Mathematics IA Algebra and Geometry (Part I)
Michaelmas Term 2002

lecturer: Professor Peter Haynes (`phh@damtp.cam.ac.uk`)

November 15, 2004

1 Complex Numbers

1.1 Introduction

Real numbers, (denoted by $\mathbb{R}$), consist of:

integers	(denoted by $\mathbb{Z}$) $\cdots -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ \ldots$
rationals	(denoted by $\mathbb{Q}$) p/q where $p,\ q$ are integers
irrationals	$\sqrt{2},\ \pi,\ e,\ \pi^2$ etc

It is often useful to visualise real numbers as lying on a line

Complex numbers (denoted by $\mathbb{C}$):
If $a,\, b \in \mathbb{R}$, then $z = a + ib \in \mathbb{C}$ ('$\in$'means belongs to), where i is such that $i^2 = -1$.

If $z = a + ib$, then write

$$a = Re\,(z) \qquad \text{(real part of } z)$$
$$b = Im\,(z) \quad \text{(imaginary part of } z)$$

Extending the number system from real ($\mathbb{R}$) to complex ($\mathbb{C}$) allows certain important generalisations. For example, in complex numbers the quadratic equation

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

always has two roots

$$x_1 = -\frac{\beta + \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \quad x_2 = -\frac{\beta - \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

where

$$x_1,\, x_2 \in \mathbb{R} \qquad \text{if } \beta^2 \geq 4\alpha\gamma$$
$$x_1,\, x_2 \in \mathbb{C} \qquad \text{if } \beta^2 < 4\alpha\gamma, \quad \text{when}$$

$$x_1 = -\frac{\beta}{2\alpha} + i\frac{\sqrt{4\alpha\gamma - \beta^2}}{2\alpha}, \quad x_2 = -\frac{\beta}{2\alpha} - i\frac{\sqrt{4\alpha\gamma - \beta^2}}{2\alpha}$$

Note: $\mathbb{C}$ contains all real numbers, i.e. if $a \in \mathbb{R}$ then $a + i.0 \in \mathbb{C}$.

A complex number $0 + i.b$ is said to be 'pure imaginary'

Algebraic manipulation for complex numbers: simply follow the rules for reals, adding the rule $i^2 = -1$.

Hence: addition/subtraction : $(a + ib) \pm (c + id)$

$= (a \pm c) + i\,(b \pm d)$

multiplication : $(a + ib)\,(c + id) =$

$ac + ibc + ida + (ib)\,(id)$

$= (ac - bd) + i\,(bc + ad)$

inverse : $(a + ib)^{-1} = \dfrac{a}{a^2 + b^2} - \dfrac{ib}{a^2 + b^2}$

[Check from the above that $z.z^{-1} = 1 + i.0$]

All these operations on elements of $\mathbb{C}$ result in new elements of $\mathbb{C}$ (This is described as 'closure': $\mathbb{C}$ is 'closed under addition' etc.)

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

New definitions

Complex conjugate of $z = a + ib$ is defined as $a - ib$, written as $\overline{z}$ (sometimes z^*).

The complex conjugate has the properties $\overline{z_1 \pm z_2} = \overline{z_1} \pm \overline{z_2}$, $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$, $\overline{(z^{-1})} = (\overline{z})^{-1}$.

Modulus of $z = a + ib$ defined as $(a^2 + b^2)^{1/2}$ and written as $|z|$.

Note that $|z|^2 = z\overline{z}$ and $z^{-1} = \overline{z}/(|z|^2)$.

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

Proof: Assume $\exists\ a, b, c, d$ real such that

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

Then $a - c = i\,(d - b)$, so $(a - c)^2 = -\,(d - b)^2$, so $a = c$ and $b = d$.

It follows that if $z_1 = z_2 \ : \ z_1, z_2 \in \mathbb{C}$, then $Re\,(z_1) = Re\,(z_2)$ and $Im\,(z_1) = Im\,(z_2)$.

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

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

For example, if $f(z) = pz^2 + qz + r$ with $p, q, r \in \mathbb{C}$ then $\overline{f}(\overline{z}) \equiv \overline{f(z)} = \overline{p}\,\overline{z}^2 + \overline{q}\,\overline{z} + \overline{r}$. Hence $\overline{f}(z) = \overline{p}\,z^2 + \overline{q}\,z + \overline{r}$.

This example generalises to any function defined by addition, subtraction, multiplication and inverse.

1.2 The Argand diagram

Consider the set of points in 2D referred to Cartesian axes.

We can represent each $z = x + iy \in \mathbb{C}$ by the point (x, y).

Label the 2D vector $\vec{OP}$ by the complex number z. This defines the Argand diagram (or the 'complex plane'). [Invented by Caspar Wessel (1797) and re-invented by Jean Robert Argand (1806)]

Call the x-axis, the 'real axis' and the y-axis, the 'imaginary axis'.

Modulus: the modulus of z corresponds to the magnitude of the vector $\vec{OP}$, $|z| = (x^2 + y^2)^{1/2}$.

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

Addition: if $z_1 = x_1 + iy_1$ associated with P_1, $z_2 = x_2 + iy_2$ associated with P_2, then $z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$.

$z_1 + z_2 = z_3$ is associated with the point P_3, obtained by completing the parallelogram $P_1\,O\,P_2 P_3$' i.e.' as vector addition $\vec{OP_3} = \vec{OP_1} + \vec{OP_2}$ (sometimes called the 'triangle law').

Theorem 1.2: If $z_1, z_2 \in \mathbb{C}$ then

(i) $|z_1 + z_2| \leq |z_1| + |z_2|$
(ii) $|z_1 - z_2| \geq |\,(|z_1| - |z_2|)\,|$

(i) is the triangle inequality.

By the cosine rule

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

.

(ii) follows from (i), putting $z_1 + z_2 = z_1'$, $z_2 = z_2'$, so $z_1 = z_1' - z_2'$. Hence, by (i), $|z_1'| \leq |z_1' - z_2'| + |z_2'|$ and $|z_1' - z_2'| \geq |z_1'| - |z_2'|$. Now interchanging z_1' and z_2', we have $|z_2'| - |z_1'| \leq |z_2' - z_1'| = |z_1' - z_2'|$, hence result.

Polar (modulus/argument) representation

Use plane polar co-ordinates to represent position in Argand diagram. $x = r\cos\theta$ and $y = r\sin\theta$, hence

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

Note that $|z| = (x^2 + y^2)^{1/2} = r$, so 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 called the 'modulus/argument form'.

The pair (r, θ) specifies z uniquely, but z does not specify (r, θ) uniquely, since adding $2n\pi$ to θ (n integer) does not change z. For each z there is a unique value of the argument θ such that $-\pi < \theta \leq \pi$, sometimes called the principal value of the argument.

Geometric interpretation of multiplication

Consider z_1, z_2 written in modulus argument form

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

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

Multiplication of z_2 by z_1, rotates z_2 by θ_1 and scales z_2 by $\mid z_1 \mid$

$|z_1 z_2| = |z_1|\,|z_2|$
$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2)\ (+2k\pi,$ with k an arbitrary integer.)

1.3 De Moivre's Theorem: complex exponentials

Theorem 1.3 (De Moivre's Theorem): $(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$
where $\theta \in \mathbb{R}$ and $n \in \mathbb{Z}$

For $n > 0$ prove by induction

Assume true for $n = p\ :\ (\cos\theta + i\sin\theta)^p = \cos p\theta + i\sin p\theta$

Then $\quad (\cos\theta + i\sin\theta)^{p+1} = (\cos\theta + i\sin\theta)\,(\cos\theta + i\sin\theta)^p$

$$= (\cos\theta + i\sin\theta)\,(\cos p\theta + i\sin p\theta)$$

$$= \cos\theta.\cos p\theta - \sin\theta.\sin p\theta + i\{\sin\theta.\cos p\theta + \cos\theta.\sin p\theta\}$$

$$= \cos(p+1)\,\theta + i\sin(p+1)\,\theta, \quad \text{hence true for } n = p+1$$

Trivially true for $n = 0$, hence true $\forall n$ by induction

Now consider $n < 0$, say $n = -p$

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

Hence true $\forall\, n \in \mathbb{Z}$

Exponential function: $\exp x = e^x$
define by power series $\exp x = 1 + x + x^2/2! \cdots = \sum_{n=0}^{\infty} x^n/n!$
(This series converges for all $x \in \mathbb{R}$ — see Analysis course.)

It follows from the series that $(\exp x)\,(\exp y) = \exp(x+y)$ for $x, y \in \mathbb{R}$ [exercise]
This, plus $\exp 1 = 1 + 1 + \frac{1}{2}\ \ldots$, may be used to justify the equivalence $\exp x = e^x$

Complex exponential defined by $\exp z = \sum_{n=0}^{\infty} z^n/n!$, $z \in \mathbb{C}$, series converges for all finite $|z|$
For short, write $\exp z = e^z$ as above.

Theorem 1.4

$$\exp(iw) = e^{iw} = \cos w + i \sin w \ , \ w \in \mathbb{C}$$

First consider w real,

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

Now define the complex functions
$\cos w = \sum_{n=0}^{\infty} (-1)^n w^{2n}/(2n)!$ and $\sin w = \sum_{n=0}^{\infty} (-1)^n w^{2n+1}/(2n+1)!$ for $w \in \mathbb{C}$.

Then $\exp(iw) = e^{iw} = \cos w + i \sin w \ , \ w \in \mathbb{C}$.
Similarly, $\exp(-iw) = e^{-iw} = \cos w - i \sin w$.

It follows that $\cos w = \frac{1}{2}\left(e^{iw} + e^{-iw}\right)$ and $\sin w = \frac{1}{2i}\left(e^{iw} - e^{-iw}\right)$.

Relation to modulus/argument form

Put $w = \theta \ , \ \theta \in \mathbb{R}$, then $e^{i\theta} = \cos\theta + i \sin\theta$.

Hence, $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$, with (again) $r = |z|$, $\theta = \arg z$.

Note that de Moivre's theorem

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

may be argued to follow from $e^{in\theta} = \left(e^{i\theta}\right)^n$.

Multiplication of two complex numbers:

$$z_1 z_2 = \left(r_1 e^{i\theta_1}\right)\left(r_2 e^{i\theta_2}\right) = r_1 r_2 e^{i(\theta_1+\theta_2)}$$

Modulus/argument expression for 1: consider solutions of $e^{i\theta} = 1$, hence $\cos\theta + i\sin\theta = 1$, $\cos\theta = 1$, $\sin\theta = 0$, hence $\theta = 2k\pi$, with $k \in \mathbb{Z}$, i.e.

$$e^{2k\pi i} = 1.$$

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

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

One solution is $z = 1$.

Seek more general solutions of the form $r\,e^{i\theta}$, $\left(r\,e^{i\theta}\right)^n = r^n e^{ni\theta} = 1$, hence $r = 1$, $e^{i\theta} = 1$, hence $n\theta = 2k\pi$, $k \in \mathbb{Z}$ with $0 \le \theta < 2\pi$.

$\theta = 2k\pi/n$ gives n distinct roots for $k = 0, 1, \ldots, n-1$, with $0 \leq \theta < 2\pi$.

Write $\omega = e^{2\pi\, i/n}$, then the roots of $z^n = 1$ are $1, \omega, \omega^2, \ldots, \omega^{n-1}$.

Note $\omega^n = 1$, also $\sum_{k=0}^{n-1} \omega^k = 1 + \omega + \cdots + \omega^{n-1} = 0$, because $\sum_{k=0}^{n-1} \omega^k = (\omega^n - 1)/(\omega - 1) = 0/(\omega - 1) = 0$.

Example: $z^5 = 1$.

Put $z = e^{i\,\theta}$, hence $e^{5i\,\theta} = e^{2\pi k i}$, hence $\theta = 2\pi k/5$, $k = 0, 1, 2, 3, 4$ and $\omega = e^{2\pi\, i/5}$.

Roots are $1, \omega, \omega^2, \omega^3, \omega^4$, with $1 + \omega + \omega^2 + \omega^3 + \omega^4 = 0$ (each root corresponds to a side of a pentagon).

1.4 Logarithms and complex powers

If $v \in \mathbb{R}$, $v > 0$, the complex equation $e^u = v$ has a unique real solution, $u = \log v$.

Definition: $\log z$ for $z \in \mathbb{C}$ is the solution w of $e^w = z$.

Set $w = u + iv$, $u, v \in \mathbb{R}$, then $e^{u+iv} = z = re^{i\theta}$

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

Thus, $w = \log\, z = \log\, |z| + i\, \arg\, z$, with $\arg\, z$, and hence $\log\, z$ a multivalued function.

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

$$-\pi < \arg z = Im(\log z) \leq \pi.$$

Example: if $z = -x\ ,\ x \in \mathbb{R}\ ,\ x > 0$ then $\log z = \log \mid -x \mid + i\ \arg(-x) = \log\ \mid x \mid + i\,\pi + 2ki\,\pi \quad k \in \mathbb{Z}$. The principal value of $\log(-x)$ is $\log|x| + i\pi$.

Powers

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

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

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

Note that since log z is multivalued so is z^w (arbitary multiple of $e^{2\pi\, ikw}\ ,\ k \in \mathbb{Z}$)

Example:

$(i)^i = e^{i \log\, i} = e^{i(\log|i| + i \arg\, i)} = e^{i(\log 1 + 2ki\,\pi + i\pi/2)} = e^{-\pi/2} \times e^{-2k\pi} \quad k \in \mathbb{Z}.$

1.5 Lines and circles in the complex plane

Line: For fixed z_0 and $c \in \mathbb{C}$, $z = z_0 + \lambda c$, $\lambda \in \mathbb{R}$ represents points on straight line through z_0 and parallel to c.

Note that $\lambda = (z - z_0)/c \in \mathbb{R}$, hence $\lambda = \bar{\lambda}$, so

$$\frac{z - z_0}{c} = \frac{\bar{z} - \bar{z}_0}{\bar{c}}$$

Hence

$$z\bar{c} - \bar{z}c = z_0\bar{c} - \bar{z}_0 c$$

is an alternative representation of the line.

Circle: circle radius r, centre a $\quad (r \in \mathbb{R},\ a \in \mathbb{C})$ is given by

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

the set of complex numbers z such that $|z - a| = r$.

If $a = p + iq$, $z = x + iy$ then $|z - a|^2 = (x - p)^2 + (y - q)^2 = r^2$, i.e. the expression for a circle with centre (p, q), radius r in Cartesian coordinates.

An alternative description of the circle comes from $|z - a|^2 = (\bar{z} - \bar{a})(z - a)$, so

$$z\bar{z} - \bar{a}z - a\bar{z} + \mid a \mid^2 - r^2 = 0.$$

1.6 Möbius transformations

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

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

where $a, b, c, d \in \mathbb{C}$ (all constant) and (i) c, d not both zero, (ii) a, c not both zero and (iii) $ad \neq bc$.

(i) ensures $f(z)$ finite for some z. (ii) and (iii) ensure different z map into different points. Combine all these conditions into $ad - bc \neq 0$.

$f(z)$ maps every point of the complex plane, except $z = -d/c$, into another.

Inverse: $z = (-dz' + b)/(cz' - a)$, which represents another Möbius transformation.

For every z' except a/c there is a corresponding z, thus f maps $\mathbb{C} \setminus \{-d/c\}$ to $\mathbb{C} \setminus \{a/c\}$.

Composition: consider a second Möbius transformation

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

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

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

The set of all Möbius maps is therefore closed under composition.

Examples:

(i) $(a = 1,\ c = 0,\ d = 1)$, $z' = z + b$ is translation. Lines map to parallel lines. Circles map to identical circles.

(ii)$(b = 0,\ c = 0,\ d = 1)$, $z' = az$, scales z by $|a|$ and rotates by $\arg a$ about O.

Line $z = z_0 + \lambda p \quad (\lambda \in \mathbb{R})$ becomes $z' = az_0 + \lambda ap = z_0' + \lambda c'$ — another line.

Circle $|z - q| = r$ becomes $|z'/a - q| = r$, hence $|z' - aq| = |a|r$, equivalently $|z' - q'| = r'$ — another circle.

(iii) $(a = 0,\ b = 1,\ c = 1,\ d = 0)$, $z' = \frac{1}{z}$, described as 'inversion' with respect to O.

Line $z = z_0 + \lambda p$ or $z\bar{p} - \bar{z}p = z_0\bar{p} - \bar{z}_0 p$, becomes

$$\frac{\bar{p}}{z'} - \frac{p}{\bar{z}'} = z_0\bar{p} - \bar{z}_0 p$$

hence

$$\bar{z}'\bar{p} - z'p = (z_0\bar{p} - \bar{z}_0 p)\, z'\bar{z}'$$

$$z'\bar{z}' - \frac{\bar{z}'\bar{p}}{z_0\bar{p} - \bar{z}_0 p} - \frac{z'p}{\bar{z}_0 p - z_0\bar{p}} = 0$$

$$\left| z' - \frac{\bar{p}}{z_0\bar{p} - \bar{z}_0 p} \right|^2 = \left| \frac{p}{\bar{z}_0 p - z_0\bar{p}} \right|^2$$

This is a circle through origin, except when $\bar{z}_0 p - z_0\bar{p} = 0$ (which is the condition that straight line passes through origin — exercise for reader). Then $\bar{z}'\bar{p} - z'p = 0$, i.e. a straight line through the origin.

Circle $\mid z - q \mid = r$ becomes $\mid \dfrac{1}{z'} - q \mid = r$, i.e. $|1 - qz'| = r|z'|$, hence $(1 - qz')\,(1 - \bar{q}\bar{z}') = r^2\bar{z}'z'$, hence $z'\bar{z}'\,\{|q|^2 - r^2\} - qz' - \bar{q}\bar{z}' + 1 = 0$, hence

$$\left| z' - \frac{\bar{q}}{|q|^2 - r^2} \right|^2 = \frac{|q|^2}{(|q|^2 - r^2)^2} - \frac{1}{|q|^2 - r^2}$$

$$= \frac{r^2}{(|q|^2 - r^2)^2}.$$

This is a circle centre $\bar{q}/(|q|^2 - r^2)$, radius $r/(|q|^2 - r^2)$, unless $|q|^2 = r^2$ (implying the original circle passed through the origin), when $qz' + \bar{q}\bar{z}' = 1$, i.e. a straight line.

Summary: under inversion in the origin circles/straight lines $\rightarrow$ circles, except circles/straight lines through origin $\rightarrow$ straight lines (to be explained later in course).

A general Möbius map can be generated by composition of translation, scaling and rotation, and inversion in origin.

Consider the sequence:

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

Then $z_5 = (az + b)/(cz + d)$. (Verify for yourself.)

This implies that a general Möbius map sends circles/straight lines to circles/straight lines (again see later in course for further discussion).

[This is a blank page]

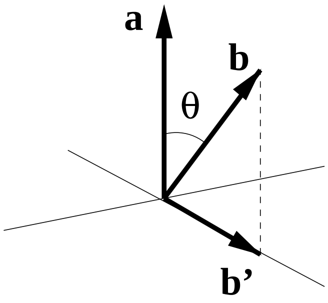
2 Vectors

2.3 Vector Product

[The printed notes are not complete for this subsection – refer to notes taken in lectures for completeness.]

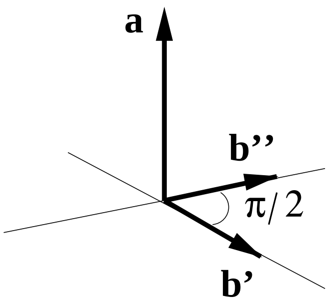
Geometrical argument for $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$.

Consider $(|\mathbf{a}|^{-1}\mathbf{a})\times\mathbf{b} = \mathbf{b}''$. This vector is the projection of $\mathbf{b}$ onto the plane perpendicular to $\mathbf{a}$, rotated by $\pi/2$ clockwise about $\mathbf{a}$. Consider this as two steps, first projection of $\mathbf{b}$ to give $\mathbf{b}'$, then rotation of $\mathbf{b}'$ to give $\mathbf{b}''$.



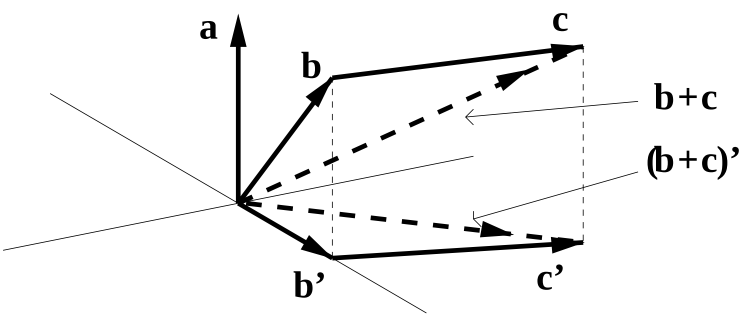
b’ is the projection of **b** onto the plane perpendicular to **a**

$|\mathbf{b}'| = |\mathbf{b}| \sin\theta$



b’’ is the result of rotating the vector **b’** through an angle $\pi/2$ clockwise about **a** (i.e looking in the direction of **a**)

Now note that if $\mathbf{x}'$ is the projection of the vector $\mathbf{x}$ onto the plane perpendicular to $\mathbf{a}$, then $\mathbf{b}'+\mathbf{c}' = (\mathbf{b}+\mathbf{c})'$. (See diagram below.)

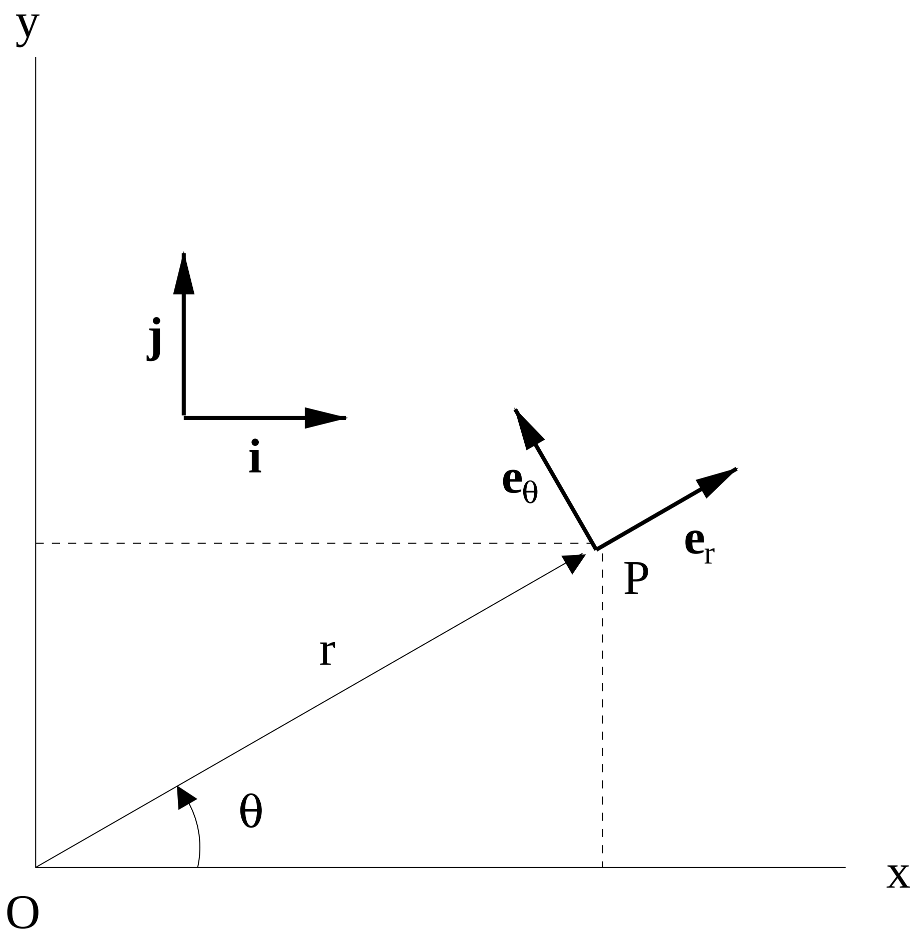


(b+c)’ is the projection of **b+c** on to the plane perpendicular to **a**

Rotating $\mathbf{b}'$, $\mathbf{c}'$ and $(\mathbf{b}+\mathbf{c})'$ by $\pi/2$ gives the required result.

2.7 Polar Coordinates

Plane polars (r, θ) **in** $\mathbb{R}^2$: $x = r\cos\theta$, $y = r\sin\theta$, with $0 \leq r < \infty$, $0 \leq \theta < 2\pi$.



$\mathbf{e}_r$ is the unit vector perpendicular to curves of constant r, in the direction of r increasing.

$\mathbf{e}_\theta$ is the unit vector perpendicular to curves of constant θ, in the direction of θ increasing.

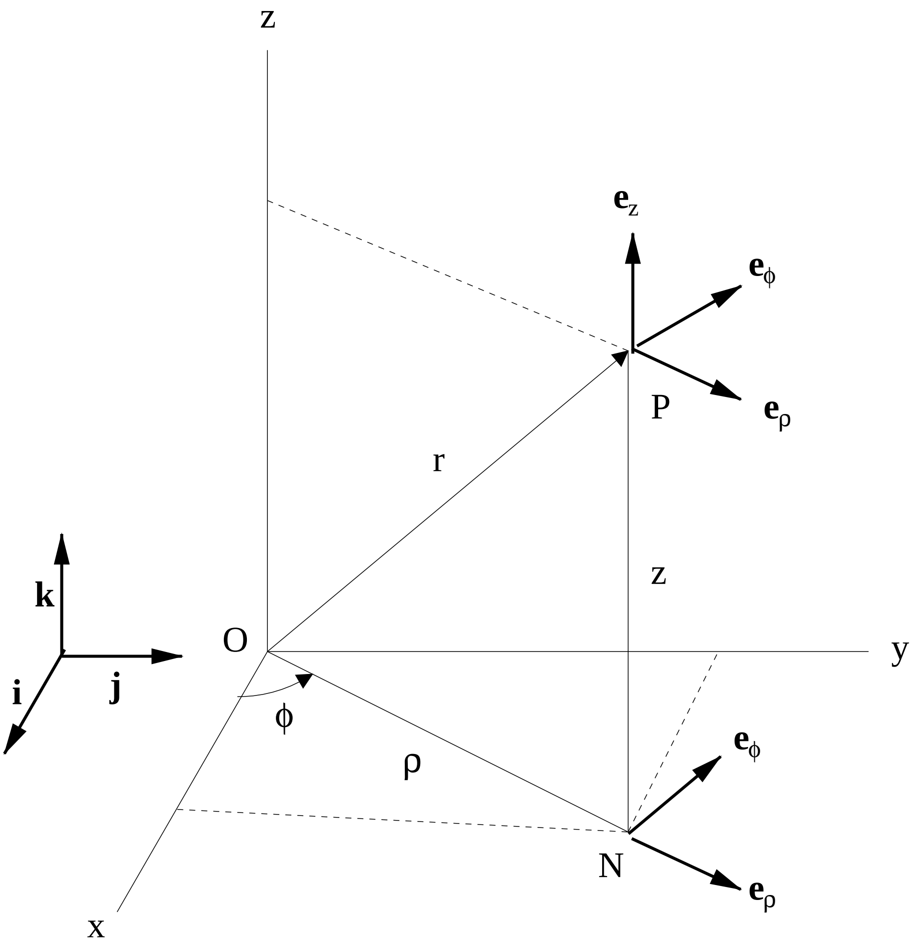
$\mathbf{e}_r = \mathbf{i}\cos\theta + \mathbf{j}\sin\theta$.

$\mathbf{e}_\theta = -\mathbf{i}\sin\theta + \mathbf{j}\cos\theta$.

$\mathbf{e}_r.\mathbf{e}_\theta = 0$.

$\mathbf{x} = \vec{OP} = x\mathbf{i} + y\mathbf{j} = r\cos\theta\mathbf{i} + r\sin\theta\mathbf{j} = r\mathbf{e}_r$.

Cylindrical polars (ρ, ϕ, z) **in** $\mathbb{R}^3$: $x = \rho\cos\phi$, $y = \rho\sin\phi$, $z = z$ with $0 \leq \rho < \infty$, $0 \leq \phi < 2\pi$, $-\infty < z < \infty$.



$\mathbf{e}_\rho$ is the unit vector perpendicular to surfaces of constant ρ, in the direction of ρ increasing.

$\mathbf{e}_\phi$ is the unit vector perpendicular to surfaces of constant ϕ, in the direction of ϕ increasing.

$\mathbf{e}_z = \mathbf{k}$

$\mathbf{e}_\rho$, $\mathbf{e}_\phi$, $\mathbf{e}_z = \mathbf{k}$ are a right-handed triad of mutually orthogonal unit vectors:

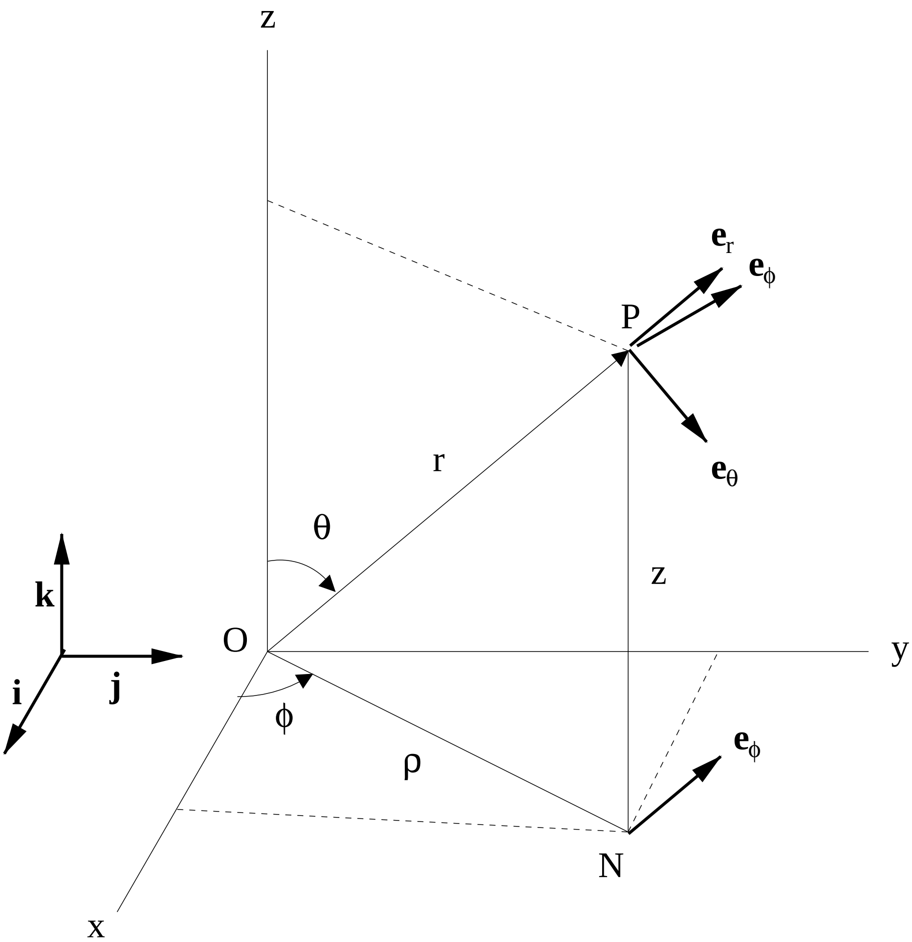
$\mathbf{e}_\rho.\mathbf{e}_\phi = 0, \qquad \mathbf{e}_z.\mathbf{e}_\rho = 0, \qquad \mathbf{e}_\phi.\mathbf{e}_z = 0.$

$\mathbf{e}_\rho \times \mathbf{e}_\phi = \mathbf{e}_z, \qquad \mathbf{e}_z \times \mathbf{e}_\rho = \mathbf{e}_\phi, \qquad \mathbf{e}_\phi \times \mathbf{e}_z = \mathbf{e}_\rho.$

$\mathbf{e}_\rho.(\mathbf{e}_\phi \times \mathbf{e}_z) = 1.$

$\mathbf{x} = \vec{ON} + \vec{NP} = \rho\mathbf{e}_\rho + z\mathbf{e}_z.$

Spherical polars (r, θ, ϕ) **in** $\mathbb{R}^3$: $x = r \sin\theta \cos\phi$, $y = r \sin\theta \sin\phi$, $z = r \cos\theta$ with $0 \le r < \infty$, $0 \le \theta \le \pi$, $0 \le \phi < 2\pi$.



$\mathbf{e}_r$ is the unit vector perpendicular to surfaces of constant r, in the direction of r increasing.

$\mathbf{e}_\theta$ is the unit vector perpendicular to surfaces of constant θ, in the direction of θ increasing.

$\mathbf{e}_\phi$ is the unit vector perpendicular to surfaces of constant ϕ, in the direction of ϕ increasing.

$\mathbf{e}_r$, $\mathbf{e}_\theta$, $\mathbf{e}_\phi$ are a right-handed triad of mutually orthogonal unit vectors:

$\mathbf{e}_r.\mathbf{e}_\theta = 0$, $\quad \mathbf{e}_\theta.\mathbf{e}_\phi = 0$, $\quad \mathbf{e}_\phi.\mathbf{e}_r = 0$.

$\mathbf{e}_r \times \mathbf{e}_\theta = \mathbf{e}_\phi$, $\quad \mathbf{e}_\theta \times \mathbf{e}_\phi = \mathbf{e}_r$, $\quad \mathbf{e}_\phi \times \mathbf{e}_r = \mathbf{e}_\theta$.

$\mathbf{e}_r.(\mathbf{e}_\theta \times \mathbf{e}_\phi) = 1$.

$\mathbf{x} = \vec{OP} = r\mathbf{e}_r$.

4 Linear Maps and Matrices

4.7 Change of basis

Consider change from standard basis of $\mathbb{R}^3$ to new basis $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3\}$, linearly independent, but not necessarily orthonormal (or even orthogonal).

Let $\mathbf{x}$ be any vector in $\mathbb{R}^3$, then

$$\mathbf{x} = \sum_{i=1}^{3} x_i \mathbf{e}_i = \sum_{k=1}^{3} \xi_k \boldsymbol{\eta}_k,$$

where $\{\xi_k\}$ are the components of $\mathbf{x}$ with respect to the new basis.

Consider $\mathbf{x}.\mathbf{e}_j$:

$$\mathbf{x}.\mathbf{e}_j = x_j = \sum_{k=1}^{3} \xi_k \boldsymbol{\eta}_k.\mathbf{e}_j = P_{jk}\xi_k$$

where P_{jk} is jth component of $\boldsymbol{\eta}_k$ (with respect to the standard basis).

We write $\mathbf{x} = P\boldsymbol{\xi}$ (where $\mathbf{x}$ and $\boldsymbol{\xi}$ are to be interpreted as column vectors whose elements are the x_i and ξ_i) where the matrix P is

$P = (\boldsymbol{\eta}_1\, \boldsymbol{\eta}_2\, \boldsymbol{\eta}_3)$ matrix with columns components of new basis vectors $\boldsymbol{\eta}_k$

Matrices are therefore a convenient way of expressing the changes in components due to a change of basis.

Since the $\boldsymbol{\eta}_k$ are a basis, there exist $E_{ki} \in \mathbb{R}$ such that $\mathbf{e}_i = \sum_{k=1}^{3} E_{ki}\boldsymbol{\eta}_k$. Hence

$$\mathbf{x} = \sum_{i=1}^{3} x_i (\sum_{k=1}^{3} E_{ki}\boldsymbol{\eta}_k) = \sum_{k=1}^{3} (\sum_{i=1}^{3} x_i E_{ki})\boldsymbol{\eta}_k = \sum_{k=1}^{3} \xi_k \boldsymbol{\eta}_k.$$

By uniqueness of components with respect to a given basis $E_{ki}x_i = \xi_k$.

Thus we have $P\boldsymbol{\xi} = \mathbf{x}$ and $E\mathbf{x} = \boldsymbol{\xi}$, for all $\mathbf{x} \in \mathbb{R}^3$, so $P\boldsymbol{\xi} = \mathbf{x} = PE\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$, hence $PE = I$. Similarly $EP = I$, and hence $E = P^{-1}$, so P is invertible.

Now consider a linear map $\mathcal{M} : \mathbb{R}^3 \to \mathbb{R}^3$ under which $\mathbf{x} \mapsto \mathbf{x}' = \mathcal{M}(\mathbf{x})$ and (in terms of column vectors) $\mathbf{x}' = M\mathbf{x}$ where $\{x_i'\}$ and $\{x_i\}$ are components with respect to the standard basis $\{\mathbf{e}_i\}$. M is the matrix of $\mathcal{M}$ with respect to the standard basis.

From above $\mathbf{x}' = P\boldsymbol{\xi}'$ and $\mathbf{x} = P\boldsymbol{\xi}$ where $\{\xi_j'\}$ and $\{\xi_j\}$ are components with respect to the new basis $\{\boldsymbol{\eta}_j\}$.

Thus $P\boldsymbol{\xi}' = MP\boldsymbol{\xi}$, hence $\boldsymbol{\xi}' = (P^{-1}MP)\boldsymbol{\xi}$.

$P^{-1}MP$ is the matrix of $\mathcal{M}$ with respect to the new basis $\{\boldsymbol{\eta}_j\}$, where $P = (\boldsymbol{\eta}_1\, \boldsymbol{\eta}_2\, \boldsymbol{\eta}_3)$, i.e. the columns of P are the components of new basis vectors with respect to old basis (and in this case the old basis is the standard basis).

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 standard basis of $\mathbb{R}^n$ and $\{\mathbf{F}_i\}$ is standard basis of $\mathbb{R}^m$, and N is matrix of $\mathcal{N}$ with respect to these two bases, so $\mathbf{X} \mapsto \mathbf{X}' = N\mathbf{X}$ (where $\mathbf{X}$ and $\mathbf{X}'$ are to be interpreted as column vectors of components).

Now consider new bases $\{\boldsymbol{\eta}_i\}$ of $\mathbb{R}^n$ and $\{\boldsymbol{\phi}_i\}$ of $\mathbb{R}^m$, with $P = (\boldsymbol{\eta}_1 \ldots \boldsymbol{\eta}_n)$ $[n \times n$ matrix$]$ and $Q = (\boldsymbol{\phi}_1 \ldots \boldsymbol{\phi}_m)$ $[m \times m$ matrix$]$.

Then $\mathbf{X} = P\boldsymbol{\xi}$, $\mathbf{X}' = Q\boldsymbol{\xi}'$, where $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ are column vectors of components with respect to bases $\{\boldsymbol{\eta}_i\}$ and $\{\boldsymbol{\phi}_i\}$ respectively.

Hence $Q\boldsymbol{\xi}' = NP\boldsymbol{\xi}$, implying $\boldsymbol{\xi}' = Q^{-1}NP\boldsymbol{\xi}$. So $Q^{-1}NP$ is matrix of transformation with respect to new bases (of $\mathbb{R}^n$ and $\mathbb{R}^m$).

Example: Consider simple shear in x_1 direction within (x_1, x_2) plane, with magnitude γ.

Matrix with respect to standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is:

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

Now consider matrix of this transformation with respect to basis $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3\}$, where

$$\begin{aligned} \boldsymbol{\eta}_1 &= \cos\psi\, \mathbf{e}_1 + \sin\psi\, \mathbf{e}_2 \\ \boldsymbol{\eta}_2 &= -\sin\psi\, \mathbf{e}_1 + \cos\psi\, \mathbf{e}_2 \\ \boldsymbol{\eta}_3 &= \mathbf{e}_3 \end{aligned}$$

Then

$$P = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{which is orthogonal, so} \quad P^{-1} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Matrix with respect to new basis is $P^{-1}MP =$

$$= \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\psi + \gamma\sin\psi & -\sin\psi + \gamma\cos\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 + \gamma\sin\psi\cos\psi & \gamma\cos^2\psi & 0 \\ -\gamma\sin^2\psi & 1 - \gamma\sin\psi\cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

[This is a blank page]

5 Determinants, Matrix Inverses and Linear Equations

5.1 Introduction

Consider linear equations in two unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= d_1 \\ a_{21}x_1 + a_{22}x_2 &= d_2 \end{aligned}$$

or equivalently, $A\mathbf{x} = \mathbf{d}$, where,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \text{ and } A = \{a_{ij}\} \ (2 \times 2 \text{ matrix}).$$

Now solve by forming suitable linear combinations of the two equations:

$$\begin{aligned} (a_{11}a_{22} - a_{21}a_{12})x_1 &= a_{22}d_1 - a_{12}d_2, \\ (a_{21}a_{12} - a_{22}a_{11})x_2 &= a_{21}d_1 - a_{11}d_2. \end{aligned}$$

We identify $a_{11}a_{22} - a_{21}a_{12}$ as $\det A$ (defined earlier).

Thus, if $\det A \neq 0$, the equations have a unique solution

$$\begin{aligned} x_1 &= (a_{22}d_1 - a_{12}d_2)/\det A, \\ x_2 &= (-a_{21}d_1 + a_{11}d_2)/\det A. \end{aligned}$$

Returning to matrix form, $A\mathbf{x} = \mathbf{d}$ implies $\mathbf{x} = A^{-1}\mathbf{d}$ (if A^{-1} exists). Thus we have that

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

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

5.2 Determinants for 3×3 and larger

For a 3×3 matrix we write

$$\begin{aligned}
\det A &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{22}a_{13}) \\
&\qquad \text{(previous definition as a triple vector product)} \\
&= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}
\end{aligned}$$

(expansion of $\det A$ in terms of elements of first column and determinants of submatrices)

We may use this as a way of defining (and evaluating) determinants of larger ($n \times n$) matrices.

Properties of determinants

(i) $\det A = \det A^T$ (follows from definition). Note that expansion of 3×3 (or larger) determinants therefore works using rows as well as columns.

(ii) We noted earlier that

$$\det \begin{pmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{pmatrix} = \epsilon_{ijk}\alpha_i\beta_j\gamma_k = \boldsymbol{\alpha}.(\boldsymbol{\beta} \times \boldsymbol{\gamma}).$$

Now write $\alpha_i = a_{i1}$, $\beta_j = a_{j2}$, $\gamma_k = a_{k3}$. Then if $A = \{a_{ij}\}$, $\det A = \epsilon_{ijk}a_{i1}a_{j2}a_{k3}$.

(iii) (Following triple product analogy) $\boldsymbol{\alpha}.(\boldsymbol{\beta} \times \boldsymbol{\gamma}) = 0$ if and only if $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are coplanar, i.e. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are linearly dependent. Similarly $\det A = 0$ if and only if there is linear dependence between the columns of A (or, from (i), the rows of A).

(iv) If we interchange any two of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ we change the sign of $\boldsymbol{\alpha}.(\boldsymbol{\beta} \times \boldsymbol{\gamma})$. Hence if we interchange any two columns of A we change the sign of $\det A$. (Similarly, from (i), if we interchange rows.)

(v) Add to any column of A linear combinations of other columns, to give $\tilde{A}$. Then $\det \tilde{A} = \det A$. [Consider $(\boldsymbol{\alpha} + \lambda\boldsymbol{\beta} + \mu\boldsymbol{\gamma}).(\boldsymbol{\beta} \times \boldsymbol{\gamma})$.] A similar result applies to rows.

(vi) Multiply any single row or column of A by λ, to give $\hat{A}$. Then $\det \hat{A} = \lambda \det A$.

(vii) $\det(\lambda A) = \lambda^3 \det A$ [or $\det(\lambda A) = \lambda^n \det A$ for $n \times n$].

Theorem 5.1: If $A = \{a_{ij}\}$ is 3×3, then $\epsilon_{pqr} \det A = \epsilon_{ijk}a_{pi}a_{qj}a_{rk}$.

Proof: (ii) above if $p = 1$, $q = 2$, $r = 3$.

If p and q are swapped then sign of left-hand side reverses and

$$\epsilon_{ijk}a_{qi}a_{pj}a_{rk} = \epsilon_{jik}a_{qj}a_{pi}a_{rk} = -\epsilon_{ijk}a_{pi}a_{qj}a_{rk},$$

so sign of right-hand side also reverses. Similarly for swaps of p and r or q and r. Hence result holds for $\{pqr\}$ any permutation of $\{123\}$.

If $p = q = 1$, say, then left-hand side is zero and

$$\epsilon_{ijk} a_{1i} a_{1j} a_{rk} = \epsilon_{jik} a_{1j} a_{1i} a_{rk} = -\epsilon_{ijk} a_{1i} a_{1j} a_{rk},$$

hence right-hand side is zero. Similarly for any case where any pair of p, q and r are equal. Hence result.

Theorem 5.2 $\det AB = (\det A)(\det B)$ (with A and B both 3×3 matrices).

Proof

$$\begin{aligned}
\det AB &= \epsilon_{ijk} (AB)_{i1} (AB)_{j2} (AB)_{k3} \\
&= \epsilon_{ijk} a_{ip} b_{p1} a_{jq} b_{q2} a_{kr} b_{r3} \\
&= \epsilon_{pqr} \det A \, b_{p1} b_{q2} b_{r3} \text{ (by Theorem 5.1)} \\
&= \det A \, \det B
\end{aligned}$$

Theorem 5.3 If A is orthogonal then $\det A = \pm 1$.

Proof: $AA^T = I$ implies $\det(AA^T) = \det I = 1$, which implies $\det A \, \det A^T = (\det A)^2 = 1$, hence $\det A = \pm 1$. (Recall earlier remarks on reflections and rotations.)

5.3 Inverse of a 3×3 matrix

Define the cofactor Δ_{ij} of the ijth element of square matrix A as

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

where M_{ij} is the (square) matrix obtained by eliminating the ith row and the jth column of A.

We have

$$\begin{aligned}
\det A &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\
&= a_{11}\Delta_{11} + a_{12}\Delta_{12} + a_{13}\Delta_{13} = a_{1j}\Delta_{1j}.
\end{aligned}$$

Similarly, noting that

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

we have

$$\begin{aligned}\det A &= a_{21}\begin{vmatrix} a_{32} & a_{33} \\ a_{12} & a_{13}\end{vmatrix} - a_{22}\begin{vmatrix} a_{31} & a_{33} \\ a_{11} & a_{13}\end{vmatrix} + a_{23}\begin{vmatrix} a_{31} & a_{32} \\ a_{11} & a_{12}\end{vmatrix} \\ &= a_{21}\Delta_{21} + a_{22}\Delta_{22} + a_{23}\Delta_{23} = a_{2j}\Delta_{2j}. \\ &= a_{31}\Delta_{31} + a_{32}\Delta_{32} + a_{33}\Delta_{33} = a_{3j}\Delta_{3j} \text{ (check).}\end{aligned}$$

Similarly $\det A = a_{j1}\Delta_{j1} = a_{j2}\Delta_{j2} = a_{j3}\Delta_{j3}$, but

$$a_{2j}\Delta_{1j} = \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{vmatrix} = 0$$

(since rows are linearly independent).

Theorem 5.4 $a_{ji}\Delta_{ki} = \det A\,\delta_{jk}$ (by above).

Theorem 5.5 Given 3×3 matrix A with $\det A \neq 0$, define B by

$$(B)_{ki} = (\det A)^{-1}\Delta_{ik},$$

then $AB = BA = I$.

Proof:

$$(AB)_{ij} = a_{ik}(B)_{kj} = (\det A)^{-1}a_{ik}\Delta_{jk} = (\det A)^{-1}\det A\,\delta_{ij} = \delta_{ij}.$$

Hence $AB = I$. Similarly $BA = I$ (check). It follows that $B = A^{-1}$ and A is invertible.

(Above is formula for inverse. A similar result holds for $n \times n$ matrices, including 2×2.

Example: consider

$$S = \begin{pmatrix} 1 & \gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1\end{pmatrix} \text{ representing simple shear.}$$

Then $\det S = 1$ and

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

Hence

$$S^{-1} = \begin{pmatrix} \Delta_{11} & \Delta_{21} & \Delta_{31} \\ \Delta_{12} & \Delta_{22} & \Delta_{32} \\ \Delta_{13} & \Delta_{23} & \Delta_{33}\end{pmatrix} = \begin{pmatrix} 1 & -\gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1\end{pmatrix}.$$

(Effect of shear is reversed by changing the sign of γ.)

5.4 Solving linear equations: Gaussian elimination

One approach to solving equations $A\mathbf{x} = \mathbf{d}$ (with A $n \times n$ matrix, $\mathbf{x}$ and $\mathbf{d}$ $n \times 1$ column vectors of unknowns and right-hand sides respectively) numerically would be to calculate A^{-1} using the method given previously (extended to $n \times n$), and then $A^{-1}\mathbf{d}$. This is actually very inefficient.

Alternative is Gaussian elimination, illustrated here for 3×3 case.

We have

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= d_1 \qquad (1) \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= d_2 \qquad (2) \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= d_3 \qquad (3) \end{aligned}$$

Assume $a_{11} \neq 0$, otherwise re-order, otherwise stop (since no unique solution). Then (1) may be used to eliminate x_1:

$$x_1 = (d_1 - a_{12}x_2 - a_{13}x_3)/a_{11}.$$

Now (2) becomes:

$$\begin{aligned} (a_{22} - \frac{a_{21}}{a_{11}}a_{12})x_2 \quad &+(a_{23} - \frac{a_{21}}{a_{11}}a_{13})x_3 &&= d_2 - \frac{a_{21}}{a_{11}}d_1 \\ a'_{22}x_2 \quad &+a'_{23}x_3 &&= d'_2 \qquad (2') \end{aligned}$$

and (3) becomes

$$\begin{aligned} (a_{32} - \frac{a_{31}}{a_{11}}a_{12})x_2 \quad &+(a_{33} - \frac{a_{31}}{a_{11}}a_{13})x_3 &&= d_3 - \frac{a_{31}}{a_{11}}d_1 \\ a'_{32}x_2 \quad &+a'_{33}x_3 &&= d'_3 \qquad (3') \end{aligned}$$

Assume $a'_{22} \neq 0$, otherwise reorder, otherwise stop. Use $(2')$ to eliminate x_2 from $(3')$ to give

$$(a'_{33} - \frac{a'_{32}}{a'_{22}}a'_{23})x_3 = a''_{33}x_3 = d'_3 - \frac{a'_{32}}{a'_{22}}d'_2 \qquad (3'')$$

Now, providing $a''_{33} \neq 0$, $(3'')$ gives x_3, then $(2')$ gives x_2, then (1) gives x_1.

This method fails only if A is not invertible, i.e. only if $\det A = 0$.

5.5 Solving linear equations

If $\det A \neq 0$ then the equations $A\mathbf{x} = \mathbf{d}$ have a unique solution $\mathbf{x} = A^{-1}\mathbf{d}$. (This is a corollary to Theorem 5.5.).

What can we say about the solution if $\det A = 0$? (As usual we consider A to be 3×3.)

$A\mathbf{x} = \mathbf{d}$ ($d \neq \mathbf{0}$) is a set of inhomogeneous equations.

$A\mathbf{x} = \mathbf{0}$ is the corresponding set of homogeneous equations (with the unique solution $A^{-1}\mathbf{0} = \mathbf{0}$ if $\det A \neq 0$.

We first consider the homogeneous equations and then return to the inhomogeneous equations.

(a) geometrical view

Write $\mathbf{r}_i$ as the vector with components equal to the elements of the ith row of A, for $i = 1, 2, 3$. Then the above equations may be expressed as

$$\mathbf{r}_i.\mathbf{x} = d_i \quad (i = 1, 2, 3) \quad \text{Inhomogeneous equations}$$

$$\mathbf{r}_i.\mathbf{x} = 0 \quad (i = 1, 2, 3) \quad \text{Homogeneous equations}$$

Each individual equation 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 the three planes each pass through O. There are three possibilities:

(i). intersection only at O.

(ii). three planes have a common line (including O).

(iii). three planes coincide.

If $\det A \neq 0$ then $\mathbf{r}_1.(\mathbf{r}_2 \times \mathbf{r}_3) \neq 0$ and the set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ is linearly independent, with span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = \mathbb{R}^3$. The intersection of the planes $\mathbf{r}_1.\mathbf{x} = 0$ and $\mathbf{r}_2.\mathbf{x} = 0$ is the line $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{k}, \lambda \in \mathbb{R}, \mathbf{k} = \mathbf{r}_1 \times \mathbf{r}_2\}$. Then $\mathbf{r}_3.\mathbf{x} = 0$ implies $\lambda = 0$, hence $\mathbf{x} = \mathbf{0}$ and the three planes intersect only at the origin, i.e. case (i).

If $\det A = 0$ then the set $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$ is linearly independent, with dim span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = 2$ or 1. Assume 2. Then without loss of generality, $\mathbf{r}_1$ and $\mathbf{r}_2$ are linearly independent and the intersection of the planes $\mathbf{r}_1.\mathbf{x} = 0$ and $\mathbf{r}_2.\mathbf{x} = 0$ is the line $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{k}, \lambda \in \mathbb{R}, \mathbf{k} = \mathbf{r}_1 \times \mathbf{r}_2\}$. Since $\mathbf{r}_1.(\mathbf{r}_2 \times \mathbf{r}_3) = 0$, all points in this line satisfy $\mathbf{r}_3.\mathbf{x} = 0$ and the intersection of the three planes is a line, i.e. case (ii).

Otherwise dim span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = 1$, $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are all parallel and $\mathbf{r}_1.\mathbf{x} = 0$ implies $\mathbf{r}_2.\mathbf{x} = 0$ and $\mathbf{r}_3.\mathbf{x} = 0$, so the intersection of the three planes is a plane, i.e. case (iii). (We may write the plane as $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \lambda\mathbf{k} + \mu\mathbf{l}, \lambda, \mu \in \mathbb{R}\}$, for any two linearly independent vectors $\mathbf{k}$ and $\mathbf{l}$ such that $\mathbf{k}.\mathbf{r}_i = \mathbf{l}.\mathbf{r}_i$, $i = 1, 2, 3$.)

(b) Linear mapping view

Consider the linear map $T_A : \mathbb{R}^3 \to \mathbb{R}^3$, such that $\mathbf{x} \mapsto \mathbf{x}' = A\mathbf{x}$. ($A$ is the matrix of T_A with respect to the standard basis.)

Kernel $K(T_A) = \{\mathbf{x} \in \mathbb{R}^3 : A\mathbf{x} = \mathbf{0}\}$, thus $K(T_A)$ is the 'solution space' of $A\mathbf{x} = \mathbf{0}$, with dimension $n(T_A)$.

Case (i) applies if $n(T_A) = 0$.
Case (ii) applies if $n(T_A) = 1$.
Case (iii) applies if $n(T_A) = 2$.

We now use the fact that if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is basis for $\mathbb{R}^3$, then $I(T_A) = \text{span}\{T_A\mathbf{u}, T_A\mathbf{v}, T_A\mathbf{w}\}$

Consider the different cases:

(i). $\lambda T_A\mathbf{u} + \mu T_A\mathbf{v} + \nu T_A\mathbf{w} = \mathbf{0}$ implies that $T_A(\lambda\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w}) = \mathbf{0}$, which implies $\lambda\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w} = \mathbf{0}$, hence $\lambda = \mu = \nu = 0$. Hence $\{T_A\mathbf{u}, T_A\mathbf{v}, T_A\mathbf{w}\}$ is linearly independent and $r(T_A) = 3$.

(ii). Without loss of generality choose $\mathbf{u} \in K(T_A)$, then $T_A\mathbf{u} = \mathbf{0}$. Consider $\text{span}\{T_A\mathbf{v}, T_A\mathbf{w}\}$. $\mu T_A\mathbf{v} + \nu T_A\mathbf{w} = \mathbf{0}$ implies $T_A(\mu\mathbf{v} + \nu\mathbf{w}) = \mathbf{0}$, hence $\exists\alpha \in \mathbb{R}$ such that $\mu\mathbf{v} + \nu\mathbf{w} = \alpha\mathbf{u}$, hence $-\alpha\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w} = \mathbf{0}$ and hence $\alpha = \mu = \nu = 0$ and $T_A\mathbf{v}$ and $T_A\mathbf{w}$ are linearly independent. Thus $\dim\ \text{span}\ \{T_A\mathbf{u}, T_A\mathbf{v}, T_A\mathbf{w}\} = r(T_A) = 2$.

(iii). Without loss of generality choose linearly independent $\mathbf{u}, \mathbf{v} \in K(T_A)$, then $A\mathbf{w} \neq \mathbf{0}$ and $\dim\ \text{span}\ \{T_A\mathbf{u}, T_A\mathbf{v}, T_A\mathbf{w}\} = r(T_A) = 1$.

Remarks: (a) In each of cases (i), (ii) and (iii) we have $n(T_A) + r(T_A) = 3$ (an example of the 'rank-nullity' formula).

(b) In each case we also have $r(T_A) = \dim\ \text{span}\ \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = r(T_A) =$ number of linearly independent rows of A ('row rank'). But $r(T_A) = \dim\ \text{span}\ \{A\mathbf{e}_1, A\mathbf{e}_2, A\mathbf{e}_3\}$ (choosing standard basis) = number of linearly independent columns of A ('column rank').

Implication for inhomogeneous equation $A\mathbf{x} = \mathbf{d}$.

If $\det A \neq 0$ then $r(T_A) = 3$ and $I(T_A) = \mathbb{R}^3$. Since $\mathbf{d} \in \mathbb{R}^3$, $\exists\mathbf{x} \in \mathbb{R}^3$ for which $\mathbf{d}$ is image under T_A, i.e. $\mathbf{x} = A^{-1}\mathbf{d}$ exists and unique.

If $\det A = 0$ then $r(T_A) < 3$ and $I(T_A)$ is a proper subspace (of $\mathbb{R}^3$). Then *either* $\mathbf{d} \notin I(T_A)$, in which there are no solutions and the equations are inconsistent *or* $\mathbf{d} \in I(T_A)$, 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 $\mathbf{d} \in I(T_A)$ then the general solution to $A\mathbf{x} = \mathbf{d}$ can be written as $\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$ where $\mathbf{x}_0$ is a particular fixed solution of $A\mathbf{x} = \mathbf{d}$ and $\mathbf{y}$ is the general solution of $A\mathbf{x} = \mathbf{0}$.

Proof: $A\mathbf{x}_0 = \mathbf{d}$ and $A\mathbf{y} = \mathbf{0}$, hence $A(\mathbf{x}_0 + \mathbf{y}) = \mathbf{d} + \mathbf{0} = \mathbf{d}$. If

(i). $n(T_A) = 0$, $r(T_A) = 3$, then $\mathbf{y} = \mathbf{0}$ and the solution is unique.

(ii). $n(T_A) = 1$, $r(T_A) = 2$, then $\mathbf{y} = \lambda\mathbf{k}$ and $\mathbf{x} = \mathbf{x}_0 + \lambda\mathbf{k}$ (representing a line).

(iii). $n(T_A) = 2$, $r(T_A) = 1$, then $\mathbf{y} = \lambda\mathbf{k} + \mu\mathbf{l}$ and $\mathbf{x} = \mathbf{x}_0 + \lambda\mathbf{k} + \mu\mathbf{l}$ (representing a plane).

Example: (2×2) case.

$$\begin{aligned} A\mathbf{x} &= \mathbf{d} \\ \begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 1 \\ b \end{pmatrix} \end{aligned}$$

$\det A = 1 - a$.

If $a \neq 1$, then $\det A \neq 0$ and so A^{-1} exists and is unique.

$$A^{-1} = \frac{1}{1-a}\begin{pmatrix} 1 & -1 \\ -a & 1 \end{pmatrix} \text{ and the unique solution is } A^{-1}\begin{pmatrix} 1 \\ b \end{pmatrix}$$

If $a = 1$, then $\det A = 0$.

$$A\mathbf{x} = \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix}, \ I(T_A) = \text{span}\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\} \text{ and } K(T_A) = \text{span}\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\}$$

so $r(T_A) = 1$ and $n(T_A) = 1$.

If $b \neq 1$ then $\begin{pmatrix} 1 \\ b \end{pmatrix} \notin I(T_A)$ and there are no solutions (equations inconsistent).

If $b = 1$ then $\begin{pmatrix} 1 \\ b \end{pmatrix} \in I(T_A)$ and solutions exist (equations consistent).

A particular solution is $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

The general solution is $\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbf{y}$ where $\mathbf{y}$ is any vector in $K(T_A)$.

Hence the general solution is $\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ where $\lambda \in \mathbb{R}$.

6 Complex vector spaces ($\mathbb{C}^n$)

6.1 Introduction

We have considered vector spaces with scalars $\in \mathbb{R}$.

Now generalise to vector spaces with scalars $\in \mathbb{C}$.

Definition $\mathbb{C}^n$ is set of n-tuples of complex numbers, i.e. for $z \in \mathbb{C}^n$, $z = (z_1, z_2, \ldots, z_n)$, $z_i \in \mathbb{C}$, $i = 1, \ldots, n$.

By extension from $\mathbb{R}^n$, define vector addition and scalar multiplication: for $z = (z_1, \ldots, z_n)$, $\zeta = (\zeta_1, \ldots, \zeta_n) \in \mathbb{C}^n$ and $c \in \mathbb{C}$.

$$z + \zeta = (z_1 + \zeta_1, z_2 + \zeta_2, \ldots, z_n + \zeta_n),$$

$$cz = (cz_1, cz_2, \ldots, cz_n).$$

Check of A1-4, B1-4 from §3 shows that $\mathbb{C}^n$ is a vector space over $\mathbb{C}$.

Note that $\mathbb{R}^n$ is a subset of $\mathbb{C}^n$, but *not* a subspace of $\mathbb{C}^n$ (as a vector space over $\mathbb{C}$) since $\mathbb{R}^n$ is not closed under multiplication by an arbitrary complex number.

$\mathbb{C}^n$ has dimension n as a vector space over $\mathbb{C}$ since the standard basis of $\mathbb{R}^n$ (as a vector space over $\mathbb{R}$) is also a basis of $\mathbb{C}^n$ (as a vector space over $\mathbb{C}$).

6.2 Linear mappings

Consider $\mathcal{M} : \mathbb{C}^n \to \mathbb{C}^m$.

Let $\{\mathbf{e}_i\}$ be the standard basis of $\mathbb{C}^n$ and $\{\mathbf{f}_i\}$ be the standard basis of $\mathbb{C}^m$.

Then under $\mathcal{M}$,

$$\mathbf{e}_j \to \mathbf{e}'_j = \mathcal{M}\mathbf{e}_j = \sum_{i=1}^{m} M_{ij}\mathbf{f}_i,$$

where $M_{ij} \in \mathbb{C}$. As before this defines matrix of $\mathcal{M}$ with respect to bases $\{\mathbf{e}_i\}$ of $\mathbb{C}^n$ and $\{\mathbf{f}_i\}$ of $\mathbb{C}^m$. M is a complex $(m \times n)$ matrix.

New definitions: square complex matrix M is *Hermitian* if $\overline{M^T} = M$, *unitary* if $\overline{M^T} = M^{-1}$.

6.3 Scalar product for $\mathbb{C}^n$

If we retain $z.\zeta = \sum_{i=1}^{n} z_i\zeta_i$ for $z, \zeta \in \mathbb{C}^n$ then we lose $z.z \in \mathbb{R}$ and hence $z.z \geq 0$.

Natural extension of scalar product is for z, $\zeta \in \mathbb{C}^n$,

$$z.\zeta = \text{ (new notation) } \langle z, \zeta \rangle = \sum_{i=1}^{n} \overline{z}_i \zeta_i.$$

Note that $\langle \zeta, z \rangle = \overline{\langle z, \zeta \rangle} \neq \langle z, \zeta \rangle$ (in general), but $\langle z, z \rangle \in \mathbb{R}$ and $\langle z, z \rangle > 0$ if $z \neq 0$.

Hence use this new scalar product as a definition of length or norm:

$$||z|| = \langle z, z\rangle^{1/2} = (\sum_{i=1}^{n} |z_i|^2)^{1/2} \text{ for } z \in \mathbb{C}^n.$$

[Exercise: Consider C^n as a vector space over $\mathbb{R}$. Note that $\mathbb{R}^n$ is subspace and $\dim \mathbb{C}^n = 2n$.]