

# Vectors and Matrices \*

Zhiyuan Bai

Compiled on October 2, 2020

This document serves as a set of revision materials for the Cambridge Mathematical Tripos Part IA course *Vectors and Matrices* in Michaelmas 2019. However, despite its primary focus, readers should note that it is NOT a verbatim recall of the lectures, since the author might have made further amendments in the content. Therefore, there should always be provisions for errors and typos while this material is being used.

## Contents

<b>1</b>	<b>The Field of Complex Numbers</b>	<b>2</b>
1.1	Basic Definitions . . . . .	2
1.2	Exponential and Trigonometric Functions . . . . .	4
1.3	Roots of Unity . . . . .	5
1.4	Logarithms and Complex Powers . . . . .	5
1.5	Lines and Circles . . . . .	5
<b>2</b>	<b>Vectors in 3-Dimensional Euclidean Space</b>	<b>6</b>
2.1	Vector addition and Scalar Multiplication . . . . .	6
2.2	Linear Combination and Span . . . . .	7
2.3	Scalar Dot Product . . . . .	7
2.4	Orthonormal Basis . . . . .	7
2.5	Vector/Cross Product . . . . .	8
2.6	Triple Products . . . . .	8
2.6.1	Scalar Triple Product . . . . .	8
2.6.2	Vector Triple Product . . . . .	9
2.7	Lines, Planes and Vector Equations . . . . .	9
2.7.1	Lines . . . . .	9
2.7.2	Plane . . . . .	9
2.7.3	Vector equations . . . . .	10
2.8	Index/Suffix Notation and the Summation Convention . . . . .	10
2.8.1	Index Notation . . . . .	10
2.8.2	Summation Convention . . . . .	11

---

\*Based on the lectures under the same name taught by Dr. J. M. Evans in Michaelmas 2019.

<b>3</b>	<b>Vectors in Euclidean Spaces of High Dimensions</b>	<b>12</b>
3.1	Vectors in the Real Vector spaces . . . . .	12
3.2	Axioms of Real Vector Spaces . . . . .	13
3.3	Basis and Dimension . . . . .	14
3.4	Vectors in the Complex Vector Space . . . . .	15
<b>4</b>	<b>Matrices and Linear Maps</b>	<b>16</b>
4.1	Definitions . . . . .	16
4.2	Matrices as Linear Maps in Real Vector Spaces . . . . .	17
4.3	Geometrical Examples . . . . .	18
4.4	Matrices in General; Matrix Algebra . . . . .	19
4.4.1	Transpose and Hermitian Conjugate . . . . .	20
4.4.2	Trace . . . . .	20
4.4.3	Decomposition . . . . .	20
4.4.4	Orthogonal and Unitary Matrices . . . . .	20
<b>5</b>	<b>Determinants and Inverses</b>	<b>21</b>
5.1	Introduction . . . . .	21
5.2	Alternating forms . . . . .	22
5.3	Properties of Determinants . . . . .	23
5.4	Minors, Cofactors and Inverses . . . . .	25
5.5	System of Linear Equations . . . . .	25
5.6	Gaussian Elimination and Echelon Form . . . . .	27
<b>6</b>	<b>Eigenvalues and Eigenvectors</b>	<b>27</b>
6.1	Diagonalizability . . . . .	29
6.2	Diagonalisation of Hermitian and Symmetric Matrices . . . . .	31
6.3	Quadratic Forms . . . . .	33
6.4	Cayley-Hamilton Theorem . . . . .	33
<b>7</b>	<b>Change of Bases, Canonical Forms and Symmetries</b>	<b>34</b>
7.1	Change of Bases in General . . . . .	34
7.2	Jordan Canonical/Normal Form . . . . .	35
7.3	Quadrics and Conics . . . . .	36
7.4	Symmetries and Transformation Groups . . . . .	37
7.5	2-Dimensional Minkowski Space and Lorentz Transformations . . . . .	38

# 1 The Field of Complex Numbers

## 1.1 Basic Definitions

We construct the complex numbers in the following way:

**Definition 1.1.** Consider the plane  $\mathbb{R}^2$ , we equip it with the *complex multiplication*  $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$  in a way that:

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc)$$

If we denote  $(x, y)$  by  $x + iy$ , the resulting field  $(\mathbb{R}^2, +, \cdot, 1, 0)$  is called the complex numbers, and is denoted by  $\mathbb{C}$ .

**Proposition 1.1.**  $\mathbb{C}$  is indeed a field.

*Proof.* Trivial. □

Note that  $i^2 = -1$

**Definition 1.2.** The conjugate  $\bar{z}$  or  $z^*$  of  $z = x + iy \in \mathbb{C}$  is defined as  $x - iy$ . The modulus  $|z|$  of  $z$  is defined as  $\sqrt{z\bar{z}} = \sqrt{x^2 + y^2}$ . The argument  $\arg z$  of  $z$  is the angle  $\theta$  such that  $z = |z|(\cos \theta + i \sin \theta)$ , taken mod  $2\pi$ .

The last expression is called the polar form of a complex number. It is obvious that  $|zw| = |z||w|$  and that  $|z| = |\bar{z}|$ .

**Claim.** The argument of any complex number  $z$  is defined.

*Proof.* Trivial. □

It is worth to note that  $\tan \theta = y/x$ . Although there are infinitely many angles that can make the equality in the polar form, we often take the principal value, i.e. within  $(-\pi, \pi]$ . But it is the most convenient to take it as a value in  $\mathbb{R}/2\pi\mathbb{Z}$ .

Of course, the complex numbers inherits the geometrical meanings of the plane  $\mathbb{R}^2$ . The geometric representation of it is called the Argand diagram. On the Argand diagram, the argument is essentially the (anticlockwise) angle between the vector representing the complex number and the positive real axis. The modulus, at the same time, is the length of that vector. The addition and subtraction of the complex numbers are the same as what we did it with 2D vectors (i.e. parallelogram law).

There is an important theorem associated with the polynomial in the complex numbers.

**Theorem 1.2** (Fundamental Theorem of Algebra). *Any nonconstant polynomial in  $\mathbb{C}$  has a root.*

One should note easily that it is equivalent to say that a nonconstant complex polynomial of order  $n$  has exactly  $n$  roots. The theorem means that the process of field extensions ends at  $\mathbb{C}$ .

*Proof.* Later. □

**Proposition 1.3** (Triangle Inequality).  $|z + w| \leq |z| + |w|$

*Proof.* Since both sides are positive, it is equivalent to its squared form:

$$(z + w)(\bar{z} + \bar{w}) \leq z\bar{z} + w\bar{w} + 2|z||w| \iff \frac{1}{2}(z\bar{w} + \bar{z}w) \leq |z\bar{w}|$$

But this is just to say that  $\Re(z\bar{w}) \leq |z\bar{w}|$ , which is true. □

**Corollary 1.4.** Replacing  $w$  by  $w - z$  gives  $|w - z| \geq |w| - |z|$ . By symmetry  $|w - z| \geq |z| - |w|$ , so we have the general form

$$|w - z| \geq ||z| - |w||$$

**Proposition 1.5.** Let  $z_1 = r_1(\cos \theta_1 + i \sin \theta_1)$  and  $z_2 = r_2(\cos \theta_2 + i \sin \theta_2)$ , then

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

That is,  $\arg z_1 + \arg z_2 = \arg z_1 z_2 \pmod{2\pi}$

*Proof.* Just compound angle formula suffices. It is known to be the De Moivre's Theorem.  $\square$

**Corollary 1.6.**  $(\cos \theta + i \sin \theta)^n = \cos(n\theta) + i \sin(n\theta)$

*Proof.* Induction shows the case  $n \in \mathbb{N}$ , for negative  $n = -m$ , we have

$$\text{LHS} = (\cos(m\theta) + i \sin(m\theta))^{-1} = \cos(m\theta) - i \sin(m\theta) = \text{RHS}$$

That establishes it  $\square$

## 1.2 Exponential and Trigonometric Functions

**Definition 1.3.**

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

The series converges for all  $z$  since it absolutely converges. Also due to absolute convergence, we can multiply and arrange the series, which gives

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

Note as well that  $e^0 = 1$  and  $(e^z)^n = e^{nz}$  for  $n \in \mathbb{Z}$ .

**Definition 1.4.**  $\cos(z) = (e^{iz} + e^{-iz})/2$ , which gives the series

$$\sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

Similarly  $\sin(z) = (e^{iz} - e^{-iz})/(2i)$ , so its series expansion is

$$\sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

By differentiating the series term by term,

$$(\sin z)' = \cos z, (\cos z)' = -\sin z, (e^z)' = e^z$$

**Theorem 1.7.**

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

Note that  $\cos z$  is not necessarily real, same for  $\sin z$ . But if  $z$  is real, they are.

**Lemma 1.8.**  $e^z = 1 \iff z = 2in\pi$  for some  $n \in \mathbb{Z}$ .

*Proof.* Write  $z = x + iy$ , then we have

$$e^{x+iy} = e^x e^{iy} = e^x (\cos y + i \sin y), x, y \in \mathbb{R}$$

So  $e^x \cos y = 1$  and  $e^x \sin y = 0$ . Solving it gives  $x = 0, y = 0 \pmod{2\pi i}$   $\square$

We have the following general form of complex number

$$z = |z| e^{i \arg z}$$

### 1.3 Roots of Unity

**Definition 1.5.**  $z \in \mathbb{C}$  is called an  $n^{\text{th}}$  root of unity if  $z^n = 1$

To find all solutions to  $z^n = 1$ , we write

$$z = re^{i\theta}$$

so  $r^n = 1$  and  $iN\theta = 2\pi in$  for some  $n \in \mathbb{Z}$ . This gives  $n$  distinct solutions:

$$z = e^{2\pi in/N}$$

where  $n \in \{0, 1, 2, \dots, n-1\}$ . The roots of unity lie on the unit circle on the Argand diagram. They are the vertices of a regular  $n$ -gon.

### 1.4 Logarithms and Complex Powers

**Definition 1.6.** Define  $w = \log z$  by  $e^w = z$  since we want  $\log$  to be the inverse of  $\exp$ , which is not injective. So  $\log$  is multi-valued.

$$\log z = \log |z| + i \arg z \pmod{2\pi i}$$

We can, of course, make it single-valued by restricting  $\arg z$  to the principal branch  $(-\pi, \pi]$  or  $[0, 2\pi)$ . But in this case, we do not have  $\log(ab) = \log a + \log b$

In fact, one can prove that it is impossible to choose a  $\log$  in the complex plane that lives up to every one of our expectations.

**Example 1.1.** If  $z = -1$ , then  $\log z = i\pi \pmod{2\pi i}$ .

**Definition 1.7.** We define  $z^\alpha = e^{\alpha \log z}$  for any  $\alpha, z \in \mathbb{C}$  where  $z \neq 0$ . Note that since  $\log$  is multi-valued, the complex powers are multi-valued in general. They differ by a multiplicative factor in the form  $e^{2n\pi i \alpha}$ ,  $n \in \mathbb{Z}$ .

If  $\alpha \in \mathbb{Z}$ , then the power is single-valued. If  $\alpha \in \mathbb{Q}$ , it is finite-valued. But in general, a complex power admits infinitely many values.

**Example 1.2.** 1. We want to calculate  $i^i$ .  $\log i = \pi/2 + 2\pi ni$ , so

$$i^i = e^{i \log i} = e^{-\pi/2 + 2\pi n}$$

where  $n \in \mathbb{Z}$ .

2. We want to calculate  $(1+i)^{1/2}$ , so it equals

$$e^{1/2 \log \sqrt{2} + i(\pi/4 + 2n\pi)} = 2^{1/4} e^{i\pi/8} / 2$$

### 1.5 Lines and Circles

For a fixed  $w \in \mathbb{C}$  such that  $w \neq 0$ , the set of points  $z = \lambda w$  is a line through the origin in the direction of  $w$ .

By shifting  $z = z_0 + \lambda w$  is a line parallel to  $z = \lambda w$  through  $z_0$ .

To write this in the form without the real parameter  $\lambda$ , we take the conjugate

$$\bar{z} = \bar{z}_0 + \lambda \bar{w}$$

Combining the two equations and eliminate  $\lambda$ , we have  $\bar{w}z - w\bar{z} = \bar{w}z_0 - w\bar{z}_0$ . This is one standard form of the equation of a line (there are others though). The equation of a circle centered at  $c \in \mathbb{C}$  with radius  $\rho$  is given by  $|z - c| = \rho$ , which is to say

$$(z - c)(\bar{z} - \bar{c}) = \rho^2 \iff z\bar{z} - c\bar{z} - \bar{c}z = \rho^2 - c\bar{c}$$

The general form of a point on the circle is  $c + \rho e^{i\theta}$  where  $\theta$  is a real parameter. Note that in the geometrical viewpoint of  $\mathbb{C}$ ,  $z \mapsto z + z_0$  is a translation,  $z \mapsto \lambda z$  where  $\lambda \in \mathbb{R}$  is a scaling,  $z \mapsto ze^{i\theta}$  where  $\theta \in \mathbb{R}$  is a rotation,  $z \mapsto \bar{z}$  is a translation,  $z \mapsto 1/z$  is an inversion. Comparing with groups, translation, rotation, scaling, inversion generates the Mobius group which normally acts on  $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\} \cong S^2$  by stereographic projection. On  $\mathbb{C}_\infty$ , lines in  $\mathbb{C}$  are circles.

## 2 Vectors in 3-Dimensional Euclidean Space

A vector is a quantity with magnitude (length/size) and direction. Examples: Force, momentum, electric & magnetic fields, etc. All these examples are modelled on position. In this chapter, we adopt a geometric approach to position vectors in  $\mathbb{R}^3$  based on the Euclidean notions of geometry (points, lines, planes, lengths, angles, etc.)

We choose a point  $O$  as origin, then points  $A$  and  $B$  have position vectors

$$\underline{a} = \overrightarrow{OA}, \underline{b} = \overrightarrow{OB}$$

We define the magnitude of a vector by  $|\underline{a}| = |\overrightarrow{OA}|$  and we denote  $O$  by  $\underline{0}$ .

### 2.1 Vector addition and Scalar Multiplication

Given vectors  $\underline{a}, \underline{b}$  denoting points  $A, B$ , we construct the parallelogram  $OACB$ , so we define  $\underline{a} + \underline{b} = \underline{c}$  where  $\underline{c} = \overrightarrow{OC}$ . One should note that we have already used the parallel postulate here.

Given  $\underline{a}$  which is the position vector of point  $A$  and a scalar  $\lambda \in \mathbb{R}$ ,  $\lambda\underline{a}$  is the position of point on the straight line  $OA$  where  $|\lambda\underline{a}| = |\lambda||\underline{a}|$ . It changes direction if and only if  $\lambda$  is negative. Note that  $\{\lambda\underline{a} : \lambda \in \mathbb{R}\}$  is the set of all points on the straight line  $OA$ .

**Proposition 2.1.** For any  $\underline{a}$ ,  $\underline{a} + \underline{0} = \underline{0} + \underline{a} = \underline{a}$ .

For any  $\underline{a}$ ,  $\underline{a} + (-\underline{a}) = (-\underline{a}) + \underline{a} = \underline{0}$ .

Also, vector addition is associative and commutative. That is, the set of vectors gives an abelian group under vector addition.

*Proof.* Trivial. Consider a parallelepiped for associativity. □

**Proposition 2.2.**  $\lambda(\underline{a} + \underline{b}) = \lambda\underline{a} + \lambda\underline{b}$ .

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

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

*Proof.* Trivial. □

## 2.2 Linear Combination and Span

$\alpha \underline{a} + \beta \underline{b}$  where  $\alpha, \beta \in \mathbb{R}$  is called a linear combination of  $\underline{a}, \underline{b}$ . We can easily extend this definition to an arbitrary set of vectors.

**Definition 2.1.** We define  $\text{span}(S)$  to be the collection of all linear combinations of vectors in  $S$ .

We say  $\underline{a}, \underline{b}$  are parallel if  $\underline{a} = \lambda \underline{b}$  or  $\underline{b} = \lambda \underline{a}$  for some  $\lambda \in \mathbb{R}$ . In this case, we write  $\underline{a} \parallel \underline{b}$ .  
If  $\underline{a} \not\parallel \underline{b}$ , then  $\text{span}(\{\underline{a}, \underline{b}\})$  is a plane.

## 2.3 Scalar Dot Product

**Definition 2.2.** Let  $\theta$  be the angle between  $\underline{a}, \underline{b}$ , then

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

If either of them is  $\underline{0}$ , then their dot product is  $0$ .

**Proposition 2.3.**  $\underline{a} \cdot \underline{b} = \underline{b} \cdot \underline{a}$ .

$\underline{a} \cdot \underline{a} = |\underline{a}|^2 \geq 0$  and the equality sign applies if and only if  $\underline{a} = \underline{0}$ .

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

*Proof.* Trivial. □

**Definition 2.3.** We say  $\underline{a}$  and  $\underline{b}$  are perpendicular, written as  $\underline{a} \perp \underline{b}$  if  $\underline{a} \cdot \underline{b} = 0$

Considering the case  $\underline{a} \neq \underline{0}$ , then

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

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

## 2.4 Orthonormal Basis

Choose vectors  $\underline{e}_1, \underline{e}_2, \underline{e}_3$  that are orthonormal, which means that any two of them are perpendicular to each other and that their magnitudes are all 1. Note that this is equivalent to the choice of Cartesian axes along the three directions.. Also  $\{\underline{e}_i\}$  is a basis, so any vector can be expressed as some linear combination of them.

$$\underline{a} = a_1 \underline{e}_1 + a_2 \underline{e}_2 + a_3 \underline{e}_3$$

Also, each component is uniquely determined.

$$\underline{a} \cdot \underline{e}_i = a_i$$

So we can describe each vector as a row or column vector  $\underline{a} = (a_1, a_2, a_3)$ . Note as well that  $\underline{a} \cdot \underline{b} = \sum_i (a_i b_i) \underline{e}_i$ , so  $|\underline{a}|^2 = \sum_i a_i^2$ . We write  $\underline{e}_1 = \underline{i}, \underline{e}_2 = \underline{j}, \underline{e}_3 = \underline{k}$ .

## 2.5 Vector/Cross Product

**Definition 2.4.** Given nonparallel vectors  $\underline{a}, \underline{b}$ . Let  $\theta$  be the angle between them, measured in the sense that it is relative to a unit normal  $\underline{n}$  to the plane they span by the “right-handed sense”, then we define

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

This is called the vector product or cross product of  $\underline{a}$  and  $\underline{b}$ .

Here  $\underline{n}$  can only be defined when  $\underline{a} \nparallel \underline{b}$ . We define  $\underline{a} \times \underline{b} = 0$  otherwise, since neither  $\theta$  nor  $\underline{n}$  is well defined in this case. Note as well that changing sign of  $\underline{n}$  does not change  $\sin \theta$ .

**Proposition 2.4** (Properties of cross product). 1.  $\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$ .

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

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

4.  $\underline{a} \times \underline{b} = 0 \iff \underline{a} \parallel \underline{b}$ .

5.  $\underline{b} \perp (\underline{a} \times \underline{b}) \perp \underline{a}$ .

*Proof.* Trivial. □

*Remark.*  $\underline{a} \times \underline{b}$  is the vector area of the parallelogram  $0, \underline{a}, \underline{b}, \underline{a} + \underline{b}$ . Its magnitude is the scalar area of that parallelogram. The direction of it then specified the orientation of this parallelogram in space. This gives the significance of vector product.

We fix  $\underline{a}, \underline{x}$  such that they are perpendicular.  $\underline{a} \times \underline{x}$  scales  $\underline{x}$  by a factor of  $|\underline{a}|$  in a different direction.

For easier calculation, we introduce a component expression of the vector product. We consider an orthogonal basis  $\underline{i}, \underline{j}, \underline{k}$  which by ordering we have

$$\underline{i} \times \underline{j} = \underline{k}, \underline{j} \times \underline{k} = \underline{i}, \underline{k} \times \underline{i} = \underline{j}$$

This is called an orthonormal right-hand set (basis).

Now, for  $\underline{a} = a_1\underline{i} + a_2\underline{j} + a_3\underline{k}, \underline{b} = b_1\underline{i} + b_2\underline{j} + b_3\underline{k}$ , we have

$$\underline{a} \times \underline{b} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

## 2.6 Triple Products

### 2.6.1 Scalar Triple Product

**Definition 2.5.** The scalar triple product of  $\underline{a}, \underline{b}, \underline{c}$  is defined as  $\underline{a} \cdot (\underline{b} \times \underline{c})$ .

The interpretation of the scalar triple product is that  $|\underline{c} \cdot (\underline{a} \times \underline{b})|$  is the volume of parallelepiped constructed by the frame  $\underline{a}, \underline{b}, \underline{c}$ . This can be shown easily by consider the geometric definitions of dot and cross products. The sign of taht expression removing the modulus sign could be interpreted as the sign of volume. We say  $\underline{a}, \underline{b}, \underline{c}$  is a right-handed set if  $\underline{c} \cdot (\underline{a} \times \underline{b}) > 0$ . Note that  $\underline{c} \cdot (\underline{a} \times \underline{b}) = 0$  if and only if  $\underline{a}, \underline{b}, \underline{c}$  are coplanar.



Again, we can write the triple product by the components. That is, if  $\underline{a} = a_1\underline{i} + a_2\underline{j} + a_3\underline{k}$ ,  $\underline{b} = b_1\underline{i} + b_2\underline{j} + b_3\underline{k}$ ,  $\underline{c} = c_1\underline{i} + c_2\underline{j} + c_3\underline{k}$ , then

$$\underline{a} \cdot (\underline{b} \times \underline{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

## 2.6.2 Vector Triple Product

**Definition 2.6.**

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

This is called the vector triple product of vectors  $\underline{a}, \underline{b}, \underline{c}$ .

We can check the above works by components by brute force, but we will establish a notation which make things easier later.

**Example 2.1.** Suppose  $\underline{a} = (2, 0, -1)$ ,  $\underline{b} = (7, -3, 5)$ . So  $\underline{a} \times \underline{b} = (-3, -17, -6)$ . We can use the scalar triple product to design a test for coplanarity. Let  $\underline{c} = (3, -3, 7)$  then we do have  $\underline{c} \cdot (\underline{a} \times \underline{b}) = 0$

The application of vector triple products, however, will be seen below.

## 2.7 Lines, Planes and Vector Equations

### 2.7.1 Lines

We can think of vectors as displacements between points. The displacement between vectors can be found by vector subtraction. So the general point on a line through  $\underline{a}$  in the direction  $\underline{u} \neq \underline{0}$  have the form

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

This called the parametric form of a line.

The alternative form of a straight line without a parameter is taken by a cross product

$$\underline{u} \times (\underline{r} - \underline{a}) = \underline{0}$$

Which literally means that  $\underline{r} - \underline{a} = \lambda\underline{u}$  for some real  $\lambda$ .

Now consider  $\underline{u} \times \underline{r} = \underline{c}$ . So we have  $\underline{u} \cdot \underline{c} = 0$ . So if we have  $\underline{u}$  and  $\underline{c}$  where  $\underline{u} \cdot \underline{c} \neq 0$ , then it is a contradiction. Conversely, if it is really 0, then we can consider  $\underline{u} \times (\underline{u} \times \underline{c}) = -|\underline{u}|^2\underline{c}$ .

So it is just a line.

### 2.7.2 Plane

The general form of a plane through some position  $\underline{a}$  to directions  $\underline{u}, \underline{v}$  (with  $\underline{u} \times \underline{v} \neq \underline{0}$ ) is

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

This is called the parametric form of a plane.

Alternatively, we can ditch the parameters:

$$(\underline{r} - \underline{a}) \cdot (\underline{u} \times \underline{v}) = 0$$

Or

$$\underline{r} \cdot (\underline{u} \times \underline{v}) = \kappa$$

where  $\kappa = \underline{a} \cdot (\underline{u} \times \underline{v})$  is a constant.

This form of equation is actually saying that the vector  $\underline{r} - \underline{a}$  is always perpendicular to some given normal  $\underline{n} = \underline{u} \times \underline{v}$ .

### 2.7.3 Vector equations

We can have some more general vector equations. More generally, a vector equation is some equation of the form  $f(\underline{r}) = 0$ .

These can often be solved by dotting or crossing with constant vectors.

**Example 2.2.**  $\underline{r} + \underline{a} \times (\underline{b} \times \underline{r}) = \underline{c}$  We first expand the vector triple product there,

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

Also we have

$$\underline{a} \cdot \underline{r} = \underline{a} \cdot \underline{c}$$

Note that this is weaker here. Putting it back in the previous equation,

$$(1 - \underline{a} \cdot \underline{b})\underline{r} + (\underline{a} \cdot \underline{c})\underline{b} = \underline{c}$$

So if  $\underline{a} \cdot \underline{b} \neq 1$ ,

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

Otherwise, there is no solution if  $\underline{c} - (\underline{a} \cdot \underline{c})\underline{b} \neq 0$ . But if it equals 0, then there are infinitely many solutions.

Sometimes more systematically, we look for a solution of the form  $\alpha\underline{a} + \beta\underline{b} + \gamma\underline{c}$  where  $\underline{a}, \underline{b}, \underline{c}$  are linearly independent. However, sometimes the dotting and crossing method is quicker.

There are many cases when the vector equation may not be linear.

**Example 2.3.**  $\underline{r} \cdot \underline{r} + \underline{r} \cdot \underline{a} = k$ . We can try completing the square to get  $(\underline{r} + \underline{a}/2)^2 = k + \underline{a}^2/4$

## 2.8 Index/Suffix Notation and the Summation Convention

### 2.8.1 Index Notation

Write vectors  $\underline{a}, \underline{b}, \underline{c}, \dots$  in terms of components  $a_i, b_i, c_i, \dots$  wrt an orthonormal right-handed basis

$$\underline{e}_1, \underline{e}_2, \underline{e}_3$$

The indices  $i, j, k, l, p, q, \dots$  will take value 1, 2, 3. So

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

$$\underline{x} = \underline{a} + (\underline{b} \cdot \underline{c})\underline{d} \iff x_i = a_j + \sum_k b_k c_k d_j$$

**Definition 2.7.** The Kronecker  $\delta$  is defined as

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

Then note that  $\underline{e}_i \cdot \underline{e}_j = \delta_{ij}$  and that  $\underline{a} \cdot \underline{b} = (\sum_i a_i \underline{e}_i) \cdot (\sum_j a_j \underline{e}_j) = \sum_{i,j} a_i b_j \delta_{ij} = \sum_i a_i b_i$ .  
Now, cross products,

**Definition 2.8.** The Levi-Civita  $\epsilon$  is defined as

$$\epsilon_{ijk} = \begin{cases} 1, & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3) \\ -1, & \text{if } (i, j, k) \text{ is an odd permutation of } (1, 2, 3) \\ 0, & \text{otherwise} \end{cases}$$

$\epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj} = -\epsilon_{kji}$ . Note that  $\underline{e}_i \times \underline{e}_j = \sum_k \epsilon_{ijk} \underline{e}_k$ .  
Now we consider  $\underline{a} \times \underline{b}$ , we have

$$\begin{aligned} \underline{a} \times \underline{b} &= \left( \sum_i a_i \underline{e}_i \right) \times \left( \sum_j a_j \underline{e}_j \right) \\ &= \sum_{i,j} a_i b_j \underline{e}_i \times \underline{e}_j \\ &= \sum_{i,j,k} a_i b_j \epsilon_{i,j,k} \underline{e}_k \end{aligned}$$

One can check that it is equivalent to our original form of cross product.

### 2.8.2 Summation Convention

Indices that appear twice are usually summed. In the summation convention, we could just ignore the sum signs. That is, the sum is understood.

**Example 2.4.**  $a_i \delta_{ij} = \sum_i a_i \delta_{ij} = a_j$ , so  $a_i \delta_{ij} = a_j$

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

$$\underline{a} \times \underline{b} = a_i b_j \epsilon_{ijk} \underline{e}_k$$

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

$$\delta_{ii} = 3.$$

Genuine rules of the convention are as follows.

1. An index appear exactly once must appear in every term of the expression, and every value of it appears. This is called a free index.
2. An index occuring exactly twice in a given term is summed over, and is called a repeated/contracted/dummy index.
3. No index appear more than twice.

One application of this notation is to prove the vector triple product identity.

**Lemma 2.5** ( $\epsilon - \epsilon$  identity).  $\epsilon_{ijk} \epsilon_{pqk} = \delta_{ip} \delta_{jq} - \delta_{iq} \delta_{jp}$

*Proof.* Both sides are anti-symmetric, and if any two are the same, then it vanishes.

So it is enough to consider a particular case.

If  $i = p = 1, j = q = 2$ , the identity hold.

If  $i = q = 1, j = p = 2$ , the identity hold.

All other index combinations giving nonzero results work similarly.  $\square$

**Proposition 2.6.**

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

*Proof.* Note that  $\epsilon_{ijk}\epsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$

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

So the theorem is proved.  $\square$

We notice one other thing: If we simplify things by letting two indices matching up, then  $\epsilon_{ijk}\epsilon_{pj k} = 2\delta_{ip}$ . Further,  $\epsilon_{ijk}\epsilon_{ijk} = 6$ .

## 3 Vectors in Euclidean Spaces of High Dimensions

### 3.1 Vectors in the Real Vector spaces

If we regard  $\mathbb{R}^n$  algebraically, i.e. see vectors as just a set of components, then it is easy to generalize from 3 to  $n$  dimensions.

**Definition 3.1.** Let  $\mathbb{R}^n$  be the set of real  $n$ -tuples. We define addition by  $(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$  and scalar multiplication by  $\lambda(x_1, x_2, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$ . So we can define linear combinations and the notion of parallel similarly.

For any  $\underline{x} \in \mathbb{R}^n$ , we can write  $\underline{x} = \sum_i x_i \underline{e}_i$  where  $\underline{e}_i$  is a set of orthogonal basis. For example, we can take  $\underline{e}_1 = (1, 0, \dots, 0), \underline{e}_2 = (0, 1, \dots, 0), \dots, \underline{e}_n = (0, 0, \dots, 1)$ . This is called the standard basis of  $\mathbb{R}^n$ . We can define the inner (dot) product in a similar way:

**Definition 3.2.** The inner product (aka scalar product, dot product) is defined by

$$\underline{x} \cdot \underline{y} = \sum_i x_i y_i$$

We have a few properties for the inner products, which is basically analogous to the case in the 3-dimensional case.

**Proposition 3.1.** 1. The inner product is symmetric.

2. The inner product is bilinear.

3.  $\underline{x} \cdot \underline{x} \geq 0$  and the equality hold if and only if  $\underline{x} = \underline{0}$ . That is, it is positive definite. We thus define the length or norm  $|\underline{x}|$  to be  $\sqrt{\underline{x} \cdot \underline{x}}$ . 4. The standard basis in  $\mathbb{R}^n$  is an orthonormal basis.

*Proof.* Trivial.  $\square$

**Theorem 3.2** (Cauchy-Schwartz Inequality).

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

The equality hold if and only if  $\underline{x}, \underline{y}$  are parallel to each other.

The deduction from this is that we can define the angle between two vectors by examining the ratio between the right and left hand side of the inequality. We can also have the triangle inequality, as one may expect norms to satisfy.

*Proof.* Consider

$$0 \leq |\underline{x} - \lambda \underline{y}|^2 = |\underline{x}|^2 - 2\lambda \underline{x} \cdot \underline{y} + \lambda^2 |\underline{y}|^2$$

Taking this as a quadratic in  $\lambda$ , we have

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

The equality holds if and only if  $\delta = 0 \iff 0 = |\underline{x} - \lambda \underline{y}|^2 \iff \underline{x} \parallel \underline{y}$ .  $\square$

**Theorem 3.3.**  $|\underline{x} + \underline{y}| \leq |\underline{x}| + |\underline{y}|$

*Proof.* Square both sides and use Cauchy-Schwartz.  $\square$

Note that by  $\underline{x} \cdot \underline{y}$ , we can think of  $\underline{x}, \underline{y}$  as both row and column vectors, since it doesn't matter. But still if we take them as for example column vectors, then their transposes are row vectors, so  $\underline{x} \cdot \underline{y} = \underline{x}^\top \underline{y}$ .

The algebraic definition of the inner product can be also written by  $\delta_{ij}$ , so it gives us the initiative to generalize the summation convention. In  $\mathbb{R}^3$ , we also have an algebraic definition of cross product. but we cannot really generalize it to  $\mathbb{R}^n$ . We have a generalization of  $\epsilon$  though, which is also antisymmetric.

But in  $\mathbb{R}^2$ , this generalization gives  $\epsilon_{ij}$ , where we can define a product by

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

Geometrically, this gives the signed area of the parallelogram that  $\underline{a}, \underline{b}$  defined. In comparison,  $[\underline{a}, \underline{b}, \underline{c}] = \underline{a} \cdot (\underline{b} \times \underline{c}) = \epsilon_{ijk} a_i b_j c_k$  is the volume of the size of a parallelepiped that these three vectors construct.

## 3.2 Axioms of Real Vector Spaces

**Definition 3.3.** Let  $V$  be a set of objects called vectors with operations:

1.  $\underline{v} + \underline{w} \in V$ .
2.  $\lambda \underline{v} \in V$ .

For  $\underline{v}, \underline{w} \in V, \lambda \in \mathbb{R}$ .

Then  $V$  is called a real vector space if  $(V, +, \underline{0})$  is an abelian group addition

1.  $\lambda(\underline{v} + \underline{w}) = \lambda \underline{v} + \lambda \underline{w}$ .
2.  $(\lambda + \mu) \underline{v} = \lambda \underline{v} + \mu \underline{v}$ .
3.  $(\lambda \mu) \underline{v} = \lambda(\mu \underline{v})$ .
4.  $1 \underline{v} = \underline{v}$ .

where  $\lambda, \mu \in \mathbb{R}, \underline{v}, \underline{w} \in V$

**Definition 3.4.** For any vectors  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_r \in V$ , we can form a linear combination

$$\sum_{i=1}^r a_i \underline{v}_i$$

where  $a_i \in \mathbb{R}$ . The span of these vectors  $\text{span}\{v_1, v_2, \dots, v_r\}$  consists of all linear combinations of these vectors.

The span is a subspace, that is, a subset of the vector space that is itself a vector space under the same way of vector addition and scalar multiplication.

It is immediate that nonempty subset  $U \subseteq V$  is a subspace if and only if  $\text{span} U = U$ .

**Example 3.1.** Take  $V = \mathbb{R}^3$ , then any line or plane through the origin is a subspace, but a line or plane that does not contain the origin is not since  $\underline{0} = \underline{v} + (-1)\underline{v} \in \text{span}\{\underline{v}\}$ .

**Definition 3.5.** A set of vectors  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_r$ , a linear relation of them is an equation

$$\sum_{i=1}^r \lambda_i \underline{v}_i = \underline{0}$$

If the equation is true only if  $\lambda_i = 0$  for every  $i$ , then the vectors are called linearly independent and that they obey only the trivial linear relation. And we say this set of vectors is an independent set.

Otherwise, we say they are linearly dependent, and the set of vectors is a dependent set.

**Example 3.2.** 1. So for example, if we take  $V = \mathbb{R}^2$  and consider the set  $(0, 1), (1, 0), (0, 2)$ . It is a dependent set since  $2(0, 1) - (0, 2) = (0, 0)$ .

2. Any set containing  $\underline{0}$  is dependent.

3.  $\{\underline{a}\}$  is independent if and only if  $\underline{a} \neq \underline{0}$ .

4.  $\{\underline{a}, \underline{b}\}$  is independent if and only if  $\underline{a} \nparallel \underline{b}$ .

**Definition 3.6.** A function  $\cdot : V \times V \rightarrow \mathbb{R}$  is called an inner product on  $V$  if and only if:

1.  $\underline{v} \cdot \underline{w} = \underline{w} \cdot \underline{v}$ .

2. It is bilinear.

3.  $\underline{v} \cdot \underline{v} \geq 0$  and the equality holds if and only if  $\underline{v} = \underline{0}$ .

### 3.3 Basis and Dimension

For general vector spaces, a basis  $\mathcal{B}$  is an independent set of vectors such that  $\text{span } \mathcal{B} = V$ . Given this property, it is trivial that the coefficients of any vector as a linear combination of elements in  $\mathcal{B}$  are unique.

**Example 3.3.** The standard basis for  $\mathbb{R}^n$  consisting of

$$\underline{e}_1 = (1, 0, \dots, 0), \underline{e}_2 = (0, 1, \dots, 0), \dots, \underline{e}_n = (0, 0, \dots, 1)$$

is a basis. There are many other bases that can be chosen though. For example, in  $\mathbb{R}^2$ ,  $\{(1, 0), (1, 1)\}$ ,  $\{(1, -1), (1, 1)\}$  or simply any  $\{\underline{a}, \underline{b}\}$  that are not parallel would be a basis.

**Theorem 3.4.** If  $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$  and  $\{\underline{f}_1, \underline{f}_2, \dots, \underline{f}_m\}$ , then  $m = n$ .

It follows that we can define the following

**Definition 3.7.** The number of vectors in a basis in a vector space is called its dimension. So with this definition,  $\mathbb{R}^n$  has dimension  $n$ .

Now we can prove the theorem.

*Proof.* Note that we can find coefficients  $A_{ai}$  such that  $\underline{f}_a = \sum_i A_{ai} \underline{e}_i$  for each  $a$ . Similarly,  $\underline{e}_i = \sum_a B_{ia} \underline{f}_a$ . Then  $\underline{f}_a = \sum_b (\sum_i A_{ai} B_{ib}) \underline{f}_b$  and  $\underline{e}_i = \sum_j (\sum_a B_{ia} A_{aj}) \underline{e}_j$ . So  $\sum_i A_{ai} B_{ib} = \delta_{ij}$  and  $\sum_a B_{ia} A_{aj} = \delta_{ij}$ .  $\sum_{i,a} A_{ai} B_{ia} = \sum_i \delta_{ii}$  and  $\sum_{i,a} A_{ai} B_{ia} = \sum_a \delta_{aa}$ , thus  $n = m$ .  $\square$

In fact we did secretly use traces, but we have presented it within the scope of this course. The proof in the general case can be done elegantly using matrices.

We can also apply our notions to the following:

**Proposition 3.5.** Let  $V$  be a vector space of dimension  $n$ .

1. If  $Y = \{\underline{w}_1, \underline{w}_2, \dots, \underline{w}_m\}$  spans  $V$ , then  $m \geq n$  and if  $m > n$  we can remove an element from  $Y$  such that the new set of vectors still spans  $V$ . Thus we can continue doing it till it becomes a basis.
2. If  $X = \{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k\}$  be an independent set of vectors, then  $k \leq n$  and if  $k < n$  we can add vector to  $X$  until we have a basis.

*Proof.* 1. If  $Y$  is linearly independent, then we are done and  $m = n$ . Otherwise, there is some linear relation  $\sum_i \lambda_i \underline{w}_i = 0$  such that  $\exists i, \lambda_i \neq 0$ . WLOG  $i = m$ , then  $\underline{w}_m = \lambda_m^{-1} (\sum_{i, i \neq m} \lambda_i \underline{w}_i)$ . So  $Y' = Y \setminus \{\underline{w}_m\}$  spans  $V$ . We can repeat this till it becomes independent, i.e. we obtain a basis.

2. If  $X$  spans  $V$ , we are done and then  $k = n$ . Otherwise, there is some  $\underline{u}_{k+1} \in V \setminus \text{span } X$ , then  $X \cup \{\underline{u}_{k+1}\}$  is still independent. Indeed, if there is some nontrivial linear relation  $\sum_i \mu_i \underline{u}_i = 0$ , then  $\mu_{k+1} \neq 0$  since  $X$  is independent, but then  $\underline{u}_{k+1}$  can be written as a linear combination of  $\underline{u}_i$  for  $i \in \{1, 2, \dots, k\}$ , contradiction.

So we can do it over again and obtain a basis at last.  $\square$

Note that the basis does not use nor depend on the inner product structure of the vector space. But we can make use of it to obtain an orthogonal basis.

**Definition 3.8.** A basis is called orthogonal if they are pairwise orthogonal.

**Proposition 3.6.** Any set of pairwise orthogonal vectors is independent.

*Proof.* Easy.  $\square$

### 3.4 Vectors in the Complex Vector Space

**Definition 3.9.** Let  $\mathbb{C}^n$  consist of all complex  $n$ -tuples. We define the vector space by taking the vector addition and scalar multiplication analogously to the  $\mathbb{R}^n$  case.

Taking real scalars in scalar multiplication, then  $\mathbb{C}^n$  is just a real vector space of dimension  $2n$ . Taking complex scalars,  $\mathbb{C}^n$  is a complex vector space. The definitions of linear combinations, linear independence, basis, dimensions are all analogous. Consider

$$(z_1, \dots, z_n) = (x_1 + iy_1, \dots, x_n + iy_n) = (x_1, \dots, x_n) + i(y_1, \dots, y_n)$$

then  $e_j$  and  $ie_j$  gives a basis for  $\mathbb{C}^n$  as a real vector space. And the  $e_j$  is a basis for  $\mathbb{C}^n$  as a complex vector space. We view  $\mathbb{C}^n$  to be over  $\mathbb{C}$  unless otherwise stated.

**Definition 3.10.** The inner product on  $\mathbb{C}^n$  is defined by  $(\underline{z}, \underline{w}) = \sum_j \bar{z}_j w_j$

**Proposition 3.7.** *The complex inner product is:*

1. Hermitian,  $(\underline{z}, \underline{w}) = \overline{(\underline{w}, \underline{z})}$ .
2. Anti-linear,  $(\underline{z}, \lambda \underline{w} + \lambda' \underline{w}') = \lambda(\underline{z}, \underline{w}) + \lambda'(\underline{z}, \underline{w}')$  and  $(\lambda \underline{z} + \lambda' \underline{z}', \underline{w}) = \bar{\lambda}(\underline{z}, \underline{w}) + \bar{\lambda}'(\underline{z}', \underline{w})$ .
3. Positive definite.

The geometric content of this sort of inner product is a little more subtle than in the real case.

**Example 3.4.** Consider the complex inner product on  $\mathbb{C}$ , that is  $n = 1$ , since there is only one component,  $(z, w) = \bar{z}w$ . Suppose  $z = a_1 + ia_2, w = b_1 + ib_2$  where  $a_1, a_2, b_1, b_2 \in \mathbb{R}$ . Let  $\underline{a} = (a_1, a_2), \underline{b} = (b_1, b_2)$ , then

$$(z, w) = a_1 b_1 + a_2 b_2 + i(a_1 b_2 - a_2 b_1) = \underline{a} \cdot \underline{b} + i[\underline{a}, \underline{b}]$$

Given the positive definite property, we can define the length or norm  $|\underline{z}|$  for any  $\underline{z} \in \mathbb{C}^n$  by  $|\underline{z}| = \sqrt{(\underline{z}, \underline{z})}$ . We still say that two complex vectors are orthogonal if their (complex) inner product vanishes. In this way, the standard basis for  $\mathbb{C}^n$  is orthogonal.

If  $\underline{z}_1, \underline{z}_2, \dots, \underline{z}_k$  are nonzero and orthogonal, then they are linearly independent. Notationally, we think of the vectors in  $\mathbb{C}^n$  as column vector, and the Hermitian conjugate  $\underline{z}^\dagger$  the row vector consisting of the conjugates of the entries.

## 4 Matrices and Linear Maps

### 4.1 Definitions

**Definition 4.1.** A linear map is a function that preserves linear combination. That is, for vector spaces  $V, W$  over the same field  $k$ , a linear map  $T : V \rightarrow W$  satisfies

$$\forall \lambda, \mu \in k, \underline{v}, \underline{v}' \in V, T(\lambda \underline{v} + \mu \underline{v}') = \lambda T(\underline{v}) + \mu T(\underline{v}')$$

**Definition 4.2.** The image of the entire vector space  $V$  under  $T$ ,  $\text{Im } T$ , is the collection of all images, that is,  $\{\underline{w} \in W : \exists \underline{v} \in V, T(\underline{v}) = \underline{w}\}$ . Also, the kernel  $\ker T$  is the set  $\{\underline{v} \in V : T(\underline{v}) = \underline{0}\}$

**Proposition 4.1.** *The kernel is a subspace of  $V$  and the image a subspace of  $W$ .*

**Definition 4.3.** The dimension of the image is called the rank of  $T$ ,  $\text{rank } T$ , and the dimension of the kernel is the nullity of  $T$ ,  $\text{null } T$ .



- Example 4.1.** 1. The zero linear map  $T$  mapping each vector to the zero vector is a linear. Its rank is 0 and the nullity of  $T$  is the dimension of  $V$ .  
 2. The identity map on  $V$  is linear with kernel  $\{0\}$  and the image  $V$ .  
 3. Suppose  $V = W = \mathbb{R}^3$  and the map  $T$  given by

$$T\underline{x} = \begin{pmatrix} 3 & 1 & 5 \\ -1 & 0 & -2 \\ 2 & 1 & 3 \end{pmatrix} \underline{x}$$

is linear. Note that this matrix is singular, so its kernel is nontrivial since it contains at least one vector, say  $(2, -1, -1)$ . One can show that the nullity is 1 (that is, it is entirely generated by this vector) and the rank is 2.

**Definition 4.4.** If  $T, S : V \rightarrow W$  are both linear maps, then  $\alpha T + \beta S$  for any  $\alpha, \beta \in F$  is obviously also a linear map. We say it is the linear combination of the linear maps.

**Definition 4.5.** If  $T : V \rightarrow W, S : U \rightarrow V$  are both linear, then easily  $T \circ S : U \rightarrow W$  is linear and is called the composition of linear maps.

The third example shown above triggers the following theorem:

**Theorem 4.2** (Rank-Nullity Theorem). *Suppose  $T : V \rightarrow W$  is linear, then  $\text{rank } T + \text{null } T = \dim V$ .*

*Proof.* Let  $\underline{e}_1, \dots, \underline{e}_k$  be a basis for  $\ker T$ . We can extend it to a basis  $\underline{e}_1, \dots, \underline{e}_n$  of  $V$ . Now  $\underline{T}(\underline{e}_{k+1}), \dots, \underline{T}(\underline{e}_n)$  is a basis for  $\text{Im } T$ , following which the theorem is proved. Indeed,

$$T \left( \sum_{i=1}^n a_i \underline{e}_i \right) = \sum_{i=1}^n a_i T(\underline{e}_i) = \sum_{i=k+1}^n a_i T(\underline{e}_i)$$

so this set indeed spans  $\text{Im } T$ . It is also linearly independent since

$$\sum_{i=k+1}^n a_i T(\underline{e}_i) = 0 \implies T \left( \sum_{i=k+1}^n a_i \underline{e}_i \right) = 0 \implies \sum_{i=k+1}^n a_i \underline{e}_i = \sum_{i=1}^k a_i \underline{e}_i$$

But that would imply that we have found a nontrivial relation between our  $\underline{e}_i$  which are independent, which is false. So it is independent, hence the proof is done.  $\square$

## 4.2 Matrices as Linear Maps in Real Vector Spaces

Let  $M$  be an  $n \times n$  array with entries  $M_{ij}$  where  $i$  labels the rows and  $j$  labels the columns. For example, if  $n = 3$ ,

$$M = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}$$

We define the map  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $T(\underline{x})_i = M_{ij}x_j$ . It is obviously a linear map. Note that if  $\underline{x} = x_i \underline{e}_i$ , then  $T(\underline{x}) = x_i T(\underline{e}_i) = x_i \underline{c}_i$  where  $\underline{c}_i$  is the column  $i$  of the matrix. Therefore the image of  $T$  is the span of the columns of  $M$ .

Given that, it is useful to consider the rows  $\underline{r}_i$  and  $\underline{c}_i$  that are the rows and columns of  $M$ . We can write  $(\underline{R}_i)_j = M_{ij} = (\underline{C}_j)_i$ . So  $(M\underline{x})_i = \underline{R}_i \cdot \underline{x}$ , the kernel of  $M$  is the set of all  $\underline{x}$  that vanishes under the linear map.

- Example 4.2.** 1.  $V = W = \mathbb{R}^n$ , the zero map corresponds to the zero matrix.  
 2. The identity map corresponds to the identity matrix  $I_{ij} = \delta_{ij}$ .  
 3.  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  corresponding to the matrix

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

The image of it is the span of columns, that is, the 2-dimensional subspace spanned by  $\underline{C}_1, \underline{C}_2$  (since  $\underline{C}_3$  is in this space). The kernel then is the 1-dimensional subspace spanned by the vector  $(2, -1, -1)$ .

### 4.3 Geometrical Examples

We first think about rotations  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$  by an angle  $\theta$  can be described as

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

which one can easily derive from either polar coordinate or the its behaviour on basis vectors.

Things go more interesting when we get to dimension 3. We consider  $\underline{x} = \underline{x}_{\parallel} + \underline{x}_{\perp}$  as the decomposition of  $\underline{x}$  along the  $\underline{n}$  direction, i.e. such that  $\underline{x}_{\parallel} \parallel \underline{n}$  and  $\underline{x}_{\perp} \perp \underline{n}$ . Then,

$$|\underline{x}_{\parallel}| = |\underline{x}| \cos \phi, |\underline{x}_{\perp}| = |\underline{x}| \sin \phi$$

So under the rotation along the axis  $\underline{n}$ ,  $\underline{x}_{\parallel}$  stays the same while the  $\underline{x}_{\perp}$  changes. Assuming the angle is  $\theta$ , we can reassemble things to obtain

$$\underline{x} \mapsto \underline{x}_{\parallel} + \cos \theta \underline{x}_{\perp} + \sin \theta \underline{n} \times \underline{x}$$

So in components

$$\begin{aligned} (M\underline{x})_i &= M_{ij}x_j = x_i \cos \theta + (1 - \cos \theta)n_j x_j n_i + \sin \theta \epsilon_{ijk} n_j x_k \\ &= (\delta_{ij} \cos \theta + (1 - \cos \theta)n_i n_j + \sin \theta \epsilon_{ijk} n_k)x_j \end{aligned}$$

We can also have reflection across the plane with normal  $\underline{n}$  which, in matrix form, would be  $M_{ij} = \delta_{ij} - 2n_i n_j$ .

Dilation and scaling are linear maps as well. If we set the scaling factors to be  $\alpha, \beta, \gamma$  along  $\underline{e}_1, \underline{e}_2, \underline{e}_3$ , then the corresponding matrix would be

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

There is another kind of linear transformation called shears. Given unit vectors  $\underline{a}, \underline{b}$  perpendicular, then a shear with parameter  $\lambda$  is defined by  $\underline{x} \mapsto \underline{x} + \lambda \underline{a}(\underline{b} \cdot \underline{x})$ . So  $\underline{a} \mapsto \underline{a}$  (in general  $\underline{u} \perp \underline{b} \implies \underline{u} \mapsto \underline{u}$ ),  $\underline{b} \mapsto \underline{b} + \lambda \underline{a}$ . In component,  $T(\underline{x})_i = (\delta_{ij} + \lambda a_i b_j)x_j$ .

## 4.4 Matrices in General; Matrix Algebra

**Definition 4.6.** Consider a linear map  $T : V \rightarrow W$  where  $V, W$  are real or complex vector spaces of dimensions  $n, m$  respectively. Assume that we have obtained a basis  $\{\underline{e}_i\}$  for  $V$  and a basis  $\{\underline{f}_a\}$  for  $W$ . So a matrix representation of  $T$  with respect to these bases is an array  $M_{ai}$  with entries in  $\mathbb{R}, \mathbb{C}$  as appropriate, with  $a \in \{1, 2, \dots, m\}$  ('rows'),  $i \in \{1, 2, \dots, n\}$  ('columns') with

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

which extends to any vectors in  $V$  by linearity.

So we have  $M\underline{x} = T(\underline{x}) = M_{ai}x_i\underline{f}_a$  where the summation convention is being used. The moral of the story is that by choice with basis we can identify the vector spaces as  $\mathbb{R}^n, \mathbb{R}^m$  or  $\mathbb{C}^n, \mathbb{C}^m$  and  $T$  as a matrix.

**Definition 4.7.** Suppose  $T : V \rightarrow W, S : V \rightarrow W, R : W \rightarrow Z$ , then given choices of bases on  $V, W, Z$ , and hence the matrices  $M$  of  $T$ ,  $N$  of  $S$ ,  $L$  of  $R$ , then  $(M + N)_{ij} = M_{ij} + N_{ij}$  and  $LM$  the matrix of  $R \circ T$ , or (as one can check) equivalently  $(LM)_{ij} = L_{ia}M_{aj}$ .

One can also observe that the products of two matrices consists of exactly the dot products of the rows of one and the columns of the other. This follows immediately from definitions. Note that the matrix multiplication cannot be defined in two arbitrary matrices. To multiply two matrices  $A, B$  to get  $AB$ , we must have  $\text{dom } A = \text{cod } B$ .

**Proposition 4.3.** For matrices where matrix multiplication is defined, it is associative and distributive over matrix addition.

*Proof.* Trivial. □

**Definition 4.8.** For  $m \times n$  matrices  $A$  and  $n \times m$  matrices  $B, C$ ,  $B$  is a left inverse of  $A$  if  $BA = I$  and  $C$  a right inverse of  $A$  if  $AC = I$  where  $I$  is the identity.

Note that not every matrices has inverse(s), but if it does, we say that it is invertible, or non-singular.

**Proposition 4.4.** If  $n = m$ , then if  $A$  is invertible, then it has both an unique left inverse and an unique right inverse, and they are the same.

*Proof.* Obvious enough. □

**Example 4.3.** 1. The rotation matrix  $R(\theta)$  with respect to some (hyper-)axis has an inverse, since  $R(\theta) \circ R(-\theta) = I$ .

2. Consider  $n \times n$  matrix  $M$ . If  $\underline{x}' = M\underline{x}$ , then  $M^{-1}\underline{x}' = \underline{x}$ . This shows the uniqueness criterion of a system of linear equations.

One can check that a  $2 \times 2$  matrix  $M$  is uniquely solvable if and only if its determinant  $\det M = [M\underline{e}_1, M\underline{e}_2]$  is nonzero. One realize that the determinant is the area of the parallelogram spanned by the images of the bases.

#### 4.4.1 Transpose and Hermitian Conjugate

If  $M$  is an  $m \times n$  matrix, then transpose, written  $M^\top$  is the  $n \times m$  defined by  $(M^\top)_{ij} = M_{ji}$ . Note that for any two  $n \times m$  matrices  $M, N$  and scalars  $\lambda, \mu$ , then  $(\lambda A + \mu B)^\top = \lambda A^\top + \mu B^\top$ . If  $A$  is a  $m \times n$  and  $B$  is  $n \times p$ , then  $(AB)^\top = B^\top A^\top$ . For square matrix  $A$ , we say  $A$  is symmetric if  $A^\top = A$ , antisymmetric if  $A^\top = -A$ .

For complex matrices, the Hermitian conjugate of an  $m \times n$  matrix  $M$  is defined by  $(M^\dagger)_{ij} = \overline{M_{ji}}$ . We can define Hermitian and anti-Hermitian matrices in the same way in square complex matrices.

#### 4.4.2 Trace

Consider a complex  $n \times n$  matrix  $M$ , the trace of  $M$ ,  $\text{tr}(M)$ , is  $M_{ii}$  (where the summation convention is being used). Immediately,  $\text{tr}(\alpha M + \beta N) = \alpha \text{tr}(M) + \beta \text{tr}(N)$ , and  $\text{tr}(MN) = \text{tr}(NM)$  and  $\text{tr}(M) = \text{tr}(M^\top)$ , and  $\text{tr}(I) = n$ .

**Example 4.4.** The reflection across the plane with normal  $\underline{n}$  can be represented by the matrix  $H$  defined by  $H_{ij} = \delta_{ij} - 2n_i n_j$ . Now  $\text{tr}(H) = H_{ii} = \delta_{ii} - 2n_i n_i = 3 - 2 = 1$ .

Note as well that antisymmetric matrices always have zero trace.

#### 4.4.3 Decomposition

An  $n \times n$  matrix  $M$  can be written as  $M = S + A$  where  $S = (M + M^\top)/2$ ,  $A = (M - M^\top)/2$ . Note that  $S$  and  $A$  are symmetric and antisymmetric respectively. Consider the matrix  $T$  defined by  $T_{ij} = \delta_{ij} - \text{tr}(S)\delta_{ij}/n$ , so  $T$  is traceless as well. But  $\text{tr} S = \text{tr} M$  since  $A$  is antisymmetric hence traceless. Thus  $M_{ij} = T_{ij} + A_{ij} + \text{tr}(M)\delta_{ij}/n$  decompose the matrix  $M$  into symmetric traceless, antisymmetric and pure trace part.

**Example 4.5.** For  $n = 3$ , suppose  $T_{ij} = 0$ , set  $A_{ij} = \epsilon_{ijk} a_k$  and  $\text{tr} M = 3\lambda$ , so  $M\underline{x} = \underline{x} \times \underline{a} + \lambda\underline{x}$ .

#### 4.4.4 Orthogonal and Unitary Matrices

A square matrix  $U$  is orthogonal if  $UU^\top = U^\top U = I$ . So the column vectors of  $U$  actually are orthonormal, same for the rows. For example, rotation matrices are orthogonal.

**Proposition 4.5.**  $U$  is orthogonal if and only if it preserves inner products.

*Proof.* For any square matrix  $U$ , we have

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

So if  $U$  is orthogonal, the last expression equals  $\underline{x} \cdot \underline{y}$ , hence  $U$  preserves dot product. Conversely, for any square matrix  $A$ ,  $\underline{e}_j^\top A \underline{e}_i = A_{ji}$ , so if  $U$  preserves inner product, then by taking  $A = U^\top U$  we have  $U^\top U = I$   $\square$

Note that in  $\mathbb{R}^n$ , if  $\{\underline{e}_i\}$  is an orthonormal basis, then  $\{U\underline{e}_i\}$  would also be an orthonormal basis.

**Example 4.6.** The general  $2 \times 2$  orthogonal matrices are all rotational or reflectional (from an axis through the origin) matrices. This can be trivially checked. In particular, reflections have determinant  $-1$  while rotations always have determinant  $1$ .

**Definition 4.9.** A complex  $n \times n$  matrix  $U$  is unitary if  $U^\dagger U = UU^\dagger = I$ . Equivalently,  $U^\dagger = U^{-1}$ .

**Proposition 4.6.**  $U$  is unitary if and only if it preserves complex inner product.

*Proof.*  $(U\underline{x}, U\underline{y}) = \underline{x}^\dagger U^\dagger U \underline{y}$ . Necessity is implied, and sufficiency is by  $\underline{x} = \underline{e}_i, \underline{y} = \underline{e}_j$ .  $\square$

## 5 Determinants and Inverses

### 5.1 Introduction

Consider a linear map  $\mathbb{R}^n \rightarrow \mathbb{R}^n$  given by the matrix  $M$ . We want to define an  $n \times n$  matrix  $\tilde{M} = \text{adj } M$  and a scalar  $\det M$  with

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

Furthermore,  $\det M$  is the factor by which an area in  $\mathbb{R}^2$  or a volume in  $\mathbb{R}^3$  is scaled. If  $\det M \neq 0$ , then we will have

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

For  $n = 2$ , we know that

$$\tilde{M} = \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}, \det M = \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} = M_{11}M_{22} - M_{12}M_{21}$$

works. Note that  $\det M \neq 0$  if and only if  $M\underline{e}_1, M\underline{e}_2$  are linearly independent if and only if  $\text{Im } M = \mathbb{R}^2$  if and only if  $\text{rank } M = 2$ .

For  $n = 3$ , we recall that given any  $\underline{a}, \underline{b}, \underline{c} \in \mathbb{R}^3$ , the scalar  $[\underline{a}, \underline{b}, \underline{c}]$  is the volume of a parallelepiped spanned by  $\underline{a}, \underline{b}, \underline{c}$ . We can also note that the standard basis vectors obey  $[\underline{e}_i, \underline{e}_j, \underline{e}_k] = \epsilon_{ijk}$ . Note that for  $M$  a real  $3 \times 3$  matrix, its columns are  $M\underline{e}_i = M_{ji}\underline{e}_j$ . So the volume is scaled by a factor

$$[M\underline{e}_1, M\underline{e}_2, M\underline{e}_3] = M_{i1}M_{j2}M_{k3}[\underline{e}_i, \underline{e}_j, \underline{e}_k] = M_{i1}M_{j2}M_{k3}\epsilon_{ijk}$$

We define this to be  $\det M$ . We can also define  $\tilde{M}$  by

$$\underline{R}_1(\tilde{M}) = \underline{C}_2(M) \times \underline{C}_3(M)$$

$$\underline{R}_2(\tilde{M}) = \underline{C}_3(M) \times \underline{C}_1(M)$$

$$\underline{R}_3(\tilde{M}) = \underline{C}_1(M) \times \underline{C}_2(M)$$

So one can see immediately that

$$(\tilde{M}M)_{ij} = \underline{R}_i(\tilde{M}) \cdot \underline{C}_j(M) = \det M \delta_{ij}$$

as desired.

How about when we consider  $n$  in general?

## 5.2 Alternating forms

We first want to generalize our  $\epsilon$  symbol to higher dimensions by considering the permutation.

**Definition 5.1.** A permutation  $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is a bijection of  $\{1, 2, \dots, n\}$  to itself.

So  $\{1, 2, \dots, n\} = \{\sigma 1, \sigma 2, \dots, \sigma n\}$ . It is immediate that the permutations on  $n$  letters form a group  $S_n$  under composition, and it is easy that  $|S_n| = n!$ .

**Definition 5.2.** A permutation  $\tau \in S_n$  is called a transposition of  $i, j \in \{1, 2, \dots, n\}$  if  $\tau(i) = j, \tau(j) = i, \forall k \neq i, j, \tau(k) = k$ . We denote  $\tau$  by  $(p\ q)$ .

**Proposition 5.1.** Any permutation is a product of transpositions.

*Proof.* Trivial. □

The way we write it is not unique, but the number of transpositions is unique modulo 2.

**Proposition 5.2.** If some permutation  $\sigma$  can be written as the product of  $k$  transpositions and the product of  $l$  transpositions, then  $k \equiv l \pmod{2}$ .

*Proof.* Will see in groups (actually quite trivial). □

**Definition 5.3.** We say the permutation  $\sigma$  is even if it can be written as the product of an even number of transpositions, odd if otherwise. We define the sign, or signature function  $\epsilon : S_n \rightarrow \{1, -1\}$  by

$$\epsilon(\sigma) = \begin{cases} 1, & \text{if } \sigma \text{ is even} \\ -1, & \text{otherwise} \end{cases}$$

Note that  $\epsilon(\text{id}) = 1$  and more generally  $\epsilon(\sigma \circ \pi) = \epsilon(\sigma)\epsilon(\pi)$  for permutations  $\sigma, \pi$ .

**Definition 5.4.** The  $\epsilon$  symbol (or tensor) on  $n$  letters is defined as

$$\epsilon_{ij\dots kl} = \begin{cases} \epsilon(\sigma), & \text{if } ij\dots kl \text{ is a permutation } \sigma \text{ of } \{1, 2, \dots, n\} \\ 0, & \text{otherwise. That is, some indices coincide.} \end{cases}$$

So we now can define the alternating forms.

**Definition 5.5.** Given vectors  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$  in  $\mathbb{R}^n$  or  $\mathbb{C}^n$ . We define the alternating form to be the scalar

$$[\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n] = \epsilon_{ij\dots kl} (v_1)_i (v_2)_j \cdots (v_{n-1})_k (v_n)_l$$

One can check that the alternating forms when  $n = 2, 3$  are exactly the same as we have defined them before.

Note that the alternating form is multilinear, thus a tensor. Also, it is totally antisymmetric: interchanging any two vectors changes the sign. Equivalently,

$$[\underline{v}_{\sigma(1)}, \underline{v}_{\sigma(2)}, \dots, \underline{v}_{\sigma(n)}] = \epsilon(\sigma) [\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n]$$

Moreover,  $[\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n] = 1$ .

One can see immediately that

**Proposition 5.3.** *If the function  $f : (F^n)^n \rightarrow F$  where  $F = \mathbb{R}$  or  $\mathbb{C}$  is multilinear, totally antisymmetric and  $f(\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n) = 1$ , then  $f$  is uniquely determined.*

*Proof.* Trivial. □

**Proposition 5.4.** *If some vector(s) is  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$  is a linear combination of others, then  $[\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n] = 0$ .*

**Example 5.1.** In  $\mathbb{C}^4$ , let  $\underline{v}_1 = (i, 0, 0, 2)$ ,  $\underline{v}_2 = (0, 0, 5i, 0)$ ,  $\underline{v}_3 = (3, 2i, 0, 0)$ ,  $\underline{v}_4 = (0, 0, -i, 1)$ , then  $[\underline{v}_1, \underline{v}_2, \underline{v}_3, \underline{v}_4] = 10i$

**Proposition 5.5.**  $[\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n] \neq 0$  if and only if  $\{\underline{v}_i\}$  is an independent set.

*Proof.* We have already shown the “only if” part, so it remains to show the other direction. If  $\{\underline{v}_i\}$  is independent, then it constituted a basis, hence if  $[\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n] = 0$ , then by multilinearity the alternating form will be zero everywhere, which is a contradiction. □

**Definition 5.6** (Definition of determinant). Consider  $M \in M_{n \times n}(F)$  where  $F = \mathbb{R}$  or  $\mathbb{C}$  with columns  $\underline{C}_a$ , then the determinant of  $M$  is

$$\begin{aligned} \det M &= [\underline{C}_1, \underline{C}_2, \dots, \underline{C}_n] \\ &= [M\underline{e}_1, M\underline{e}_2, \dots, M\underline{e}_n] \\ &= \epsilon_{ij\dots kl} M_{i1} M_{j2} \cdots M_{k(n-1)} M_{ln} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n} \end{aligned}$$

**Example 5.2.** 1. The definition of determinant here in general coincides with the cases in 2 and 3 dimensional cases.

2. If  $M$  is diagonal, then  $\det M$  is the product of all diagonal entries. So  $\det I = 1$ .

3. If we have

$$M = \left( \begin{array}{ccc|c} & & & 0 \\ & A & & \vdots \\ & & & 0 \\ \hline 0 & \dots & 0 & 1 \end{array} \right)$$

Then  $\det M = \det A$

### 5.3 Properties of Determinants

If we multiply one of the columns (or rows) of the matrix  $M$  by a scalar  $\lambda$  to produce  $M'$ , we have  $\det M' = \lambda \det M$ . Furthermore, if we interchange two adjacent columns, the determinant is negated. We can see these directly from the definitive property of determinants. In addition,  $\det M \neq 0$  if and only if the columns are linearly independent.

**Proposition 5.6.** *For any  $n \times n$  matrix  $M$ ,  $\det M = \det M^\top$ . Equivalently,  $[\underline{C}_1, \underline{C}_2, \dots, \underline{C}_n] = [\underline{R}_1, \underline{R}_2, \dots, \underline{R}_n]$ .*

*Proof.* We have

$$\begin{aligned}
[\underline{C}_1, \underline{C}_2, \dots, \underline{C}_n] &= \sum_{\sigma \in S_n} \epsilon(\sigma) M_{\sigma(1)1} M_{\sigma(2)2} \cdots M_{\sigma(n)n} \\
&= \sum_{\sigma \in S_n} \epsilon(\sigma) M_{1\sigma^{-1}(1)} M_{2\sigma^{-1}(2)} \cdots M_{n\sigma^{-1}(n)} \\
&= \sum_{\sigma' \in S_n} \epsilon(\sigma') M_{1\sigma'(1)} M_{2\sigma'(2)} \cdots M_{n\sigma'(n)} \\
&= [\underline{R}_1, \underline{R}_2, \dots, \underline{R}_n]
\end{aligned}$$

Since the map  $\sigma \rightarrow \sigma' = \sigma^{-1}$  is an automorphism on  $S_n$  and  $\epsilon(\sigma) = \epsilon(\sigma')$ .  $\square$

We can evaluate determinants by expanding rows or columns.

**Definition 5.7.** For  $M$  an  $n \times n$  matrix, Define  $M^{ia}$  be the determinant of the  $(n-1) \times (n-1)$  matrix obtained by deleting the row  $i$  and column  $a$  of  $M$ . This is called a minor.

**Proposition 5.7.**

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

*Proof.* Trivial but see later sections for the proof.  $\square$

By some trivial computation, we discover <sup>1</sup> that matrices having many zeros would be easier to calculate, so it brings us to the ways to simplify the determinants.

The first thing we could do is row/column operations. If we modify  $M$  by mapping  $\underline{C}_i \mapsto \underline{C}_i + \lambda \underline{C}_j, i \neq j$  (or equivalently on rows), then the determinant is not changed. This follows immediate from multilinearity and total antisymmetry. Plus, as we stated above, if we interchange  $\underline{C}_i, \underline{C}_j, i \neq j$  (same for rows), then the determinant changes sign. This can help us simplify the calculation greatly since we can produce a lot of 0's from there.

**Theorem 5.8.** For any  $n \times n$  matrices  $M, N$ ,  $\det(MN) = \det(M) \det(N)$ .

**Lemma 5.9.**

$$\epsilon_{i_1 i_2 \dots i_n} M_{i_1 a_1} M_{i_2 a_2} \cdots M_{i_n a_n} = \epsilon_{a_1 a_2 \dots a_n} \det M$$

*Proof.* Trivial.  $\square$

*Proof of the multiplicativity of determinants.* By the preceding lemma,

$$\begin{aligned}
\det(MN) &= \epsilon_{i_1 i_2 \dots i_n} (MN)_{i_1 1} \cdots (MN)_{i_n n} \\
&= \epsilon_{i_1 i_2 \dots i_n} M_{i_1 k_1} N_{k_1 1} \cdots M_{i_n k_n} N_{k_n n} \\
&= \epsilon_{i_1 i_2 \dots i_n} M_{i_1 k_1} \cdots M_{i_n k_n} N_{k_1 1} \cdots N_{k_n n} \\
&= \det M \epsilon_{k_1 k_2 \dots k_n} N_{k_1 1} \cdots N_{k_n n} \\
&= \det M \det N
\end{aligned}$$

As desired.  $\square$

<sup>1</sup>We definitely did not see that coming



Note that this would mean that  $\det$  is a group homomorphism. There are a few consequences of the multiplicative property:

- Proposition 5.10.** 1. If  $M$  is invertible, then  $\det(M^{-1}) = \det(M)^{-1}$ .  
 2. If  $M$  is orthogonal, then  $\det(M) = \pm 1$ .  
 3. If  $M$  is unitary, then  $|\det(M)| = 1$ .

*Proof.* Trivial. □

## 5.4 Minors, Cofactors and Inverses

**Definition 5.8.** Select a column  $\underline{C}_a$  of matrix  $M$ , we can write  $\underline{C}_a = M_{ia}e_i$ , so

$$\begin{aligned} \det M &= [\underline{C}_1, \underline{C}_2, \dots, \underline{C}_a, \dots, \underline{C}_n] \\ &= [\underline{C}_1, \underline{C}_2, \dots, M_{ia}e_i, \dots, \underline{C}_n] \\ &= M_{ia}[\underline{C}_1, \underline{C}_2, \dots, \underline{e}_i, \dots, \underline{C}_n] \\ &=: \sum_i M_{ia}\Delta_{ia} \end{aligned}$$

So

$$\Delta_{ia} = [\underline{C}_1, \underline{C}_2, \dots, \underline{e}_i, \dots, \underline{C}_n]$$

is called the cofactor. Note that the cofactor is exactly the determinant of the matrix removing the row  $i$  and column  $a$ .

*Proof of Proposition 5.7.* We know, then, that

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

by shuffling the rows and columns. Therefore we have

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

since columns and rows do not have real difference (we can do a transpose anyways), the two statements are proved. □

Now we want to find our adjugate (or adjoint)  $\tilde{M}$ . Note that in our above proof we can observe that  $M_{ib}\Delta_{ia} = \det(M)\delta_{ab}$ , so we can easily define  $\tilde{M}_{ij} = \Delta_{ji}$ . So  $\tilde{M}$  is the transpose of the matrix of cofactors, then the relation above becomes

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

as desired. This justifies the existence of  $\tilde{M}$  and  $\det(M)$  in the full generality of  $\mathbb{C}^N$  from beginning of this section.

## 5.5 System of Linear Equations

Consider a system of  $n$  linear equations in  $n$  unknowns  $x_i$ , written in vector-matrix form  $A\underline{x} = \underline{b}$  where  $A$  is some  $n \times n$  matrix. If  $\det A$  is nonzero, then  $A^{-1}$  exists, which implies a unique solution  $\underline{x} = A^{-1}\underline{b}$ . But what if  $\det A$  is zero?

We know that if  $\underline{b} \notin \text{Im } A$ , then by definition there is no solution. But if

$\underline{b} \in \text{Im } A$ , then the entire (shifted) space  $\underline{x}_p + \ker A$  where  $A\underline{x}_p = \underline{b}$  is exactly all the solutions. This can be seen from the linear superposition of the solutions. Note that the formula  $\underline{x}_p + \ker A$  also applied to the case for  $\det A \neq 0$  since in that case  $\ker A = \{\underline{0}\}$ . If  $\underline{u}_i$  is a basis for  $\ker A$ , then the general solution is  $\underline{x}_p + a_i \underline{u}_i$ .

**Example 5.3.** Consider  $A\underline{x} = B$  where

$$A = \begin{pmatrix} 1 & 1 & a \\ a & 1 & 1 \\ 1 & a & 1 \end{pmatrix}, \underline{b} = \begin{pmatrix} 1 \\ c \\ 1 \end{pmatrix}$$

Now  $\det A = (a-1)^2(a+2)$ . So if  $a \notin \{1, -2\}$ , then

$$A^{-1} = \frac{1}{(a-1)(a+2)} \begin{pmatrix} -1 & a+1 & -1 \\ -1 & -1 & a+1 \\ a+1 & -1 & -1 \end{pmatrix}$$

So

$$\underline{x} = A^{-1}\underline{b} = \frac{1}{(1-a)(a+2)} \begin{pmatrix} 2-c-ca \\ c-a \\ c-a \end{pmatrix}, c \in \mathbb{R}$$

Geometrically, this solves to give a point. If  $a = 1$ , then

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \implies \text{Im } A = \left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R} \right\}$$

So there is no solution if  $c \neq 1$ . For  $c = 1$ , since  $(1, 0, 0)^\top$  would be a particular solution, the solutions form the plane  $(1, 0, 0)^\top + \ker A$ , i.e. the general solution is of the form

$$\begin{pmatrix} 1 - \lambda - \mu \\ \lambda \\ \mu \end{pmatrix}, \lambda, \mu \in \mathbb{R}$$

For  $c = -2$ , by again looking at the image we conclude that  $c$  must be  $-2$ , in which case the same analysis gives us the general solution

$$\begin{pmatrix} 1 + \lambda \\ \lambda \\ \lambda \end{pmatrix}, \lambda \in \mathbb{R}$$

Let  $\underline{R}_1, \underline{R}_2, \underline{R}_3$  be the rows of  $A$ , then

$$A\underline{u} = \underline{0} \iff \forall i \in \{1, 2, 3\}, \underline{R}_i \cdot \underline{u} = 0$$

Note that the latter system of equations described planes through the origin. We know that the solution of the homogeneous problem is equivalent to finding the kernel of  $A$ . If  $\text{rank } A = 3$ , then we must have  $\underline{u} = \underline{0}$  since  $\{\underline{R}_i\}$  would be independent. If  $\text{rank } A = 2$ , then  $\{\underline{R}_i\}$  spans a plane, so the kernel is living along the line of normal to the plane. If  $\text{rank } A = 1$ , then the normals to the pairwise spanned planes are parallel, so the kernel is a plane.

Now consider instead  $A\underline{u} = \underline{b}$ , then it happens iff  $\underline{R}_i \cdot \underline{u} = b_i$ . In this case,

if  $\text{rank } A = 3$ , then the normals of the planes described by the system intersect at a point, thus there is a unique solution. If  $\text{rank } A = 2$ , then the planes may intersect, in which case the solution is a line, but it might not be the case. If  $\text{rank } A = 1$ , then the planes may coincide, in which case they are all the same, thus we have a plane of solution. But again this might not be the case. Two of them may coincide but that is not enough.

But how do you solve the equations and find the kernels systematically?

## 5.6 Gaussian Elimination and Echelon Form

Consider  $A\underline{x} = \underline{b}$  where  $\underline{x} \in \mathbb{R}^n$ ,  $\underline{b} \in \mathbb{R}^m$  and  $A$  is an  $m \times n$  matrix, then we can solve it by Gaussian Elimination. In general, we can reorder rows by row operations to rewrite original system in simpler form. Note that when we do row operations, we have to do it simultaneously on the matrix and on the vector. Our aim is to finally transform the matrix to the following form

$$M = \begin{pmatrix} M_{11} & \star & \star & \dots & \star & \star & \dots & \star \\ 0 & M_{22} & \star & \dots & \star & \star & \dots & \star \\ 0 & 0 & M_{33} & \dots & \star & \star & \dots & \star \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & M_{kk} & \star & \dots & \star \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}$$

Note that the row rank equals the column rank. We can use induction to prove that we can indeed use a way to obtain  $M$  from  $A$  in a finitely many number of row operations. Note that  $\det A = \pm \det M$

## 6 Eigenvalues and Eigenvectors

For a linear map  $T : V \rightarrow V$  where  $V$  is a real or complex vector space, a vector  $\underline{v} \neq \underline{0}$  is called an eigenvector of  $T$  with eigenvalue  $\lambda$  if  $T\underline{v} = \lambda\underline{v}$ . Note that this happens if and only if

$$A\underline{v} = \lambda\underline{v} \iff (A - \lambda I)\underline{v} = \underline{0}$$

where  $A$  is the matrix of  $T$ . So  $\lambda$  is an eigenvalue if and only if  $\det(A - \lambda I) = 0$  if and only if it is a root of the polynomial

$$\chi_A(t) = \det(A - tI)$$

which is a degree  $n$  polynomial.

**Definition 6.1.** For  $A$  an  $n \times n$  matrix, the characteristic polynomial  $\chi_A(t)$  is defined as  $\det(A - tI)$ .

This gives us a way to find the eigenvalues by looking at the characteristic polynomial and to find the eigenvectors by looking at the kernel of  $A - \lambda I$ .

**Example 6.1.** 1. Let  $V = \mathbb{C}^3$  and

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}$$

So  $\chi_A(t) = (2-t)^2 - 1$ , so  $\lambda$  is 1 or 3. For the eigenvalue  $\lambda = 1$ , we find that the eigenvectors are spanned by  $(1, i)$ , and for  $\lambda = 3$ , the eigenvectors are spanned by  $(1, -i)$ .

2. Let  $V = \mathbb{R}^2$ , consider the shear

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

So  $\chi_A(t) = (1-t)^2$ , so  $\lambda = 1$ . The eigenvectors are then spanned by  $(1, 0)$

3. Let  $V = \mathbb{C}^2$ , and

$$U = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

So  $\chi_U(t) = t^2 - 2t \cos \theta + 1$ , so  $\lambda = e^{\pm i\theta}$ , so the eigenvectors are spanned by  $(1, \pm i)$  respectively.

4. Consider  $V = \mathbb{C}^n$  and  $A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , then the eigenvalues are  $\lambda_i$  and the corresponding eigenvectors are spanned by  $\underline{e}_i$ .

**Proposition 6.1.** 1. *There exists at least one eigenvalue. In fact, there exists  $n$  eigenvalues counting multiplicity.*

2. *The trace of the matrix equals the sum of the eigenvalues (counting multiplicity as well).*

3.  $\det A = \chi_A(0)$ .

4. *If  $A$  is real, then non-real eigenvalues occur in conjugates.*

*Proof.* Trivial. □

**Definition 6.2.** For an eigenvalue  $\lambda$  of a matrix  $A$ , define the eigenspace  $E_\lambda$  is the set of all  $\underline{v}$  with  $A\underline{v} = \lambda\underline{v}$ , so  $E_\lambda = \ker(A - \lambda I)$ . The geometrical multiplicity of  $\lambda$  is  $m_\lambda = \dim E_\lambda$ , and the algebraic multiplicity  $M_\lambda$  is the multiplicity of  $\lambda$  as a root of  $\chi_A(t)$ . So  $\chi_A(t) = (t - \lambda)f(t)$  for  $f(\lambda) \neq 0$ .

**Proposition 6.2.**  $M_\lambda \geq m_\lambda$ .

*Proof.* Postponed to later. □

Note that the strict inequality can happen in many non-trivial cases.

**Example 6.2.** 1. Take

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

Then  $\chi_A(t) = (5-t)(t+3)^2$ , so 5, -3 are the eigenvalues. By calculation we find that  $E_5 = \text{span}\{(1, 2, -1)^\top\}$  and  $E_{-3} = \text{span}\{(-2, 1, 0)^\top, (3, 0, 1)^\top\}$ , so  $M_5 = m_5 = 1, M_{-3} = m_{-3} = 2$ .

2. Take

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

Then  $\chi_A(t) = -(t+2)^3$ , so the only eigenvalue is  $-2$ , but  $E_{-2}$  only has dimension 2, therefore  $M_{-2} = 3 \neq 2 = m_{-2}$ .

3. Consider the reflection across some plane through the origin. It reflects every point on the normal without changing direction and fixes every point on the plane. So the only eigenvalues would be  $\pm 1$  and  $M_{-1} = m_{-1} = 1, M_1 = m_1 = 2$ .
4. Consider the generic rotation through some normal. So the only real eigenvalue would be 1 and  $M_1 = m_1 = 1$  since the transformation changes every vector's direction except those on the normal, which it fixes.

**Proposition 6.3.** *Let  $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_r$  be eigenvectors of a matrix  $A$  with eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_r$ . If the eigenvalues are distinct, then the eigenvectors are linearly independent.*

*Proof.* Suppose  $\underline{w} = \sum_i a_i \underline{v}_i$ , then  $(A - \lambda I)\underline{w} = \sum_i a_i (\lambda_j - \lambda) \underline{v}_i$ . Let  $\underline{w} = \underline{0}$ , and if we can find a linear combination where not all  $a_i$ 's is 0, then we pick such a linear combination  $\underline{0} = \sum_i a_i \underline{v}_i$  such that the number of nonzero  $a_i$ 's is the least. WLOG in this choice  $a_1 \neq 0$ , then  $\underline{0} = (A + \lambda I)\underline{0} = \sum_i a_i (\lambda_j - \lambda_1) \underline{v}_i$ , which contradicts the minimality of the number of nonzero  $a_i$  of our choice.  $\square$

*Alternative proof.* Given a linear relation  $\underline{0} = \sum_i a_i \underline{v}_i$ , then for any  $j$ , the consider

$$\underline{0} = \prod_{i \neq j} (A - \lambda_i I) \sum_{i=1}^r a_i \underline{v}_i = a_j \underline{v}_j \prod_{i \neq j} (\lambda_j - \lambda_i)$$

So  $a_j = 0$ .  $\square$

## 6.1 Diagonalizability

**Proposition 6.4.** *For an  $n \times n$  matrix  $A$  acting on  $V = \mathbb{R}^n$  or  $\mathbb{C}^n$ , the followings are equivalent:*

1. *There exists a basis for  $V$  consisting of eigenvectors of  $A$ .*
2. *There exists an  $n \times n$  invertible matrix  $P$  such that  $P^{-1}AP = D$  where  $D$  is a diagonal matrix whose entries are the eigenvalues.*

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

*Proof.* For any matrix  $P$ ,  $AP$  has columns  $\underline{AC}_i^{(P)}$ , and  $DP$  has columns

$$\lambda_i \underline{C}_i^{(P)}$$

so what it is saying is that every column of  $P$  is an eigenvector of  $A$ . So if there is a basis of eigenvectors, we can just take  $P$  to be the matrix whose columns are the basis eigenvectors. Conversely, given such a matrix  $P$ , its columns are a basis for  $V$ .  $\square$

**Example 6.3.** 1. Consider

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

So we only have one eigenvector, hence  $A$  is not diagonalizable.

2. Consider

$$U = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

So it does not have real eigenvalues, but it has 2 complex eigenvalues with linearly independent eigenvectors, thus it is diagonalizable over  $\mathbb{C}^n$ . Indeed, we can let

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

Then

$$P^{-1}UP = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

**Proposition 6.5.** *If  $A$  has  $n$  distinct eigenvalues, then it is diagonalizable.*

*Proof.* Different eigenvalues give linearly independent eigenvectors.  $\square$

**Proposition 6.6.**  *$A$  is diagonalizable if and only if  $M_\lambda = m_\lambda$  for each eigenvalue  $\lambda$  of  $A$ .*

*Proof.* The union of the basis vectors for the eigenspaces would be a basis for the entire space since we would have  $\sum_\lambda M_\lambda = \sum_\lambda m_\lambda = n$ .  $\square$

**Example 6.4.** 1. Again use

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

We already know from previous example that  $\lambda = 5, 3$  and  $M_5 = m_5 = 1, M_{-3} = m_{-3} = 2$ , so  $A$  is diagonalizable. Indeed, we have

$$P^{-1}AP = \begin{pmatrix} 5 & & \\ & -3 & \\ & & -3 \end{pmatrix}, P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

2. Use the previous example

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

Then  $\lambda = -2$  and  $M_{-2} = 3 > m_{-2} = 2$ , so  $A$  is not diagonalizable. Indeed, if it is, then there is some invertible  $P$  with  $A = P(-2I)P^{-1} = -2I$  which is a contradiction.

**Definition 6.3.** Two  $n \times n$  matrices  $A, B$  are similar if there is some invertible  $n \times n$  matrix  $P$  such that  $A = PBP^{-1}$ .

This is trivially an equivalence relation.

**Proposition 6.7.** *If  $A, B$  are similar, then  $\text{tr } B = \text{tr } A$  and  $\det B = \det A$ . Also  $\chi_B = \chi_A$ .*

In fact, if two matrices are similar, then we could get from one to the other by a change of basis. The results here then are obvious.

*Proof.*

$$\begin{aligned}\operatorname{tr}(B) &= \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr}(A) \\ \det(B) &= \det(P^{-1}AP) = \det(P)^{-1} \det(A) \det(P) = \det(A) \\ \chi_A(t) &= \det(A - tI) = \det(P^{-1}) \det(A - tI) \det(P) = \det(B - tI) = \chi_B(t)\end{aligned}$$

As desired.  $\square$

Note that if  $A$  is diagonalizable, then so is  $B$  since they are all in the same equivalence class containing some diagonal matrix.

*Proof of Theorem 6.2.* Choose an eigenvalue  $\lambda$  of an  $n \times n$  matrix  $A$  with  $m_\lambda = r$  and a basis  $\underline{v}_1, \dots, \underline{v}_r$  for  $E_\lambda$ . Extend this basis to the entire space by adding vectors  $\underline{w}_{r+1}, \dots, \underline{w}_n$ . Consider the matrix  $P$  with columns  $\underline{C}_i(P) = \underline{v}_i$  for  $i = 1, \dots, r$  and  $\underline{C}_a(P) = \underline{w}_a$  for  $a = r+1, \dots, n$ , then  $AP = PB$  where  $B$  is of the form

$$B = \left( \begin{array}{c|c} \lambda I & * \\ \hline 0 & \tilde{B} \end{array} \right)$$

Note that  $P^{-1}AP = B$ , so  $\chi_A(t) = \chi_B(t) = (\lambda - t)^r \det(\tilde{B} - tI)$ , so  $M_\lambda \geq m_\lambda$ .  $\square$

## 6.2 Diagonalisation of Hermitian and Symmetric Matrices

Observe that if  $A$  is Hermitian, then  $(A\underline{v})^\dagger \underline{w} = \underline{v}^\dagger A \underline{w}$  for any complex vectors  $\underline{v}, \underline{w}$ .

**Proposition 6.8.** *Hermitian matrices have real eigenvalues and orthogonal eigenvectors for distinct eigenvalues. In addition, if the matrix is actually real (hence symmetric), then for each eigenvalue stated above, we can choose real eigenvectors.*

*Proof.* Let  $A$  be a Hermitian matrix and  $\lambda$  an eigenvalue of it with eigenvector  $\underline{v}$ , then

$$\bar{\lambda} \underline{v}^\dagger \underline{v} = (A\underline{v})^\dagger \underline{v} = \underline{v}^\dagger A \underline{v} = \lambda \underline{v}^\dagger \underline{v}$$

So  $\lambda = \bar{\lambda}$ , thus  $\lambda$  is real.

If  $\lambda, \mu$  are distinct eigenvalues with eigenvectors  $\underline{v}, \underline{w}$ , then

$$\mu \underline{v}^\dagger \underline{w} = \underline{v}^\dagger (A \underline{w}) = (A \underline{v})^\dagger \underline{w} = \lambda \underline{v}^\dagger \underline{w} \implies (\lambda - \mu) \underline{v}^\dagger \underline{w} = 0$$

So  $\underline{v}^\dagger \underline{w} = 0$  since  $\mu \neq \lambda$ , thus  $\underline{v} \perp \underline{w}$ .

For the last part, just choose the real or imaginary part (at least one of them is nonzero) of the eigenvectors.  $\square$

**Proposition 6.9** (Gram-Schmidt Orthogonalization). *Given a linearly independent set of vectors  $\underline{w}_1, \underline{w}_2, \dots, \underline{w}_k$  in  $\mathbb{C}^n$ , we can construct a set  $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k$  of orthonormal vectors such that  $\operatorname{span}(\{\underline{u}_i\}) = \operatorname{span}(\{\underline{w}_i\})$ .*

*Proof.* We do it recursively by

$$\underline{u}_1 = \frac{1}{\|\underline{w}_1\|} \underline{w}_1, \forall j > 1, \underline{w}_j^{(1)} = \underline{w}_j - (\underline{u}_1^\dagger \underline{w}_j) \underline{u}_1$$

and

$$\underline{u}_r = \frac{1}{\|\underline{w}_r^{(r-1)}\|} \underline{w}_r^{(r-1)}, \forall j > r, \underline{w}_j^{(r)} = \underline{w}_j^{(r-1)} - (\underline{u}_r^\dagger \underline{w}_j^{(r-1)}) \underline{u}_r$$

Which, as one can easily verify, is valid.  $\square$

Using Gram-Schmidt, we can convert any basis for some eigenspace to an orthonormal basis  $\mathcal{B}_\lambda$ . If  $\lambda_i$  are distinct eigenvalues of some Hermitian basis, then by choosing this orthonormal basis for all eigenspaces, we have an orthonormal set of eigenvectors.

**Example 6.5.** 1. Take

$$A = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}, A^\dagger = A$$

So the eigenvalues are 1, 3 with eigenvectors  $2^{-1/2}(1, i)^\top, 2^{-1/2}(1, -i)^\top$ . Note that these two eigenvectors is an orthonormal basis for  $\mathbb{R}^2$ . So we can choose

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

Which is unitary and  $P^{-1}AP = \text{diag}(1, 3)$ .

2. Consider

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

which is real and symmetric thus Hermitian and have eigenvalues  $-1, 2$  where the algebraic multiplicity of  $-1$  is 2. By calculation we find that the geometric multiplicity of  $-1$  is also 2. We can find an orthonormal basis for  $E_{-1}$  by the Gram-Schmidt process. For example we can have

$$E_{-1} = \text{span}\{2^{-1/2}(1, -1, 0)^\top, 6^{-1/2}(1, 1, -2)^\top\}$$

Also  $E_2 = \text{span}\{3^{-1/2}(1, 1, 1)^\top\}$ , thus we have found an orthonormal eigenbasis for  $\mathbb{R}^3$ . We can also find the corresponding (unitary hence orthogonal)  $P$  which diagonalizes  $A$ .

**Theorem 6.10.** *Any  $n \times n$  Hermitian matrix is diagonalizable by an unitary matrix. Equivalently, there is an orthonormal eigenbasis. In particular, for real and symmetric case, we can choose such that the orthonormal basis is real.*

*Proof.* We have already established all of the theorem but the part of the matrix being diagonalizable, so it suffices to show that, which is sadly not examinable, but here goes:

Consider a Hermitian  $A = \mathbb{C}^n \rightarrow \mathbb{C}^n$ . If  $V$  is any subspace of  $\mathbb{C}^n$  that is invariant under  $A$ , then for any  $\underline{v}_1, \underline{v}_2 \in V$  we have  $\underline{v}_1^\dagger(A\underline{v}_2) = (A\underline{v}_1)^\dagger \underline{v}_2$ . We shall prove by induction on  $m = \dim V \leq n$  that such a subspace has an orthonormal basis of eigenvectors of  $A$ , which shall establish the result by taking  $m = n$ . As the initial step is trivial, we will show the induction step.

Given  $V$  as above, since we are working in  $\mathbb{C}$ , the characteristic equation always



has a root, so there is some  $\underline{v} \in V$  such that  $A\underline{v} = \lambda\underline{v}$  for some  $\lambda$ . Let  $W$  be the subspace  $\{\underline{w} \in V : \underline{v}^\dagger \underline{w} = 0\}$ , then  $\dim W = m - 1 < m$ . Now

$$\underline{w} \in W \implies \underline{v}^\dagger(A\underline{w}) = (A\underline{v})^\dagger \underline{w} = \lambda \underline{v}^\dagger \underline{w} = 0 \implies A\underline{w} \in W$$

So  $W$  is invariant under  $A$ , so by induction hypothesis, there is an orthonormal basis for  $W$  of eigenvectors of  $A$ , say  $\{\underline{u}_1, \dots, \underline{u}_{m-1}\}$ , then we deem  $\underline{u}_m = |\underline{v}|^{-1} \underline{v}$ , then  $\{\underline{u}_1, \dots, \underline{u}_m\}$  gives an orthonormal basis of  $V$ , and the proof is done.  $\square$

### 6.3 Quadratic Forms

**Definition 6.4.** A quadratic form is a function  $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R}$  given by  $\mathcal{F}(\underline{x}) = \underline{x}^\top A \underline{x}$  where  $A$  is a real symmetric  $n \times n$  matrix. So  $\mathcal{F}(\underline{x}) = x_i A_{ij} x_j$ .

So by Theorem 6.10, we can diagonalize it by an orthogonal matrix  $P$ . Then by letting  $\underline{x}' = P^\top \underline{x}$ , we have  $\mathcal{F}(\underline{x}) = \underline{x}'^\top D \underline{x}'$  where  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , so we saw that  $\mathcal{F}$  is diagonalized. Note that  $\underline{x} = x'_i \underline{u}_i$  where  $\underline{u}_i$  is the orthonormal eigenbasis. Thus we can interpret what we have done as not changing the vector, but instead changing the basis in which we are working under. This basis (and the corresponding coordinate axes) is called the principle axes of the quadratic form.

**Example 6.6.** 1. In  $\mathbb{R}^2$ , take

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

So we can diagonalize it by  $\lambda_1 = \alpha + \beta$ ,  $\lambda_2 = \alpha - \beta$  and  $\underline{u}_1 = 2^{-1/2}(1, 1)^\top$ ,  $\underline{u}_2 = 2^{-1/2}(1, -1)^\top$ . Hence  $\underline{x}' = (2^{-1/2}(x_1 + x_2), 2^{-1/2}(x_1 - x_2))^\top$ . For example, if we take  $\alpha = 3/2, \beta = -1/2$ , then the quadratic form defines an ellipse. If we let  $\beta = 3/2, \alpha = -1/2$ , then the quadratic form would be a hyperbola.

2. In  $\mathbb{R}^3$ , by a suitable choice of basis,  $\mathcal{F}(\underline{x}) = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2$ . If  $\lambda_i > 0$ , the graph is an ellipsoid (a sphere stretched towards the directions of the axes). Otherwise, for example, the matrix that we have introduced in a former example

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

has eigenvalues  $-1, 2$  where the algebraic multiplicity of  $-1$  is 2. So on the principle axes  $\mathcal{F}_A(\underline{x}) = -x_1^2 - x_2^2 + 2x_3^2$ . Hence it gives a hyperboloid (obtained by rotating a hyperbola).

It has important application when we use the Hessians as the quadratic Form.

### 6.4 Cayley-Hamilton Theorem

**Theorem 6.11** (Cayley-Hamilton).  $\chi_A(A) = 0$ .

*Remark.* If  $c_0 I + c_1 A + \dots + c_n A^n$ , then  $A(c_1 I + C_2 A + \dots + c_n A^{n-1}) = -c_0 I$ . So if  $c_0 = \det A \neq 0$ , we have

$$A^{-1} = -\frac{1}{c_0}(c_1 I + C_2 A + \dots + c_n A^{n-1})$$

*Proof.* The general  $2 \times 2$  case is immediate. The case for diagonalizable matrices is also obvious. The general case can be dealt with a continuity argument or using modules. A proof by basic algebra is also presented in lecture but is deemed not elegant enough by the author, hence is omitted.  $\square$

## 7 Change of Bases, Canonical Forms and Symmetries

### 7.1 Change of Bases in General

Recall that given a linear operator  $T : V \rightarrow W$ , we can write its matrix form once we have specified the bases  $\{\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n\}$  and  $\{\underline{f}_1, \underline{f}_2, \dots, \underline{f}_m\}$  for  $V, W$ . So  $T(\underline{e}_i) = A_{ai}\underline{f}_a$  for some matrix  $A$ . For other bases  $\{\underline{e}'_i\}$  and  $\{\underline{f}'_i\}$  for  $V, W$ , we have a matrix  $A'$  with  $T(\underline{e}'_i) = A'_{ai}\underline{f}'_a$ . Suppose the bases are related by  $\underline{e}'_i = P_{ji}\underline{e}_j, \underline{f}'_a = Q_{ba}\underline{f}_b$ . So  $P, Q$  are invertible and

**Proposition 7.1.**  $A' = Q^{-1}AP$ .

We say  $P, Q$  constitutes a change of bases. The column  $i$  of  $A$  consists of components of  $T(\underline{e}_i)$  with respect to the basis  $\{\underline{f}_a\}$ , column  $i$  of  $P$  consists of components of new basis vectors with respect to the old basis, similar for  $Q$ .

**Example 7.1.** Let  $\dim V = n = 2$ ,  $\dim W = m = 3$ , and consider the linear map  $T$  with  $T(\underline{e}_1) = \underline{f}_1 + 2\underline{f}_2 - \underline{f}_3$  and  $T(\underline{e}_2) = -\underline{f}_1 + 2\underline{f}_2 + \underline{f}_3$ , so  $T$  has matrix

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

Under these choice of bases. Now consider the new basis  $\underline{e}'_i$  for  $V$  by

$$\underline{e}'_1 = \underline{e}_1 - \underline{e}_2, \underline{e}'_2 = \underline{e}_1 + \underline{e}_2$$

which yields a changes of basis matrix

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

And a new basis  $\underline{f}'_i$  for  $W$  is defined by

$$\underline{f}'_1 = \underline{f}_1 - \underline{f}_3, \underline{f}'_2 = \underline{f}_2, \underline{f}'_3 = \underline{f}_1 + \underline{f}_3$$

which has a matrix

$$Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

So the change of basis formula gives

$$A' = Q^{-1}AP = \begin{pmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix}$$

which can be verified by direct calculation.

*Proof.* For vectors  $\underline{y}, \underline{x}$  with  $\underline{y} = A\underline{x}$  for some invertible matrix (representing a basis)  $A$ , we denote the components in the respective components by

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

to avoid confusion. So for a change of basis by matrix  $P, Q$ , the formula  $Y = AX$  turns to  $Y' = A'X'$ , hence, we have  $\underline{x} = x_i \underline{e}_i = x'_j \underline{e}'_j = x'_j (\underline{e}_i P_{ij}) = P_{ij} x'_j \underline{e}_i$ . Therefore  $X = PX'$ , similarly  $Y = QY'$ , so  $A'X' = \bar{Y}' = Q^{-1}Y = Q^{-1}AX = (Q^{-1}AP)X'$ , so  $A' = Q^{-1}AP$   $\square$

In the special cases where  $V = W$  and the bases are the same for  $V, W$  before and after the change of basis, then  $A$  is transformed to a matrix similar to  $A$  by that change-of-basis matrix. Note that this can justify immediately that  $\text{tr } A' = \text{tr } A$ ,  $\det A' = \det A$ ,  $\chi_{A'} = \chi_A$  for similar matrices  $A, A'$ . If  $V = W = \mathbb{F}^n$  where  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{C}$ . For a diagonalizable matrix  $M$ , then by changing the standard basis into the eigenbasis,  $M$  becomes diagonal.

*Alternative proof.* Take the linear map  $T$ , then

$$f_a Q_{ab} A'_{bi} = f'_b A'_{bi} = T(\underline{e}'_i) = T(\underline{e}_j P_{ji}) = P_{ji} T(\underline{e}_j) = f_a A_{aj} P_{ji}$$

So  $AP = QA'$  by considering the coefficient of  $f_a$ .  $\square$

## 7.2 Jordan Canonical/Normal Form

This result classifies complex  $n \times n$  matrices up to similarity (i.e. conjugacy classes).

**Proposition 7.2.** *Any  $2 \times 2$  complex matrix  $A$  is similar to one of the followings:*

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

Where  $\lambda_1, \lambda_2, \lambda \in \mathbb{C}$ .

*Proof.* If  $A$  is diagonalizable (which includes the case where  $\chi_A$  has two distinct roots) then the proposition is immediate. Otherwise,  $\chi_A(t) = c(t - \lambda)^2$ ,  $c \neq 0$  and  $\text{null}(A - \lambda I) = 1$  and we shall show that  $A$  is similar to a matrix of the third form. Indeed, let  $\underline{v}$  be an eigenvector for  $\lambda$  and  $\underline{w}$  any vector that is independent from  $\underline{v}$ . Then  $A\underline{v} = \lambda\underline{v}$ ,  $A\underline{w} = \alpha\underline{v} + \beta\underline{w}$ , then the matrix of this transformation under the basis  $\underline{v}, \underline{w}$  is

$$\begin{pmatrix} \lambda & \alpha \\ 0 & \beta \end{pmatrix}$$

But then  $\beta = \lambda$  by looking at the eigenvalue of matrices of this form, and  $\alpha \neq 0$  by assumption. So we can set  $\underline{u} = \alpha\underline{v}$ , then under the basis  $\underline{u}, \underline{w}$ , we get the matrix of  $A$  to be

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

As claimed.  $\square$

*Alternative approach for the third case.* If  $A$  has characteristic polynomial of the form  $\chi_A(t) = c(t - \lambda)^2$  with  $c \neq 0$  and  $A \neq \lambda I$ , then there is some vector  $\underline{w}$  with

$$(A - \lambda I)\underline{w} = \underline{u} \neq \underline{0}$$

but by Cayley-Hamilton, we have

$$(A - \lambda I)\underline{u} = (A - \lambda I)^2 \underline{w} = \underline{0}$$

so  $A$  is of the form under the basis  $\underline{u}, \underline{w}$ . □

**Example 7.2.** Consider

$$A = \begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix}$$

Then  $\lambda = 3$  and we can have  $\underline{w} = (1, 0)^\top$ ,  $\underline{u} = (-2, -1)^\top$ , therefore

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}$$

**Theorem 7.3.** Any  $n \times n$  complex matrix is similar to a matrix of the form

$$A' = \begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_r}(\lambda_r) \end{pmatrix}$$

where  $\lambda_1, \dots, \lambda_r$  are the eigenvalues,  $n_1 + n_2 + \dots + n_r = n$  and  $J_p(\lambda)$ , the Jordan Block, is of the form

$$J_p(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}$$

### 7.3 Quadrics and Conics

A quadric in  $\mathbb{R}^n$  is a hypersurface defined by

$$Q(\underline{x}) = \underline{x}^\top A \underline{x} + \underline{b}^\top \underline{x} + c = 0$$

For some real symmetric  $n \times n$  matrix  $A$ ,  $\underline{b} \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$ . So  $Q(\underline{x}) = A_{ij}x_i x_j + b_i x_i + c$ .

Consider classifying solutions up to geometrical equivalence, i.e. no distinction up to isometries.

**Theorem 7.4.** Any isometry in  $\mathbb{R}^n$  is a composition of translation and orthogonal transformation.

*Proof.* Trivial. □

If  $A$  is invertible, we can complete the square. We set  $\underline{y} = \underline{x} + A^{-1}\underline{b}$ , so  $\mathcal{F}(\underline{y}) = \underline{y}^\top A \underline{y} = Q(\underline{x}) - k$  for some constant  $k$ . So  $Q(\underline{x}) = 0 \iff \mathcal{F}(\underline{y}) = k$  for some  $k$ , but the quadratic form is with respect to a new origin. In this case, we can diagonalize  $\mathcal{F}$  and change the basis as appropriate and think of them as new coordinate (bases). The case in  $\mathbb{R}^3$  has been discussed in the previous sections and we see that the quadrics are ellipsoids and hyperboloids (in 2 directions).

If  $A$  is singular, that is, if  $A$  has one or more 0 eigenvalue, things change.

Back in  $\mathbb{R}^2$ , if  $A$  is invertible, then by our analysis above, we can write it as a (diagonalized) quadratic form  $k = \lambda_1 x_1^2 + \lambda_2 x_2^2$ . For  $\lambda_1, \lambda_2 > 0$ ,  $k > 0$  gives an ellipse,  $k = 0$  gives a point and  $k < 0$  no solution. For  $\lambda_1 \lambda_2 < 0$ , so WLOG  $\lambda_1 > 0 > \lambda_2$ , then we get a hyperbola for any  $k \neq 0$ , and a pair of lines for  $k = 0$ .

If  $\det A = 0$ , then unless  $A$  is zero, we have  $\lambda_1 \neq 0$  and  $\lambda_2 = 0$ , so we diagonalize  $A$  in the original formula to get

$$\lambda_1 x_1'^2 + b_1' x_1' + b_2' x_2' + c = 0 \iff \lambda_1 x_1''^2 + b_2' x_2' + c' = 0$$

By a shift in  $x_1$ . If  $b_2' = 0$ , then we again get a pair of lines for  $c' < 0$ , a single line for  $c' = 0$  and no solution for  $c' > 0$ . Otherwise  $b_2' \neq 0$ , it gives a parabola. Note that all these changes of coordinates is prepresented by an isometry. They are called conics because these shapes can be obtained by slicing the cone. The general form for conics can be expressed in terms of  $a, b$  semi-major and semi-minor axes or in terms of length unit  $\ell$  and eccentricity  $e$ .

For ellipses, we have the general form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

WLOG  $b > a$ , then  $b^2 = (1 - e^2)a^2$  for some  $0 \leq e < 1$ . So we take this  $e$  to be the eccentricity. So the foci are at  $x = \pm ae$ .

And for parabola in the form  $y^2 = 4ax$ , the focus is at  $x = a$  and  $e = 1$ .

For hyperbola, i.e.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

So for  $b^2 = a^2(e^2 - 1)$  for  $e > 1$  the eccentricity. The foci are  $x = \pm ae$ .

## 7.4 Symmetries and Transformation Groups

Recall from a previous section that  $R$  is orthogonal iff  $R^\top R = RR^\top = I$  iff it preserves dot products. The set of such matrices is a group  $O(n)$ . Also, given the properties above and the multiplicativity of  $\det$ ,  $\det R = \pm 1$ . And  $SO(n) \leq O(n)$  is the set of orthogonal matrices with determinant 1. While orthogonal matrices preserves lengths, angles and volumes (by alternating forms), special orthogonal matrices also preserves orientation. Reflections belongs in  $O(n) \setminus SO(n)$ . For a specific  $H \in O(n) \setminus SO(n)$ , any element in  $O(n)$  is of the form  $R$  or  $RH$  for some  $R \in SO(n)$ . For example, if  $n$  is odd, then we can take  $H = -I$ . We can regard the transformation  $x'_i = R_{ij}x_j$  in two ways:

The “active” way is to say the rotations transform vectors. Then  $SO(n) \star \underline{x}$  would be a hypersphere. The components  $x'_i$  are components of new vector.

By contrast, the “passive” point of view is to think about the basis vectors and the vectors that they are mapped to, so  $x'_i$  are the components of the same vector  $\underline{x}$  wrt a new (rotated) orthonormal basis.

## 7.5 2-Dimensional Minkowski Space and Lorentz Transformations

Consider a new “inner product” in  $\mathbb{R}^2$  given by

$$(\underline{x}, \underline{y}) = \underline{x}^\top J \underline{y}, J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

And also we write  $\underline{x} = (x_0, x_1)^\top$  and  $\underline{y} = (y_0, y_1)^\top$ . Note that all original properties of inner products hold except positive definiteness. We can choose basis vectors that are orthonormal, which are the standard basis  $\underline{e}_0, \underline{e}_1$  with  $(\underline{e}_0, \underline{e}_0) = 1 = -(\underline{e}_1, \underline{e}_1), (\underline{e}_0, \underline{e}_1) = 0$ .

**Definition 7.1.** This “new inner product” is called the Minkowski metric, and  $\mathbb{R}^2$  equipped with it is called Minkowski space.

Consider

$$M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$$

which preserves Minkowski metric if and only if  $(M\underline{x}, M\underline{y}) = (\underline{x}, \underline{y})$  if and only if

$$\underline{x}^\top M^\top J M \underline{y} = \underline{x}^\top J \underline{y}$$

for any  $\underline{x}, \underline{y} \in \mathbb{R}^2$ , which happens iff  $M^\top J M = J$ . By taking determinant,  $\det(M) = \pm 1$ . So all such  $M$  constitutes a group and has a subgroup with  $\det M = 1$  and  $M_{00} > 0$  is the Lorentz group.

To determine the general form of  $M$  in Lorentz group, we can find it by requiring  $M\underline{e}_0, M\underline{e}_1$  to be orthonormal in this generalized sense, so  $M$  is in the general form

$$M(\theta) = \begin{pmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{pmatrix}$$

Fixed by  $(M\underline{e}_0, M\underline{e}_0) = 1$  and  $M_{00} > 0$ . Note that  $M(\theta)M(\phi) = M(\theta + \phi)$ .

Curves with  $(\underline{x}, \underline{x}) = k$  with  $k$  constant are simply hyperbolic. Since lengths are preserved,  $\underline{x}' = M\underline{x}$  must lie on the same hyperbola.

Physically, we want to set

$$M(\theta) = \gamma(v) \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}$$

where  $v = \tanh \theta$ , then  $|v| < 1$  and  $\gamma(v) = (1 - v^2)^{-1/2}$ . Then (with  $x_0 = t$  the time dimension and  $x_1 = x$  the space parameter)

$$\begin{cases} t' = \gamma(v)(t + vx) \\ x' = \gamma(v)(x + vt) \end{cases}$$

This is the Lorentz Transformation, or boost, relating time and space coordinates for observers moving with relative velocity  $v < 1$ . The  $\gamma$  factor here implies time dilation and length contraction. Note that  $M(\theta_3) = M_{\theta_1} M_{\theta_2}$  and  $v_i = \tanh(\theta_i)$  gives  $v_3 = \frac{v_1 + v_2}{1 + v_1 v_2}$  which is consistent with  $v < 1$ .