$\mathbf{e}_r\times\mathbf{e}_\theta = \mathbf{e}_\phi\,,\quad \mathbf{e}_\theta\times\mathbf{e}_\phi = \mathbf{e}_r\,,\quad \mathbf{e}_\phi\times\mathbf{e}_r = \mathbf{e}_\theta\,, \tag{2.58c}$$

$$[\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi] = 1\,. \tag{2.58d}$$

Remark. Since $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi\}$[13] satisfy analogous relations to those satisfied by $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, i.e. (2.29a), (2.29b), (2.29c) and (2.29d), we can show that analogous vector component identities to (2.38), (2.39), (2.40) and (2.41) also hold.

[13] The ordering of $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_\phi$ is important since $\{\mathbf{e}_r, \mathbf{e}_\phi, \mathbf{e}_\theta\}$ would form a left-handed triad.

Component form. Since $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi\}$ form a basis we can write any vector $\mathbf{v}$ in component form as

$$\mathbf{v} = v_r \mathbf{e}_r + v_\theta \mathbf{e}_\theta + v_\phi \mathbf{e}_\phi \,. \tag{2.59}$$

Example. With respect to $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi\}$ the position vector $\mathbf{r}$ can be expressed as

$$\begin{aligned} \mathbf{r} \;=\; \overrightarrow{OP} &= r\,\mathbf{e}_r && (2.60\text{a})\\ &= (r, 0, 0)\,. && (2.60\text{b}) \end{aligned}$$

2.10 Suffix Notation

So far we have used *dyadic notation* for vectors. *Suffix notation* is an alternative means of expressing vectors (and tensors). Once familiar with suffix notation, it is generally easier to manipulate vectors using suffix notation.[14]

In (2.30) we introduced the notation

$$\mathbf{v} = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} = (v_x, v_y, v_z)\,.$$

An alternative is to write

$$\begin{aligned} \mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k} &= (v_1, v_2, v_3) && (2.61\text{a})\\ &= \{v_i\} \quad \text{for} \quad i = 1, 2, 3\,. && (2.61\text{b}) \end{aligned}$$

Suffix notation. Refer to $\mathbf{v}$ as $\{v_i\}$, with the $i = 1, 2, 3$ understood. i is then termed a *free suffix*.

Example: the position vector. Write the position vector $\mathbf{r}$ as

$$\mathbf{r} = (x, y, z) = (x_1, x_2, x_3) = \{x_i\}\,. \tag{2.62}$$

Remark. The use of $\mathbf{x}$, rather than $\mathbf{r}$, for the position vector in dyadic notation possibly seems more understandable given the above expression for the position vector in suffix notation. Henceforth we will use $\mathbf{x}$ and $\mathbf{r}$ interchangeably.

2.10.1 Dyadic and suffix equivalents

If two vectors $\mathbf{a}$ and $\mathbf{b}$ are equal, we write

$$\mathbf{a} = \mathbf{b}\,, \tag{2.63a}$$

or equivalently in component form

$$\begin{aligned} a_1 &= b_1\,, && (2.63\text{b})\\ a_2 &= b_2\,, && (2.63\text{c})\\ a_3 &= b_3\,. && (2.63\text{d}) \end{aligned}$$

In suffix notation we express this equality as

$$a_i = b_i \quad \text{for} \quad i = 1, 2, 3\,. \tag{2.63e}$$

This is a vector equation where, when we omit the for $i = 1, 2, 3$, it is understood that the one free suffix i ranges through $1, 2, 3$ so as to give three component equations. Similarly

$$\begin{aligned} \mathbf{c} = \lambda\mathbf{a} + \mu\mathbf{b} \quad &\Leftrightarrow \quad c_i = \lambda a_i + \mu b_i\\ &\Leftrightarrow \quad c_j = \lambda a_j + \mu b_j\\ &\Leftrightarrow \quad c_\alpha = \lambda a_\alpha + \mu b_\alpha\\ &\Leftrightarrow \quad c_{¥} = \lambda a_{¥} + \mu b_{¥}\,, \end{aligned}$$

where is is assumed that i, j, α and ¥, respectively, range through $(1, 2, 3)$.[15]

[14] Although there are dissenters to that view.

[15] In higher dimensions the suffices would be assumed to range through the number of dimensions.

Remark. It does not matter what letter, or symbol, is chosen for the free suffix, but it must be the same in each term.

Dummy suffices. In suffix notation the scalar product becomes

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= a_1b_1 + a_2b_2 + a_3b_3\\ &= \sum_{i=1}^{3} a_ib_i\\ &= \sum_{k=1}^{3} a_kb_k\,, \quad \text{etc.,}\end{aligned}$$

where the i, k, etc. are referred to as *dummy suffices* since they are 'summed out' of the equation. Similarly

$$\mathbf{a}\cdot\mathbf{b} = \lambda \quad\Leftrightarrow\quad \sum_{\alpha=1}^{3} a_\alpha b_\alpha = \lambda\,,$$

where we note that the equivalent equation on the right hand side has no free suffices since the dummy suffix (in this case α) has again been summed out.

Further examples.

(i) As another example consider the equation $(\mathbf{a}\cdot\mathbf{b})\mathbf{c} = \mathbf{d}$. In suffix notation this becomes

$$\sum_{k=1}^{3} (a_kb_k)\,c_i = \sum_{k=1}^{3} a_kb_kc_i = d_i\,, \tag{2.64}$$

where k is the dummy suffix, and i is the free suffix that is assumed to range through $(1, 2, 3)$. It is essential that we used different symbols for both the dummy and free suffices!

(ii) In suffix notation the expression $(\mathbf{a}\cdot\mathbf{b})(\mathbf{c}\cdot\mathbf{d})$ becomes

$$\begin{aligned}(\mathbf{a}\cdot\mathbf{b})(\mathbf{c}\cdot\mathbf{d}) &= \left(\sum_{i=1}^{3} a_ib_i\right)\left(\sum_{j=1}^{3} c_jd_j\right)\\ &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_ib_ic_jd_j\,,\end{aligned}$$

where, especially after the rearrangement, it is essential that the dummy suffices are different.

2.10.2 Summation convention

In the case of free suffices we are assuming that they range through $(1, 2, 3)$ without the need to explicitly say so. Under Einstein's summation the explicit sum, $\sum$, can be omitted for dummy suffices. In particular

- if a suffix appears once it is taken to be a free suffix and ranged through,
- if a suffix appears twice it is taken to be a dummy suffix and summed over,
- if a suffix appears more than twice in one term of an equation, something has gone wrong (unless there is an explicit sum).

Remark. This notation is powerful because it is highly abbreviated (and so aids calculation, especially in examinations), but the above rules *must* be followed, and remember to check your answers (e.g. the free suffices should be identical on each side of an equation).

Examples. Under suffix notation and the summation convection

$$\begin{aligned} \mathbf{a}+\mathbf{b}=\mathbf{c} &\quad \text{is written as} \quad a_i+b_i=c_i \\ (\mathbf{a}\cdot\mathbf{b})\mathbf{c}=\mathbf{d} &\quad \text{is written as} \quad a_i\,b_i\,c_j=d_j\,. \end{aligned}$$

The following equations make no sense

$$\begin{aligned} a_k &= b_j \quad \text{because the free suffices are different} \\ a_k\,b_k\,c_k &= d_k \quad \text{because } k \text{ is repeated more than twice on the left-hand side.} \end{aligned}$$

8/03

2.10.3 Kronecker delta

The Kronecker delta, δ_{ij}, $i,j=1,2,3$, is a set of nine numbers defined by

$$\delta_{11}=1\,,\quad \delta_{22}=1\,,\quad \delta_{33}=1\,, \tag{2.65a}$$

$$\delta_{ij}=0 \quad \text{if} \quad i\neq j\,. \tag{2.65b}$$

This can be written as a matrix equation:

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\,. \tag{2.65c}$$

Properties.

1. Using the definition of the delta function:

$$\begin{aligned} a_i\delta_{i1} &= \sum_{i=1}^{3} a_i\delta_{i1} \\ &= a_1\delta_{11}+a_2\delta_{21}+a_3\delta_{31} \\ &= a_1\,. \end{aligned} \tag{2.66a}$$

Similarly

$$a_i\delta_{ij} = a_j\,. \tag{2.66b}$$

2.

$$\delta_{ij}\delta_{jk}=\sum_{j=1}^{3}\delta_{ij}\delta_{jk}=\delta_{ik}\,. \tag{2.66c}$$

3.

$$\delta_{ii}=\sum_{i=1}^{3}\delta_{ii}=\delta_{11}+\delta_{22}+\delta_{33}=3\,. \tag{2.66d}$$

4.

$$a_p\,\delta_{pq}\,b_q=a_p\,b_p=a_q\,b_q=\mathbf{a}\cdot\mathbf{b}\,. \tag{2.66e}$$

2.10.4 More on basis vectors

Now that we have introduced suffix notation, it is more convenient to write $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ for the Cartesian unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. An alternative notation is $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$ and $\mathbf{e}^{(3)}$, where the use of superscripts may help emphasise that the 1, 2 and 3 are labels rather than components.

Then in terms of the superscript notation

$$\mathbf{e}^{(i)} \cdot \mathbf{e}^{(j)} = \delta_{ij}\,, \tag{2.67a}$$
$$\mathbf{a} \cdot \mathbf{e}^{(i)} = a_i\,, \tag{2.67b}$$

and hence

$$\mathbf{e}^{(j)} \cdot \mathbf{e}^{(i)} = \left(\mathbf{e}^{(j)}\right)_i \qquad \text{(the } i\text{th component of } \mathbf{e}^{(j)}) \tag{2.67c}$$
$$= \left(\mathbf{e}^{(i)}\right)_j = \delta_{ij}\,. \tag{2.67d}$$

Or equivalently

$$\left(\mathbf{e}_j\right)_i = \left(\mathbf{e}_i\right)_j = \delta_{ij}\,. \tag{2.67e}$$

8/06

2.10.5 The alternating tensor or Levi-Civita symbol

Definition. We define ε_{ijk} $(i, j, k = 1, 2, 3)$ to be the set of 27 quantities such that

$$\varepsilon_{ijk} = 1 \qquad \text{if } i\,j\,k \text{ is an even permutation of } 1\,2\,3; \tag{2.68a}$$
$$= -1 \qquad \text{if } i\,j\,k \text{ is an odd permutation of } 1\,2\,3; \tag{2.68b}$$
$$= 0 \qquad \text{otherwise.} \tag{2.68c}$$

An ordered sequence is an even/odd permutation if the number of pairwise swaps (or exchanges or transpositions) necessary to recover the original ordering, in this case $1\,2\,3$, is even/odd. Hence the non-zero components of ε_{ijk} are given by

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1 \tag{2.69a}$$
$$\varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} = -1 \tag{2.69b}$$

Further

$$\varepsilon_{ijk} = \varepsilon_{jki} = \varepsilon_{kij} = -\varepsilon_{ikj} = -\varepsilon_{kji} = -\varepsilon_{jik}\,. \tag{2.69c}$$

8/02

Example. For a symmetric tensor s_{ij}, $i, j = 1, 2, 3$, such that $s_{ij} = s_{ji}$ evalulate $\epsilon_{ijk}s_{ij}$. Since

$$\epsilon_{ijk}s_{ij} = \epsilon_{jik}s_{ji} = -\epsilon_{ijk}s_{ij}\,, \tag{2.70}$$

we conclude that $\epsilon_{ijk}s_{ij} = 0$.

2.10.6 The vector product in suffix notation

We claim that

$$\left(\mathbf{a} \times \mathbf{b}\right)_i = \sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{ijk}\, a_j\, b_k = \varepsilon_{ijk}\, a_j\, b_k\,, \tag{2.71}$$

where we note that there is one free suffix and two dummy suffices.

Check.

$$\left(\mathbf{a} \times \mathbf{b}\right)_1 = \varepsilon_{123}\, a_2\, b_3 + \varepsilon_{132}\, a_3\, b_2 = a_2\, b_3 - a_3\, b_2\,,$$

as required from (2.40). Do we need to do more?

Example.

$$\begin{aligned}\left(\mathbf{e}^{(j)} \times \mathbf{e}^{(k)}\right)_i &= \varepsilon_{ilm}\left(\mathbf{e}^{(j)}\right)_l\left(\mathbf{e}^{(k)}\right)_m \\ &= \varepsilon_{ilm}\,\delta_{jl}\,\delta_{km} \\ &= \varepsilon_{ijk}\,. \end{aligned}\tag{2.72}$$

2.10.7 An identity

Theorem 2.1.

$$\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\,\delta_{kq} - \delta_{jq}\,\delta_{kp}\,. \tag{2.73}$$

Remark. There are four free suffices/indices on each side, with i as a dummy suffix on the left-hand side. Hence (2.73) represents 3^4 equations.

Proof. If $j = k = 1$, say; then

$$\begin{aligned} \text{LHS} &= \varepsilon_{i11}\varepsilon_{ipq} = 0\,, \\ \text{RHS} &= \delta_{1p}\,\delta_{1q} - \delta_{1q}\,\delta_{1p} = 0\,. \end{aligned}$$

Similarly whenever $j = k$ (or $p = q$). Next suppose $j = 1$ and $k = 2$, say; then

$$\begin{aligned} \text{LHS} &= \varepsilon_{i12}\,\varepsilon_{ipq} \\ &= \varepsilon_{312}\,\varepsilon_{3pq} \\ &= \begin{cases} 1 & \text{if } p = 1, q = 2 \\ -1 & \text{if } p = 2, q = 1 \\ 0 & \quad\text{otherwise} \end{cases}, \end{aligned}$$

while

$$\begin{aligned} \text{RHS} &= \delta_{1p}\,\delta_{2q} - \delta_{1q}\,\delta_{2p} \\ &= \begin{cases} 1 & \text{if } p = 1, q = 2 \\ -1 & \text{if } p = 2, q = 1 \\ 0 & \quad\text{otherwise} \end{cases}. \end{aligned}$$

Similarly whenever $j \neq k$. □

Example. Take $j = p$ in (2.73) as an example of a repeated suffix; then

$$\begin{aligned} \varepsilon_{ipk}\varepsilon_{ipq} &= \delta_{pp}\,\delta_{kq} - \delta_{pq}\,\delta_{kp} \\ &= 3\delta_{kq} - \delta_{kq} \;=\; 2\delta_{kq}\,. \end{aligned} \tag{2.74}$$

2.10.8 Scalar triple product

In suffix notation the scalar triple product is given by

$$\begin{aligned} \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= a_i(\mathbf{b}\times\mathbf{c})_i \\ &= \varepsilon_{ijk}\,a_i\,b_j\,c_k\,. \end{aligned} \tag{2.75}$$

2.10.9 Vector triple product

Using suffix notation for the vector triple product we recover

$$\begin{aligned} \big(\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\big)_i &= \varepsilon_{ijk}\,a_j\,(\mathbf{b}\times\mathbf{c})_k \\ &= \varepsilon_{ijk}\,a_j\,\varepsilon_{klm}\,b_l\,c_m \\ &= -\varepsilon_{kji}\,\varepsilon_{klm}\,a_j\,b_l\,c_m \\ &= -(\delta_{jl}\delta_{im} - \delta_{jm}\delta_{il})\,a_j\,b_l\,c_m \\ &= a_j\,b_i\,c_j - a_j\,b_j\,c_i \\ &= \big((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\big)_i\,, \end{aligned}$$

in agreement with (2.42).

2.11 Vector Equations

When presented with a vector equation one approach might be to write out the equation in components, e.g. $(\mathbf{x}-\mathbf{a})\cdot\mathbf{n}=0$ would become

$$x_1\,n_1 + x_2\,n_2 + x_3\,n_3 = a_1\,n_1 + a_2\,n_2 + a_3\,n_3\,.$$

For given $\mathbf{a},\mathbf{n}\in\mathbb{R}^3$ this is a single equation for three unknowns $\mathbf{x}=(x_1,x_2,x_3)\in\mathbb{R}^3$, and hence we might expect two arbitrary parameters in the solution (as we shall see is the case in (2.81) below). An alternative, and often better, way forward is to use vector manipulation to make progress.

Worked Exercise. For given $\mathbf{a},\mathbf{b},\mathbf{c}\in\mathbb{R}^3$ find solutions $\mathbf{x}\in\mathbb{R}^3$ to

$$\mathbf{x}-(\mathbf{x}\times\mathbf{a})\times\mathbf{b}=\mathbf{c}\,. \tag{2.76}$$

Solution. First expand the vector triple product using (2.42):

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

then dot this with $\mathbf{b}$:

$$\mathbf{b}\cdot\mathbf{x}=\mathbf{b}\cdot\mathbf{c}\ ;$$

then substitute this result into the previous equation to obtain:

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

now rearrange to deduce that

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

Remark. For the case when $\mathbf{a}$ and $\mathbf{c}$ are not parallel, we could have alternatively sought a solution using $\mathbf{a}$, $\mathbf{c}$ and $\mathbf{a}\times\mathbf{c}$ as a basis.

2.12 Lines and Planes

Certain geometrical objects can be described by vector equations.

2.12.1 Lines

Consider the line through a point A parallel to a vector $\mathbf{t}$, and let P be a point on the line. Then the vector equation for a point on the line is given by

$$\overrightarrow{OP} = \overrightarrow{OA}+\overrightarrow{AP}$$

or equivalently

$$\mathbf{x} = \mathbf{a}+\lambda\mathbf{t}\,, \tag{2.77a}$$

for some $\lambda\in\mathbb{R}$.

We may eliminate λ from the equation by noting that $\mathbf{x}-\mathbf{a}=\lambda\mathbf{t}$, and hence

$$(\mathbf{x}-\mathbf{a})\times\mathbf{t}=\mathbf{0}\,. \tag{2.77b}$$

This is an equivalent equation for the line since the solutions to (2.77b) for $\mathbf{t}\neq\mathbf{0}$ are either $\mathbf{x}=\mathbf{a}$ or
9/03 $(\mathbf{x}-\mathbf{a})$ parallel to $\mathbf{t}$.

Remark. Equation (2.77b) has many solutions; the multiplicity of the solutions is represented by a single arbitrary scalar.

Worked Exercise. For given $\mathbf{u}, \mathbf{t} \in \mathbb{R}^3$ find solutions $\mathbf{x} \in \mathbb{R}^3$ to

$$\mathbf{u} = \mathbf{x} \times \mathbf{t}\,. \tag{2.78}$$

Solution. First 'dot' (2.78) with $\mathbf{t}$ to obtain

$$\mathbf{t} \cdot \mathbf{u} = \mathbf{t} \cdot (\mathbf{x} \times \mathbf{t}) = 0\,.$$

Thus there are no solutions unless $\mathbf{t} \cdot \mathbf{u} = 0$. Next 'cross' (2.78) with $\mathbf{t}$ to obtain

$$\mathbf{t} \times \mathbf{u} = \mathbf{t} \times (\mathbf{x} \times \mathbf{t}) = (\mathbf{t} \cdot \mathbf{t})\mathbf{x} - (\mathbf{t} \cdot \mathbf{x})\mathbf{t}\,.$$

Hence

$$\mathbf{x} = \frac{\mathbf{t} \times \mathbf{u}}{|\mathbf{t}|^2} + \frac{(\mathbf{t} \cdot \mathbf{x})\mathbf{t}}{|\mathbf{t}|^2}\,.$$

Finally observe that if $\mathbf{x}$ is a solution to (2.78) so is $\mathbf{x} + \mu\mathbf{t}$ for any $\mu \in \mathbb{R}$, i.e. solutions of (2.78) can only be found up to an arbitrary multiple of $\mathbf{t}$. Hence the general solution to (2.78), assuming that $\mathbf{t} \cdot \mathbf{u} = 0$, is

$$\mathbf{x} = \frac{\mathbf{t} \times \mathbf{u}}{|\mathbf{t}|^2} + \mu\mathbf{t}\,, \tag{2.79}$$

9/02 i.e. a straight line in direction $\mathbf{t}$ through $(\mathbf{t} \times \mathbf{u})/|\mathbf{t}|^2$.

2.12.2 Planes

Consider a plane that goes through a point A and that is orthogonal to a unit vector $\mathbf{n}$; $\mathbf{n}$ is the normal to the plane. Let P be any point in the plane. Then

$$\begin{aligned} \overrightarrow{AP} \cdot \mathbf{n} &= 0\,, \\ \left(\overrightarrow{AO} + \overrightarrow{OP}\right) \cdot \mathbf{n} &= 0\,, \\ (\mathbf{x} - \mathbf{a}) \cdot \mathbf{n} &= 0\,. \end{aligned} \tag{2.80a}$$

Let Q be the point in the plane such that $\overrightarrow{OQ}$ is parallel to $\mathbf{n}$. Suppose that $\overrightarrow{OQ} = d\mathbf{n}$, then d is the distance of the plane to O. Further, since Q is in the plane, it follows from (2.80a) that

$$(d\mathbf{n} - \mathbf{a}) \cdot \mathbf{n} = 0\,, \quad \text{and hence} \quad \mathbf{a} \cdot \mathbf{n} = d\mathbf{n}^2 = d\,.$$

The equation of the plane is thus

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n} = d\,. \tag{2.80b}$$

Remarks.

1. If $\mathbf{l}$ and $\mathbf{m}$ are two linearly independent vectors such that $\mathbf{l} \cdot \mathbf{n} = 0$ and $\mathbf{m} \cdot \mathbf{n} = 0$ (so that both vectors lie in the plane), then any point $\mathbf{x}$ in the plane may be written as

$$\mathbf{x} = \mathbf{a} + \lambda\mathbf{l} + \mu\mathbf{m}\,, \tag{2.81}$$

   where $\lambda, \mu \in \mathbb{R}$.

2. (2.81) is a solution to equation (2.80b). The arbitrariness in the *two* independent arbitrary scalars λ and μ means that the equation has [uncountably] many solutions.

Worked exercise. Under what conditions do the two lines $L_1 : (\mathbf{x}-\mathbf{a})\times\mathbf{t} = 0$ and $L_2 : (\mathbf{x}-\mathbf{b})\times\mathbf{u} = 0$ intersect?

Solution. If the lines are to intersect they cannot be parallel, hence $\mathbf{t}$ and $\mathbf{u}$ must be linearly independent. L_1 passes through $\mathbf{a}$; let L_2' be the line passing through $\mathbf{a}$ parallel to $\mathbf{u}$. Let Π be the plane containing L_1 and L_2', with normal $\mathbf{t}\times\mathbf{u}$. Hence from (2.80a) the equation specifying points on the plane Π is

$$\Pi : \qquad (\mathbf{x}-\mathbf{a})\cdot(\mathbf{t}\times\mathbf{u}) = 0\,. \tag{2.82}$$

Because L_2 is parallel to L_2' and thence Π, either L_2 intersects Π nowhere (in which case L_1 does not intersect L_2), or L_2 lies in Π (in which case L_1 intersects L_2). If the latter case, then $\mathbf{b}$ lies in Π and we deduce that a *necessary* condition for the lines to intersect is that

$$(\mathbf{b}-\mathbf{a})\cdot(\mathbf{t}\times\mathbf{u}) = 0\,. \tag{2.83}$$

Further, we can show that (2.83) is also a *sufficient* condition for the lines to intersect. For if (2.83) holds, then $(\mathbf{b}-\mathbf{a})$ must lie in the plane through the origin that is normal to $(\mathbf{t}\times\mathbf{u})$. This plane is spanned by $\mathbf{t}$ and $\mathbf{u}$, and hence there exists $\lambda,\mu\in\mathbb{R}$ such that

$$\mathbf{b}-\mathbf{a} = \lambda\mathbf{t}+\mu\mathbf{u}\,.$$

Let

$$\mathbf{x} = \mathbf{a}+\lambda\mathbf{t} = \mathbf{b}-\mu\mathbf{u}\,;$$

then from the equation of a line, (2.77a), we deduce that $\mathbf{x}$ is a point on both L_1 and L_2 (as required).

2.13 Cones and Conic Sections

A right circular cone is a surface on which every point P is such that OP makes a fixed angle, say α $(0<\alpha<\frac{1}{2}\pi)$, with a given axis that passes through O. The point O is referred to as the vertex of the cone.

Let $\mathbf{n}$ be a unit vector parallel to the axis. Then from the above description

$$\mathbf{x}\cdot\mathbf{n} = |\mathbf{x}|\cos\alpha\,. \tag{2.84a}$$

The vector equation for a cone with its vertex at the origin is thus

$$(\mathbf{x}\cdot\mathbf{n})^2 = \mathbf{x}^2\cos^2\alpha\,, \tag{2.84b}$$

where by squaring the equation we have included the 'reverse' cone.

By means of a translation we can now generalise (2.84b) to the equation for a cone with a vertex at a general point $\mathbf{a}$:

$$\big[(\mathbf{x}-\mathbf{a})\cdot\mathbf{n}\big]^2 = (\mathbf{x}-\mathbf{a})^2\cos^2\alpha\,. \tag{2.84c}$$

Component form. Suppose that in terms of a standard Cartesian basis

$$\mathbf{x} = (x, y, z)\,, \quad \mathbf{a} = (a, b, c)\,, \quad \mathbf{n} = (l, m, n)\,,$$

then (2.84c) becomes

$$\big[(x-a)l + (y-b)m + (z-c)n\big]^2 = \big[(x-a)^2 + (y-b)^2 + (z-c)^2\big]\cos^2\alpha\,. \tag{2.85}$$

10/03

Intersection of a cone and a plane. Let us consider the intersection of this cone with the plane $z = 0$. In that plane

$$\big[(x-a)l + (y-b)m - cn\big]^2 = \big[(x-a)^2 + (y-b)^2 + c^2\big]\cos^2\alpha\,. \tag{2.86}$$

This is a curve defined by a quadratic polynomial in x and y. In order to simplify the algebra suppose that, wlog, we choose Cartesian axes so that the axis of the cone is in the yz plane. In that case $l = 0$ and we can express $\mathbf{n}$ in component form as

$$\mathbf{n} = (l, m, n) = (0, \sin\beta, \cos\beta)\,.$$

Further, translate the axes by the transformation

$$X = x - a\,, \qquad Y = y - b + \frac{c\sin\beta\cos\beta}{\cos^2\alpha - \sin^2\beta}\,,$$

so that (2.86) becomes

$$\left[\left(Y - \frac{c\sin\beta\cos\beta}{\cos^2\alpha - \sin^2\beta}\right)\sin\beta - c\cos\beta\right]^2 = \left[X^2 + \left(Y - \frac{c\sin\beta\cos\beta}{\cos^2\alpha - \sin^2\beta}\right)^2 + c^2\right]\cos^2\alpha\,,$$

which can be simplified to

$$X^2\cos^2\alpha + Y^2\left(\cos^2\alpha - \sin^2\beta\right) = \frac{c^2\sin^2\alpha\cos^2\alpha}{\cos^2\alpha - \sin^2\beta}\,. \tag{2.87}$$

10/02

There are now three cases that need to be considered: $\sin^2\beta < \cos^2\alpha$, $\sin^2\beta > \cos^2\alpha$ and $\sin^2\beta = \cos^2\alpha$. To see why this is, consider graphs of the intersection of the cone with the $X = 0$ plane, and for definiteness suppose that $0 \leqslant \beta \leqslant \frac{\pi}{2}$.

First suppose that

$$\beta + \alpha < \frac{\pi}{2}$$

i.e.

$$\beta < \tfrac{1}{2}\pi - \alpha$$

i.e.

$$\sin\beta < \sin\left(\tfrac{1}{2}\pi - \alpha\right) = \cos\alpha\,.$$

In this case the intersection of the cone with the $z = 0$ plane will yield a closed curve.

Next suppose that

$$\beta + \alpha > \frac{\pi}{2}$$

i.e.

$$\sin\beta > \cos\alpha\,.$$

In this case the intersection of the cone with the $z = 0$ plane will yield two open curves, while if $\sin\beta = \cos\alpha$ there will be one open curve.

Define

$$\frac{c^2 \sin^2 \alpha}{|\cos^2 \alpha - \sin^2 \beta|} = A^2 \quad \text{and} \quad \frac{c^2 \sin^2 \alpha \cos^2 \alpha}{\left(\cos^2 \alpha - \sin^2 \beta\right)^2} = B^2 . \tag{2.88}$$

$\sin\beta < \cos\alpha$. In this case (2.87) becomes

$$\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1 . \tag{2.89a}$$

This is the equation of an *ellipse* with *semi-minor* and *semi-major* axes of lengths A and B respectively (from (2.88) it follows that $A < B$).

$\sin\beta > \cos\alpha$. In this case

$$-\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1 . \tag{2.89b}$$

This is the equation of a *hyperbola*, where B is one half of the distance between the two *vertices*.

$\sin\beta = \cos\alpha$. In this case (2.86) becomes

$$X^2 = -2c \cot\beta \, Y , \tag{2.89c}$$

where

$$X = x - a \quad \text{and} \quad Y = y - b - c\cot 2\beta . \tag{2.89d}$$

This is the equation of a *parabola*.

Remarks.

(i) The above three curves are known collectively as *conic sections*.

(ii) The identification of the general quadratic polynomial as representing one of these three types will be discussed later in the Tripos.

2.14 Maps: Isometries and Inversions

2.14.1 Isometries

Definition. An *isometry* of $\mathbb{R}^n$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that distances are preserved (where $n = 2$ or $n = 3$ for the time being). In particular, suppose $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ are mapped to $\mathbf{x}'_1, \mathbf{x}'_2 \in \mathbb{R}^n$, i.e. $\mathbf{x}_1 \mapsto \mathbf{x}'_1$ and $\mathbf{x}_2 \mapsto \mathbf{x}'_2$, then

$$|\mathbf{x}_1 - \mathbf{x}_2| = |\mathbf{x}'_1 - \mathbf{x}'_2| . \tag{2.90}$$

Examples.

Translation. Suppose that $\mathbf{b} \in \mathbb{R}^n$, and that

$$\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} + \mathbf{b} .$$

Then

$$|\mathbf{x}'_1 - \mathbf{x}'_2| = |(\mathbf{x}_1 + \mathbf{b}) - (\mathbf{x}_2 + \mathbf{b})| = |\mathbf{x}_1 - \mathbf{x}_2| .$$

Translation is an isometry.

Reflection. Consider reflection in a plane Π, where $\Pi = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{n} = 0\}$ and $\mathbf{n}$ is a constant unit vector. For a point P, let N be the foot of the perpendicular from P to the plane. Suppose also that

$$\mathbf{x} = \overrightarrow{OP} \mapsto \mathbf{x}' = \overrightarrow{OP'} \ .$$

Then

$$\overrightarrow{NP'} = \overrightarrow{PN} = -\overrightarrow{NP} \,,$$

and so

$$\overrightarrow{OP'} = \overrightarrow{OP} + \overrightarrow{PN} + \overrightarrow{NP'} = \overrightarrow{OP} - 2\,\overrightarrow{NP} \ .$$

But $|NP| = |\mathbf{x} \cdot \mathbf{n}|$ and

$$\overrightarrow{NP} = \begin{cases} \ \ |NP|\mathbf{n} & \text{if } \overrightarrow{NP} \text{ has the same sense as } \mathbf{n}, \text{ i.e. } \mathbf{x} \cdot \mathbf{n} > 0 \\ -|NP|\mathbf{n} & \text{if } \overrightarrow{NP} \text{ has the opposite sense as } \mathbf{n}, \text{ i.e. } \mathbf{x} \cdot \mathbf{n} < 0 \,. \end{cases}$$

Hence $\overrightarrow{NP} = (\mathbf{x} \cdot \mathbf{n})\mathbf{n}$, and

$$\overrightarrow{OP'} = \mathbf{x}' = \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{n})\mathbf{n} \,. \tag{2.91}$$

Remark. $\mathbf{x}'$ is linear in $\mathbf{x}$, i.e. the mapping is a linear function of the components of $\mathbf{x}$.

Having derived the mapping, we now need to show that distances are preserved by the mapping. Suppose

$$\mathbf{x}_j \mapsto \mathbf{x}'_j = \mathbf{x}_j - 2(\mathbf{x}_j \cdot \mathbf{n})\mathbf{n} \quad (j = 1, 2) \,.$$

Let $\mathbf{x}_{12} = \mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{x}'_{12} = \mathbf{x}'_1 - \mathbf{x}'_2$, then since $\mathbf{n}^2 = 1$,

$$\begin{aligned} |\mathbf{x}'_{12}|^2 &= |\mathbf{x}_1 - 2(\mathbf{x}_1 \cdot \mathbf{n})\mathbf{n} - \mathbf{x}_2 + 2(\mathbf{x}_2 \cdot \mathbf{n})\mathbf{n}|^2 \\ &= |\mathbf{x}_{12} - 2(\mathbf{x}_{12} \cdot \mathbf{n})\mathbf{n}|^2 \\ &= \mathbf{x}_{12}\mathbf{x}_{12} - 4(\mathbf{x}_{12} \cdot \mathbf{n})(\mathbf{x}_{12} \cdot \mathbf{n}) + 4(\mathbf{x}_{12} \cdot \mathbf{n})^2\mathbf{n}^2 \\ &= |\mathbf{x}_{12}|^2 \,, \end{aligned}$$

as required for isometry.

2.15 Inversion in a Sphere

Let $\Sigma = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| = k \in \mathbb{R}(k > 0)\}$. Then Σ represents a sphere with centre O and radius k.

Definition. For each point P, the inverse point P' with respect to Σ lies on OP and satisfies

$$OP' = \frac{k^2}{OP} \,. \tag{2.92}$$

Let $\overrightarrow{OP} = \mathbf{x}$ and $\overrightarrow{OP'} = \mathbf{x}'$, then

$$\mathbf{x}' = |\mathbf{x}'|\frac{\mathbf{x}}{|\mathbf{x}|} = \frac{k^2}{|\mathbf{x}|}\frac{\mathbf{x}}{|\mathbf{x}|} = \frac{k^2}{|\mathbf{x}|^2}\mathbf{x} \,. \tag{2.93}$$

Exercise. Show that P' has inverse point P.

Remark. Inversion in a sphere is not an isometry, e.g. two points close to the origin map to far from the sphere, and far from each other. 10/06

Example: inversion of a sphere. Suppose that the sphere is given by

$$|\mathbf{x}-\mathbf{a}| = r\,, \quad \text{i.e.} \quad \mathbf{x}^2 - 2\mathbf{a}\cdot\mathbf{x} + \mathbf{a}^2 - r^2 = 0\,. \tag{2.94}$$

From (2.93)

$$|\mathbf{x}'|^2 = \frac{k^4}{|\mathbf{x}|^2} \quad \text{and so} \quad \mathbf{x} = \frac{k^2}{|\mathbf{x}'|^2}\mathbf{x}'\,.$$

(2.94) can thus be rewritten

$$\frac{k^4}{|\mathbf{x}'|^2} - 2\mathbf{a}\cdot\mathbf{x}'\frac{k^2}{|\mathbf{x}'|^2} + \mathbf{a}^2 - r^2 \quad = \quad 0\,,$$

or equivalently

$$k^4 - 2\mathbf{a}\cdot\mathbf{x}'k^2 + (\mathbf{a}^2 - r^2)|\mathbf{x}'|^2 \quad = \quad 0\,,$$

or equivalently if $\mathbf{a}^2 \neq r^2$ we can complete the square to obtain

$$\left|\mathbf{x}' - \frac{k^2\mathbf{a}}{\mathbf{a}^2 - r^2}\right|^2 = \frac{r^2k^4}{(\mathbf{a}^2 - r^2)^2}\,,$$

which is the equation of another sphere. Hence inversion of a sphere in a sphere gives another sphere (if $\mathbf{a}^2 \neq r^2$).

Exercise. What happens if $\mathbf{a}^2 = r^2$?

11/02

3 Vector Spaces

3.0 Why Study This?

One of the strengths of mathematics, indeed one of the ways that mathematical progress is made, is by linking what seem to be disparate subjects/results. Often this is done by taking something that we are familiar with (e.g. real numbers), identifying certain key properties on which to focus (e.g. in the case of real numbers, say, closure, associativity, identity and the existence of an inverse under addition or multiplication), and then studying all mathematical objects with those properties (groups in the example just given). The aim of this section is to 'abstract' the previous section. Up to this point you have thought of vectors as a set of 'arrows' in 3D (or 2D) space. However, they can also be sets of polynomials, matrices, functions, etc.. Vector spaces occur throughout mathematics, science, engineering, finance, etc..

3.1 What is a Vector Space?

We will consider vector spaces *over* the real numbers. There are generalisations to vector spaces over the complex numbers, or indeed over any *field* (see *Groups, Rings and Modules* in Part IB for the definition of a field).[16]

Sets. We have referred to sets already, and I hope that you have covered them in *Numbers and Sets.* To recap, a *set* is a collection of objects considered as a whole. The objects of a set are called elements or members. Conventionally a set is listed by placing its elements between braces, e.g. $\{x : x \in \mathbb{R}\}$ is the set of real numbers.

The empty set. The empty set, $\{\,\}$ or $\emptyset$, is the unique set which contains no elements. It is a subset of every set.

3.1.1 Definition

A *vector space over the real numbers* is a set V of elements, or 'vectors', together with two *binary operations*

- *vector addition* denoted for $\mathbf{x}, \mathbf{y} \in V$ by $\mathbf{x} + \mathbf{y}$, where $\mathbf{x} + \mathbf{y} \in V$ so that there is closure under vector addition;
- *scalar multiplication* denoted for $a \in \mathbb{R}$ and $\mathbf{x} \in V$ by $a\mathbf{x}$, where $a\mathbf{x} \in V$ so that there is closure under scalar multiplication;

satisfying the following eight axioms or rules:[17]

A(i) addition is *associative*, i.e. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$

$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}\,; \tag{3.1a}$$

A(ii) addition is *commutative*, i.e. for all $\mathbf{x}, \mathbf{y} \in V$

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}\,; \tag{3.1b}$$

A(iii) there exists an element $\mathbf{0} \in V$, called the *null or zero vector*, such that for all $\mathbf{x} \in V$

$$\mathbf{x} + \mathbf{0} = \mathbf{x}\,, \tag{3.1c}$$

i.e. vector addition has an identity element;

[16] Pedants may feel they have reached nirvana at the start of this section; normal service will be resumed towards the end of the section.
[17] The first four mean that V is an *abelian group* under addition.

A(iv) for all $\mathbf{x} \in V$ there exists an additive *negative or inverse vector* $\mathbf{x}' \in V$ such that

$$\mathbf{x} + \mathbf{x}' = \mathbf{0}\,; \tag{3.1d}$$

B(v) scalar multiplication of vectors is 'associative',[18] i.e. for all $\lambda, \mu \in \mathbb{R}$ and $\mathbf{x} \in V$

$$\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}\,, \tag{3.1e}$$

B(vi) scalar multiplication has an identity element, i.e. for all $\mathbf{x} \in V$

$$1\,\mathbf{x} = \mathbf{x}\,, \tag{3.1f}$$

where 1 is the multiplicative identity in $\mathbb{R}$;

B(vii) scalar multiplication is distributive over vector addition, i.e. for all $\lambda \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$

$$\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}\,; \tag{3.1g}$$

B(viii) scalar multiplication is distributive over scalar addition, i.e. for all $\lambda, \mu \in \mathbb{R}$ and $\mathbf{x} \in V$

11/03

$$(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}\,. \tag{3.1h}$$

3.1.2 Properties.

(i) The zero vector $\mathbf{0}$ is unique because if $\mathbf{0}$ and $\mathbf{0}'$ are both zero vectors in V then from (3.1b) and (3.1c) $\mathbf{0} + \mathbf{x} = \mathbf{x}$ and $\mathbf{x} + \mathbf{0}' = \mathbf{x}$ for all $\mathbf{x} \in V$, and hence

$$\mathbf{0}' = \mathbf{0} + \mathbf{0}' = \mathbf{0}\,.$$

(ii) The additive inverse of a vector $\mathbf{x}$ is unique, for suppose that both $\mathbf{y}$ and $\mathbf{z}$ are additive inverses of $\mathbf{x}$ then

$$\begin{aligned} \mathbf{y} &= \mathbf{y} + \mathbf{0} \\ &= \mathbf{y} + (\mathbf{x} + \mathbf{z}) \\ &= (\mathbf{y} + \mathbf{x}) + \mathbf{z} \\ &= \mathbf{0} + \mathbf{z} \\ &= \mathbf{z}\,. \end{aligned}$$

We denote the unique inverse of $\mathbf{x}$ by $-\mathbf{x}$.

(iii) The existence of a unique negative/inverse vector allows us to *subtract* as well as add vectors, by defining

$$\mathbf{y} - \mathbf{x} \equiv \mathbf{y} + (-\mathbf{x})\,. \tag{3.2}$$

(iv) Scalar multiplication by 0 yields the zero vector, i.e. for all $\mathbf{x} \in V$,

$$0\mathbf{x} = \mathbf{0}\,, \tag{3.3}$$

since

$$\begin{aligned} 0\mathbf{x} &= 0\mathbf{x} + \mathbf{0} \\ &= 0\mathbf{x} + (\mathbf{x} + (-\mathbf{x})) \\ &= (0\mathbf{x} + \mathbf{x}) + (-\mathbf{x}) \\ &= (0\mathbf{x} + 1\mathbf{x}) + (-\mathbf{x}) \\ &= (0 + 1)\mathbf{x} + (-\mathbf{x}) \\ &= \mathbf{x} + (-\mathbf{x}) \\ &= \mathbf{0}\,. \end{aligned}$$

[18] Strictly we are not asserting the associativity of an operation, since there are two different operations in question, namely scalar multiplication of vectors (e.g. $\mu\mathbf{x}$) and multiplication of real numbers (e.g. $\lambda\mu$).

(v) Scalar multiplication by -1 yields the additive inverse of the vector, i.e. for all $\mathbf{x} \in V$,

$$(-1)\mathbf{x} = -\mathbf{x}\,, \tag{3.4}$$

since

$$\begin{aligned}
(-1)\mathbf{x} &= (-1)\mathbf{x} + \mathbf{0} \\
&= (-1)\mathbf{x} + (\mathbf{x} + (-\mathbf{x})) \\
&= ((-1)\mathbf{x} + \mathbf{x}) + (-\mathbf{x}) \\
&= (-1+1)\mathbf{x} + (-\mathbf{x}) \\
&= 0\mathbf{x} + (-\mathbf{x}) \\
&= \mathbf{0} + (-\mathbf{x}) \\
&= -\mathbf{x}\,.
\end{aligned}$$

(vi) Scalar multiplication with the zero vector yields the zero vector, i.e. for all $\lambda \in \mathbb{R}$, $\lambda\mathbf{0} = \mathbf{0}$. To see this we first observe that $\lambda\mathbf{0}$ is a zero vector since

$$\begin{aligned}
\lambda\mathbf{0} + \lambda\mathbf{x} &= \lambda(\mathbf{0} + \mathbf{x}) \\
&= \lambda\mathbf{x}\,,
\end{aligned}$$

and then appeal to the fact that the zero vector is unique to conclude that

$$\lambda\mathbf{0} = \mathbf{0}\,. \tag{3.5}$$

(vii) Suppose that $\lambda\mathbf{x} = \mathbf{0}$, where $\lambda \in \mathbb{R}$ and $\mathbf{x} \in V$. One possibility is that $\lambda = 0$. However suppose that $\lambda \neq 0$, in which case there exists λ^{-1} such that $\lambda^{-1}\lambda = 1$. Then we conclude that

$$\begin{aligned}
\mathbf{x} &= 1\,\mathbf{x} \\
&= (\lambda^{-1}\lambda)\mathbf{x} \\
&= \lambda^{-1}(\lambda\mathbf{x}) \\
&= \lambda^{-1}\mathbf{0} \\
&= \mathbf{0}\,.
\end{aligned}$$

So if $\lambda\mathbf{x} = \mathbf{0}$ then either $\lambda = 0$ or $\mathbf{x} = \mathbf{0}$.

(viii) Negation commutes freely since for all $\lambda \in \mathbb{R}$ and $\mathbf{x} \in V$

$$\begin{aligned}
(-\lambda)\mathbf{x} &= (\lambda(-1))\mathbf{x} \\
&= \lambda((-1)\mathbf{x}) \\
&= \lambda(-\mathbf{x})\,,
\end{aligned} \tag{3.6a}$$

and

$$\begin{aligned}
(-\lambda)\mathbf{x} &= ((-1)\lambda)\mathbf{x} \\
&= (-1)(\lambda\mathbf{x}) \\
&= -(\lambda\mathbf{x})\,.
\end{aligned} \tag{3.6b}$$

3.1.3 Examples

(i) Let $\mathbb{R}^n$ be the set of all $n-$tuples $\{\mathbf{x} = (x_1, x_2, \ldots, x_n) : x_j \in \mathbb{R} \text{ with } j = 1, 2, \ldots, n\}$, where n is any strictly positive integer. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with $\mathbf{x}$ as above and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, define

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)\,, \tag{3.7a}$$
$$\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\,, \tag{3.7b}$$
$$\mathbf{0} = (0, 0, \ldots, 0)\,, \tag{3.7c}$$
$$-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n)\,. \tag{3.7d}$$

It is straightforward to check that A(i), A(ii), A(iii) A(iv), B(v), B(vi), B(vii) and B(viii) are satisfied. Hence $\mathbb{R}^n$ is a vector space over $\mathbb{R}$.

(ii) Consider the set F of real-valued functions $f(x)$ of a real variable $x \in [a, b]$, where $a, b \in \mathbb{R}$ and $a < b$. For $f, g \in F$ define

$$\begin{aligned}(f+g)(x) &= f(x)+g(x)\,, &(3.8a)\\ (\lambda f)(x) &= \lambda f(x)\,, &(3.8b)\\ O(x) &= 0\,, &(3.8c)\\ (-f)(x) &= -f(x)\,, &(3.8d)\end{aligned}$$

where the function O is the zero 'vector'. Again it is straightforward to check that A(i), A(ii), A(iii) A(iv), B(v), B(vi), B(vii) and B(viii) are satisfied, and hence that F is a vector space (where each
11/06 vector element is a function).

3.2 Subspaces

Definition. A subset U of the elements of a vector space V is called a *subspace* of V if U is a vector space under the same operations (i.e. vector addition and scalar multiplication) as are used to define V.

Proper subspaces. Strictly V and $\{\mathbf{0}\}$ (i.e. the set containing the zero vector only) are subspaces of V. A *proper subspace* is a subspace of V that is not V or $\{\mathbf{0}\}$.

Theorem 3.1. *A subset U of a vector space V is a subspace of V if and only if under operations defined on V*

(i) *for each $\mathbf{x}, \mathbf{y} \in U$, $\mathbf{x}+\mathbf{y} \in U$,*

(ii) *for each $\mathbf{x} \in U$ and $\lambda \in \mathbb{R}$, $\lambda\mathbf{x} \in U$,*

i.e. if and only if U is closed under vector addition and scalar multiplication.

Remark. We can combine (i) and (ii) as the single condition

$$\text{for each } \mathbf{x}, \mathbf{y} \in U \text{ and } \lambda, \mu \in \mathbb{R},\ \lambda\mathbf{x}+\mu\mathbf{y} \in U\,.$$

Proof. Only if. If U is a subspace then it is a vector space, and hence (i) and (ii) hold.

If. It is straightforward to show that A(i), A(ii), B(v), B(vi), B(vii) and B(viii) hold since the elements of U are also elements of V. We need demonstrate that A(iii) (i.e. $\mathbf{0}$ is an element of U), and A(iv) (i.e. the every element has an inverse in U) hold.

A(iii). For each $\mathbf{x} \in U$, it follows from (ii) that $0\mathbf{x} \in U$; but since also $\mathbf{x} \in V$ it follows from (3.3) that $0\mathbf{x} = \mathbf{0}$. Hence $\mathbf{0} \in U$.

A(iv). For each $\mathbf{x} \in U$, it follows from (ii) that $(-1)\mathbf{x} \in U$; but since also $\mathbf{x} \in V$ it follows from (3.4) that $(-1)\mathbf{x} = -\mathbf{x}$. Hence $-\mathbf{x} \in U$. □

3.2.1 Examples

(i) For $n \geqslant 2$ let $U = \{\mathbf{x} : \mathbf{x} = (x_1, x_2, \ldots, x_{n-1}, 0) \text{ with } x_j \in \mathbb{R} \text{ and } j = 1, 2, \ldots, n-1\}$. Then U is a subspace of $\mathbb{R}^n$ since, for $\mathbf{x}, \mathbf{y} \in U$ and $\lambda, \mu \in \mathbb{R}$,

$$\lambda(x_1, x_2, \ldots, x_{n-1}, 0) + \mu(y_1, y_2, \ldots, y_{n-1}, 0) = (\lambda x_1 + \mu y_1, \lambda x_2 + \mu y_2, \ldots, \lambda x_{n-1} + \mu y_{n-1}, 0)\,.$$

Thus U is closed under vector addition and scalar multiplication, and is hence a subspace.

(ii) For $n \geqslant 2$ consider the set $W = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \alpha_i x_i = 0\}$ for given scalars $\alpha_j \in \mathbb{R}$ $(j = 1, 2, \ldots, n)$. W is a subspace of V since, for $\mathbf{x}, \mathbf{y} \in W$ and $\lambda, \mu \in \mathbb{R}$,

$$\lambda\mathbf{x} + \mu\mathbf{y} = (\lambda x_1 + \mu y_1, \lambda x_2 + \mu y_2, \ldots, \lambda x_n + \mu y_n)\,,$$

and

$$\sum_{i=1}^n \alpha_i(\lambda x_i + \mu y_i) = \lambda \sum_{i=1}^n \alpha_i x_i + \mu \sum_{i=1}^n \alpha_i y_i = 0\,.$$

Thus W is closed under vector addition and scalar multiplication, and is hence a subspace. Later we shall see that W is a hyper-plane through the origin (see (3.24)).

(iii) For $n \geqslant 2$ consider the set $\widetilde{W} = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \alpha_i x_i = 1\}$ for given scalars $\alpha_j \in \mathbb{R}$ $(j = 1, 2, \ldots, n)$ not all of which are zero (wlog $\alpha_1 \neq 0$, if not reorder the numbering of the axes). $\widetilde{W}$, which is a hyper-plane that does not pass through the origin, is *not* a subspace of $\mathbb{R}^n$. To see this *either* note that $\mathbf{0} \notin \widetilde{W}$, *or* consider $\mathbf{x} \in \widetilde{W}$ such that

$$\mathbf{x} = (\alpha_1^{-1}, 0, \ldots, 0)\,.$$

Then $\mathbf{x} \in \widetilde{W}$ but $\mathbf{x}+\mathbf{x} \notin \widetilde{W}$ since $\sum_{i=1}^n \alpha_i(x_i+x_i) = 2$. Thus $\widetilde{W}$ is not closed under vector addition,

12/02 and so $\widetilde{W}$ cannot be a subspace of $\mathbb{R}^n$.

3.3 Spanning Sets, Dimension and Bases

3.3.1 Linear independence

Extend ideas from vectors in $\mathbb{R}^3$.

Definition. A set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots\ \mathbf{v}_n\}$, $\mathbf{v}_j \in V$ $(j = 1, 2, \ldots, n)$, is *linearly independent* if for all scalars $\lambda_j \in \mathbb{R}$ $(j = 1, 2, \ldots, n)$,

$$\sum_{i=1}^n \lambda_i \mathbf{v}_i = \mathbf{0} \quad\Rightarrow\quad \lambda_i = 0 \quad\text{for}\quad i = 1, 2, \ldots, n\,. \tag{3.9}$$

Otherwise, the vectors are said to be *linearly dependent* since there exist scalars $\lambda_j \in \mathbb{R}$ $(j = 1, 2, \ldots, n)$, not all of which are zero, such that

$$\sum_{i=1}^n \lambda_i \mathbf{v}_i = \mathbf{0}\,.$$

12/03 *Remark.* With this definition $\{\mathbf{0}\}$ is a linearly dependent set.

3.3.2 Spanning sets

Definition. A subset of $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots\ \mathbf{u}_n\}$ of vectors in V is a *spanning set* for V if for every vector $\mathbf{v} \in V$, there exist scalars $\lambda_j \in \mathbb{R}$ $(j = 1, 2, \ldots, n)$, such that

$$\mathbf{v} = \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 + \ldots + \lambda_n\mathbf{u}_n\,. \tag{3.10}$$

Remark. The λ_j are not necessarily unique (but see below for when the spanning set is a basis). Further, we are implicitly assuming here (and in much that follows) that the vector space V can be spanned by a finite number of vectors (this is not always the case).

Definition. The set of all linear combinations of the $\mathbf{u}_j$ $(j = 1, 2, \ldots, n)$, i.e.

$$U = \left\{\mathbf{u} : \mathbf{u} = \sum_{i=1}^n \lambda_i \mathbf{u}_i,\ \lambda_j \in \mathbb{R}, j = 1, 2, \ldots, n\right\}$$

is closed under vector addition and scalar multiplication, and hence is a subspace of V. The vector space U is called the *span* of S (written $U = \operatorname{span} S$), and we say that S *spans* U.

Example. Consider $\mathbb{R}^3$, and let

$$\mathbf{u}_1 = (1,0,0)\,, \quad \mathbf{u}_2 = (0,1,0)\,, \quad \mathbf{u}_3 = (0,0,1)\,, \quad \mathbf{u}_4 = (1,1,1)\,, \quad \mathbf{u}_5 = (0,1,1)\,. \tag{3.11}$$

- The set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$ is linearly dependent since

$$\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 - \mathbf{u}_4 = \mathbf{0}\,.$$

  This set spans $\mathbb{R}^3$ since for any $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$,

$$\sum_{i=1}^{3} x_i \mathbf{u}_i = (x_1, x_2, x_3)\,.$$

- The set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent since if

$$\sum_{i=1}^{3} \lambda_i \mathbf{u}_i = (\lambda_1, \lambda_2, \lambda_3) = \mathbf{0}\,,$$

  then $\lambda_1 = \lambda_2 = \lambda_3 = 0$. This set spans $\mathbb{R}^3$ (as above).
- The set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4\}$ is linearly independent since if

$$\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \lambda_4 \mathbf{u}_4 = (\lambda_1 + \lambda_4, \lambda_2 + \lambda_4, \lambda_4) = \mathbf{0}\,,$$

  then $\lambda_1 = \lambda_2 = \lambda_4 = 0$. This set spans $\mathbb{R}^3$ since for any $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$,

$$(x_1 - x_3)\mathbf{u}_1 + (x_2 - x_3)\mathbf{u}_2 + x_3 \mathbf{u}_4 = (x_1, x_2, x_3)\,.$$

- The set $\{\mathbf{u}_1, \mathbf{u}_4, \mathbf{u}_5\}$ is linearly dependent since

$$\mathbf{u}_1 - \mathbf{u}_4 + \mathbf{u}_5 = \mathbf{0}\,.$$

  This set does *not* span $\mathbb{R}^3$, e.g. the vector $(0, 1, -1)$ cannot be expressed as a linear combination of $\mathbf{u}_1$, $\mathbf{u}_4$, $\mathbf{u}_5$. (The set does span the plane containing $\mathbf{u}_1$, $\mathbf{u}_4$ and $\mathbf{u}_5$.)

Theorem 3.2. *If a set $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linearly dependent, and spans the vector space V, we can reduce S to a linearly independent set also spanning V.*

Proof. If S is linear dependent then there exists λ_j, $j = 1, \ldots, n$, not all zero such that

$$\sum_{i=1}^{n} \lambda_i \mathbf{u}_i = \mathbf{0}\,.$$

Suppose that $\lambda_n \neq 0$ (if not reorder the vectors). Then

$$\mathbf{u}_n = -\sum_{i=1}^{n-1} \frac{\lambda_i}{\lambda_n} \mathbf{u}_i\,.$$

Since S spans V, and $\mathbf{u}_n$ can be expressed in terms of $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$, the set $S_{n-1} = \{\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ spans V. If S_{n-1} is linearly independent then we are done. If not repeat until $S_p = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, which spans V, is linearly independent. □

Example. Consider the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$, for $\mathbf{u}_j$ as defined in (3.11). This set spans $\mathbb{R}^3$, but is linearly dependent since $\mathbf{u}_4 = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$. There are a number of ways by which this set can be reduced to a linearly independent one that still spans $\mathbb{R}^3$, e.g. the sets $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_4\}$, $\{\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_4\}$ and $\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ all span R^3 and are linearly independent.

12/06 **Definition.** A *basis* for a vector space V is a linearly independent spanning set of vectors in V.

Lemma 3.3. *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis for a vector space V, then so is $\{\mathbf{w}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for*

$$\mathbf{w}_1 = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \tag{3.12}$$

if $\lambda_1 \neq 0$.

Proof. Let $\mathbf{x} \in V$, then since $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ spans V there exists μ_i such that

$$\mathbf{x} = \sum_{i=1}^{n} \mu_i \mathbf{u}_i \,.$$

It follows from (3.12) that

$$\mathbf{x} = \frac{\mu_1}{\lambda_1}\left(\mathbf{w}_1 - \sum_{i=2}^{n} \lambda_i \mathbf{u}_i\right) + \sum_{i=2}^{n} \mu_i \mathbf{u}_i = \frac{\mu_1}{\lambda_1}\mathbf{w}_1 + \sum_{i=2}^{n}\left(\mu_i - \frac{\lambda_i \mu_1}{\lambda_1}\right)\mathbf{u}_i \,,$$

and hence that $\{\mathbf{w}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ spans V. Further, suppose that

$$\nu_1 \mathbf{w}_1 + \sum_{i=2}^{n} \nu_i \mathbf{u}_i = \mathbf{0} \,.$$

Then from (3.12)

$$\lambda_1 \nu_1 \mathbf{u}_1 + \sum_{i=2}^{n} (\nu_1 \lambda_i + \nu_i) \mathbf{u}_i = \mathbf{0} \,.$$

But $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis so

$$\lambda_1 \nu_1 = 0 \,, \quad \nu_1 \lambda_i + \nu_i = 0 \,, \quad i = 2, \ldots, n \,.$$

Since $\lambda_1 \neq 0$, we deduce that $\nu_i = 0$ for $i = 1, \ldots, n$, and hence that $\{\mathbf{w}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is linearly independent. □

Theorem 3.4. *Let V be a finite-dimensional vector space, then every basis has the same number of elements.*

Proof. Suppose that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are two bases for V with $m \geqslant n$ (wlog). Since the set $\{\mathbf{u}_i, i = 1, \ldots, n\}$ is a basis there exist λ_i, $i = 1, \ldots, n$ such that

$$\mathbf{w}_1 = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i$$

with $\lambda_1 \neq 0$ (if not reorder the $\mathbf{u}_i$, $i = 1, \ldots, n$). The set $\{\mathbf{w}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an alternative basis. Hence, there exists μ_i, $i = 1, \ldots, n$ (at least one of which is non-zero) such that

$$\mathbf{w}_2 = \mu_1 \mathbf{w}_1 + \sum_{i=2}^{n} \mu_i \mathbf{u}_i \,.$$

Moreover, at least one of the μ_i, $i = 2, \ldots, n$ must be non-zero, otherwise the subset $\{\mathbf{w}_1, \mathbf{w}_2\}$ would be linearly dependent and so violate the hypothesis that the set $\{\mathbf{w}_i, i = 1, \ldots, m\}$ forms a basis. Assume $\mu_2 \neq 0$ (otherwise reorder the $\mathbf{u}_i$, $i = 2, \ldots, n$). We can then deduce that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n\}$ forms a basis. Continue to deduce that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ forms a basis, and hence spans V. If $m > n$ this would mean that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ was linearly dependent, contradicting the original assumption. Hence $m = n$. □

Remark. An analogous argument can be used to show that no linearly independent set of vectors can have more members than a basis.

Definition. The number of vectors in a basis of V is the *dimension* of V, written $\dim V$.

Remarks.

- The vector space $\{\mathbf{0}\}$ is said to have dimension 0.
- We shall restrict ourselves to vector spaces of finite dimension (although we have already encountered one vector space of infinite dimension, i.e. the vector space of real functions defined on $[a, b]$).

Exercise. Show that any set of $\dim V$ linearly independent vectors of V is a basis.

Examples.

(i) The set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, with $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, 0, \ldots, 1)$, is a basis for $\mathbb{R}^n$, and thus $\dim \mathbb{R}^n = n$.

(ii) The subspace $U \subset \mathbb{R}^n$ consisting of vectors $(x, x, \ldots, x)$ with $x \in \mathbb{R}$ is spanned by $\mathbf{e} = (1, 1, \ldots, 1)$. Since $\{\mathbf{e}\}$ is linearly independent, $\{\mathbf{e}\}$ is a basis and thus $\dim U = 1$.

13/02

Theorem 3.5. *If U is a proper subspace of V, any basis of U can be extended into a basis for V.*

Proof. (*Unlectured.*) Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell\}$ be a basis of U ($\dim U = \ell$). If U is a proper subspace of V then there are vectors in V but not in U. Choose $\mathbf{w}_1 \in V$, $\mathbf{w}_1 \notin U$. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell, \mathbf{w}_1\}$ is linearly independent. If it spans V, i.e. ($\dim V = \ell + 1$) we are done, if not it spans a proper subspace, say U_1, of V. Now choose $\mathbf{w}_2 \in V$, $\mathbf{w}_2 \notin U_1$. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell, \mathbf{w}_1, \mathbf{w}_2\}$ is linearly independent. If it spans V we are done, if not it spans a proper subspace, say U_2, of V. Now repeat until $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\ell, \mathbf{w}_1, \ldots, \mathbf{w}_m\}$, for some m, spans V (which must be possible since we are only studying finite dimensional vector spaces). □

Example. Let $V = \mathbb{R}^3$ and suppose $U = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 = 0\}$. U is a subspace (see example (ii) on page 46 with $\alpha_1 = \alpha_2 = 1,\ \alpha_3 = 0$). Since $x_1 + x_2 = 0$ implies that $x_2 = -x_1$ with no restriction on x_3, for $\mathbf{x} \in U$ we can write

$$\begin{aligned} \mathbf{x} &= (x_1, -x_1, x_3) \qquad x_1, x_3 \in \mathbb{R} \\ &= x_1(1, -1, 0) + x_3(0, 0, 1)\,. \end{aligned}$$

Hence a basis for U is $\{(1, -1, 0), (0, 0, 1)\}$.

Now choose any $\mathbf{y}$ such that $\mathbf{y} \in \mathbb{R}^3$, $\mathbf{y} \notin U$, e.g. $\mathbf{y} = (1, 0, 0)$, or $(0, 1, 0)$, or $(1, 1, 0)$. Then $\{(1, -1, 0), (0, 0, 1)\}, \mathbf{y}\}$ is linearly independent and forms a basis for span $\{(1, -1, 0), (0, 0, 1)\}, \mathbf{y}\}$, which has dimension three. But $\dim \mathbb{R}^3 = 3$, hence

$$\operatorname{span}\{(1, -1, 0), (0, 0, 1), \mathbf{y}\} = \mathbb{R}^3\,,$$

and $\{(1, -1, 0), (0, 0, 1), \mathbf{y}\}$ is an extension of a basis of U to a basis of $\mathbb{R}^3$.

3.4 Components

Theorem 3.6. *Let V be a vector space and let $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis of V. Then each $\mathbf{v} \in V$ can be expressed as a linear combination of the $\mathbf{e}_i$, i.e.*

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i\,, \tag{3.13}$$

where the v_i are unique for each $\mathbf{v}$.

Proof. If S is a basis then, from the definition of a basis, for all $\mathbf{v} \in V$ there exist $v_i \in \mathbb{R}$, $i = 1, \ldots, n$, such that

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

Suppose also there exist $v'_i \in \mathbb{R}$, $i = 1, \ldots, n$, such that

$$\mathbf{v} = \sum_{i=1}^{n} v'_i \mathbf{e}_i \,.$$

Then

$$\sum_{i=1}^{n} v_i \mathbf{e}_i - \sum_{i=1}^{n} v'_i \mathbf{e}_i = \sum_{i=1}^{n} (v_i - v'_i) \mathbf{e}_i = \mathbf{0} \,.$$

But the $\mathbf{e}_i$ are linearly independent, hence $(v_i - v'_i) = 0$, $i = 1, \ldots, n$, and hence the v_i are unique. □

Definition. We call the v_i the *components* of $\mathbf{v}$ with respect to the basis S.

Remark. For a given basis, we conclude that each vector $\mathbf{x}$ in a vector space V of dimension n is associated with a unique set of real numbers, namely its components $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Further, if $\mathbf{y}$ is associated with components $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ then $(\lambda\mathbf{x} + \mu\mathbf{y})$ is associated with components $(\lambda\mathbf{x} + \mu\mathbf{y}) = (\lambda x_1 + \mu y_1, \ldots, \lambda x_n + \mu y_n) \in \mathbb{R}^n$. A little more work then demonstrates that that all dimension n vector spaces over the real numbers are the 'same as' $\mathbb{R}^n$. To be more precise, it is possible to show that every real n-dimensional vector space V is *isomorphic* to $\mathbb{R}^n$. Thus $\mathbb{R}^n$ is the
13/03 prototypical example of a real n-dimensional vector space.

3.5 Intersection and Sum of Subspaces

Let U and W be subspaces of a vector space V over $\mathbb{R}$.

Definition. $U \cap W$ is the *intersection* of U and W and consists of all vectors $\mathbf{v}$ such that $\mathbf{v} \in U$ and $\mathbf{v} \in W$.

Remark. $U \cap W$ contains at least $\mathbf{0}$ so is not empty.

Theorem 3.7. *$U \cap W$ is a subspace of V.*

Proof. If $\mathbf{x}, \mathbf{y} \in U \cap W$, then $\mathbf{x}, \mathbf{y} \in U$ and $\mathbf{x}, \mathbf{y} \in W$, hence for any $\lambda, \mu \in \mathbb{R}$, $\lambda\mathbf{x} + \mu\mathbf{y} \in U$ and $\lambda\mathbf{x} + \mu\mathbf{y} \in W$ (since both U and W are subspaces). Hence $\lambda\mathbf{x} + \mu\mathbf{y} \in U \cap W$, and $U \cap W$ is a subspace of V. □

Remark. $U \cap W$ is also a subspace of U and a subspace of W. 13/06

Definition. $U + W$ is called the *sum* of subspaces U and W, and consists of all vectors $\mathbf{u} + \mathbf{w} \in V$ where $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Remark. $U + W$ is not the same as $U \cup W$. If $U = \text{span}\{(1, 0)\}$ and $W = \text{span}\{(0, 1)\}$, then $U \cup W$ is the abscissa and the ordinate, while $U + W$ is the 2D-plane.

Theorem 3.8. *$U + W$ is a subspace of V.*

Proof. If $\mathbf{v}_1, \mathbf{v}_2 \in U + W$, then there exist $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$ such that $\mathbf{v}_1 = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{v}_2 = \mathbf{u}_2 + \mathbf{w}_2$. For any $\lambda, \mu \in \mathbb{R}$, $\lambda\mathbf{u}_1 + \mu\mathbf{u}_2 \in U$ and $\lambda\mathbf{w}_1 + \mu\mathbf{w}_2 \in W$. Hence

$$\lambda\mathbf{v}_1 + \mu\mathbf{v}_2 = \lambda(\mathbf{u}_1 + \mathbf{w}_1) + \mu(\mathbf{u}_2 + \mathbf{w}_2) = (\lambda\mathbf{u}_1 + \mu\mathbf{u}_2) + (\lambda\mathbf{w}_1 + \mu\mathbf{w}_2) \in U + W \,,$$

and thus $U + W$ is a subspace. □

Remark. Since $U \subseteq U + W$ and $W \subseteq U + W$, both U and W are subspaces of $U + W$.

3.5.1 $\dim(U+W)$

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be bases for U and W respectively. Then

$$U + W = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{w}_1, \ldots, \mathbf{w}_m\},$$

and hence $\dim(U+W) \leqslant \dim U + \dim W$. However, we can say more than this.

Theorem 3.9.

$$\dim(U+W) + \dim(U \cap W) = \dim U + \dim W. \tag{3.14}$$

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ be a basis for $U \cap W$ (OK if empty). Extend this to a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s\}$ for U and a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{g}_1, \ldots, \mathbf{g}_t\}$ for W. Then $\{\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s, \mathbf{g}_1, \ldots, \mathbf{g}_t\}$ spans $U+W$. For $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, r$, $\mu_i \in \mathbb{R}$, $i = 1, \ldots, s$ and $\nu_i \in \mathbb{R}$, $i = 1, \ldots, t$, seek solutions to

$$\sum_{i=1}^{r} \lambda_i \mathbf{e}_i + \sum_{i=1}^{s} \mu_i \mathbf{f}_i + \sum_{i=1}^{t} \nu_i \mathbf{g}_i = \mathbf{0}. \tag{3.15}$$

Define $\mathbf{v}$ by

$$\mathbf{v} = \sum_{i=1}^{s} \mu_i \mathbf{f}_i = -\sum_{i=1}^{r} \lambda_i \mathbf{e}_i - \sum_{i=1}^{t} \nu_i \mathbf{g}_i,$$

where the first sum is a vector in U, while the RHS is a vector in W. Hence $\mathbf{v} \in U \cap W$, and so

$$\mathbf{v} = \sum_{i=1}^{r} \alpha_i \mathbf{e}_i \quad \text{for some } \alpha_i \in \mathbb{R},\ i = 1, \ldots, r.$$

We deduce that

$$\sum_{i=1}^{r} (\alpha_i + \lambda_i) \mathbf{e}_i + \sum_{i=1}^{t} \nu_i \mathbf{g}_i = \mathbf{0}.$$

But $\{\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{g}_1, \ldots, \mathbf{g}_t\}$ is a basis (for W) and hence linearly independent, and so $\nu_i = 0$, $i = 1, \ldots, t$. Similarly $\mu_i = 0$, $i = 1, \ldots, s$. Further $\lambda_i = 0$, $i = 1, \ldots, r$ from (3.15) since $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ is a basis (for $U \cap W$).

It follows that $\{\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s, \mathbf{g}_1, \ldots, \mathbf{g}_t\}$ is linearly independent and is thus a basis for $U + W$. Thus

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

14/02

□

3.5.2 Examples

(i) Suppose that $U = \{\mathbf{x} \in \mathbb{R}^4 : x_1 = 0\}$ and $W = \{\mathbf{x} \in \mathbb{R}^4 : x_1 + 2x_2 = 0\}$ (both are subspaces of $\mathbb{R}^4$ from example (ii) on page 46). Then, say,

$$\begin{aligned} U &= \operatorname{span}\{(0,1,0,0),(0,0,1,0),(0,0,0,1)\} \quad \Rightarrow \quad \dim U = 3, \\ W &= \operatorname{span}\{(-2,1,0,0),(0,0,1,1),(0,0,1,-1)\} \quad \Rightarrow \quad \dim W = 3. \end{aligned}$$

If $\mathbf{x} \in U \cap W$ then

$$x_1 = 0 \quad \text{and} \quad x_1 + 2x_2 = 0, \quad \text{and hence} \quad x_1 = x_2 = 0.$$

Thus

$$U \cap W = \operatorname{span}\{(0,0,1,0),(0,0,0,1)\} = \operatorname{span}\{(0,0,1,1),(0,0,1,-1)\} \quad \text{and} \quad \dim U \cap W = 2.$$

On the other hand

$$\begin{aligned} U + W &= \operatorname{span}\{(0,1,0,0),(0,0,1,0),(0,0,0,1),(-2,1,0,0),(0,0,1,1),(0,0,1,-1)\} \\ &= \operatorname{span}\{(0,1,0,0),(0,0,1,0),(0,0,0,1),(-2,1,0,0)\}, \end{aligned}$$

after reducing the linearly dependent spanning set to a linearly independent spanning set. Hence $\dim U+W=4$. Hence, in agreement with (3.14),

$$\begin{aligned}\dim U+\dim W &= 3+3 \;=\; 6\,,\\ \dim(U\cap W)+\dim(U{+}W) &= 2+4 \;=\; 6\,.\end{aligned}$$

(ii) Let V be the vector space of real-valued functions on $\{1,2,3\}$, i.e. $V=\{f:(f(1),f(2),f(3))\in\mathbb{R}^3\}$. Define addition, scalar multiplication, etc. as in (3.8a), (3.8b), (3.8c) and (3.8d). Let e_1 be the function such that $e_1(1)=1$, $e_1(2)=0$ and $e_1(3)=0$; define e_2 and e_3 similarly (where we have deliberately not used **bold** notation). Then $\{e_i, i=1,2,3\}$ is a basis for V, and $\dim V=3$.

Let

$$\begin{aligned} U &= \{f:(f(1),f(2),f(3))\in\mathbb{R}^3 \quad\text{with}\quad f(1)=0\}, \quad\text{then}\quad \dim U=2\,,\\ W &= \{f:(f(1),f(2),f(3))\in\mathbb{R}^3 \quad\text{with}\quad f(2)=0\}, \quad\text{then}\quad \dim W=2\,.\end{aligned}$$

Then

$$\begin{aligned} U+W &= V \quad\text{so}\quad \dim(U{+}W)=3\,,\\ U\cap W &= \{f:(f(1),f(2),f(3))\in\mathbb{R}^3 \quad\text{with}\quad f(1)=f(2)=0\}, \quad\text{so}\quad \dim U\cap W=1\,;\end{aligned}$$

again verifying (3.14).

14/03

3.6 Scalar Products (a.k.a. Inner Products)

3.6.1 Definition of a scalar product

The three-dimensional linear vector space $V=\mathbb{R}^3$ has the additional property that any two vectors $\mathbf{u}$ and $\mathbf{v}$ can be combined to form a *scalar* $\mathbf{u}\cdot\mathbf{v}$. This can be generalised to an n-dimensional vector space V over the reals by assigning, for every pair of vectors $\mathbf{u}$, $\mathbf{v}\in V$, a *scalar product*, or *inner product*, $\mathbf{u}\cdot\mathbf{v}\in\mathbb{R}$ with the following properties.

(i) *Symmetry*, i.e.
$$\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}\,. \tag{3.16a}$$

(ii) *Linearity* in the second argument, i.e. for $\lambda,\mu\in\mathbb{R}$
$$\mathbf{u}\cdot(\lambda\mathbf{v}_1+\mu\mathbf{v}_2)=\lambda\,\mathbf{u}\cdot\mathbf{v}_1+\mu\,\mathbf{u}\cdot\mathbf{v}_2\,. \tag{3.16b}$$

(iii) *Non-negativity*, i.e. a scalar product of a vector with itself should be positive, i.e.
$$\mathbf{v}\cdot\mathbf{v}\geqslant 0\,. \tag{3.16c}$$

This allows us to write $\mathbf{v}\cdot\mathbf{v}=|\mathbf{v}|^2$, where the real positive number $|\mathbf{v}|$ is a *norm* (cf. length) of the vector $\mathbf{v}$.

(iv) *Non-degeneracy*, i.e. the only vector of zero norm should be the zero vector, i.e.
$$|\mathbf{v}|=0 \quad\Rightarrow\quad \mathbf{v}=\mathbf{0}\,. \tag{3.16d}$$

Remark. Properties (3.16a) and (3.16b) imply linearity in the first argument, i.e. for $\lambda,\mu\in\mathbb{R}$

$$\begin{aligned}(\lambda\mathbf{u}_1+\mu\mathbf{u}_2)\cdot\mathbf{v} &= \mathbf{v}\cdot(\lambda\mathbf{u}_1+\mu\mathbf{u}_2)\\ &= \lambda\mathbf{v}\cdot\mathbf{u}_1+\mu\mathbf{v}\cdot\mathbf{u}_2\\ &= \lambda\mathbf{u}_1\cdot\mathbf{v}+\mu\mathbf{u}_2\cdot\mathbf{v}\,.\end{aligned} \tag{3.17}$$

Alternative notation. An alternative notation for scalar products and norms is

$$\langle\,\mathbf{u}\mid\mathbf{v}\,\rangle\equiv\mathbf{u}\cdot\mathbf{v}\,, \tag{3.18a}$$
$$\|\mathbf{v}\|\equiv|\mathbf{v}|=(\mathbf{v}\cdot\mathbf{v})^{\frac{1}{2}}\,. \tag{3.18b}$$

3.6.2 Schwarz's inequality

Schwarz's inequality states that

$$|\mathbf{u}\cdot\mathbf{v}| \leqslant \|\mathbf{u}\|\,\|\mathbf{v}\|\,, \tag{3.19}$$

with equality only when $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

Proof. For $\lambda \in \mathbb{R}$ consider

$$\begin{aligned}
0 \leqslant \|\mathbf{u}+\lambda\mathbf{v}\|^2 &= (\mathbf{u}+\lambda\mathbf{v})\cdot(\mathbf{u}+\lambda\mathbf{v}) && \text{from (3.18b)}\\
&= \mathbf{u}\cdot\mathbf{u}+\lambda\mathbf{u}\cdot\mathbf{v}+\lambda\mathbf{v}\cdot\mathbf{u}+\lambda^2\mathbf{v}\cdot\mathbf{v} && \text{from (3.16b) and (3.17)}\\
&= \|\mathbf{u}\|^2+2\lambda\mathbf{u}\cdot\mathbf{v}+\lambda^2\|\mathbf{v}\|^2 && \text{from (3.16a).}
\end{aligned}$$

We have two cases to consider: $\mathbf{v}=\mathbf{0}$ and $\mathbf{v}\neq\mathbf{0}$. First, suppose that $\mathbf{v}=\mathbf{0}$, so that $\|\mathbf{v}\|=0$. The right-hand-side then simplifies from a quadratic in λ to an expression that is linear in λ. If $\mathbf{u}\cdot\mathbf{v}\neq 0$ we have a contradiction since for certain choices of λ this simplified expression can be negative. Hence we conclude that

$$\mathbf{u}\cdot\mathbf{v}=0 \quad \text{if} \quad \mathbf{v}=\mathbf{0}\,,$$

in which case (3.19) is satisfied as an equality.

Second, suppose that $\mathbf{v}\neq\mathbf{0}$. The right-hand-side is then a quadratic in λ that, since it is not negative, has at most one real root. Hence '$b^2 \leqslant 4ac$', i.e.

$$(2\mathbf{u}\cdot\mathbf{v})^2 \leqslant 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2\,.$$

Schwarz's inequality follows on taking the positive square root, with equality only if $\mathbf{u}=-\lambda\mathbf{v}$ for some λ.

□

Remark. A more direct way of proving that $\mathbf{u}\cdot\mathbf{0}=0$ is to set $\lambda=\mu=0$ in (3.16b). Then, since $\mathbf{0}=\mathbf{0}+\mathbf{0}$ from (3.1c) and $\mathbf{0}=0\mathbf{v}_j$ from (3.3),

$$\mathbf{u}\cdot\mathbf{0}=\mathbf{u}\cdot(0\mathbf{v}_1+0\mathbf{v}_2)=0\,(\mathbf{u}\cdot\mathbf{v}_1)+0\,(\mathbf{u}\cdot\mathbf{v}_2)=0\,.$$

3.6.3 Triangle inequality

This triangle inequality is a generalisation of (1.13a) and (2.5) and states that

$$\|\mathbf{u}+\mathbf{v}\| \leqslant \|\mathbf{u}\|+\|\mathbf{v}\|\,. \tag{3.20}$$

Proof. This follows from taking square roots of the following inequality:

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2+2\,\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2 && \text{from above with } \lambda=1\\
&\leqslant \|\mathbf{u}\|^2+2|\mathbf{u}\cdot\mathbf{v}|+\|\mathbf{v}\|^2\\
&\leqslant \|\mathbf{u}\|^2+2\|\mathbf{u}\|\,\|\mathbf{v}\|+\|\mathbf{v}\|^2 && \text{from (3.19)}\\
&\leqslant (\|\mathbf{u}\|+\|\mathbf{v}\|)^2\,.
\end{aligned}$$

□

Remark and exercise. In the same way that (1.13a) can be extended to (1.13b) we can similarly deduce that

14/06

$$\|\mathbf{u}-\mathbf{v}\| \geqslant \big|\,\|\mathbf{u}\|-\|\mathbf{v}\|\,\big|\,. \tag{3.21}$$

3.6.4 The scalar product for $\mathbb{R}^n$

In example (i) on page 44 we saw that the space of all n-tuples of real numbers forms an n-dimensional vector space over $\mathbb{R}$, denoted by $\mathbb{R}^n$ (and referred to as real coordinate space).

We define the *scalar product* on $\mathbb{R}^n$ as (cf. (2.39))

$$\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1y_1 + x_2y_2 + \ldots + x_ny_n\,. \tag{3.22}$$

Exercise. Confirm that (3.22) satisfies (3.16a), (3.16b), (3.16c) and (3.16d), i.e. for $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathbb{R}^n$ and $\lambda,\mu\in\mathbb{R}$,

$$\begin{aligned}
\mathbf{x}\cdot\mathbf{y} &= \mathbf{y}\cdot\mathbf{x}\,,\\
\mathbf{x}\cdot(\lambda\mathbf{y}+\mu\mathbf{z}) &= \lambda\,\mathbf{x}\cdot\mathbf{y}+\mu\,\mathbf{x}\cdot\mathbf{z}\,,\\
\|\mathbf{x}\|^2 = \mathbf{x}\cdot\mathbf{x} &\geqslant 0\,,\\
\|\mathbf{x}\| = 0 &\Rightarrow \mathbf{x}=\mathbf{0}\,.
\end{aligned}$$

Remarks.

- The *length*, or *Euclidean norm*, of a vector $\mathbf{x}\in\mathbb{R}^n$ is defined as

  $$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}\,,$$

  while the *interior angle* θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined to be

  $$\theta = \arccos\left(\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right)\,.$$

- We need to be a little careful with the definition (3.22). It is important to appreciate that the scalar product for $\mathbb{R}^n$ as defined by (3.22) is consistent with the scalar product for $\mathbb{R}^3$ defined in (2.13) only when the x_i and y_i are components with respect to an orthonormal basis. To this end we might view (3.22) as the scalar product with respect to the *standard basis*

  $$\mathbf{e}_1 = (1,0,\ldots,0),\ \mathbf{e}_2 = (0,1,\ldots,0),\ \ldots,\ \mathbf{e}_n = (0,0,\ldots,1).$$

  For the case when the x_i and y_i are components with respect to a non-orthonormal basis (e.g. a non-orthogonal basis), the scalar product in $\mathbb{R}^n$ equivalent to (2.13) has a more complicated form than (3.22) (and for which we need matrices). The good news is that for *any* scalar product defined
15/02 on a vector space over $\mathbb{R}$, orthonormal bases always exist.

3.6.5 Examples

Hyper-sphere in $\mathbb{R}^n$. The (hyper-)sphere in $\mathbb{R}^n$ with centre $\mathbf{a}\in\mathbb{R}^n$ and radius $r\in\mathbb{R}$ is given by

$$\Sigma = \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}-\mathbf{a}\| = r > 0,\ r\in\mathbb{R},\ \mathbf{a}\in\mathbb{R}^n\}. \tag{3.23}$$

Remark. Σ is not a subspace of $\mathbb{R}^n$.

Hyper-plane in $\mathbb{R}^n$. The (hyper-)plane in $\mathbb{R}^n$ that passes through $\mathbf{b}\in\mathbb{R}^n$ and has normal $\mathbf{n}\in\mathbb{R}^n$, is given by

$$\Pi = \{\mathbf{x}\in\mathbb{R}^n : (\mathbf{x}-\mathbf{b})\cdot\mathbf{n} = 0,\ \mathbf{b},\mathbf{n}\in\mathbb{R}^n \text{ with } \|\mathbf{n}\|=1\}. \tag{3.24}$$

Remark. If $\mathbf{b}\cdot\mathbf{n}=0$ so that $\Pi = \{\mathbf{x}\in\mathbb{R}^n : \sum_{i=1}^n x_in_i = 0\}$, i.e. so that the hyper-plane passes through the origin, then Π is a subspace of dimension $(n-1)$ (see example (ii) of §3.2.1).

4 Linear Maps and Matrices

4.0 Why Study This?

Many problems in the real world are linear, e.g. electromagnetic waves satisfy linear equations, and almost all sounds you hear are 'linear' (exceptions being sonic booms). Moreover, many computational approaches to solving nonlinear problems involve 'linearisations' at some point in the process (since computers are good at solving linear problems). The aim of this section is to construct a general framework for viewing such linear problems.

4.1 General Definitions

Definition. Let A, B be sets. A *map* f of A into B is a rule that assigns to each $x \in A$ a unique $x' \in B$. We write

$$f : A \to B \quad \text{and/or} \quad x \mapsto x' = f(x)$$

Examples. Möbius maps of the complex plane, translation, inversion with respect to a sphere.

Definition. A is the *domain* of f.

Definition. B is the *range*, or *codomain*, of f.

Definition. $f(x) = x'$ is the *image of x under f*.

Definition. $f(A)$ is the *image of A under f*, i.e. the set of all image points $x' \in B$ of $x \in A$.

Remark. $f(A) \subseteq B$, but there may be elements of B that are not images of any $x \in A$.

4.2 Linear Maps

We shall consider maps from a vector space V to a vector space W, focussing on $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, for $m, n \in \mathbb{Z}^+$.

Remark. This is not unduly restrictive since every real n-dimensional vector space V is isomorphic to $\mathbb{R}^n$ (recall that any element of a vector space V with $\dim V = n$ corresponds to a unique element of $\mathbb{R}^n$).

Definition. Let V, W be vector spaces over $\mathbb{R}$. The map $T : V \to W$ is a *linear map* or *linear transformation* if

(i)
$$T(\mathbf{a} + \mathbf{b}) = T(\mathbf{a}) + T(\mathbf{b}) \quad \text{for all} \quad \mathbf{a}, \mathbf{b} \in V\,, \tag{4.1a}$$

(ii)
$$T(\lambda\mathbf{a}) = \lambda T(\mathbf{a}) \quad \text{for all} \quad \mathbf{a} \in V \text{ and } \lambda \in \mathbb{R}\,, \tag{4.1b}$$

or equivalently if

$$T(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda T(\mathbf{a}) + \mu T(\mathbf{b}) \quad \text{for all} \quad \mathbf{a}, \mathbf{b} \in V \text{ and } \lambda, \mu \in \mathbb{R}\,. \tag{4.2}$$

Property. $T(V)$ is a subspace of W, since for $T(\mathbf{a}), T(\mathbf{b}) \in T(V)$

$$\lambda T(\mathbf{a}) + \mu T(\mathbf{b}) = T(\lambda\mathbf{a} + \mu\mathbf{b}) \in T(V) \quad \text{for all} \quad \lambda, \mu \in \mathbb{R}. \tag{4.3}$$

Now apply Theorem 3.1 on page 45.

The zero element. Since $T(V)$ is a subspace, it follows that $\mathbf{0} \in T(V)$. However, we can say more than that. Set $\mathbf{b} = \mathbf{0} \in V$ in (4.1a), to deduce that

$$T(\mathbf{a}) = T(\mathbf{a} + \mathbf{0}) = T(\mathbf{a}) + T(\mathbf{0}) \quad \text{for all} \quad \mathbf{a} \in V\,.$$

Thus from the uniqueness of the zero element it follows that $T(\mathbf{0}) = \mathbf{0} \in W$.

15/03 *Remark.* $T(\mathbf{b}) = \mathbf{0} \in W$ does **not** imply $\mathbf{b} = \mathbf{0} \in V$.

4.2.1 Examples

(i) Consider translation in $\mathbb{R}^3$ (an isometry), i.e. consider

$$\mathbf{x} \mapsto \mathbf{x}' = T(\mathbf{x}) = \mathbf{x} + \mathbf{a}\,, \quad \text{where} \quad \mathbf{a} \in \mathbb{R}^3 \quad \text{and} \quad \mathbf{a} \neq \mathbf{0}\,.$$

This is *not* a linear map by the *strict* definition of a linear map since

$$T(\mathbf{x}) + T(\mathbf{y}) = \mathbf{x} + \mathbf{a} + \mathbf{y} + \mathbf{a} = T(\mathbf{x} + \mathbf{y}) + \mathbf{a} \neq T(\mathbf{x} + \mathbf{y})\,.$$

(ii) Next consider the isometric map $\mathcal{H}_\Pi : \mathbb{R}^3 \to \mathbb{R}^3$ consisting of reflections in the plane $\Pi : \mathbf{x}.\mathbf{n} = 0$ where $\mathbf{x}, \mathbf{n} \in \mathbb{R}^3$ and $|\mathbf{n}| = 1$; then from (2.91)

$$\mathbf{x} \mapsto \mathbf{x}' = \mathcal{H}_\Pi(\mathbf{x}) = \mathbf{x} - 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n}\,.$$

Hence for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ under the map $\mathcal{H}_\Pi$

$$\mathbf{x}_1 \mapsto \mathbf{x}_1' = \mathbf{x}_1 - 2(\mathbf{x}_1\cdot\mathbf{n})\mathbf{n} = \mathcal{H}_\Pi(\mathbf{x}_1)\,,$$

$$\mathbf{x}_2 \mapsto \mathbf{x}_2' = \mathbf{x}_2 - 2(\mathbf{x}_2\cdot\mathbf{n})\mathbf{n} = \mathcal{H}_\Pi(\mathbf{x}_2)\,,$$

and so for all $\lambda, \mu \in \mathbb{R}$

$$\begin{aligned}\mathcal{H}_\Pi(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) &= (\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) - 2\left((\lambda\mathbf{x}_1 + \mu\mathbf{x}_2)\cdot\mathbf{n}\right)\mathbf{n} \\ &= \lambda(\mathbf{x}_1 - 2(\mathbf{x}_1\cdot\mathbf{n})\mathbf{n}) + \mu(\mathbf{x}_2 - 2(\mathbf{x}_2\cdot\mathbf{n})\mathbf{n}) \\ &= \lambda\mathcal{H}_\Pi(\mathbf{x}_1) + \mu\mathcal{H}_\Pi(\mathbf{x}_2)\,.\end{aligned}$$

Thus (4.2) is satisfied, and so $\mathcal{H}_\Pi$ is a linear map.

(iii) As another example consider projection onto a line with direction $\mathbf{t} \in \mathbb{R}^3$ as defined by (cf. (2.17b))

$$P : \mathbb{R}^3 \to \mathbb{R}^3 \quad \mathbf{x} \mapsto \mathbf{x}' = P(\mathbf{x}) = (\mathbf{x}\cdot\mathbf{t})\mathbf{t} \quad \text{where} \quad \mathbf{t}\cdot\mathbf{t} = 1\,.$$

From the observation that

$$\begin{aligned}P(\lambda\mathbf{x}_1 + \mu\mathbf{x}_2) &= ((\lambda\mathbf{x}_1 + \mu\mathbf{x}_2)\cdot\mathbf{t})\mathbf{t} \\ &= \lambda(\mathbf{x}_1\cdot\mathbf{t})\mathbf{t} + \mu(\mathbf{x}_2\cdot\mathbf{t})\mathbf{t} \\ &= \lambda P(\mathbf{x}_1) + \mu P(\mathbf{x}_2)\end{aligned}$$

we conclude that this is a linear map.

Remarks.

(i) The range of P is given by $P(\mathbb{R}^3) = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{t} \text{ for } \lambda \in \mathbb{R}\}$, which is a 1-dimensional subspace of $\mathbb{R}^3$.

(ii) A projection is not an isometry.

(iv) As another example of a map where the domain of T has a higher dimension than the range of T consider $T:\mathbb{R}^3\to\mathbb{R}^2$ where

$$(x,y,z)\mapsto(u,v)=T(\mathbf{x})=(x+y,2x-z)\,.$$

We observe that T is a linear map since

$$\begin{aligned}T(\lambda\mathbf{x}_1+\mu\mathbf{x}_2) &= ((\lambda x_1+\mu x_2)+(\lambda y_1+\mu y_2),2(\lambda x_1+\mu x_2)-(\lambda z_1+\mu z_2))\\ &= \lambda(x_1+y_1,2x_1-z_1)+\mu(x_2+y_2,2x_2-z_2)\\ &= \lambda T(\mathbf{x}_1)+\mu T(\mathbf{x}_2)\,.\end{aligned}$$

Remarks.

- The standard basis vectors for $\mathbb{R}^3$, i.e. $\mathbf{e}_1=(1,0,0)$, $\mathbf{e}_2=(0,1,0)$ and $\mathbf{e}_3=(0,0,1)$ are mapped to the vectors

$$\left.\begin{aligned}T(\mathbf{e}_1) &= (1,2)\\ T(\mathbf{e}_2) &= (1,0)\\ T(\mathbf{e}_3) &= (0,-1)\end{aligned}\right\}\quad\text{which are linearly dependent and span }\mathbb{R}^2.$$

  Hence $T(\mathbb{R}^3)=\mathbb{R}^2$.

- We observe that

$$\mathbb{R}^3=\operatorname{span}\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}\quad\text{and}\quad T(\mathbb{R}^3)=\mathbb{R}^2=\operatorname{span}\{T(\mathbf{e}_1),T(\mathbf{e}_2),T(\mathbf{e}_3)\}.$$

- We also observe that

$$T(\mathbf{e}_1)-T(\mathbf{e}_2)+2T(\mathbf{e}_3)=\mathbf{0}\in\mathbb{R}^2\quad\text{which means that}\quad T(\lambda(\mathbf{e}_1-\mathbf{e}_2+2\mathbf{e}_3))=\mathbf{0}\in\mathbb{R}^2\,,$$

  for all $\lambda\in\mathbb{R}$. Thus the whole of the subspace of $\mathbb{R}^3$ spanned by $\{\mathbf{e}_1-\mathbf{e}_2+2\mathbf{e}_3\}$, i.e. the 1-dimensional line specified by $\lambda(\mathbf{e}_1-\mathbf{e}_2+2\mathbf{e}_3)$, is mapped onto $\mathbf{0}\in\mathbb{R}^2$.

(v) As an example of a map where the domain of T has a lower dimension than the range of T consider $T:\mathbb{R}^2\to\mathbb{R}^4$ where

$$(x,y)\mapsto(s,t,u,v)=T(x,y)=(x+y,x,y-3x,y)\,.$$

T is a linear map since

$$\begin{aligned}T(\lambda\mathbf{x}_1+\mu\mathbf{x}_2) &= ((\lambda x_1+\mu x_2)+(\lambda y_1+\mu y_2),\lambda x_1+\mu x_2,\lambda y_1+\mu y_2-3(\lambda x_1+\mu x_2),\lambda y_1+\mu y_2)\\ &= \lambda(x_1+y_1,x_1,y_1-3x_1,y_1)+\mu(x_2+y_2,x_2,y_2-3x_2,y_2)\\ &= \lambda T(\mathbf{x}_1)+\mu T(\mathbf{x}_2)\,.\end{aligned}$$

Remarks.

- In this case we observe that the standard basis vectors of $\mathbb{R}^2$ are mapped to the vectors

$$\left.\begin{aligned}T(\mathbf{e}_1) &= T((1,0)) &= (1,1,-3,0)\\ T(\mathbf{e}_2) &= T((0,1)) &= (1,0,1,1)\end{aligned}\right\}\quad\text{which are linearly independent,}$$

  and which form a basis for $T(\mathbb{R}^2)$.

- The subspace $T(\mathbb{R}^2)=\operatorname{span}\{(1,1,-3,0),(1,0,1,1)\}$ of $\mathbb{R}^4$ is a 2-D hyper-plane through the origin.

4.3 Rank, Kernel and Nullity

Let $T : V \to W$ be a linear map. Recall that $T(V)$ is the image of V under T and that $T(V)$ is a subspace of W.

Definition. The *rank* of T is the dimension of the image, i.e.

$$r(T) = \dim T(V)\,. \tag{4.4}$$

Examples. In both example (iv) and example (v) of §4.2.1, $r(T) = 2$.

Definition. The subset of V that maps to the zero element in W is call the *kernel*, or *null space*, of T, i.e.

15/06

$$K(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0} \in W\}\,. \tag{4.5}$$

Theorem 4.1. *$K(T)$ is a subspace of V.*

Proof. Suppose that $\mathbf{u}, \mathbf{v} \in K(T)$ and $\lambda, \mu \in \mathbb{R}$ then from (4.2)

$$\begin{aligned} T(\lambda\mathbf{u} + \mu\mathbf{v}) &= \lambda T(\mathbf{u}) + \mu T(\mathbf{v}) \\ &= \lambda\mathbf{0} + \mu\mathbf{0} \\ &= \mathbf{0}\,, \end{aligned}$$

and hence $\lambda\mathbf{u} + \mu\mathbf{v} \in K(T)$. The proof then follows from invoking Theorem 3.1 on page 45. □

Remark. Since $T(\mathbf{0}) = \mathbf{0} \in W$, $\mathbf{0} \in K(T) \subseteq V$, so $K(T)$ contains at least $\mathbf{0}$.

Definition. The *nullity of* T is defined to be the dimension of the kernel, i.e.

$$n(T) = \dim\, K(T)\,. \tag{4.6}$$

Examples. In example (iv) on page 57 $n(T) = 1$, while in example (v) on page 57 $n(T) = 0$.

4.3.1 Examples

(i) Consider the map $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that

$$(x, y) \mapsto T(x, y) = (2x + 3y, 4x + 6y, -2x - 3y) = (2x + 3y)(1, 2, -1)\,.$$

Hence $T(\mathbb{R}^2)$ is the line $\mathbf{x} = \lambda(1, 2, -1) \in \mathbb{R}^3$, and so the rank of the map is such that

$$r(T) = \dim\, T(\mathbb{R}^2) = 1\,.$$

Further, $\mathbf{x} = (x, y) \in K(T)$ if $2x + 3y = 0$, so

$$K(T) = \{\mathbf{x} = (-3s, 2s) : s \in \mathbb{R}\},$$

which is a line in $\mathbb{R}^2$. Thus

$$n(T) = \dim K(T) = 1 .$$

For future reference we note that

$$r(T) + n(T) = 2 = \text{dimension of domain, i.e. } \mathbb{R}^2.$$

(ii) Next consider projection onto a line, $P : \mathbb{R}^3 \to \mathbb{R}^3$, where as in (2.17b)

$$\mathbf{x} \mapsto \mathbf{x}' = P(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{n})\mathbf{n},$$

where $\mathbf{n}$ is a fixed unit vector. Then

$$P(\mathbb{R}^3) = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{n},\ \lambda \in \mathbb{R}\}$$

which is a line in $\mathbb{R}^3$; thus $r(P) = 1$. Further, the kernel is given by

$$K(P) = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{n} = 0\},$$

which is a plane in $\mathbb{R}^3$; thus $n(P) = 2$. Again we note that

$$r(P) + n(P) = 3 = \text{dimension of domain, i.e. } \mathbb{R}^3. \tag{4.7}$$

Theorem 4.2 (The Rank-Nullity Theorem). *Let V and W be real vector spaces and let $T : V \to W$ be a linear map, then*

$$r(T) + n(T) = \dim V . \tag{4.8}$$

Proof. See *Linear Algebra* in Part IB. □

4.4 Composition of Maps

Suppose that $S : U \to V$ and $T : V \to W$ are linear maps such that

$$\mathbf{u} \mapsto \mathbf{v} = S(\mathbf{u}), \quad \mathbf{v} \mapsto \mathbf{w} = T(\mathbf{v}). \tag{4.9}$$

Definition. The *composite* or *product* map TS is the map $TS : U \to W$ such that

$$\mathbf{u} \mapsto \mathbf{w} = T(S(\mathbf{u})),$$

where we note that S acts first, then T. For the map to be well-defined the domain of T must include the image of S.

16/03

4.4.1 Examples

(i) Let $\mathcal{H}_\Pi$ be reflection in a plane

$$\mathcal{H}_\Pi : \mathbb{R}^3 \to \mathbb{R}^3 .$$

Since the range $\subseteq$ domain, we may apply map twice. Then by geometry (or exercise and algebra) it follows that

$$\mathcal{H}_\Pi \mathcal{H}_\Pi = \mathcal{H}_\Pi^2 = I , \tag{4.10}$$

where I is the identity map.

(ii) Let P be projection onto a line

$$P : \mathbb{R}^3 \to \mathbb{R}^3 .$$

Again we may apply the map twice, and in this case we can show (by geometry or algebra) that

$$P\,P = P^2 = P . \tag{4.11}$$

(iii) Let $S : \mathbb{R}^2 \to \mathbb{R}^3$ and $T : \mathbb{R}^3 \to \mathbb{R}$ be linear maps such that

$$(u, v) \quad \mapsto \quad S(u, v) = (-v, u, u + v)$$

and

$$(x, y, z) \quad \mapsto \quad T(x, y, z) = x + y + z ,$$

respectively. Then $TS : \mathbb{R}^2 \to \mathbb{R}$ is such that

$$(u, v) \mapsto T(-v, u, u + v) = 2u .$$

Remark. ST not well defined because the range of T is not the domain of S.

16/02

4.5 Bases and the Matrix Description of Maps

Let $\mathbf{e}_i$, $i = 1, 2, 3$, be a basis for $\mathbb{R}^3$ (it may help to think of it as the standard orthonormal basis $\mathbf{e}_1 = (1, 0, 0)$, etc.). Using Theorem 3.6 on page 49 we note that any $\mathbf{x} \in \mathbb{R}^3$ has a unique expansion in terms of this basis, namely

$$\mathbf{x} = \sum_{j=1}^{3} x_j \mathbf{e}_j ,$$

where the x_j are the components of $\mathbf{x}$ with respect to the given basis.

Let $\mathcal{M} : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map (where for the time being we take the domain and range to be the same). From the definition of a linear map, (4.2), it follows that

$$\mathcal{M}(\mathbf{x}) = \mathcal{M}\left(\sum_{j=1}^{3} x_j \mathbf{e}_j\right) = \sum_{j=1}^{3} x_j \mathcal{M}(\mathbf{e}_j) . \tag{4.12}$$

Now consider the action of $\mathcal{M}$ on a basis vector, say $\mathbf{e}_j$. Then

$$\mathcal{M}(\mathbf{e}_j) \equiv \mathbf{e}'_j = \sum_{i=1}^{3} (\mathbf{e}'_j)_i \, \mathbf{e}_i = \sum_{i=1}^{3} M_{ij} \, \mathbf{e}_i , \quad j = 1, 2, 3, \tag{4.13}$$

where $\mathbf{e}'_j$ is the image of $\mathbf{e}_j$, $(\mathbf{e}'_j)_i$ is the i^{th} component of $\mathbf{e}'_j$ with respect to the basis $\{\mathbf{e}_i\}$ ($i = 1, 2, 3$, $j = 1, 2, 3$), and we have set

$$M_{ij} = (\mathbf{e}'_j)_i = (\mathcal{M}(\mathbf{e}_j))_i . \tag{4.14}$$

It follows that for general $\mathbf{x} \in \mathbb{R}^3$

$$\begin{aligned} \mathbf{x} \mapsto \mathbf{x}' = \mathcal{M}(\mathbf{x}) \quad &= \quad \sum_{j=1}^{3} x_j \mathcal{M}(\mathbf{e}_j) \\ &= \quad \sum_{j=1}^{3} x_j \sum_{i=1}^{3} M_{ij} \mathbf{e}_i \\ &= \quad \sum_{i=1}^{3} \left(\sum_{j=1}^{3} M_{ij} x_j\right) \mathbf{e}_i . \end{aligned} \tag{4.15}$$

Thus in component form

$$\mathbf{x}' = \sum_{i=1}^{3} x_i' \mathbf{e}_i \quad \text{where} \quad x_i' = \sum_{j=1}^{3} M_{ij} x_j \,, \quad i = 1,2,3\,. \tag{4.16a}$$

Alternatively, in terms of the suffix notation and summation convention introduced earlier

$$x_i' = M_{ij} x_j \,. \tag{4.16b}$$

Since $\mathbf{x}$ was an arbitrary vector, what this means is that once we know the M_{ij} we can calculate the results of the mapping $\mathcal{M} : \mathbb{R}^3 \to \mathbb{R}^3$ for all elements of $\mathbb{R}^3$. In other words, the mapping $\mathcal{M} : \mathbb{R}^3 \to \mathbb{R}^3$ is, once the standard basis (or any other basis) has been chosen, completely specified by the 9 quantities M_{ij}, $i = 1,2,3$, $j = 1,2,3$.

The explicit relation between x_i', $i = 1,2,3$, and x_j, $j = 1,2,3$ is

$$\begin{aligned} x_1' &= M_{11}x_1 + M_{12}x_2 + M_{13}x_3\,, \\ x_2' &= M_{21}x_1 + M_{22}x_2 + M_{23}x_3\,, \\ x_3' &= M_{31}x_1 + M_{32}x_2 + M_{33}x_3\,. \end{aligned}$$

4.5.1 Matrix notation

The above equations can be written in a more convenient form by using matrix notation. Let x and x' be the *column matrices*, or *column vectors*,

$$\mathsf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathsf{x}' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \tag{4.17a}$$

respectively, and let M be the 3×3 *square matrix*

$$\mathsf{M} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}\,. \tag{4.17b}$$

Remarks.

- We call the M_{ij}, $i, j = 1,2,3$ the *elements* of the matrix M.
- Sometimes we write (a) $\mathsf{M} = \{M_{ij}\}$ and (b) $M_{ij} = (\mathsf{M})_{ij}$.
- The first suffix i is the *row* number, while the second suffix j is the *column* number.
- We now have **bold x** denoting a vector, *italic* x_i denoting a component of a vector, and sans serif x denoting a column matrix of components.
- To try and avoid confusion we have introduced for a short while a specific notation for a column matrix, i.e. x. However, in the case of a column matrix of vector components, i.e. a column vector, an accepted convention is to use the standard notation for a vector, i.e. $\mathbf{x}$. Hence we now have

  $$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (x_1, x_2, x_3)\,,$$

  where we draw attention to the commas on the RHS.

Equation (4.16a), or equivalently (4.16b), can now be expressed in matrix notation as

$$\mathsf{x}' = \mathsf{M}\mathsf{x} \qquad \text{or equivalently} \qquad \mathbf{x}' = \mathsf{M}\mathbf{x}\,, \tag{4.18a}$$

where a matrix multiplication rule has been defined in terms of matrix elements as

$$x_i' = \sum_{j=1}^{3} M_{ij} x_j \,. \tag{4.18b}$$

Remarks.

(i) For $i = 1, 2, 3$ let $\mathbf{r}_i$ be the vector with components equal to the elements of the ith row of M, i.e.

$$\mathbf{r}_i = (M_{i1}, M_{i2}, M_{i3}), \quad \text{for } i = 1, 2, 3.$$

Then in terms of the scalar product for vectors

$$i^{\text{th}} \text{ row of } \mathsf{x}' = x'_i = \mathbf{r}_i \cdot \mathbf{x}, \quad \text{for } i = 1, 2, 3.$$

(ii) From (4.14)

$$\mathsf{M} = \begin{pmatrix} (\mathbf{e}'_1)_1 & (\mathbf{e}'_2)_1 & (\mathbf{e}'_3)_1 \\ (\mathbf{e}'_1)_2 & (\mathbf{e}'_2)_2 & (\mathbf{e}'_3)_2 \\ (\mathbf{e}'_1)_3 & (\mathbf{e}'_2)_3 & (\mathbf{e}'_3)_3 \end{pmatrix} = \begin{pmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{pmatrix}, \tag{4.19}$$

where the $\mathbf{e}'_i$ on the RHS are to be interpreted as column vectors.

(iii) The elements of M depend on the choice of basis. Hence when specifying a matrix M associated with a map $\mathcal{M}$, it is necessary to give the basis with respect to which it has been constructed.

4.5.2 Examples (including some important new definitions of maps)

(i) *Reflection.* Consider reflection in the plane $\Pi = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{n} = 0 \text{ and } |\mathbf{n}| = 1\}$. From (2.91) we have that

$$\mathbf{x} \mapsto \mathcal{H}_\Pi(\mathbf{x}) = \mathbf{x}' = \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{n})\mathbf{n}.$$

We wish to construct matrix H, with respect to the standard basis, that represents $\mathcal{H}_\Pi$. To this end consider the action of $\mathcal{H}_\Pi$ on each member of the standard basis. Then, recalling that $\mathbf{e}_j \cdot \mathbf{n} = n_j$, it follows that

$$\mathcal{H}_\Pi(\mathbf{e}_1) = \mathbf{e}'_1 = \mathbf{e}_1 - 2n_1\mathbf{n} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 2n_1 \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} 1 - 2n_1^2 \\ -2n_1n_2 \\ -2n_1n_3 \end{pmatrix}.$$

This is the first column of H. Similarly we obtain

$$\mathsf{H} = \begin{pmatrix} 1 - 2n_1^2 & -2n_1n_2 & -2n_1n_3 \\ -2n_1n_2 & 1 - 2n_2^2 & -2n_2n_3 \\ -2n_1n_3 & -2n_2n_3 & 1 - 2n_3^2 \end{pmatrix}. \tag{4.20a}$$

Alternatively we can obtain this same result using suffix notation since from (2.91) 16/06

$$\begin{aligned} (\mathcal{H}_\Pi(\mathbf{x}))_i &= x_i - 2x_jn_jn_i \\ &= \delta_{ij}x_j - 2x_jn_jn_i \\ &= (\delta_{ij} - 2n_in_j)x_j \\ &\equiv H_{ij}x_j. \end{aligned}$$

Hence

$$(\mathsf{H})_{ij} = H_{ij} = \delta_{ij} - 2n_in_j, \quad i, j = 1, 2, 3. \tag{4.20b}$$

(ii) Consider the map $\mathcal{P}_\mathbf{b} : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\mathbf{x} \mapsto \mathbf{x}' = \mathcal{P}_\mathbf{b}(\mathbf{x}) = \mathbf{b} \times \mathbf{x}. \tag{4.21}$$

In order to construct the map's matrix P with respect to the standard basis, first note that

$$\mathcal{P}_\mathbf{b}(\mathbf{e}_1) = (b_1, b_2, b_3) \times (1, 0, 0) = (0, b_3, -b_2).$$

Now use formula (4.19) and similar expressions for $\mathcal{P}_\mathbf{b}(\mathbf{e}_2)$ and $\mathcal{P}_\mathbf{b}(\mathbf{e}_3)$ to deduce that

$$\mathsf{P}_\mathbf{b} = \begin{pmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{pmatrix}. \tag{4.22a}$$

The elements $\{P_{ij}\}$ of $\mathsf{P}_\mathbf{b}$ could also be derived as follows. From (4.16b) and (4.21)

$$P_{ij}x_j = x'_i = \varepsilon_{ijk}b_j x_k = (\varepsilon_{ikj}b_k)x_j\,.$$

Hence, in agreement with (4.22a),

$$P_{ij} = \varepsilon_{ikj}b_k = -\varepsilon_{ijk}b_k\,. \tag{4.22b}$$

(iii) *Rotation.* Consider rotation by an angle θ about the x_3 axis.

Under such a rotation

$$\begin{aligned}\mathbf{e}_1 \mapsto \mathbf{e}'_1 &= \mathbf{e}_1\cos\theta + \mathbf{e}_2\sin\theta\,,\\ \mathbf{e}_2 \mapsto \mathbf{e}'_2 &= -\mathbf{e}_1\sin\theta + \mathbf{e}_2\cos\theta\,,\\ \mathbf{e}_3 \mapsto \mathbf{e}'_3 &= \mathbf{e}_3\,.\end{aligned}$$

Thus the rotation matrix, $\mathsf{R}(\theta)$, is given by

$$\mathsf{R}(\theta) = \begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}. \tag{4.23}$$

(iv) *Dilatation.* Consider the mapping $\mathbb{R}^3 \to \mathbb{R}^3$ defined by $\mathbf{x} \mapsto \mathbf{x}'$ where

$$x'_1 = \lambda x_1\,,\quad x'_2 = \mu x_2\,,\quad x'_3 = \nu x_3 \quad \text{where } \lambda,\mu,\nu\in\mathbb{R} \text{ and } \lambda,\mu,\nu>0.$$

Then

$$\mathbf{e}'_1 = \lambda\mathbf{e}_1\,,\quad \mathbf{e}'_2 = \mu\mathbf{e}_2\,,\quad \mathbf{e}'_3 = \nu\mathbf{e}_3\,,$$

and so the map's matrix with respect to the standard basis, say D, is given by

$$\mathsf{D} = \begin{pmatrix}\lambda & 0 & 0\\ 0 & \mu & 0\\ 0 & 0 & \nu\end{pmatrix}. \tag{4.24}$$

The effect on the unit cube $0\leqslant x_1\leqslant 1$, $0\leqslant x_2\leqslant 1$, $0\leqslant x_3\leqslant 1$, of this map is to send it to $0\leqslant x'_1\leqslant\lambda$, $0\leqslant x'_2\leqslant\mu$, $0\leqslant x'_3\leqslant\nu$, i.e. to a cuboid that has been stretched or contracted by different factors along the different Cartesian axes. If $\lambda=\mu=\nu$ then the transformation is called a pure dilatation. 17/03

(v) *Shear.* A simple shear is a transformation in the plane, e.g. the x_1x_2-plane, that displaces points in one direction, e.g. the x_1 direction, by an amount proportional to the distance in that plane from, say, the x_1-axis. Under this transformation

$$\mathbf{e}_1 \mapsto \mathbf{e}'_1 = \mathbf{e}_1\,,\quad \mathbf{e}_2 \mapsto \mathbf{e}'_2 = \mathbf{e}_2 + \lambda\mathbf{e}_1\,,\quad \mathbf{e}_3 \mapsto \mathbf{e}'_3 = \mathbf{e}_3\,,\quad \text{where } \lambda\in\mathbb{R}. \tag{4.25}$$

For this example the map's matrix (with respect to the standard basis), say S_λ, is given by

$$\mathsf{S}_\lambda = \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{4.26}$$

17/02

4.5.3 dim(**domain**) $\neq$ dim(**range**)

So far we have considered matrix representations for maps where the domain and the range are the same. For instance, we have found that a map $\mathcal{M} : \mathbb{R}^3 \to \mathbb{R}^3$ leads to a 3×3 matrix M. Consider now a linear map $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}^m$ where $m, n \in \mathbb{Z}^+$, i.e. a map $\mathbf{x} \mapsto \mathbf{x}' = \mathcal{N}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x}' \in \mathbb{R}^m$.

Let $\{\mathbf{e}_k\}$ be a basis of $\mathbb{R}^n$, so

$$\mathbf{x} \;=\; \sum_{k=1}^{n} x_k \mathbf{e}_k \,, \tag{4.27a}$$

and

$$\mathcal{N}(\mathbf{x}) \;=\; \sum_{k=1}^{n} x_k \, \mathcal{N}(\mathbf{e}_k) \,. \tag{4.27b}$$

Let $\{\mathbf{f}_j\}$ be a basis of $\mathbb{R}^m$, then there exist $N_{jk} \in \mathbb{R}$ $(j = 1, \ldots, m, k = 1, \ldots, n)$ such that

$$\mathcal{N}(\mathbf{e}_k) \;=\; \sum_{j=1}^{m} N_{jk} \, \mathbf{f}_j \,. \tag{4.28a}$$

Hence

$$\begin{aligned} \mathbf{x}' \;=\; \mathcal{N}(\mathbf{x}) \;&=\; \sum_{k=1}^{n} x_k \left(\sum_{j=1}^{m} N_{jk} \, \mathbf{f}_j \right) \\ &=\; \sum_{j=1}^{m} \left(\sum_{k=1}^{n} N_{jk} \, x_k \right) \mathbf{f}_j \,, \end{aligned} \tag{4.28b}$$

and thus

$$x'_j \;=\; (\mathcal{N}(\mathbf{x}))_j \;=\; \sum_{k=1}^{n} N_{jk} \, x_k \;=\; N_{jk} \, x_k \quad \text{(s.c.)} \,. \tag{4.28c}$$

Using the same rules of multiplication as before we write

$$\underbrace{\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{pmatrix}}_{\substack{m \times 1 \text{ matrix} \\ \text{(column vector} \\ \text{with } m \text{ rows)}}} = \underbrace{\begin{pmatrix} N_{11} & N_{12} & \ldots & N_{1n} \\ N_{21} & N_{22} & \ldots & N_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ N_{m1} & N_{m2} & \ldots & N_{mn} \end{pmatrix}}_{\substack{m \times n \text{ matrix} \\ (m \text{ rows, } n \text{ columns})}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\substack{n \times 1 \text{ matrix} \\ \text{(column vector} \\ \text{with } n \text{ rows)}}} \tag{4.29a}$$

i.e.

$$\mathbf{x}' \;=\; \mathsf{N}\mathbf{x} \quad \text{where} \quad \mathsf{N} = \{N_{ij}\} \,. \tag{4.29b}$$

4.6 Algebra of Matrices

4.6.1 Multiplication by a scalar

Let $\mathcal{A}:\mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then for given $\lambda \in \mathbb{R}$ define $(\lambda\mathcal{A})$ such that

$$(\lambda\mathcal{A})(\mathbf{x}) = \lambda(\mathcal{A}(\mathbf{x}))\,. \tag{4.30}$$

This is also a linear map. Let $\mathsf{A} = \{a_{ij}\}$ be the matrix of $\mathcal{A}$ (with respect to given bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, or a given basis of $\mathbb{R}^n$ if $m = n$). Then from (4.28b)

$$\begin{aligned}(\lambda\mathcal{A})(\mathbf{x}) \;=\; \lambda\mathcal{A}(\mathbf{x}) \;&=\; \lambda\left(\sum_{j=1}^{m}\sum_{k=1}^{n} a_{jk}x_k\mathbf{f}_j\right)\\ &=\; \sum_{j=1}^{m}\sum_{k=1}^{n}(\lambda a_{jk})x_k\mathbf{f}_j\,.\end{aligned}$$

Hence, for consistency the matrix of $\lambda\mathcal{A}$ must be

$$\lambda\mathsf{A} = \{\lambda a_{ij}\}\,, \tag{4.31}$$

which we use as the definition of a matrix multiplied by a scalar.

4.6.2 Addition

Similarly, if $\mathsf{A} = \{a_{ij}\}$ and $\mathsf{B} = \{b_{ij}\}$ are both $m \times n$ matrices associated with maps $\mathbb{R}^n \to \mathbb{R}^m$, then for consistency we define

$$\mathsf{A} + \mathsf{B} = \{a_{ij} + b_{ij}\}\,. \tag{4.32}$$

4.6.3 Matrix multiplication

Let $\mathcal{S}:\mathbb{R}^n \to \mathbb{R}^m$ and $\mathcal{T}:\mathbb{R}^m \to \mathbb{R}^\ell$ be linear maps. Let

$\mathsf{S} = \{S_{ij}\}$ be the $m \times n$ matrix of $\mathcal{S}$ with respect to a given basis,
$\mathsf{T} = \{T_{ij}\}$ be the $\ell \times m$ matrix of $\mathcal{T}$ with respect to a given basis.

Now consider the composite map $\mathcal{W} = \mathcal{T}\mathcal{S}:\mathbb{R}^n \to \mathbb{R}^\ell$, with associated $\ell \times n$ matrix $\mathsf{W} = \{W_{ij}\}$. If

$$\mathbf{x}' = \mathcal{S}(\mathbf{x}) \quad\text{and}\quad \mathbf{x}'' = \mathcal{T}(\mathbf{x}')\,, \tag{4.33a}$$

then from (4.28c), and using the summation convention,

$$x'_j = S_{jk}x_k \quad\text{and}\quad x''_i = T_{ij}x'_j \quad\text{(s.c.)}, \tag{4.33b}$$

and thus

$$x''_i = T_{ij}(S_{jk}x_k) = (T_{ij}S_{jk})x_k \quad\text{(s.c.)}. \tag{4.34a}$$

However,

$$\mathbf{x}'' = \mathcal{T}\mathcal{S}(\mathbf{x}) = \mathcal{W}(\mathbf{x}) \quad\text{and}\quad x''_i = W_{ik}x_k \quad\text{(s.c.)}, \tag{4.34b}$$

and hence because (4.34a) and (4.34b) must identical for arbitrary $\mathbf{x}$, it follows that

$$W_{ik} = T_{ij}S_{jk} \quad\text{(s.c.)}\,; \tag{4.35}$$

We interpret (4.35) as defining the elements of the matrix product TS. In words,

the ik^{th} element of TS is equal to the scalar product of the i^{th} row of T with the k^{th} column of S.

Remarks.

(i) For the scalar product to be well defined, the number of columns of T must equal the number of rows of S; this is the case above since T is a $\ell \times m$ matrix, while S is a $m \times n$ matrix.

(ii) The above definition of matrix multiplication is consistent with the $n = 1$ special case considered in (4.18a) and (4.18b), i.e. the special case when S is a column matrix (or column vector).

(iii) If A is a $p \times q$ matrix, and B is a $r \times s$ matrix, then

AB exists only if $q = r$, and is then a $p \times s$ matrix;
BA exists only if $s = p$, and is then a $r \times q$ matrix.

For instance

$$\begin{pmatrix} p & q & r \\ s & t & u \end{pmatrix} \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} = \begin{pmatrix} pa + qc + re & pb + qd + rf \\ sa + tc + ue & sb + td + uf \end{pmatrix},$$

while

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} p & q & r \\ s & t & u \end{pmatrix} = \begin{pmatrix} ap + bs & aq + bt & ar + bu \\ cp + ds & cq + dt & cr + du \\ ep + fs & eq + ft & er + fu \end{pmatrix}.$$

(iv) Even if $p = q = r = s$, so that both AB and BA exist and have the same number of rows and columns,

$$\mathsf{AB} \neq \mathsf{BA} \quad \text{in general.} \tag{4.36}$$

For instance

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

while

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

Lemma 4.3. *The multiplication of matrices is associative, i.e. if* $\mathsf{A} = \{a_{ij}\}$, $\mathsf{B} = \{b_{ij}\}$ *and* $\mathsf{C} = \{c_{ij}\}$ *are matrices such that* AB *and* BC *exist, then*

$$\mathsf{A}(\mathsf{BC}) = (\mathsf{AB})\mathsf{C}. \tag{4.37}$$

Proof. In terms of suffix notation (and the summation convention)

$$\begin{aligned} (\mathsf{A}(\mathsf{BC}))_{ij} &= a_{ik}(\mathsf{BC})_{kj} = a_{ik}b_{k\ell}c_{\ell j} = a_{i£}b_{£¥}c_{¥j}\,, \\ ((\mathsf{AB})\mathsf{C})_{ij} &= (\mathsf{AB})_{ik}c_{kj} = a_{i\ell}b_{\ell k}c_{kj} = a_{i£}b_{£¥}c_{¥j}\,. \end{aligned}$$

□

18/03

4.6.4 Transpose

Definition. If $\mathsf{A} = \{a_{ij}\}$ is a $m \times n$ matrix, then its *transpose* A^{T} is defined to be a $n \times m$ matrix with elements

$$(\mathsf{A}^{\mathrm{T}})_{ij} = (\mathsf{A})_{ji} = a_{ji}\,. \tag{4.38a}$$

Remark.

$$(\mathsf{A}^{\mathrm{T}})^{\mathrm{T}} = \mathsf{A}\,. \tag{4.38b}$$

Examples.

(i)

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} .$$

(ii)

$$\text{If } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ is a column vector, } \mathbf{x}^{\mathrm{T}} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \text{ is a row vector.}$$

Remark. Recall that commas are sometimes important:

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2, \dots, x_n) , \\ \mathbf{x}^{\mathrm{T}} &= \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} . \end{aligned}$$

Lemma 4.4. *If* $\mathsf{A} = \{a_{ij}\}$ *and* $\mathsf{B} = \{b_{ij}\}$ *are matrices such that* AB *exists, then*

$$(\mathsf{AB})^{\mathrm{T}} = \mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}} . \tag{4.39}$$

Proof.

$$\begin{aligned} ((\mathsf{AB})^{\mathrm{T}})_{ij} &= (\mathsf{AB})_{ji} \\ &= a_{jk}b_{ki} \\ &= (\mathsf{B})_{ki}(\mathsf{A})_{jk} \\ &= (\mathsf{B}^{\mathrm{T}})_{ik}(\mathsf{A}^{\mathrm{T}})_{kj} \\ &= (\mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}})_{ij} . \end{aligned}$$

□

Example. Let $\mathbf{x}$ and $\mathbf{y}$ be 3×1 column vectors, and let $\mathsf{A} = \{a_{ij}\}$ be a 3×3 matrix. Then

$$\mathbf{x}^{\mathrm{T}}\mathsf{A}\mathbf{y} = x_i a_{ij} y_j = x_{£} a_{£¥} y_{¥} .$$

is a 1×1 matrix, i.e. a scalar. It follows that

$$(\mathbf{x}^{\mathrm{T}}\mathsf{A}\mathbf{y})^{\mathrm{T}} = \mathbf{y}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}}\mathbf{x} = y_i a_{ji} x_j = y_{¥} a_{£¥} x_{£} = \mathbf{x}^{\mathrm{T}}\mathsf{A}\mathbf{y} .$$

4.6.5 Symmetry

Definition. A square $n \times n$ matrix $\mathsf{A} = \{a_{ij}\}$ is *symmetric* if

18/02

$$\mathsf{A}^{\mathrm{T}} = \mathsf{A} , \quad \text{i.e.} \quad a_{ji} = a_{ij} . \tag{4.40}$$

Definition. A square $n \times n$ matrix $\mathsf{A} = \{a_{ij}\}$ is *antisymmetric* if

$$\mathsf{A}^{\mathrm{T}} = -\mathsf{A} , \quad \text{i.e.} \quad a_{ji} = -a_{ij} . \tag{4.41a}$$

Remark. For an antisymmetric matrix, $a_{11} = -a_{11}$, i.e. $a_{11} = 0$. Similarly we deduce that all the *diagonal* elements of an antisymmetric matrix are zero, i.e.

$$a_{11} = a_{22} = \dots = a_{nn} = 0 . \tag{4.41b}$$

Examples.

(i) A symmetric 3×3 matrix S has the form

$$\mathsf{S} = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix},$$

i.e. it has six independent elements.

(ii) An antisymmetric 3×3 matrix A has the form

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

i.e. it has three independent elements.

Remark. Let $a = v_3$, $b = v_2$ and $c = v_1$, then (cf. (4.22a) and (4.22b))

$$\mathsf{A} = \{a_{ij}\} = \{\varepsilon_{ijk} v_k\}. \tag{4.42}$$

Thus each antisymmetric 3×3 matrix corresponds to a unique vector $\mathbf{v}$ in $\mathbb{R}^3$.

4.6.6 Trace

Definition. The *trace* of a square $n \times n$ matrix $\mathsf{A} = \{a_{ij}\}$ is equal to the sum of the diagonal elements, i.e.

$$Tr(\mathsf{A}) = a_{ii} \quad \text{(s.c.)}. \tag{4.43}$$

Remark. Let $\mathsf{B} = \{b_{ij}\}$ be a $m \times n$ matrix and $\mathsf{C} = \{c_{ij}\}$ be a $n \times m$ matrix, then BC and CB both exist, but are not usually equal (even if $m = n$). However

$$\begin{aligned} Tr(\mathsf{BC}) &= (\mathsf{BC})_{ii} = b_{ij} c_{ji}, \\ Tr(\mathsf{CB}) &= (\mathsf{CB})_{ii} = c_{ij} b_{ji} = b_{ij} c_{ji}, \end{aligned}$$

and hence $Tr(\mathsf{BC}) = Tr(\mathsf{CB})$ (even if $m \neq n$ so that the matrices are of different sizes).

4.6.7 The unit or identity matrix

Definition. The *unit* or *identity* $n \times n$ matrix is defined to be

$$\mathsf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \tag{4.44}$$

i.e. all the elements are 0 except for the diagonal elements that are 1.

Example. The 3×3 identity matrix is given by

$$\mathsf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \{\delta_{ij}\}. \tag{4.45}$$

Property. Define the Kronecker delta in $\mathbb{R}^n$ such that

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \text{for } i, j = 1, 2, \ldots, n. \tag{4.46}$$

Let $\mathsf{A} = \{a_{ij}\}$ be a $n \times n$ matrix, then

$$\begin{aligned} (\mathsf{IA})_{ij} &= \delta_{ik} a_{kj} = a_{ij}\,, \\ (\mathsf{AI})_{ij} &= a_{ik} \delta_{kj} = a_{ij}\,, \end{aligned}$$

i.e.

$$\mathsf{IA} = \mathsf{AI} = \mathsf{A}\,. \tag{4.47}$$

4.6.8 The inverse of a matrix

Definition. Let A be a square $n \times n$ matrix. B is a *left inverse* of A if $\mathsf{BA} = \mathsf{I}$. C is a *right inverse* of A if $\mathsf{AC} = \mathsf{I}$.

Lemma 4.5. *If* B *is a left inverse of* A *and* C *is a right inverse of* A *then* $\mathsf{B} = \mathsf{C}$ *and we write* $\mathsf{B} = \mathsf{C} = \mathsf{A}^{-1}$.

Proof. From (4.37), (4.47), $\mathsf{BA} = \mathsf{I}$ and $\mathsf{AC} = \mathsf{I}$ it follows that

$$\mathsf{B} = \mathsf{BI} = \mathsf{B}(\mathsf{AC}) = (\mathsf{BA})\mathsf{C} = \mathsf{IC} = \mathsf{C}\,.$$

□

Remark. The lemma is based on the premise that both a left inverse and right inverse exist. In general, the existence of a left inverse does not necessarily guarantee the existence of a right inverse, or vice versa. However, in the case of a square matrix, the existence of a left inverse does imply the existence of a right inverse, and vice versa (see Part IB *Linear Algebra* for a general proof). The above lemma then implies that they are the same matrix.

Definition. Let A be a $n \times n$ matrix. A is said to be *invertible* if there exists a $n \times n$ matrix B such that

$$\mathsf{BA} = \mathsf{AB} = \mathsf{I}\,. \tag{4.48}$$

The matrix B is called the *inverse* of A, is unique (see above) and is denoted by A^{-1} (see above).

Property. From (4.48) it follows that $\mathsf{A} = \mathsf{B}^{-1}$ (in addition to $\mathsf{B} = \mathsf{A}^{-1}$). Hence

$$\mathsf{A} = (\mathsf{A}^{-1})^{-1}\,. \tag{4.49}$$

Lemma 4.6. *Suppose that* A *and* B *are both invertible* $n \times n$ *matrices. Then*

$$(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}\,. \tag{4.50}$$

Proof. From using (4.37), (4.47) and (4.48) it follows that

$$\begin{aligned} \mathsf{B}^{-1}\mathsf{A}^{-1}(\mathsf{AB}) &= \mathsf{B}^{-1}(\mathsf{A}^{-1}\mathsf{A})\mathsf{B} = \mathsf{B}^{-1}\mathsf{IB} = \mathsf{B}^{-1}\mathsf{B} = \mathsf{I}\,, \\ (\mathsf{AB})\mathsf{B}^{-1}\mathsf{A}^{-1} &= \mathsf{A}(\mathsf{BB}^{-1})\mathsf{A}^{-1} = \mathsf{AIA}^{-1} = \mathsf{AA}^{-1} = \mathsf{I}\,. \end{aligned}$$

□

4.6.9 Determinants (for 3×3 and 2×2 matrices)

Recall that the signed volume of the $\mathbb{R}^3$ parallelepiped defined by $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ (positive if $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are right-handed, negative if left-handed).

Consider the effect of a linear map, $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$, on the volume of the unit cube defined by standard orthonormal basis vectors $\mathbf{e}_i$. Let $\mathsf{A} = \{a_{ij}\}$ be the matrix associated with $\mathcal{A}$, then the volume of the mapped cube is, with the aid of (2.67e) and (4.28c), given by

$$\begin{aligned} \mathbf{e}'_1 \cdot \mathbf{e}'_2 \times \mathbf{e}'_3 &= \varepsilon_{ijk} (\mathbf{e}'_1)_i (\mathbf{e}'_2)_j (\mathbf{e}'_3)_k \\ &= \varepsilon_{ijk} a_{i\ell} (\mathbf{e}_1)_\ell \, a_{jm} (\mathbf{e}_2)_m \, a_{kn} (\mathbf{e}_3)_n \\ &= \varepsilon_{ijk} a_{i\ell} \delta_{1\ell} \, a_{jm} \delta_{2m} \, a_{km} \delta_{3n} \\ &= \varepsilon_{ijk} a_{i1} a_{j2} a_{k3} \, . \end{aligned} \tag{4.51}$$

Definition. The *determinant* of a 3×3 matrix A is given by

$$\begin{aligned} \det \mathsf{A} &= \varepsilon_{ijk} a_{i1} a_{j2} a_{k3} && (4.52a) \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) + a_{21}(a_{32}a_{13} - a_{12}a_{33}) + a_{31}(a_{12}a_{23} - a_{22}a_{13}) && (4.52b) \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) && (4.52c) \\ &= \varepsilon_{ijk} a_{1i} a_{2j} a_{3k} && (4.52d) \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \, . && (4.52e) \end{aligned}$$

Alternative notation. Alternative notations for the determinant of the matrix A include

$$\det \mathsf{A} = |\,\mathsf{A}\,| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{vmatrix} \, . \tag{4.53}$$

Remarks.

- A linear map $\mathbb{R}^3 \to \mathbb{R}^3$ is *volume preserving* if and only if the determinant of its matrix with respect to an orthonormal basis is ± 1 (strictly 'an' should be replaced by 'any', but we need some extra machinery before we can prove that).
- If $\{\mathbf{e}_i\}$ is a standard right-handed orthonormal basis then the set $\{\mathbf{e}'_i\}$ is right-handed if $\begin{vmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{vmatrix} > 0$, and left-handed if $\begin{vmatrix} \mathbf{e}'_1 & \mathbf{e}'_2 & \mathbf{e}'_3 \end{vmatrix} < 0$.

18/06

Exercises.

(i) Show that the determinant of the rotation matrix R defined in (4.23) is $+1$.

(ii) Show that the determinant of the reflection matrix H defined in (4.20a), or equivalently (4.20b), is -1 (since reflection sends a right-handed set of vectors to a left-handed set).

2×2 *matrices.* A map $\mathbb{R}^3 \to \mathbb{R}^3$ is effectively two-dimensional if $a_{33} = 1$ and $a_{13} = a_{31} = a_{23} = a_{32} = 0$ (cf. (4.23)). Hence for a 2×2 matrix A we define the determinant to be given by (see (4.52b) or (4.52c))

$$\det \mathsf{A} = |\,\mathsf{A}\,| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \, . \tag{4.54}$$

A map $\mathbb{R}^2 \to \mathbb{R}^2$ is area preserving if $\det \mathsf{A} = \pm 1$.

An abuse of notation. This is not for the faint-hearted. If you are into the abuse of notation show that

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \, , \tag{4.55}$$

where we are treating the vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, as 'components'.

19/03

4.7 Orthogonal Matrices

Definition. An $n \times n$ matrix $\mathsf{A} = \{a_{ij}\}$ is *orthogonal* if

$$\mathsf{A}\mathsf{A}^{\mathrm{T}} = \mathsf{I} = \mathsf{A}^{\mathrm{T}}\mathsf{A}\,, \tag{4.56a}$$

i.e. if A is invertible and $\mathsf{A}^{-1} = \mathsf{A}^{\mathrm{T}}$.

Property: orthogonal rows and columns. In components (4.56a) becomes

$$(\mathsf{A})_{ik}(\mathsf{A}^{\mathrm{T}})_{kj} = a_{ik}a_{jk} = \delta_{ij}\,. \tag{4.56b}$$

Thus the scalar product of the i^{th} and j^{th} rows of A is zero unless $i = j$ in which case it is 1. This implies that the rows of A form an orthonormal set. Similarly, since $\mathsf{A}^{\mathrm{T}}\mathsf{A} = \mathsf{I}$,

$$(\mathsf{A}^{\mathrm{T}})_{ik}(\mathsf{A})_{kj} = a_{ki}a_{kj} = \delta_{ij}\,, \tag{4.56c}$$

and so the columns of A also form an orthonormal set.

Property: map of orthonormal basis. Suppose that the map $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$ has a matrix A with respect to an orthonormal basis. Then from (4.19) we recall that

$$\begin{array}{lcll} \mathbf{e}_1 & \mapsto & \mathsf{A}\mathbf{e}_1 & \text{the first column of } \mathsf{A}, \\ \mathbf{e}_2 & \mapsto & \mathsf{A}\mathbf{e}_2 & \text{the second column of } \mathsf{A}, \\ \mathbf{e}_3 & \mapsto & \mathsf{A}\mathbf{e}_3 & \text{the third column of } \mathsf{A}. \end{array}$$

Thus if A is an orthogonal matrix the $\{\mathbf{e}_i\}$ transform to an orthonormal set (which may be right-handed or left-handed depending on the sign of $\det \mathsf{A}$).

Examples.

(i) We have already seen that the application of a reflection map $\mathcal{H}_\Pi$ twice results in the identity map, i.e.

$$\mathcal{H}_\Pi^2 = I\,. \tag{4.57}$$

From (4.20b) the matrix of $\mathcal{H}_\Pi$ with respect to a standard basis is specified by

$$(\mathsf{H})_{ij} = \{\delta_{ij} - 2n_in_j\}\,.$$

From (4.57), or a little manipulation, it follows that

$$\mathsf{H}^2 = \mathsf{I}\,. \tag{4.58a}$$

Moreover H is symmetric, hence

$$\mathsf{H} = \mathsf{H}^{\mathrm{T}}\,, \quad \text{and so} \quad \mathsf{H}^2 = \mathsf{H}\mathsf{H}^{\mathrm{T}} = \mathsf{H}^{\mathrm{T}}\mathsf{H} = \mathsf{I}\,. \tag{4.58b}$$

Thus H is orthogonal.

(ii) With respect to the standard basis, rotation by an angle θ about the x_3 axis has the matrix (see (4.23))

$$\mathsf{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\,. \tag{4.59a}$$

R is orthogonal since both the rows and the columns are orthogonal vectors, and thus

19/02

$$\mathsf{R}\mathsf{R}^{\mathrm{T}} = \mathsf{R}^{\mathrm{T}}\mathsf{R} = \mathsf{I}\,. \tag{4.59b}$$

Preservation of the scalar product. Under a map represented by an orthogonal matrix with respect to an orthonormal basis, a scalar product is preserved. For suppose that, in component form, $\mathsf{x} \mapsto \mathsf{x}' = \mathsf{A}\mathsf{x}$ and $\mathsf{y} \mapsto \mathsf{y}' = \mathsf{A}\mathsf{y}$ (note the use of sans serif), then

$$\begin{aligned} \mathbf{x}' \cdot \mathbf{y}' &= \mathsf{x}'^{\mathrm{T}}\mathsf{y}' && \text{(since the basis is orthonormal)} \\ &= (\mathsf{x}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}})(\mathsf{A}\mathsf{y}) \\ &= \mathsf{x}^{\mathrm{T}}\,\mathsf{I}\,\mathsf{y} \\ &= \mathsf{x}^{\mathrm{T}}\mathsf{y} \\ &= \mathbf{x} \cdot \mathbf{y} && \text{(since the basis is orthonormal)}. \end{aligned}$$

Isometric maps. If a linear map is represented by an orthogonal matrix A with respect to an orthonormal basis, then the map is an isometry since

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

Hence $|\mathbf{x}' - \mathbf{y}'| = |\mathbf{x} - \mathbf{y}|$, i.e. lengths are preserved.

Remark. The only length preserving maps of $\mathbb{R}^3$ are translations (which are not linear maps) and reflections and rotations (which we have already seen are associated with orthogonal matrices).

4.8 Change of basis

We wish to consider a change of basis. To fix ideas it may help to think of a change of basis from the standard orthonormal basis in $\mathbb{R}^3$ to a new basis which is not necessarily orthonormal. However, we will work in $\mathbb{R}^n$ and will not assume that either basis is orthonormal (unless stated otherwise).

4.8.1 Transformation matrices

Let $\{\mathbf{e}_i : i = 1, \dots, n\}$ and $\{\widetilde{\mathbf{e}}_i : i = 1, \dots, n\}$ be two sets of basis vectors for an n-dimensional real vector space V. Since the $\{\mathbf{e}_i\}$ is a basis, the individual basis vectors of the basis $\{\widetilde{\mathbf{e}}_i\}$ can be written as

$$\widetilde{\mathbf{e}}_j = \sum_{i=1}^{n} \mathbf{e}_i\, A_{ij} \quad (j = 1, \dots, n)\,, \tag{4.60}$$

for some numbers A_{ij}, where A_{ij} is the ith component of the vector $\widetilde{\mathbf{e}}_j$ in the basis $\{\mathbf{e}_i\}$. The numbers A_{ij} can be represented by a square $n \times n$ *transformation matrix* A

$$\mathsf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} = \begin{pmatrix} \widetilde{\mathbf{e}}_1 & \widetilde{\mathbf{e}}_2 & \dots & \widetilde{\mathbf{e}}_n \end{pmatrix}\,, \tag{4.61}$$

where in the final matrix the $\widetilde{\mathbf{e}}_j$ are to be interpreted as column vectors of the components of the $\widetilde{\mathbf{e}}_j$ in the $\{\mathbf{e}_i\}$ basis (cf. (4.19)).

Similarly, since the $\{\widetilde{\mathbf{e}}_i\}$ is a basis, the individual basis vectors of the basis $\{\mathbf{e}_i\}$ can be written as

$$\mathbf{e}_i = \sum_{k=1}^{n} \widetilde{\mathbf{e}}_k B_{ki} \quad (i = 1, 2, \dots, n)\,, \tag{4.62}$$

for some numbers B_{ki}, where B_{ki} is the kth component of the vector $\mathbf{e}_i$ in the basis $\{\widetilde{\mathbf{e}}_k\}$. Again the B_{ki} can be viewed as the entries of a matrix B

$$\mathsf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{pmatrix}\,, \tag{4.63}$$

where in the final matrix the $\mathbf{e}_j$ are to be interpreted as column vectors of the components of the $\mathbf{e}_j$ in the $\{\widetilde{\mathbf{e}}_i\}$ basis.

4.8.2 Properties of transformation matrices

From substituting (4.62) into (4.60) we have that

$$\widetilde{\mathbf{e}}_j = \sum_{i=1}^{n}\left[\sum_{k=1}^{n}\widetilde{\mathbf{e}}_k B_{ki}\right] A_{ij} = \sum_{k=1}^{n}\widetilde{\mathbf{e}}_k\left[\sum_{i=1}^{n} B_{ki}A_{ij}\right]. \tag{4.64}$$

However, the set $\{\widetilde{\mathbf{e}}_j\}$ is a basis and so linearly independent. Thus, from noting that

$$\widetilde{\mathbf{e}}_j = \sum_{k=1}^{n}\widetilde{\mathbf{e}}_k\,\delta_{kj}\,,$$

or otherwise, it follows that

$$\sum_{i=1}^{n} B_{ki}A_{ij} = \delta_{kj}\,. \tag{4.65}$$

Hence in matrix notation, $\mathsf{BA} = \mathsf{I}$, where I is the identity matrix. Conversely, substituting (4.60) into (4.62) leads to the conclusion that $\mathsf{AB} = \mathsf{I}$ (alternatively argue by a relabeling symmetry). Thus

$$\mathsf{B} = \mathsf{A}^{-1}\,. \tag{4.66}$$

4.8.3 Transformation law for vector components

Consider a vector $\mathbf{v}$, then in the $\{\mathbf{e}_i\}$ basis it follows from (3.13) that

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

Similarly, in the $\{\widetilde{\mathbf{e}}_i\}$ basis we can write

$$\begin{aligned}
\mathbf{v} &= \sum_{j=1}^{n}\widetilde{v}_j\widetilde{\mathbf{e}}_j && (4.67)\\
&= \sum_{j=1}^{n}\widetilde{v}_j\sum_{i=1}^{n}\mathbf{e}_i A_{ij} && \text{from (4.60)}\\
&= \sum_{i=1}^{n}\mathbf{e}_i\sum_{j=1}^{n}(A_{ij}\widetilde{v}_j) && \text{swap summation order.}
\end{aligned}$$

Since a basis representation is unique it follows that

$$v_i = \sum_{j=1}^{n} A_{ij}\widetilde{v}_j\,, \tag{4.68a}$$

i.e. that

$$\mathsf{v} = \mathsf{A}\widetilde{\mathsf{v}}\,, \tag{4.68b}$$

where we have deliberately used a sans serif font to indicate the column matrices of components (since otherwise there is serious ambiguity). By applying A^{-1} to either side of (4.68b) it follows that

$$\widetilde{\mathsf{v}} = \mathsf{A}^{-1}\mathsf{v}\,. \tag{4.68c}$$

Equations (4.68b) and (4.68c) relate the components of $\mathbf{v}$ with respect to the $\{\mathbf{e}_i\}$ basis to the components with respect to the $\{\widetilde{\mathbf{e}}_i\}$ basis.

Worked Example. Let $\{\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1)\}$ and $\{\widetilde{\mathbf{e}}_1 = (1,1), \widetilde{\mathbf{e}}_2 = (-1,1)\}$ be two sets of basis vectors in $\mathbb{R}^2$. Find the transformation matrix A that connects them. Verify the transformation law for the components of an arbitrary vector $\mathbf{v}$ in the two coordinate systems.

Answer. We have that

$$\begin{aligned}
\widetilde{\mathbf{e}}_1 &= (\ \ 1,1) = \quad\ (1,0) + (0,1) = \ \ \mathbf{e}_1 + \mathbf{e}_2 \,, \\
\widetilde{\mathbf{e}}_2 &= (-1,1) = -1\,(1,0) + (0,1) = -\mathbf{e}_1 + \mathbf{e}_2 \,.
\end{aligned}$$

Hence from comparison with (4.60)

$$A_{11} = 1\,, \quad A_{21} = 1\,, \quad A_{12} = -1 \quad \text{and} \quad A_{22} = 1\,,$$

i.e.

$$\mathsf{A} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Similarly, since

$$\begin{aligned}
\mathbf{e}_1 &= (1,0) = \tfrac{1}{2}(\ (1,1) - (-1,1)\ ) = \tfrac{1}{2}(\widetilde{\mathbf{e}}_1 - \widetilde{\mathbf{e}}_2)\,, \\
\mathbf{e}_2 &= (0,1) = \tfrac{1}{2}(\ (1,1) + (-1,1)\ ) = \tfrac{1}{2}(\widetilde{\mathbf{e}}_1 + \widetilde{\mathbf{e}}_2)\,,
\end{aligned}$$

it follows from (4.62) that

$$\mathsf{B} = \mathsf{A}^{-1} = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Moreover $\mathsf{A}^{-1}\mathsf{A} = \mathsf{A}\mathsf{A}^{-1} = \mathsf{I}$.

Now consider an arbitrary vector $\mathbf{v}$. Then

$$\begin{aligned}
\mathbf{v} &= v_1\mathbf{e}_1 + v_2\mathbf{e}_2 \\
&= \tfrac{1}{2}v_1(\widetilde{\mathbf{e}}_1 - \widetilde{\mathbf{e}}_2) + \tfrac{1}{2}v_2(\widetilde{\mathbf{e}}_1 + \widetilde{\mathbf{e}}_2) \\
&= \tfrac{1}{2}(v_1 + v_2)\,\widetilde{\mathbf{e}}_1 - \tfrac{1}{2}(v_1 - v_2)\,\widetilde{\mathbf{e}}_2\,.
\end{aligned}$$

Thus

$$\widetilde{v}_1 = \tfrac{1}{2}(v_1 + v_2) \quad \text{and} \quad \widetilde{v}_2 = -\tfrac{1}{2}(v_1 - v_2)\,,$$

and thus from (4.68c), i.e. $\widetilde{\mathsf{v}} = \mathsf{A}^{-1}\mathsf{v}$, we deduce that (as above)

$$\mathsf{A}^{-1} = \tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

4.8.4 Transformation law for matrices representing linear maps

Now consider a linear map $\mathcal{M} : \mathbb{R}^n \to \mathbb{R}^n$ under which $\mathbf{v} \mapsto \mathbf{v}' = \mathcal{M}(\mathbf{v})$ and (in terms of column vectors)

$$\mathsf{v}' = \mathsf{M}\mathsf{v}\,,$$

where v' and v are the component column matrices of $\mathbf{v}'$ and $\mathbf{v}$, respectively, with respect to the basis $\{\mathbf{e}_i\}$, and M is the matrix of $\mathcal{M}$ with respect to this basis.

Let $\widetilde{\mathsf{v}}'$ and $\widetilde{\mathsf{v}}$ be the component column matrices of $\mathbf{v}'$ and $\mathbf{v}$ with respect to the alternative basis $\{\widetilde{\mathbf{e}}_i\}$. Then it follows from from (4.68b) that

$$\mathsf{A}\widetilde{\mathsf{v}}' = \mathsf{M}\mathsf{A}\widetilde{\mathsf{v}}\,, \quad \text{i.e.} \quad \widetilde{\mathsf{v}}' = (\mathsf{A}^{-1}\mathsf{M}\mathsf{A})\,\widetilde{\mathsf{v}}\,.$$

We deduce that the matrix of $\mathcal{M}$ with respect to the alternative basis $\{\widetilde{\mathbf{e}}_i\}$ is given by

$$\widetilde{\mathsf{M}} = \mathsf{A}^{-1}\mathsf{M}\mathsf{A}\,. \tag{4.69}$$

Maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ (unlectured). A similar approach may be used to deduce the matrix of the map $\mathcal{N} : \mathbb{R}^n \to \mathbb{R}^m$ (where $m \neq n$) with respect to new bases of both $\mathbb{R}^n$ and $\mathbb{R}^m$.

Suppose $\{\mathbf{e}_i\}$ is a basis of $\mathbb{R}^n$, $\{\mathbf{f}_i\}$ is a basis of $\mathbb{R}^m$, and N is matrix of $\mathcal{N}$ with respect to these two bases. then from (4.29b)

$$\mathsf{v} \mapsto \mathsf{v}' = \mathsf{N}\,\mathsf{v}$$

where v and v' are component column matrices of $\mathbf{v}$ and $\mathbf{v}'$ with respect to bases $\{\mathbf{e}_i\}$ and $\{\mathbf{f}_i\}$ respectively.

Now consider new bases $\{\widetilde{\mathbf{e}}_i\}$ of $\mathbb{R}^n$ and $\{\widetilde{\mathbf{f}}_i\}$ of $\mathbb{R}^m$, and let

$$\mathsf{A} = \begin{pmatrix} \widetilde{\mathbf{e}}_1 & \dots & \widetilde{\mathbf{e}}_n \end{pmatrix} \quad \text{and} \quad \mathsf{C} = \begin{pmatrix} \widetilde{\mathbf{f}}_1 & \dots & \widetilde{\mathbf{f}}_m \end{pmatrix}.$$

where A is a $n \times n$ matrix of components (see (4.61)) and C is a $m \times m$ matrix of components. Then

$$\mathsf{v} = \mathsf{A}\widetilde{\mathsf{v}}\,, \quad \text{and} \quad \mathsf{v}' = \mathsf{C}\widetilde{\mathsf{v}}'\,,$$

where $\widetilde{\mathsf{v}}$ and $\widetilde{\mathsf{v}}'$ are component column matrices of $\mathbf{v}$ and $\mathbf{v}'$ with respect to bases $\{\widetilde{\mathbf{e}}_i\}$ and $\{\widetilde{\mathbf{f}}_i\}$ respectively. Hence

$$\mathsf{C}\widetilde{\mathsf{v}}' = \mathsf{N}\mathsf{A}\widetilde{\mathsf{v}}\,, \quad \text{and so} \quad \widetilde{\mathsf{v}}' = \mathsf{C}^{-1}\mathsf{N}\mathsf{A}\widetilde{\mathsf{v}}\,.$$

It follows that $\mathsf{C}^{-1}\mathsf{N}\mathsf{A}$ is map's matrix with respect to the new bases $\{\widetilde{\mathbf{e}}_i\}$ and $\{\widetilde{\mathbf{f}}_i\}$, i.e.

$$\widetilde{\mathsf{N}} = \mathsf{C}^{-1}\mathsf{N}\mathsf{A}\,. \tag{4.70}$$

19/06

Example. Consider a simple shear with magnitude γ in the x_1 direction within the (x_1, x_2) plane. Then from (4.26) the matrix of this map with respect to the standard basis $\{\mathbf{e}_i\}$ is

$$\mathsf{S}_\gamma = \begin{pmatrix} 1 & \gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let $\{\widetilde{\mathbf{e}}\}$ be the basis obtained by rotating the standard basis by an angle θ about the x_3 axis. Then

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

and thus

$$\mathsf{A} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have already deduced that rotation matrices are orthogonal, see (4.59a) and (4.59b), and hence

$$\mathsf{A}^{-1} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix of the shear map with respect to new basis is thus given by

$$\begin{aligned} \widetilde{\mathsf{S}}_\gamma &= \mathsf{A}^{-1}\mathsf{S}_\gamma\mathsf{A} \\ &= \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta + \gamma\sin\theta & -\sin\theta + \gamma\cos\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 + \gamma\sin\theta\cos\theta & \gamma\cos^2\theta & 0 \\ -\gamma\sin^2\theta & 1 - \gamma\sin\theta\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

20/03

5 Determinants, Matrix Inverses and Linear Equations

5.0 Why Study This?

This section continues our study of linear mathematics. The 'real' world can often be described by equations of various sorts. Some models result in linear equations of the type studied here. However, even when the real world results in more complicated models, the solution of these more complicated models often involves the solution of linear equations. We will concentrate on the case of three linear equations in three unknowns, but the systems in the case of the real world are normally much larger (often by many orders of magnitude).

5.1 Solution of Two Linear Equations in Two Unknowns

Consider two linear equations in two unknowns:

$$a_{11}x_1 + a_{12}x_2 = d_1 \tag{5.1a}$$
$$a_{21}x_1 + a_{22}x_2 = d_2 \tag{5.1b}$$

or equivalently

$$\mathsf{A}\mathbf{x} = \mathbf{d} \tag{5.2a}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad \text{and} \quad \mathsf{A} = \{a_{ij}\} \quad (\text{a } 2\times 2 \text{ matrix}). \tag{5.2b}$$

Now solve by forming suitable linear combinations of the two equations (e.g. $a_{22} \times (5.1a) - a_{12} \times (5.1b)$)

$$\begin{aligned}(a_{11}a_{22} - a_{21}a_{12})x_1 &= a_{22}d_1 - a_{12}d_2\,, \\ (a_{21}a_{12} - a_{22}a_{11})x_2 &= a_{21}d_1 - a_{11}d_2\,.\end{aligned}$$

From (4.54) we have that

$$(a_{11}a_{22} - a_{21}a_{12}) = \det \mathsf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Thus, if $\det \mathsf{A} \neq 0$, the equations have a *unique* solution

$$\begin{aligned}x_1 &= (a_{22}d_1 - a_{12}d_2)/\det \mathsf{A}\,, \\ x_2 &= (-a_{21}d_1 + a_{11}d_2)/\det \mathsf{A}\,,\end{aligned}$$

i.e.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\det \mathsf{A}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}. \tag{5.3a}$$

However, from left multiplication of (5.2a) by A^{-1} (if it exists) we have that

$$\mathbf{x} = \mathsf{A}^{-1}\mathbf{d}\,. \tag{5.3b}$$

We therefore conclude that

$$\mathsf{A}^{-1} = \frac{1}{\det \mathsf{A}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}. \tag{5.4}$$

Exercise. Check that $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}$.

5.2 Determinants for 3×3 Matrices

For a 3×3 matrix A we already have that (see (4.52a) and following)

$$\det \mathsf{A} \equiv |\,\mathsf{A}\,| \equiv \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \tag{5.5a}$$
$$= \varepsilon_{ijk} a_{i1} a_{j2} a_{k3} \tag{5.5b}$$
$$= \varepsilon_{ijk} a_{1i} a_{2j} a_{3k} \tag{5.5c}$$
$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \tag{5.5d}$$
$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{22}a_{13})\,. \tag{5.5e}$$

We observe that the last line, i.e. (5.5e), can be rewritten

$$\det \mathsf{A} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} . \tag{5.6}$$

Remarks.

- Note the sign pattern in (5.6).
- (5.6) is an expansion of $\det \mathsf{A}$ in terms of elements of the first column of A and determinants of 2×2 sub-matrices. This observation can be used to generalise the definition (using recursion), and
20/02 evaluation, of determinants to larger $(n \times n)$ matrices (but is left undone)

5.2.1 Properties of 3×3 determinants

(i) Since $\varepsilon_{ijk} a_{i1} a_{j2} a_{k3} = \varepsilon_{ijk} a_{1i} a_{2j} a_{3k}$

$$\det \mathsf{A} = \det \mathsf{A}^{\mathrm{T}} . \tag{5.7}$$

This means that an expansion equivalent to (5.6) in terms of the elements of the first row of A and determinants of 2×2 sub-matrices also exists. Hence from (5.5d), or otherwise,

$$\det \mathsf{A} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} . \tag{5.8}$$

Remark. (5.7) generalises to $n \times n$ determinants (but is left unproved).

(ii) Let $a_{i1} = \alpha_i$, $a_{j2} = \beta_j$, $a_{k3} = \gamma_k$, then from (5.5b)

$$\det \begin{pmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{pmatrix} = \epsilon_{ijk} \alpha_i \beta_j \gamma_k = \boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) .$$

Now $\boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) = 0$ if and only if $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are coplanar, i.e. if $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are linearly dependent. Similarly $\det \mathsf{A} = 0$ if and only if there is linear dependence between the columns of A, or from (5.7) if and only there is linear dependence between the rows of A.

Remark. This property generalises to $n \times n$ determinants (but is left unproved).

(iii) If we interchange any two of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ we change the sign of $\boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma})$. Hence if we interchange any two columns of A we change the sign of $\det \mathsf{A}$; similarly if we interchange any two rows of A we change the sign of $\det \mathsf{A}$.

Remark. This property generalises to $n \times n$ determinants (but is left unproved).

(iv) Suppose that we construct a matrix, say $\widetilde{\mathsf{A}}$, by adding to a given column of A linear combinations of the other columns. Then

$$\det \widetilde{\mathsf{A}} = \det \mathsf{A} .$$

This follows from the fact that

$$\begin{aligned} \begin{vmatrix} \alpha_1 + \lambda\beta_1 + \mu\gamma_1 & \beta_1 & \gamma_1 \\ \alpha_2 + \lambda\beta_2 + \mu\gamma_2 & \beta_2 & \gamma_2 \\ \alpha_3 + \lambda\beta_3 + \mu\gamma_3 & \beta_3 & \gamma_3 \end{vmatrix} &= (\boldsymbol{\alpha} + \lambda\boldsymbol{\beta} + \mu\boldsymbol{\gamma}) \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) \\ &= \boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) + \lambda\boldsymbol{\beta} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) + \mu\boldsymbol{\gamma} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) \\ &= \boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) . \end{aligned}$$

From invoking (5.7) we can deduce that a similar result applies to rows.

Remarks.

- This property generalises to $n \times n$ determinants (but is left unproved).

- This property is useful in evaluating determinants. For instance, if $a_{11} \neq 0$ add multiples of the first row to subsequent rows to make $a_{i1} = 0$ for $i = 2, \ldots, n$; then $\det \mathsf{A} = a_{11}\Delta_{11}$ (see (5.13b)).

(v) If we multiply any single row or column of A by λ, to give $\widehat{\mathsf{A}}$, then

$$\det \widehat{\mathsf{A}} = \lambda \det \mathsf{A}\,,$$

since

$$(\lambda\boldsymbol{\alpha}) \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}) = \lambda(\boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}))\,.$$

Remark. This property generalises to $n \times n$ determinants (but is left unproved).

(vi) If we multiply the matrix A by λ, then

$$\det(\lambda\mathsf{A}) = \lambda^3 \det \mathsf{A}\,, \tag{5.9a}$$

since

$$\lambda\boldsymbol{\alpha} \cdot (\lambda\boldsymbol{\beta} \times \lambda\boldsymbol{\gamma}) = \lambda^3(\boldsymbol{\alpha} \cdot (\boldsymbol{\beta} \times \boldsymbol{\gamma}))\,.$$

Remark. This property generalises for $n \times n$ determinants to

$$\det(\lambda\mathsf{A}) = \lambda^n \det \mathsf{A}\,, \tag{5.9b}$$

but is left unproved.

Theorem 5.1. *If* $\mathsf{A} = \{a_{ij}\}$ *is* 3×3*, then*

$$\begin{aligned} \epsilon_{pqr} \det \mathsf{A} &= \epsilon_{ijk} a_{pi} a_{qj} a_{rk}\,, && (5.10a) \\ \epsilon_{pqr} \det \mathsf{A} &= \epsilon_{ijk} a_{ip} a_{jq} a_{kr}\,. && (5.10b) \end{aligned}$$

Proof. Start with (5.10a), and suppose that $p = 1$, $q = 2$, $r = 3$. Then (5.10a) is just (5.5c). Next suppose that p and q are swapped. Then the sign of the left-hand side of (5.10a) reverses, while the right-hand side becomes

$$\epsilon_{ijk} a_{qi} a_{pj} a_{rk} = \epsilon_{jik} a_{qj} a_{pi} a_{rk} = -\epsilon_{ijk} a_{pi} a_{qj} a_{rk}\,,$$

so the sign of right-hand side also reverses. Similarly for swaps of p and r, or q and r. It follows that (5.10a) holds for any $\{pqr\}$ that is a permutation of $\{123\}$.

Suppose now that two (or more) of p, q and r in (5.10a) are equal. Wlog take $p = q = 1$, say. Then the left-hand side is zero, while the right-hand side is

$$\epsilon_{ijk} a_{1i} a_{1j} a_{rk} = \epsilon_{jik} a_{1j} a_{1i} a_{rk} = -\epsilon_{ijk} a_{1i} a_{1j} a_{rk}\,,$$

which is also zero. Having covered all cases we conclude that (5.10a) is true.

Similarly for (5.10b) starting from (5.5b) □

This is a supervisor's copy of the notes. It is not to be distributed to students.

Remark. This theorem generalises to $n \times n$ determinants (but is left unproved). 20/06

Theorem 5.2. *If* A *and* B *are both square matrices, then*

$$\det \mathsf{AB} = (\det \mathsf{A})(\det \mathsf{B})\,. \tag{5.11}$$

Proof. We prove the result only for 3×3 matrices. From (5.5b) and (5.10b)

$$\begin{aligned} \det \mathsf{AB} &= \epsilon_{ijk}\, (\mathsf{AB})_{i1}\, (\mathsf{AB})_{j2}\, (\mathsf{AB})_{k3} \\ &= \epsilon_{ijk}\, a_{ip} b_{p1}\, a_{jq} b_{q2}\, a_{kr} b_{r3} \\ &= \epsilon_{ijk}\, a_{ip} a_{jq} a_{kr}\, b_{p1} b_{q2} b_{r3} \\ &= \epsilon_{pqr} \det \mathsf{A}\, b_{p1} b_{q2} b_{r3} \\ &= \det \mathsf{A} \det \mathsf{B}\,. \end{aligned}$$

□

Theorem 5.3. *If* A *is orthogonal then*

$$\det \mathsf{A} = \pm 1\,. \tag{5.12}$$

Proof. If A is orthogonal then $\mathsf{A}\mathsf{A}^{\mathrm{T}} = \mathsf{I}$. It follows from (5.7) and (5.11) that

$$(\det \mathsf{A})^2 = (\det \mathsf{A})(\det \mathsf{A}^{\mathrm{T}}) = \det(\mathsf{A}\mathsf{A}^{\mathrm{T}}) = \det \mathsf{I} = 1\,.$$

Hence $\det \mathsf{A} = \pm 1$. □

21/03 *Remark.* You have already verified this for some reflection and rotation matrices.

5.3 The Inverse of a 3×3 Matrix

5.3.1 Minors and cofactors

For a square $n \times n$ matrix $\mathsf{A} = \{a_{ij}\}$, define A^{ij} to be the square matrix obtained by eliminating the i^{th} row and the j^{th} column of A. Hence

$$\mathsf{A}^{ij} = \begin{pmatrix} a_{11} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ a_{(i+1)1} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n(j-1)} & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix}.$$

Definition. Define the *minor*, M_{ij}, of the ij^{th} element of square matrix A to be the determinant of the square matrix obtained by eliminating the i^{th} row and the j^{th} column of A, i.e.

$$M_{ij} = \det \mathsf{A}^{ij}\,. \tag{5.13a}$$

Definition. Define the *cofactor* Δ_{ij} of the ij^{th} element of square matrix A as

$$\Delta_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det \mathsf{A}^{ij}\,. \tag{5.13b}$$

The above definitions apply for $n \times n$ matrices, but henceforth assume that A is a 3×3 matrix. Then from (5.6) and (5.13b)

$$\begin{aligned} \det \mathsf{A} &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ &= a_{11}\Delta_{11} + a_{21}\Delta_{21} + a_{31}\Delta_{31} \\ &= a_{j1}\Delta_{j1}\,. \end{aligned} \tag{5.14a}$$

Similarly, after noting from an interchanges of columns that

$$\det \mathsf{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{vmatrix},$$

we have that

$$\begin{aligned} \det \mathsf{A} &= -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ &= a_{12}\Delta_{12} + a_{22}\Delta_{22} + a_{32}\Delta_{32} \\ &= a_{j2}\Delta_{j2}\,. \end{aligned} \tag{5.14b}$$

Analogously (or by a relabelling symmetry)

$$\det \mathsf{A} = a_{j3}\Delta_{j3}\,. \tag{5.14c}$$

Similarly, but starting from (5.8) and subsequently interchanging rows,

$$\det \mathsf{A} = a_{1j}\Delta_{1j} = a_{2j}\Delta_{2j} = a_{3j}\Delta_{3j}\,. \tag{5.15}$$

Next we wish to show that

$$a_{ij}\Delta_{ik} = 0 \quad \text{if } j \neq k. \tag{5.16a}$$

To this end consider $j = 2$ and $k = 1$, then

$$\begin{aligned}
a_{i2}\Delta_{i1} &= a_{12}\Delta_{11} + a_{22}\Delta_{21} + a_{32}\Delta_{31} \\
&= a_{12}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{22}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{32}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\
&= \begin{vmatrix} a_{12} & a_{12} & a_{13} \\ a_{22} & a_{22} & a_{23} \\ a_{32} & a_{32} & a_{33} \end{vmatrix} \\
&= 0\,,
\end{aligned}$$

since two columns are linearly dependent (actually equal). Proceed similarly for other choices of j and k to obtain (5.16a). Further, we can also show that

$$a_{ji}\Delta_{ki} = 0 \quad \text{if } j \neq k, \tag{5.16b}$$

since two rows turn out to be linearly dependent.

We can combine all the above results in a lemma.

Lemma 5.4.

$$\begin{aligned}
a_{ij}\Delta_{ik} &= \delta_{jk}\,\det \mathsf{A}\,, && (5.17a)\\
a_{ji}\Delta_{ki} &= \delta_{jk}\,\det \mathsf{A}\,. && (5.17b)
\end{aligned}$$

Proof. See (5.14a), (5.14b), (5.14c), (5.15), (5.16a) and (5.16b). □

5.3.2 Construction of the inverse

Theorem 5.5. *Given a* 3×3 *matrix* A *with* $\det \mathsf{A} \neq 0$*, define* B *by*

$$(\mathsf{B})_{ij} = \frac{1}{\det \mathsf{A}}\Delta_{ji}\,, \tag{5.18a}$$

then

$$\mathsf{AB} = \mathsf{BA} = \mathsf{I}. \tag{5.18b}$$

Proof. From (5.17b) and (5.18a),

$$\begin{aligned}
(\mathsf{AB})_{ij} &= a_{ik}(\mathsf{B})_{kj} \\
&= \frac{a_{ik}\Delta_{jk}}{\det \mathsf{A}} \\
&= \frac{\delta_{ij}\det \mathsf{A}}{\det \mathsf{A}} \\
&= \delta_{ij}\,.
\end{aligned}$$

Hence $\mathsf{AB} = \mathsf{I}$. Similarly from (5.17a) and (5.18a), $\mathsf{BA} = \mathsf{I}$. It follows that $\mathsf{B} = \mathsf{A}^{-1}$ and A is invertible. □

Remark. The formula for the elements of the inverse of A, i.e.

$$(\mathsf{A}^{-1})_{ij} = \frac{1}{\det \mathsf{A}} \Delta_{ji}\,, \tag{5.19}$$

holds for $n \times n$ matrices (including 2×2 matrices) for suitably defined cofactors.

Example. Consider the simple shear

$$\mathsf{S}_\gamma = \begin{pmatrix} 1 & \gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Then $\det \mathsf{S}_\gamma = 1$, and after a little manipulation

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

Hence

$$\mathsf{S}_\gamma^{-1} = \begin{pmatrix} \Delta_{11} & \Delta_{21} & \Delta_{31} \\ \Delta_{12} & \Delta_{22} & \Delta_{32} \\ \Delta_{13} & \Delta_{23} & \Delta_{33} \end{pmatrix} = \begin{pmatrix} 1 & -\gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

which makes physical sense in that the effect of a shear γ is reversed by changing the sign of γ.

21/02

5.4 Solving Linear Equations: Gaussian Elimination

Suppose that we wish to solve the system of equations

$$\mathsf{A}\mathbf{x} = \mathbf{d}\,, \tag{5.20a}$$

where A is a $n \times n$ matrix, $\mathbf{x}$ is a $n \times 1$ column vector of unknowns, and $\mathbf{d}$ is a given $n \times 1$ column vector. Assuming that $\det \mathsf{A} \neq 0$, this has the formal solution

$$\mathbf{x} = \mathsf{A}^{-1}\mathbf{d}\,. \tag{5.20b}$$

Hence if we wished to solve (5.20a) numerically, then one method would be to calculate A^{-1} using (5.19), and then form $\mathsf{A}^{-1}\mathbf{d}$. However, this is actually very inefficient.

A better method is Gaussian elimination, which we will illustrate for 3×3 matrices. Hence suppose we wish to solve

$$a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 = d_1\,, \tag{5.21a}$$
$$a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 = d_2\,, \tag{5.21b}$$
$$a_{31}\,x_1 + a_{32}\,x_2 + a_{33}\,x_3 = d_3\,. \tag{5.21c}$$

Assume $a_{11} \neq 0$, otherwise re-order the equations so that $a_{11} \neq 0$, and if that is not possible then stop (since there is then no unique solution). Now use (5.21a) to eliminate x_1 by forming

$$(5.21\text{b}) - \frac{a_{21}}{a_{11}} \times (5.21\text{a}) \quad \text{and} \quad (5.21\text{c}) - \frac{a_{31}}{a_{11}} \times (5.21\text{a})\,,$$

so as to obtain

$$\left(a_{22} - \frac{a_{21}}{a_{11}}a_{12}\right)x_2 + \left(a_{23} - \frac{a_{21}}{a_{11}}a_{13}\right)x_3 = d_2 - \frac{a_{21}}{a_{11}}d_1\,, \tag{5.22a}$$
$$\left(a_{32} - \frac{a_{31}}{a_{11}}a_{12}\right)x_2 + \left(a_{33} - \frac{a_{31}}{a_{11}}a_{13}\right)x_3 = d_3 - \frac{a_{31}}{a_{11}}d_1\,. \tag{5.22b}$$

In order to simplify notation let

$$a'_{22} = \left(a_{22} - \frac{a_{21}}{a_{11}}a_{12}\right)\,, \quad a'_{23} = \left(a_{23} - \frac{a_{21}}{a_{11}}a_{13}\right)\,, \quad d'_2 = d_2 - \frac{a_{21}}{a_{11}}d_1\,,$$
$$a'_{32} = \left(a_{32} - \frac{a_{31}}{a_{11}}a_{12}\right)\,, \quad a'_{33} = \left(a_{33} - \frac{a_{31}}{a_{11}}a_{13}\right)\,, \quad d'_3 = d_3 - \frac{a_{31}}{a_{11}}d_1\,,$$

so that (5.22a) and (5.22b) become

$$a'_{22}x_2 + a'_{23}x_3 = d'_2\,, \tag{5.23a}$$
$$a'_{32}x_2 + a'_{33}x_3 = d'_3\,. \tag{5.23b}$$

Assume $a_{22} \neq 0$, otherwise re-order the equations so that $a_{22} \neq 0$, and if that is not possible then stop (since there is then no unique solution). Now use (5.23a) to eliminate x_2 by forming

$$(5.23\text{b}) - \frac{a'_{32}}{a'_{22}} \times (5.23\text{a})\,,$$

so as to obtain

$$\left(a'_{33} - \frac{a'_{32}}{a'_{22}}a'_{23}\right)x_3 = d'_3 - \frac{a'_{32}}{a'_{22}}d'_2\,. \tag{5.24a}$$

Now, providing that

$$\left(a'_{33} - \frac{a'_{32}}{a'_{22}}a'_{23}\right) \neq 0\,, \tag{5.24b}$$

(5.24a) gives x_3, then (5.23a) gives x_2, and finally (5.21a) gives x_1. If

$$\left(a'_{33} - \frac{a'_{32}}{a'_{22}}a'_{23}\right) = 0\,,$$

there is no unique solution (if there is a solution at all).

Remark. It is possible to show that this method fails only if A is not invertible, i.e. only if $\det\mathsf{A} = 0$.

5.5 Solving linear equations

5.5.1 Inhomogeneous and homogeneous problems

If $\det\mathsf{A} \neq 0$ (and A is a 2×2 or a 3×3 matrix) then we know from (5.4) and Theorem 5.5 on page 80 that A is invertible and A^{-1} exists. In such circumstances it follows that the system of equations $\mathsf{A}\mathbf{x} = \mathbf{d}$ has a unique solution $\mathbf{x} = \mathsf{A}^{-1}\mathbf{d}$.

This section is about what can be deduced if $\det\mathsf{A} = 0$, where as usual we specialise to the case where A is a 3×3 matrix.

Definition. If $\mathbf{d} \neq \mathbf{0}$ then the system of equations

$$\mathsf{A}\mathbf{x} = \mathbf{d} \tag{5.25a}$$

is said to be a system of *inhomogeneous* equations.

Definition. The system of equations

$$\mathsf{A}\mathbf{x} = \mathbf{0} \tag{5.25b}$$

21/06 is said to be a system of *homogeneous* equations. If $\det\mathsf{A} \neq 0$ then the unique solution is $\mathsf{A}^{-1}\mathbf{0} = \mathbf{0}$.

5.5.2 Geometrical view of $\mathsf{A}\mathbf{x} = \mathbf{0}$

For $i = 1, 2, 3$ let $\mathbf{r}_i$ be the vector with components equal to the elements of the ith row of A, in which case

$$\mathsf{A} = \begin{pmatrix} \mathbf{r}_1^{\mathrm{T}} \\ \mathbf{r}_2^{\mathrm{T}} \\ \mathbf{r}_3^{\mathrm{T}} \end{pmatrix} . \tag{5.26}$$

Equations (5.25a) and (5.25b) may then be expressed as

$$\mathbf{r}_i^{\mathrm{T}}\mathbf{x} \;=\; \mathbf{r}_i \cdot \mathbf{x} \;=\; d_i \qquad (i = 1, 2, 3)\,, \tag{5.27a}$$
$$\mathbf{r}_i^{\mathrm{T}}\mathbf{x} \;=\; \mathbf{r}_i \cdot \mathbf{x} \;=\; 0 \qquad (i = 1, 2, 3)\,, \tag{5.27b}$$

respectively. Since each of these individual equations represents a plane in $\mathbb{R}^3$, the solution of each set of 3 equations is the intersection of 3 planes.

For the homogeneous equations (5.27b) the three planes each pass through the origin, O. There are three possibilities:

(i) the three planes intersect only at O;

(ii) the three planes have a common line (including O);

(iii) the three planes coincide.

22/03

We consider each of these cases in turn.

(i) If $\det \mathsf{A} \neq 0$ then $\mathbf{r}_1 \cdot (\mathbf{r}_2 \times \mathbf{r}_3) \neq 0$ and the set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ consists of three linearly independent vectors; hence $\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = \mathbb{R}^3$. The first two equations of (5.27b) imply that $\mathbf{x}$ must lie on the intersection of the planes $\mathbf{r}_1 \cdot \mathbf{x} = 0$ and $\mathbf{r}_2 \cdot \mathbf{x} = 0$, i.e. $\mathbf{x}$ must lie on the line

$$\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{t}, \lambda \in \mathbb{R}, \mathbf{t} = \mathbf{r}_1 \times \mathbf{r}_2\}.$$

The final condition $\mathbf{r}_3 \cdot \mathbf{x} = 0$ then implies that $\lambda = 0$ (since we have assumed that $\mathbf{r}_1 \cdot (\mathbf{r}_2 \times \mathbf{r}_3) \neq 0$), and hence $\mathbf{x} = \mathbf{0}$, i.e. the three planes intersect only at the origin. The *solution space* thus has zero dimension.

(ii) Next suppose that $\det \mathsf{A} = 0$. In this case the set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ is linearly dependent with the dimension of $\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ being equal to 2 or 1; first we consider the case when it is 2. Assume wlog that $\mathbf{r}_1$ and $\mathbf{r}_2$ are the two linearly independent vectors. Then as above the first two equations of (5.27b) again imply that $\mathbf{x}$ must lie on the line

$$\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{t}, \lambda \in \mathbb{R}, \mathbf{t} = \mathbf{r}_1 \times \mathbf{r}_2\}.$$

Since $\mathbf{r}_1 \cdot (\mathbf{r}_2 \times \mathbf{r}_3) = 0$, all points in this line satisfy $\mathbf{r}_3 \cdot \mathbf{x} = 0$. Hence the intersection of the three planes is a line, i.e. the solution for $\mathbf{x}$ is a line. The solution space thus has dimension one.

(iii) Finally we need to consider the case when the dimension of $\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ is 1. The three row vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ must then all be parallel. This means that $\mathbf{r}_1 \cdot \mathbf{x} = 0$, $\mathbf{r}_2 \cdot \mathbf{x} = 0$ and $\mathbf{r}_3 \cdot \mathbf{x} = 0$ all imply each other. Thus the intersection of the three planes is a plane, i.e. solutions to (5.27b) lie on a plane. If $\mathbf{a}$ and $\mathbf{b}$ are any two linearly independent vectors such that $\mathbf{a} \cdot \mathbf{r}_1 = \mathbf{b} \cdot \mathbf{r}_1 = 0$, then we may specify the plane, and thus the solution space, by

$$\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{a} + \mu\mathbf{b} \text{ where } \lambda, \mu \in \mathbb{R}\}\,. \tag{5.28}$$

The solution space thus has dimension two.

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

Consider the linear map $\mathcal{A} : \mathbb{R}^3 \to \mathbb{R}^3$, such that $\mathbf{x} \mapsto \mathbf{x}' = \mathsf{A}\mathbf{x}$, where A is the matrix of $\mathcal{A}$ with respect to the standard basis. From our earlier definition (4.5), the kernel of $\mathcal{A}$ is given by

$$K(\mathcal{A}) = \{\mathbf{x} \in \mathbb{R}^3 : \mathsf{A}\mathbf{x} = \mathbf{0}\}\,. \tag{5.29}$$

$K(\mathcal{A})$ is the solution space of $\mathsf{A}\mathbf{x} = \mathbf{0}$, with a dimension denoted by $n(\mathcal{A})$. The three cases studied in our *geometric view* of §5.5.2 correspond to

(i) $n(\mathcal{A}) = 0$,

(ii) $n(\mathcal{A}) = 1$,

(iii) $n(\mathcal{A}) = 2$.

Remark. We do not need to consider the case $n(\mathcal{A}) = 3$ as long as we exclude the map with $\mathsf{A} = 0$.

Next recall that if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis for $\mathbb{R}^3$, then the image of $\mathcal{A}$ is spanned by $\{\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\}$, i.e.

$$\mathcal{A}(\mathbb{R}^3) = \operatorname{span}\{\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\}\,.$$

We now consider the three different possibilities for the value of the nullity in turn.

(i) If $n(\mathcal{A}) = 0$ then $\{\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\}$ is a linearly independent set since

$$\begin{aligned}\{\lambda\mathcal{A}(\mathbf{u}) + \mu\mathcal{A}(\mathbf{v}) + \nu\mathcal{A}(\mathbf{w}) = \mathbf{0}\} \quad &\Leftrightarrow \quad \{\mathcal{A}(\lambda\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w}) = \mathbf{0}\} \\ &\Leftrightarrow \quad \{\lambda\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w} = \mathbf{0}\}\,,\end{aligned}$$

and so $\lambda = \mu = \nu = 0$ since $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is basis. It follows that the rank of $\mathcal{A}$ is three, i.e. $r(\mathcal{A}) = 3$.

(ii) If $n(\mathcal{A}) = 1$ choose non-zero $\mathbf{u} \in K(\mathcal{A})$; $\mathbf{u}$ is a basis for the proper subspace $K(\mathcal{A})$. Next choose $\mathbf{v}, \mathbf{w} \notin K(\mathcal{A})$ to extend this basis to form a basis of $\mathbb{R}^3$ (recall from Theorem 3.5 on page 49 that this is always possible). We claim that $\{\mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\}$ are linearly independent. To see this note that

$$\begin{aligned}\{\mu\mathcal{A}(\mathbf{v}) + \nu\mathcal{A}(\mathbf{w}) = \mathbf{0}\} \quad &\Leftrightarrow \quad \{\mathcal{A}(\mu\mathbf{v} + \nu\mathbf{w}) = \mathbf{0}\} \\ &\Leftrightarrow \quad \{\mu\mathbf{v} + \nu\mathbf{w} = \alpha\mathbf{u}\}\,,\end{aligned}$$

for some $\alpha \in \mathbb{R}$. Hence $-\alpha\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w} = \mathbf{0}$, and so $\alpha = \mu = \nu = 0$ since $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is basis. We conclude that $\mathcal{A}(\mathbf{v})$ and $\mathcal{A}(\mathbf{w})$ are linearly independent, and that the rank of $\mathcal{A}$ is two, i.e. $r(\mathcal{A}) = 2$ (since the dimension of $\operatorname{span}\{\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\}$ is two).

(iii) If $n(\mathcal{A}) = 2$ choose non-zero $\mathbf{u}, \mathbf{v} \in K(\mathcal{A})$ such that the set $\{\mathbf{u}, \mathbf{v}\}$ is a basis for $K(\mathcal{A})$. Next choose $\mathbf{w} \notin K(\mathcal{A})$ to extend this basis to form a basis of $\mathbb{R}^3$. Since $\mathsf{A}\mathbf{u} = \mathsf{A}\mathbf{v} = \mathbf{0}$ and $\mathsf{A}\mathbf{w} \neq \mathbf{0}$, it follows that

$$\dim\operatorname{span}\{\mathcal{A}(\mathbf{u}), \mathcal{A}(\mathbf{v}), \mathcal{A}(\mathbf{w})\} = r(\mathcal{A}) = 1\,.$$

Remarks.

(a) In each of cases (i), (ii) and (iii) we have in accordance of the rank-nullity Theorem (see Theorem 4.2 on page 59)

$$r(\mathcal{A}) + n(\mathcal{A}) = 3\,.$$

(b) In each case we also have from comparison of the results in this section with those in §5.5.2,

$$\begin{aligned} r(\mathcal{A}) \quad &= \quad \dim\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} \\ &= \quad \text{number of linearly independent rows of } \mathsf{A} \ (\textit{row rank}).\end{aligned}$$

Suppose now we choose $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ to be the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Then the image of $\mathcal{A}$ is spanned by $\{\mathcal{A}(\mathbf{e}_1), \mathcal{A}(\mathbf{e}_2), \mathcal{A}(\mathbf{e}_3)\}$, and the number of linearly independent vectors in this set must equal $r(\mathcal{A})$. It follows from (4.14) and (4.19) that

$$\begin{aligned} r(\mathcal{A}) \quad &= \quad \dim\operatorname{span}\{\mathsf{A}\mathbf{e}_1, \mathsf{A}\mathbf{e}_2, \mathsf{A}\mathbf{e}_3\} \\ &= \quad \text{number of linearly independent columns of } \mathsf{A} \ (\textit{column rank}).\end{aligned}$$

5.5.4 Implications for the solution of the inhomogeneous equation $\mathsf{A}\mathbf{x} = \mathbf{d}$

If $\det \mathsf{A} \neq 0$ then $r(\mathcal{A}) = 3$ and $I(\mathcal{A}) = \mathbb{R}^3$ (where $I(\mathcal{A})$ is notation for the image of $\mathcal{A}$). Since $\mathbf{d} \in \mathbb{R}^3$, there must exist $\mathbf{x} \in \mathbb{R}^3$ for which $\mathbf{d}$ is the image under $\mathcal{A}$, i.e. $\mathbf{x} = \mathsf{A}^{-1}\mathbf{d}$ exists and unique.

If $\det \mathsf{A} = 0$ then $r(\mathcal{A}) < 3$ and $I(\mathcal{A})$ is a proper subspace of $\mathbb{R}^3$. Then

- *either* $\mathbf{d} \notin I(\mathcal{A})$, in which there are no solutions and the equations are *inconsistent*;
- *or* $\mathbf{d} \in I(\mathcal{A})$, in which case there is at least one solution and the equations are *consistent*.

The latter case is described by Theorem 5.6 below.

Theorem 5.6. *If $\mathbf{d} \in I(\mathcal{A})$ then the general solution to $\mathsf{A}\mathbf{x} = \mathbf{d}$ can be written as $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$ where $\mathbf{x}_0$ is a particular fixed solution of $\mathsf{A}\mathbf{x} = \mathbf{d}$ and $\mathbf{y}$ is the general solution of $\mathsf{A}\mathbf{x} = \mathbf{0}$.*

Proof. First we note that $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$ is a solution since $\mathsf{A}\mathbf{x}_0 = \mathbf{d}$ and $\mathsf{A}\mathbf{y} = \mathbf{0}$, and thus

$$\mathsf{A}(\mathbf{x}_0 + \mathbf{y}) = \mathbf{d} + \mathbf{0} = \mathbf{d}\,.$$

Further, from §5.5.2 and §5.5.3, if

(i) $n(\mathcal{A}) = 0$ and $r(\mathcal{A}) = 3$, then $\mathbf{y} = \mathbf{0}$ and the solution is unique.

(ii) $n(\mathcal{A}) = 1$ and $r(\mathcal{A}) = 2$, then $\mathbf{y} = \lambda\mathbf{t}$ and $\mathbf{x} = \mathbf{x}_0 + \lambda\mathbf{t}$ (representing a line).

(iii) $n(\mathcal{A}) = 2$ and $r(\mathcal{A}) = 1$, then $\mathbf{y} = \lambda\mathbf{a} + \mu\mathbf{b}$ and $\mathbf{x} = \mathbf{x}_0 + \lambda\mathbf{a} + \mu\mathbf{b}$ (representing a plane). □

Example. Consider the (2×2) inhomogeneous case of $\mathsf{A}\mathbf{x} = \mathbf{d}$ where

$$\begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ b \end{pmatrix}\,. \tag{5.30}$$

Since $\det \mathsf{A} = (1 - a)$, if $a \neq 1$ then $\det \mathsf{A} \neq 0$ and A^{-1} exists and is unique. Specifically

$$\mathsf{A}^{-1} = \frac{1}{1-a} \begin{pmatrix} 1 & -1 \\ -a & 1 \end{pmatrix}, \text{ and the unique solution is } \mathbf{x} = \mathsf{A}^{-1} \begin{pmatrix} 1 \\ b \end{pmatrix}\,. \tag{5.31}$$

If $a = 1$, then $\det \mathsf{A} = 0$, and

$$\mathsf{A}\mathbf{x} = \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix}\,.$$

Hence

$$I(\mathcal{A}) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\} \quad \text{and} \quad K(\mathcal{A}) = \operatorname{span}\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}\,,$$

and so $r(\mathcal{A}) = 1$ and $n(\mathcal{A}) = 1$. Whether there is a solution now depends on the value of b.

- If $b \neq 1$ then $\begin{pmatrix} 1 \\ b \end{pmatrix} \notin I(\mathcal{A})$, and there are no solutions because the equations are inconsistent.
- If $b = 1$ then $\begin{pmatrix} 1 \\ b \end{pmatrix} \in I(\mathcal{A})$ and solutions exist (the equations are consistent). A particular solution is

  $$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\,.$$

  The general solution is then $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$, where $\mathbf{y}$ is any vector in $K(\mathcal{A})$, i.e.

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

22/06 where $\lambda \in \mathbb{R}$.

6 Complex Vector Spaces

6.0 Why Study This?

In the same way that complex numbers are a natural extension of real numbers, and allow us to solve more problems, complex vector spaces are a natural extension of real numbers, and allow us to solve more problems. *Inter alia* they are important in quantum mechanics.

6.1 Vector Spaces Over The Complex Numbers

We have considered vector spaces with real scalars. We now generalise to vector spaces with complex scalars.

6.1.1 Definition

We just adapt the definition for a vector space over the real numbers by exchanging 'real' for 'complex' and '$\mathbb{R}$' for '$\mathbb{C}$'. Hence a vector space over the complex numbers is a set V of elements, or 'vectors', together with two binary operations

- *vector addition* denoted for $\mathbf{x}, \mathbf{y} \in V$ by $\mathbf{x}+\mathbf{y}$, where $\mathbf{x}+\mathbf{y} \in V$ so that there is closure under vector addition;
- *scalar multiplication* denoted for $a \in \mathbb{C}$ and $\mathbf{x} \in V$ by $a\mathbf{x}$, where $a\mathbf{x} \in V$ so that there is closure under scalar multiplication;

satisfying the following eight axioms or rules:

A(i) addition is *associative*, i.e. for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$

$$\mathbf{x}+(\mathbf{y}+\mathbf{z}) = (\mathbf{x}+\mathbf{y})+\mathbf{z}\,; \tag{6.1a}$$

A(ii) addition is *commutative*, i.e. for all $\mathbf{x}, \mathbf{y} \in V$

$$\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}\,; \tag{6.1b}$$

A(iii) there exists an element $\mathbf{0} \in V$, called the *null or zero vector*, such that for all $\mathbf{x} \in V$

$$\mathbf{x}+\mathbf{0} = \mathbf{x}\,, \tag{6.1c}$$

i.e. vector addition has an identity element;

A(iv) for all $\mathbf{x} \in V$ there exists an additive *negative or inverse vector* $\mathbf{x}' \in V$ such that

$$\mathbf{x}+\mathbf{x}' = \mathbf{0}\,; \tag{6.1d}$$

B(v) scalar multiplication of vectors is 'associative', i.e. for all $\lambda, \mu \in \mathbb{C}$ and $\mathbf{x} \in V$

$$\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}\,, \tag{6.1e}$$

B(vi) scalar multiplication has an identity element, i.e. for all $\mathbf{x} \in V$

$$1\,\mathbf{x} = \mathbf{x}\,, \tag{6.1f}$$

where 1 is the multiplicative identity in $\mathbb{C}$;

B(vii) scalar multiplication is distributive over vector addition, i.e. for all $\lambda \in \mathbb{C}$ and $\mathbf{x}, \mathbf{y} \in V$

$$\lambda(\mathbf{x}+\mathbf{y}) = \lambda\mathbf{x}+\lambda\mathbf{y}\,; \tag{6.1g}$$

B(viii) scalar multiplication is distributive over scalar addition, i.e. for all $\lambda, \mu \in \mathbb{C}$ and $\mathbf{x} \in V$

$$(\lambda+\mu)\mathbf{x} = \lambda\mathbf{x}+\mu\mathbf{x}\,. \tag{6.1h}$$

6.1.2 Properties

Most of the properties and theorems of §3 follow through much as before, e.g.

(i) the zero vector $\mathbf{0}$ is unique;

(ii) the additive inverse of a vector is unique;

(iii) if $\mathbf{x} \in V$ and $\lambda \in \mathbb{C}$ then

$$0\mathbf{x} = \mathbf{0}, \quad (-1)\mathbf{x} = -\mathbf{x} \quad \text{and} \quad \lambda\mathbf{0} = \mathbf{0};$$

(iv) Theorem 3.1 on when a subset U of a vector space V is a subspace of V still applies if we exchange 'real' for 'complex' and '$\mathbb{R}$' for '$\mathbb{C}$'.

6.1.3 $\mathbb{C}^n$

Definition. For fixed positive integer n, define $\mathbb{C}^n$ to be the set of n-tuples $(z_1, z_2, \ldots, z_n)$ of complex numbers $z_i \in \mathbb{C}$, $i = 1, \ldots, n$. For $\lambda \in \mathbb{C}$ and complex vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, where

$$\mathbf{u} = (u_1, \ldots, u_n) \quad \text{and} \quad \mathbf{v} = (v_1, \ldots, v_n),$$

define vector addition and scalar multiplication by

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1 + v_1, \ldots, u_n + v_n) \in \mathbb{C}^n, && (6.2a) \\ \lambda\mathbf{u} &= (\lambda u_1, \ldots, \lambda u_n) \in \mathbb{C}^n. && (6.2b) \end{aligned}$$

Lemma 6.1. *$\mathbb{C}^n$ is a vector space over $\mathbb{C}$*

Proof by exercise. Check that A(i), A(ii), A(iii), A(iv), B(v), B(vi), B(vii) and B(viii) of §6.1.1 are satisfied. □

Remarks.

(i) $\mathbb{R}^n$ is a subset of $\mathbb{C}^n$, but it is *not* a subspace of the vector space $\mathbb{C}^n$, since $\mathbb{R}^n$ is not closed under multiplication by an arbitrary complex number.

(ii) The standard basis for $\mathbb{R}^n$, i.e.

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \; \ldots \; , \mathbf{e}_n = (0, 0, \ldots, 1), \tag{6.3a}$$

also serves as a standard basis for $\mathbb{C}^n$. Hence $\mathbb{C}^n$ has dimension n when viewed as a vector space over $\mathbb{C}$. Further we can express any $\mathbf{z} \in \mathbb{C}^n$ in terms of components as

$$\mathbf{z} = (z_1, \ldots, z_n) = \sum_{i=1}^{n} z_i \mathbf{e}_i\,. \tag{6.3b}$$

6.2 Scalar Products for Complex Vector Spaces

In generalising from real vector spaces to complex vector spaces, we have to be careful with scalar products.

6.2.1 Definition

Let V be a n-dimensional vector space over the complex numbers. We will denote a scalar product, or inner product, of the ordered pair of vectors $\mathbf{u}$, $\mathbf{v} \in V$ by

$$\langle\, \mathbf{u}\,, \,\mathbf{v}\,\rangle \in \mathbb{C}\,, \tag{6.4}$$

where alternative notations are $\mathbf{u}\cdot\mathbf{v}$ and $\langle\,\mathbf{u}\,|\,\mathbf{v}\,\rangle$. A scalar product of a vector space over the complex numbers must have the following properties.

(i) *Conjugate symmetry*, i.e.
$$\langle\,\mathbf{u}\,,\,\mathbf{v}\,\rangle = \langle\,\mathbf{v}\,,\,\mathbf{u}\,\rangle^*\,, \tag{6.5a}$$
where a * is an alternative notation for a complex conjugate (we shall swap between $^-$ and * freely). Implicit in this equation is the assumption that for a complex vector space the *ordering* of the vectors in the scalar product is important (whereas for $\mathbb{R}^n$ this is not important). Further, if we let $\mathbf{u}=\mathbf{v}$, then (6.5a) implies that
$$\langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle = \langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle^*\,, \tag{6.5b}$$
i.e. $\langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle$ is *real*.

(ii) *Linearity* in the second argument, i.e. for $\lambda,\mu\in\mathbb{C}$
$$\langle\,\mathbf{u}\,,\,(\lambda\mathbf{v}_1+\mu\mathbf{v}_2)\,\rangle = \lambda\,\langle\,\mathbf{u}\,,\,\mathbf{v}_1\,\rangle + \mu\,\langle\,\mathbf{u}\,,\,\mathbf{v}_2\,\rangle\,. \tag{6.5c}$$

(iii) *Non-negativity*, i.e. a scalar product of a vector with itself should be positive, i.e.
$$\langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle \geqslant 0\,. \tag{6.5d}$$
This allows us to write $\langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle = \|\mathbf{v}\|^2$, where the real positive number $\|\mathbf{v}\|$ is the *norm* of the vector $\mathbf{v}$.

(iv) *Non-degeneracy*, i.e. the only vector of zero norm should be the zero vector, i.e.
$$\|\mathbf{v}\|^2 \equiv \langle\,\mathbf{v}\,,\,\mathbf{v}\,\rangle = 0 \quad\Rightarrow\quad \mathbf{v}=\mathbf{0}\,. \tag{6.5e}$$

6.2.2 Properties

Scalar product with $\mathbf{0}$. We can again show that
$$\langle\,\mathbf{u}\,,\,\mathbf{0}\,\rangle = \langle\,\mathbf{0}\,,\,\mathbf{u}\,\rangle = 0\,. \tag{6.6}$$

Anti-linearity in the first argument. Properties (6.5a) and (6.5c) imply so-called 'anti-linearity' in the first argument, i.e. for $\lambda,\mu\in\mathbb{C}$
$$\begin{aligned}\langle\,(\lambda\mathbf{u}_1+\mu\mathbf{u}_2)\,,\,\mathbf{v}\,\rangle &= \langle\,\mathbf{v}\,,\,(\lambda\mathbf{u}_1+\mu\mathbf{u}_2)\,\rangle^*\\ &= \lambda^*\,\langle\,\mathbf{v}\,,\,\mathbf{u}_1\,\rangle^* + \mu^*\,\langle\,\mathbf{v}\,,\,\mathbf{u}_2\,\rangle^*\\ &= \lambda^*\,\langle\,\mathbf{u}_1\,,\,\mathbf{v}\,\rangle + \mu^*\,\langle\,\mathbf{u}_2\,,\,\mathbf{v}\,\rangle\,. \end{aligned}\tag{6.7}$$

Schwarz's inequality and the triangle inequality. It is again true that
$$|\langle\,\mathbf{u}\,,\,\mathbf{v}\,\rangle| \quad\leqslant\quad \|\mathbf{u}\|\,\|\mathbf{v}\|\,, \tag{6.8a}$$
$$\|\mathbf{u}+\mathbf{v}\| \quad\leqslant\quad \|\mathbf{u}\|+\|\mathbf{v}\|\,. \tag{6.8b}$$
with equality only when $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

6.2.3 Scalar Products in Terms of Components

Suppose that we have a scalar product defined on a complex vector space with a given basis $\{\mathbf{e}_i\}$, $i=1,\dots,n$. We claim that the scalar product is determined for all pairs of vectors by its values for all pairs of basis vectors. To see this first define the complex numbers G_{ij} by
$$G_{ij} = \langle\,\mathbf{e}_i\,,\,\mathbf{e}_j\,\rangle \quad (i,j=1,\dots,n)\,. \tag{6.9}$$
Then, for *any* two vectors
$$\mathbf{v}=\sum_{i=1}^n v_i\mathbf{e}_i \quad\text{and}\quad \mathbf{w}=\sum_{j=1}^n w_j\mathbf{e}_j\,, \tag{6.10}$$

we have that

$$
\begin{aligned}
\langle\, \mathbf{v} \,,\, \mathbf{w} \,\rangle &= \left\langle \left(\sum_{i=1}^{n} v_i \mathbf{e}_i\right), \left(\sum_{j=1}^{n} w_j \mathbf{e}_j\right) \right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} v_i^* w_j \,\langle\, \mathbf{e}_i \,,\, \mathbf{e}_j \,\rangle \qquad \text{from (6.5c) and (6.7)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} v_i^* G_{ij} w_j \,. \qquad (6.11)
\end{aligned}
$$

We can simplify this expression (which determines the scalar product in terms of the G_{ij}), but first it helps to have a definition.

Hermitian conjugates.

Definition. The *Hermitian conjugate* or *conjugate transpose* or *adjoint* of a matrix $\mathsf{A} = \{a_{ij}\}$, where $a_{ij} \in \mathbb{C}$, is defined to be

$$\mathsf{A}^\dagger = (\mathsf{A}^{\mathrm{T}})^* = (\mathsf{A}^*)^{\mathrm{T}} \,, \qquad (6.12)$$

where, as before, $^{\mathrm{T}}$ denotes a transpose.

Example.

$$\text{If} \quad \mathsf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{then} \quad \mathsf{A}^\dagger = \begin{pmatrix} a_{11}^* & a_{21}^* \\ a_{12}^* & a_{22}^* \end{pmatrix}.$$

Properties. For matrices A and B recall that $(\mathsf{AB})^{\mathrm{T}} = \mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}}$. Hence $(\mathsf{AB})^{\mathrm{T}*} = \mathsf{B}^{\mathrm{T}*}\mathsf{A}^{\mathrm{T}*}$, and so

$$(\mathsf{AB})^\dagger = \mathsf{B}^\dagger \mathsf{A}^\dagger \,. \qquad (6.13a)$$

Also, from (6.12),

$$\mathsf{A}^{\dagger\dagger} = \left(\mathsf{A}^{*\mathrm{T}}\right)^{\mathrm{T}*} = \mathsf{A} \,. \qquad (6.13b)$$

Let w be the column matrix of components,

$$\mathsf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}, \qquad (6.14a)$$

and let $\mathsf{v}^\dagger$ be the adjoint of the column matrix v, i.e. $\mathsf{v}^\dagger$ is the row matrix

$$\mathsf{v}^\dagger \equiv (\mathsf{v}^*)^{\mathrm{T}} = \begin{pmatrix} v_1^* & v_2^* & \dots & v_n^* \end{pmatrix} . \qquad (6.14b)$$

Then in terms of this notation the scalar product (6.11) can be written as

$$\langle\, \mathbf{v} \,,\, \mathbf{w} \,\rangle = \mathsf{v}^\dagger \mathsf{G}\, \mathsf{w} \,, \qquad (6.15)$$

where G is the matrix, or *metric*, with entries G_{ij}.

Remark. If the $\{\mathbf{e}_i\}$ form an orthonormal basis, i.e. are such that

$$G_{ij} = \langle\, \mathbf{e}_i \,,\, \mathbf{e}_j \,\rangle = \delta_{ij} \,, \qquad (6.16a)$$

then (6.11), or equivalently (6.15), reduces to (cf. (3.22))

$$\langle\, \mathbf{v} \,,\, \mathbf{w} \,\rangle = \sum_{i=1}^{n} v_i^* w_i \,. \qquad (6.16b)$$

Exercise. Confirm that the scalar product given by (6.16b) satisfies the required properties of a scalar product, namely (6.5a), (6.5c), (6.5d) and (6.5e).

6.3 Linear Maps

Similarly, we can extend the theory of §4 on linear maps and matrices to vector spaces over complex numbers.

Example. Consider the linear map $\mathcal{N} : \mathbb{C}^n \to \mathbb{C}^m$. Let $\{\mathbf{e}_i\}$ be a basis of $\mathbb{C}^n$ and $\{\mathbf{f}_i\}$ be a basis of $\mathbb{C}^m$. Then under $\mathcal{N}$,

$$\mathbf{e}_j \to \mathbf{e}'_j = \mathcal{N}\mathbf{e}_j = \sum_{i=1}^{m} N_{ij}\mathbf{f}_i,$$

where $N_{ij} \in \mathbb{C}$. As before this defines a matrix, $\mathsf{N} = \{N_{ij}\}$, with respect to bases $\{\mathbf{e}_i\}$ of $\mathbb{C}^n$ and $\{\mathbf{f}_i\}$ of $\mathbb{C}^m$. N will in general be a complex $(m \times n)$ matrix.

Real linear transformations. We have observed that the standard basis, $\{\mathbf{e}_i\}$, is both a basis for $\mathbb{R}^n$ and $\mathbb{C}^n$. Consider a linear map $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^m$, and let T be the associated matrix with respect to the standard bases of both domain and range; T is a real matrix. Extend $\mathcal{T}$ to a map

$$\widehat{\mathcal{T}} : \mathbb{C}^n \to \mathbb{C}^m\,,$$

where $\widehat{\mathcal{T}}$ has the same effect as $\mathcal{T}$ on real vectors. If $\mathbf{e}_i \to \mathbf{e}'_i$, then as before

$$(\widehat{\mathsf{T}})_{ij} = (\mathbf{e}'_i)_j \quad \text{and hence} \quad \widehat{\mathsf{T}} = \mathsf{T}\,.$$

Further real components transform as before, but complex components are now also allowed. Thus if $\mathbf{v} \in \mathbb{C}^n$, then the components of $\mathbf{v}$ with respect to the standard basis transform to components of $\mathbf{v}'$ with respect to the standard basis according to

$$\mathsf{v}' = \mathsf{T}\mathsf{v}\,.$$

Maps such as $\widehat{\mathcal{T}}$ are referred to as *real linear transformations* of $\mathbb{C}^n \to \mathbb{C}^m$.

Change of bases. Under a change of bases $\{\mathbf{e}_i\}$ to $\{\widetilde{\mathbf{e}}_i\}$ and $\{\mathbf{f}_i\}$ to $\{\widetilde{\mathbf{f}}_i\}$ the transformation law (4.70) follows through for a linear map $\mathcal{N} : \mathbb{C}^n \to \mathbb{C}^m$, i.e.

$$\widetilde{\mathsf{N}} = \mathsf{C}^{-1}\mathsf{N}\mathsf{A}\,. \tag{6.17}$$

Remark. If $\mathcal{N}$ is a real linear transformation so that N is real, it is not necessarily true that $\widetilde{\mathsf{N}}$ is real, e.g. this will not be the case if we transform from standard bases to bases consisting of complex vectors.

Example. Consider the map $\mathcal{R} : \mathbb{R}^2 \to \mathbb{R}^2$ consisting of a rotation by θl; from (4.23)

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathsf{R}(\theta) \begin{pmatrix} x \\ y \end{pmatrix}\,.$$

Since diagonal matrices have desirable properties (e.g. they are straightforward to invert) we might ask whether there is a change of basis under which

$$\widetilde{\mathsf{R}} = \mathsf{A}^{-1}\mathsf{R}\mathsf{A} \tag{6.18}$$

is a diagonal matrix. One way (but emphatically not the best way) to proceed would be to [partially] expand out the right-hand side of (6.18) to obtain

$$\mathsf{A}^{-1}\mathsf{R}\mathsf{A} = \frac{1}{\det\mathsf{A}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} a_{11}\cos\theta - a_{21}\sin\theta & a_{12}\cos\theta - a_{22}\sin\theta \\ a_{11}\sin\theta + a_{21}\cos\theta & a_{12}\sin\theta + a_{22}\cos\theta \end{pmatrix},$$

and so

$$\begin{aligned} (\mathsf{A}^{-1}\mathsf{R}\mathsf{A})_{12} &= \frac{1}{\det\mathsf{A}}\left(a_{22}(a_{12}\cos\theta - a_{22}\sin\theta) - a_{12}(a_{12}\sin\theta + a_{22}\cos\theta)\right) \\ &= -\frac{\sin\theta}{\det\mathsf{A}}\left(a_{12}^2 + a_{22}^2\right), \end{aligned}$$

and

$$(\mathsf{A}^{-1}\mathsf{R}\mathsf{A})_{21} \quad = \quad \frac{\sin\theta}{\det\mathsf{A}}\left(a_{11}^2 + a_{21}^2\right) .$$

Hence $\widetilde{\mathsf{R}}$ is a diagonal matrix if $a_{12} = \pm i a_{22}$ and $a_{21} = \pm i a_{11}$. A convenient normalisation (that results in an orthonormal basis) is to choose

$$\mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} . \tag{6.19}$$

Thence from (4.61) it follows that

$$\widetilde{\mathbf{e}}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix} , \quad \widetilde{\mathbf{e}}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 1 \end{pmatrix} , \tag{6.20a}$$

where we note from using (6.16b) that

$$\langle\, \widetilde{\mathbf{e}}_1 \,,\, \widetilde{\mathbf{e}}_1 \,\rangle = 1 \,, \quad \langle\, \widetilde{\mathbf{e}}_1 \,,\, \widetilde{\mathbf{e}}_2 \,\rangle = 0 \,, \quad \langle\, \widetilde{\mathbf{e}}_2 \,,\, \widetilde{\mathbf{e}}_2 \,\rangle = 1 \,, \quad \text{i.e.} \quad \langle\, \widetilde{\mathbf{e}}_i \,,\, \widetilde{\mathbf{e}}_j \,\rangle = \delta_{ij} \,. \tag{6.20b}$$

Moreover, from (6.3), (6.18) and (6.19) it follows that, as required, $\widetilde{\mathsf{R}}$ is diagonal:

$$\begin{aligned} \widetilde{\mathsf{R}} &= \frac{1}{2}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} . \end{aligned} \tag{6.21}$$

Remark. We know from (4.59b) that real rotation matrices are orthogonal, i.e. $\mathsf{R}\mathsf{R}^{\mathrm{T}} = \mathsf{R}^{\mathrm{T}}\mathsf{R} = \mathsf{I}$; however, it is not true that $\widetilde{\mathsf{R}}\widetilde{\mathsf{R}}^{\mathrm{T}} = \widetilde{\mathsf{R}}^{\mathrm{T}}\widetilde{\mathsf{R}} = \mathsf{I}$. Instead we note from (6.21) that

$$\widetilde{\mathsf{R}}\widetilde{\mathsf{R}}^{\dagger} = \widetilde{\mathsf{R}}^{\dagger}\widetilde{\mathsf{R}} = \mathsf{I} \,, \tag{6.22a}$$

i.e.

$$\widetilde{\mathsf{R}}^{\dagger} = \widetilde{\mathsf{R}}^{-1} \,. \tag{6.22b}$$

Definition. A complex square matrix U is said to be *unitary* if its Hermitian conjugate is equal to its inverse, i.e. if

$$\mathsf{U}^{\dagger} = \mathsf{U}^{-1} \,. \tag{6.23}$$

Unitary matrices are to complex matrices what orthogonal matrices are to real matrices. Similarly, there is an equivalent for complex matrices to symmetric matrices for real matrices.

Definition. A complex square matrix A is said to be *Hermitian* if it is equal to its own Hermitian conjugate, i.e. if

$$\mathsf{A}^{\dagger} = \mathsf{A} \,. \tag{6.24}$$

Example. The metric G is Hermitian since from (6.5a), (6.9) and (6.12)

$$(\mathsf{G}^{\dagger})_{ij} = G_{ji}^{*} = \langle\, \mathbf{e}_j \,,\, \mathbf{e}_i \,\rangle^{*} = \langle\, \mathbf{e}_i \,,\, \mathbf{e}_j \,\rangle = (\mathsf{G})_{ij} \,. \tag{6.25}$$

Remark. As mentioned above, diagonal matrices have desirable properties. It is therefore useful to know what classes of matrices can be diagonalized by a change of basis. You will learn more about this in in the second part of this course (and elsewhere). For the time being we state that Hermitian matrices can always be diagonalized, as can all *normal matrices*, i.e. matrices such that $\mathsf{A}^{\dagger}\mathsf{A} = \mathsf{A}\,\mathsf{A}^{\dagger}$ . . . a class that includes skew-symmetric Hermitian matrices (i.e. matrices such that $\mathsf{A}^{\dagger} = -\mathsf{A}$) and unitary matrices, as well as Hermitian matrices.