Linear Algebra II - Dr AG Curnock

Large Lecture Course

I gave this set of Lectures at the Mathematical Institute, University of Oxford, HT 2007- 2009.
My thanks go to the students who attended and gave feedback

Contents

Outline

Eight lectures to First year -Mods Students

Review of a matrix of a linear transformation with respect to bases, and change of bases.

Permutations and related topics. Determinants of square matrices; properties of the determinant function; determinants and the scalar triple product.

Methods for Computation of determinant. Proof that a square matrix is invertible if and only if its determinant is non-zero. Determinant of a linear transformation of a vector space to itself.

Eigenvalues of linear transformations of a vector space to itself. The characteristic polynomial of a square matrix and its uses. The linear independence of a set of eigenvectors associated with distinct eigenvalues; diagonalisation of matrices.

Recommended Bibliography

1. C. W. Curtis, *Linear Algebra - An Introductory Approach*, Springer (4th edition), reprinted 1994

2. R. B. J. T. Allenby, *Linear Algebra*, Arnold, 1995

3. T. S. Blyth and E. F. Robertson, *Basic Linear Algebra*, Springer, 1998

4. I. N. Herstein *Topics in Algebra* 1977, Wiley International Editions

5. P.G. Kumpel and J.A. Thorpe *Linear Algebra*

6. S. Lang *Linear Algebra*, Springer Undergraduate Mathematics Series, 1999

7. B. Seymour Lipschutz, Marc Lipson, *Linear Algebra*, Third Edition 2001

8. D. A. Towers, *A Guide to Linear Algebra*, Macmillan, 1988

   For Permutations :

9. R.B.J.T. Allenby *Rings, Fields and Groups*, Second Edition, Edward Arnold, 1999.

10. D.A. Wallace *Groups, Rings, and Fields.* Springer Undergraduate Mathematics Series,

Lecture 1

1 Review

1.1 Recap of Vector Spaces and $\mathbb{R}^n$

So far you will have looked at the vector space structure of $\mathbb{R}^n$, properties such as :

- zero vector;
- addition of vectors;
- scalar multiplication of vectors (by elements in $\mathbb{R}/\mathbb{C}$, or more generally in a Field);
- multiplication by negative scalars and we also have subtraction of vectors;
- we have a basis - n vectors which span the vectors space and which are linearly independent.

$\mathbb{R}^n$ is a prototype for any vector space. In fact it's much more - it's the prototype for many algebraic, geometric and topological structures. We will look at some of these over this course of lectures.

Bases

Spanning set- generates the space and so every element of the vector space can be written as a linear combination of vectors from the basis. The basis is a maximal set - if we increase it - it's no longer linearly independent. The basis vectors don't have to be mutually perpendicular - but there is such a canonical basis in $\mathbb{R}^n$, and we denote these by $\{e_1, e_2, ..., e_n\}$ where e_i have zero entries everywhere except at the ith entry it has a 1.

Bases are used in vector spaces to provide a method of computation. Two notions related to this are how to express (uniquely) a vector in terms of a given basis (studied in the MT course) and how to relate co-ordinate systems if you have two different bases on the same vector space.

Suppose we have a basis $B = \{v_1,\ ...,\ v_n\}$ of a vector space V. This means that every $v \in V$ can be expressed as a linear combination of elements of the basis:–

$$v = \lambda_1 v_1 + \lambda_2 v_2 + ... + \lambda_n v_n$$

where $\lambda_i \in F$. (Such a representation is unique as v_i are linearly independent) and so λ_i are completely determined by v and B. For this reason we may call the scalars λ_1 ..., λ_n called the **co-ordinates of** v **wrt. basis** $\{v_1, ..., v_n\}$ and $\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$ is called the **co-ordinate vector** for v wrt this basis.

This tells us something important : to each $v \in V \longleftrightarrow n$-tuple $(\lambda_1, ..., \lambda_n) \in F^n$. Conversely if $(\lambda_1, ..., \lambda_n)$ is an n-tuple in F^n then $\exists$ a vector in V of the form $\lambda_1 v_1 + \lambda_2 v_2 + ... + \lambda_n v_n$. This tell us there is a $1-1$ correspondence between elements of V and elements of F^n, which gives us a theorem :

Theorem 1.1. *If V is a finite dimensional vector space over a field F, then V and F^n are isomorphic (they are identical except possibly with different labels).*

1.2 Linear Transformations and Representing Matrices

This section is largely based on D.A. Towers or A.O. Morris (see the bibliography for details). Let V and W we finite-dimensional vector spaces over a field F and $T : V \to W$ be a linear transformation. Consider bases $\mathcal{B}_V = \{v_1, v_2, \ldots, v_n\}$ and $\mathcal{B}_W = \{w_1, w_2, \ldots, w_m\}$ for V and W respectively.

We may express the image of each v_i under T as a linear combination of elements of $\mathcal{B}_W$, namely

$$\begin{array}{ccccccc} T(v_1) & = & a_{11}w_1 & + & \ldots & + & a_{m1}w_m \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ T(v_i) & = & a_{1i}w_1 & + & \ldots & + & a_{mi}w_m \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ T(v_n) & = & a_{1n}w_1 & + & \ldots & + & a_{mn}w_m \end{array} \tag{1.1}$$

collecting the coefficients (we choose to do this in columns - the so-called *column convention*) gives us a matrix

$$A = (a_{ij}), \qquad \text{called the } \textit{Matrix Representation of } T,$$

or the *Transition matrix of* T. Here the a_{ij} belong to F. Note that A is the transpose of the matrix given in set of equations in (1.1). Thus

$$T(v_i) = \sum_{j=1,^m} a_{ji} w_j \quad \text{for each} \quad i$$

As $\mathcal{B}_W$ is a basis, the coefficients in the equations (1.1) are uniquely determined. If we wish to emphasise the bases used to find A we write $M_{\mathcal{B}_W}^{\mathcal{B}_V}(T)$ for A.

Example 1.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T((x, y, z)^t) = (2x + y, y + z)^t$.

Take the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$ denoting them by $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively. Write down the matrix for T; we call this matrix $A = M_{\mathcal{B}_2}^{\mathcal{B}_1}(T)$.

Solution Firstly $T((1, 0, 0)^t) = (2, 0)^t = 2(1, 0)^t + 0(0, 1)^t$; $\quad T((0, 1, 0)^t) = (1, 1)^t = 1(1, 0)^t + 1(0, 1)^t$ and

$T((0, 0, 1)^t) = (0, 1)^t = 0(1, 0)^t + 1(0, 1)^t$ which gives us $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$.

Example 1.3. Let V be an n-dimensional vector space over a field $\mathbb{R}$ and $id : V \to V$ be the identity transformation. Then for any basis $\mathcal{B}$ on V the matrix $M_{\mathcal{B}}^{\mathcal{B}}(id) = I_n$.

Remark We may write id_V but for ease of notation, often write id except when we wish to emphasise the domain.

CAUTION : If we work with distinct bases $\mathcal{B}_1, \mathcal{B}_2$ on V then $M_{\mathcal{B}_2}^{\mathcal{B}_1}(id) \neq I_n$, but we will shortly have a result relating these two matrices (Corollary 1.6 below). Firstly we see how to compute $M_{\mathcal{B}_2}^{\mathcal{B}_1}(id)$ applying $T = id$ to the equations in (1.1).

Example 1.4. Let $V = \mathbb{R}^3$ and consider the identity mapping $id : \mathbb{R}^3 \to \mathbb{R}^3$.
Find $M_{\mathcal{B}_2}^{\mathcal{B}_1}(id)$ when $\mathcal{B}_1 = \{(1, 5, 0)^t, (-1, 0, 2)^t, (0, 4, 1)^t\}$ and
$\mathcal{B}_2 = \{(-2, 1, 2)^t, (2, 3, -1)^t, (-1, -1, 1)^t\}$ are bases for V.

Solution We will find coefficients a_{11}, a_{21}, a_{31} which are found by solving

$$\begin{pmatrix} 1 \\ 5 \\ 0 \end{pmatrix} = a_{11} \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix} + a_{21} \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} + a_{31} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

These give us $a_{11} = 1, a_{21} = 1$ and $a_{31} = -1$.

Continuing in this way gives us the matrix

$$M_{\mathcal{B}_2}^{\mathcal{B}_1}(id) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ -1 & 3 & 0 \end{pmatrix}.$$

In fact we draw together results from your MT course relating simple properties of matrices of the form $M_{\mathcal{B}_W}^{\mathcal{B}_V}(T)$ into one Proposition.

Proposition 1.5. *Let S and T be linear transformations and V and W vector spaces over a field F. Suppose $S : V \to W$ and $T : V \to W$ where V and W have bases $\mathcal{B}_V$ and $\mathcal{B}_W$ respectively. Then the following hold*

(i) $M_{\mathcal{B}_W}^{\mathcal{B}_V}(S + T) = M_{\mathcal{B}_W}^{\mathcal{B}_V}(S) + M_{\mathcal{B}_W}^{\mathcal{B}_V}(T)$;

(ii) $M_{\mathcal{B}_W}^{\mathcal{B}_V}(\lambda T) = \lambda M_{\mathcal{B}_W}^{\mathcal{B}_V}(T)$;

(iii) If $R : W \to U$ where U is a vector space over F with basis $\mathcal{B}_U$ then

$$M_{\mathcal{B}_U}^{\mathcal{B}_V}(R \circ T) = M_{\mathcal{B}_U}^{\mathcal{B}_W}(R) \times M_{\mathcal{B}_W}^{\mathcal{B}_V}(T);$$

(iv) If T is Invertible then $M_{\mathcal{B}_V}^{\mathcal{B}_W}(T^{-1}) = M_{\mathcal{B}_W}^{\mathcal{B}_V}(T)^{-1}$.

Proof. See your MT Notes. □

This immediately gives us the following Corollary:

Corollary 1.6. *Let V be an n-dimensional vector space over a field $\mathbb{R}$ and $\mathcal{B}_1, \mathcal{B}_2$ be distinct bases on V. Then*

$$M_{\mathcal{B}_2}^{\mathcal{B}_1}(id) \times M_{\mathcal{B}_1}^{\mathcal{B}_2}(id) = I_n.$$

and so $M_{\mathcal{B}_2}^{\mathcal{B}_1}(id)$ is always invertible.

Proof. Apply the result in part (iii) above.

□

The above Proposition also gives us our first Change of Basis Theorem.

Theorem 1.7 (Change of Basis 1)**.** *Let V be a finite-dimensional vector space over a field F (e.g. $\mathbb{R}$) with distinct bases $\mathcal{B}_1$ and $\mathcal{B}_2$ and $T : (V, \mathcal{B}_i) \to (V, \mathcal{B}_i)$ be a linear transformation for $i = 1, 2$. Then there exists an invertible matrix P such that*

$$M_{\mathcal{B}_2}^{\mathcal{B}_2}(T) = PM_{\mathcal{B}_1}^{\mathcal{B}_1}(T)P^{-1},$$

and in fact we may take $P = M_{\mathcal{B}_2}^{\mathcal{B}_1}(id)$

Proof. Simply draw a mapping diagram and construct T as a composition operator and apply (iii) in Proposition 1.5.

□

Question :What happens if we have a general linear transformation

We are interested in changing the basis and to see what effect this has on the matrix form of the linear transformation. We will look at three cases via examples before stating the results formally. Namely considering :

(i) Changing the basis on the domain;

(ii) Changing the basis on the co-domain;

(iii) Changing the bases on both the domain and co-domain.

Example 1.8. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T((x, y, z)^t) = (2x + y, y + z)^t$. Refer to Example 1.2 above where we used standard bases for the domain and co-domain arriving at a matrix A.

(i) Take a new basis for the domain given by $\mathcal{B}_1' = \{(1, 1, 0)^t, (0, 1, 0)^t, (1, 1, -1)^t\}$ but keep the standard basis for the co-domain. Find the matrix B which represents T w.r.t. these bases.

State the relationship between A, B and $M_{\mathcal{B}_1}^{\mathcal{B}_1'}(id)$;

(ii) Take a new basis for the co-domain given by $\mathcal{B}_2' = \{(1, 1), (1, 0)\}$ but keep the standard basis for the domain. Find the matrix C which represents T w.r.t. these bases.

State the relationship between A, C and $M_{\mathcal{B}_2'}^{\mathcal{B}_2}(id)$;

(iii) Take the above new bases for $\mathbb{R}^3$ and $\mathbb{R}^2$. Find the matrix D which represents T w.r.t. these bases.

State the relationship between A, D, and $M^{\mathcal{B}_1'}_{\mathcal{B}_2'}(id)$.

Solution Recall $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$.

(i) Changing the basis on the domain means that T maps the new basis vectors to $T((1,1,0)^t) = (3,1)^t = 3(1,0)^t + 1(0,1)^t; T((0,1,0)^t) = (1,1)^t = 1(1,0)^t + 1(0,1)^t$ and

$T((1,1,-1)^t) = (3,0)^t = 3(1,0)^t + 0(0,1)^t$ which gives us $B = \begin{pmatrix} 3 & 1 & 3 \\ 1 & 1 & 0 \end{pmatrix}$. To find the relationship between A and B we need to compute $M^{\mathcal{B}_1}_{\mathcal{B}_1'}(id)$. As with the above example,

$$(1,0,0)^t = 1(1,1,0)^t - 1(0,1,0)^t + 1(1,1,-1)^t$$

etc which gives us $M^{\mathcal{B}_1}_{\mathcal{B}_1'}(id) = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

We see that if we post-multiply A by $M^{\mathcal{B}_1}_{\mathcal{B}_1'}(id)^{-1}$ we will have B,

i.e. $B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 3 \\ 1 & 1 & 0 \end{pmatrix}$.

Formally

$$B = \quad \text{New matrix for } T \quad = M^{\mathcal{B}_1'}_{\mathcal{B}_2}(T) = A \times M^{\mathcal{B}_1}_{\mathcal{B}_1'}(id)^{-1}.$$

[To prove this : look at the matrices as mappings w.r.t. the given bases and use the result that composing mappings is equivalent to multiplying the matrices. Here we would be assuming Prop 1.5. In the lecture we proved this without assuming 1.5 but just using first principles.]

(ii) Computing as above, for example $T((1,0,0)^t) = (2,0)^t = 0(1,1)^t + 2(1,0)^t$ we have

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

and if we pre-multiply A by $M^{\mathcal{B}_2}_{\mathcal{B}'_2}(id)$, we will have the matrix representation for T w.r.t. the new basis on the co-domain. Thus

$$C = M^{\mathcal{B}_1}_{\mathcal{B}'_2}(T) = M^{\mathcal{B}_2}_{\mathcal{B}'_2}(id) \times A;$$

(Hint : Here you will find that $M^{\mathcal{B}_2}_{\mathcal{B}'_2}(id) = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$ and it is straightforward to prove the relationship in (ii).)

(iii) This is the most general form. Again compute $T((1,1,0)^t) = (3,1)^t = 1(1,1)^t + 2(1,0)^t$ etc, which gives us

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

and we should find that

$$D = M^{\mathcal{B}'_1}_{\mathcal{B}'_2}(T) = M^{\mathcal{B}_2}_{\mathcal{B}'_2}(id) \times A \times M^{\mathcal{B}_1}_{\mathcal{B}'_1}(id)^{-1}.$$

(or $M^{\mathcal{B}'_1}_{\mathcal{B}'_2}(T) = QAP^{-1}$.)

Exercise 1.9. Confirm these last two calculations. Further exercises and examples may be found in, for example, D.A. Towers or A.O. Morris.

We can state this latter result formally as our second Change of Basis Theorem (which is sometimes called the QAP^{-1} Theorem.)

Theorem 1.10. *(Change of Basis 2) Let V and W be vector spaces over a field F and suppose V has a basis $\mathcal{B}_1$ and W has a basis $\mathcal{B}_2$. Let $T : (V, \mathcal{B}_1) \to (V, \mathcal{B}_2)$ be a linear transformation w.r.t. these bases. Suppose $\mathcal{B}'_1$ and $\mathcal{B}'_2$ are new bases for the domain and co-domain respectively.*

Then

$$M^{\mathcal{B}'_1}_{\mathcal{B}'_2}(T) = M^{\mathcal{B}_2}_{\mathcal{B}'_2}(id_W) \times A \times M^{\mathcal{B}_1}_{\mathcal{B}'_1}(id_V)^{-1}.$$

Proof. Consider the composition of map $T = id_W \circ T \circ id_V$ with respect to the following bases:

$$\begin{array}{ccccccc} V & \xrightarrow{id_V} & V & \xrightarrow{T} & W & \xrightarrow{id_W} & W. \\ \text{basis } \mathcal{B}_1' & & \text{basis } \mathcal{B}_1 & & \text{basis } \mathcal{B}_2 & & \text{basis } \mathcal{B}_2' \end{array}.$$

Then

$$\begin{aligned} M_{\mathcal{B}_2'}^{\mathcal{B}_1'}(T) &= M_{\mathcal{B}_2'}^{\mathcal{B}_1'}(id_W \circ T \circ id_V) \\ &= M_{\mathcal{B}_2}^{\mathcal{B}_2'}(id_W) \circ M_{\mathcal{B}_2}^{\mathcal{B}_1}(T) \circ M_{\mathcal{B}_1}^{\mathcal{B}_1'}(id_V) \end{aligned}$$

and by Proposition 1.5 and its Corollary, the required form appears. (i.e. by interchanging the bases on the final righthand matrix and taking its inverse.)

□

Lecture 2

2 Permutations

We wish to study the Determinant of a linear transformation and to do so we will need the concept of a Permutation, its sign and some simple properties. Permutations will be re-visited in the next course on Groups, Rings and Fields; we follow the conventions of Herstein and Dr W.B.S. Stewart (see booklist).

Let X be the set $\{x_1, x_2, \ldots, x_n\}$ then a *permuatation* σ *of* X is a bijection from X onto itself. Such a mapping may be written as

$$\sigma = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ (x_1)\sigma & (x_2)\sigma & \ldots & (x_n)\sigma \end{pmatrix}.$$

N.B. For functions and linear transformations you are used to writing $x \to f(x)$ or $x \to T(x)$ however in abstract algebra the convention is to write $(x)f$ when no confusion is likely (so if you are reading books : check what convention is used).

The set of all permutations on X is given by S_X. We will only consider finite sets X and so we may as well take $X = \{1, 2, \ldots, n\}$ (as there is always a bijection from X onto this set) and we write S_n in place of S_X, noting that S_n has precisely $n!$ elements.

We look at some examples to help build the theory of permutations.

Example 2.1. Let σ be a permutation in S_3 given by $1 \mapsto 2, 2 \mapsto 1$ and $3 \mapsto 3$. One notation for σ is to write the inputs in a row with the outputs in a row below each input.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

We tend to write the top row in increasing order but equally the order in which we write each column can be changed around and still give the same permutation. The important thing that matters is that we keep the input with its output. So we could equally write

$$\sigma = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The images of $1, 2$ and 3 are the same in each.

Composition and Inverses of Permutations

We have seen that we write $(1)\sigma$ to denote the image of an element under the permutation σ. We omit the brackets from now on unless we need them for clarity.

Definition 2.2. Let $\pi, \sigma \in S_n$. If $\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ 1\pi & 2\pi & \dots & n\pi \end{pmatrix}$, and $\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ 1\sigma & 2\sigma & \dots & n\sigma \end{pmatrix}$, then the *composition* $\pi \circ \sigma$ which means do π followed by σ, is denoted by

$$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ (1\pi)\sigma & (2\pi)\sigma & \dots & (n\pi)\sigma \end{pmatrix}.$$

By the *Identity mapping* we mean the mapping ι which sends each element to itself.

If π is defined as above, then the inverse π^{-1} is found by reading the columns upwards, i.e. $1\pi \mapsto 1, 2\pi \mapsto 2$ etc, giving us $\pi^{-1} = \begin{pmatrix} 1\pi & 2\pi & \dots & n\pi \\ 1 & 2 & \dots & n \end{pmatrix}$ and $\pi \circ \pi^{-1} = \iota$.

We would normally re-arrange the columns of π^{-1} so that elements in the top row are in increasing order (keeping the input with its output).

Example 2.3. If $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$,

find $\pi \circ \sigma$, $\sigma \circ \pi$, and π^{-1}, σ^{-1}. Check that $\pi \circ \pi^{-1} = \iota$. What is $(\pi \circ \sigma)^{-1}$ in terms of π and σ?

Solution $\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Starting from the left, $1 \mapsto 2$ in π and $2 \mapsto 3$ in σ, so $1 \mapsto 3$. So we have

$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. Similarly $\sigma \circ \pi = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$.

As we would expect, permutations **do not** generally commute under composition.

$$\pi^{-1} = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

[Check : $\pi^{-1} \circ \pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$.]

Exercise 2.4. Complete this example, in particular, see that $(\pi \circ \sigma)^{-1} = \sigma^{-1} \circ \pi^{-1}$. Prove this for permutations in S_n.

This two-row notation is limited : it does not help us to see the structure of the permutation, so we now look at *cycle notation* and *cycle structure*.

Structure of Permutations : Orbits and Cycles

Example 2.5. Consider the permutation $\sigma \in S_7$ given by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 7 & 5 & 4 & 1 & 2 & 6 \end{pmatrix}.$$

Writing a large dot for each input and an arrow for its output, represent this on a diagram. We can see some loops or *cycles* appearing and we need to be able to extract this structure. First let's define this more fully.

Definition 2.6. Let σ be a permutation in S_n and k be an element in $X = \{1, 2, \ldots, n\}$. The *orbit of k under σ* is the set of all elements

$$k(= (k)(\sigma)^0), (k)\sigma, (k)\sigma^2, \ldots, (k)\sigma^{\ell-1}.$$

[**Note** : Here in the above definition, as X is finite, there is a smallest integer ℓ such that $k\sigma^\ell = k$. More generally if X is an infinite set, the orbit of k is $\{k\sigma^i\}$, where $i = 0, \pm1, \pm2, \ldots$.] A *cycle* is the ordered list of the orbit of an element, i.e., a cycle of k is $(k, (k)\sigma, (k)\sigma^2, \ldots, (k)\sigma^{\ell-1})$. A cycle of length ℓ is called an *ℓ-cycle*. A cycle of length 2 is called a *transposition*.

Let's now find the orbits and cycles for each element in the above example. Consider 1. Then $1 = (1)\sigma^0, (1)\sigma = 3, (1)\sigma^2 = (3)\sigma = 5$, and finally $(1)\sigma^3 = 1$. Thus the orbit of 1 is $\{1, 3, 5\}$. The orbit of 2 is as follows : $2(\sigma^0) = 2, (2)\sigma = 7, (2)\sigma^2 = 6$ (and again $(2)\sigma^3 = 2$) thus the orbit of $2 = \{2, 7, 6\}$. The orbit of 4 is just itself. Thus the cycles are : cycle of $1 = (1, 3, 5)$ which is a 3-cycle; cycle of $2 = (2, 7, 6)$, also is a 3-cycle and finally the cycle of $4 = (4)$ which is a 1-cycle. Thus σ may be written as $\sigma = ((1, 3, 5), (2, 7, 6), (4))$. It is usual to suppress the commas giving us

$$\sigma = (135)(276)(4). \tag{2.1}$$

We saw in our diagram (and in the above notation) that these cycles were disjoint - here we have a permutation which is written as a "**Product of Disjoint Cycles**." This is a property we can prove more generally.

[**Caution** : removing commas as in 2.1 might lead to ambiguity. For example, how do we distinguish between $(1, 2, 3)$ and $(1, 23)$? We avoid confusion by pre-fixing a blank space in front of each numeral in a cycle. Thus writing the above as (1 2 3) and (1 23).]

Our next example gives permutations which are one ℓ-cycles.

Example 2.7. (i) Let $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 5 & 6 & 1 & 3 \end{pmatrix}$.

We can see that $1 \mapsto 2 \mapsto 4 \mapsto 6 \mapsto 3, 3 \mapsto 5$ and $5 \mapsto 1$. It is now clear that when looking for cycles, the first row of the permutation is somewhat redundant. What needs to be recorded is this *chain order* and so we adopt the following *cycle notation* : Thus $\sigma = (\ 1\ 2\ 4\ 6\ 3\ 5)$ and σ is a 6-cycle.

(ii) Let $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 1 & 4 & 3 & 6 \end{pmatrix}$. Notice that elements $2, 4$ and 6 are fixed under π and so these are not listed in the cycle notation. The cycle flows from the following map : $1 \mapsto 5, 5 \mapsto 3, 3 \mapsto 1$ and all the elements have been used. Thus $\pi = (\ 1\ 5\ 3)$, a 3-cycle.

(iii) List all the permutations in S_3 and express each as a cycle. (Our convention will be to omit the 1-cycles).

$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, the identity ι, by convention we write (1), a 1-cycle,

$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (12)$, $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = (23)$, and $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = (13)$ these are transpositions - meaning two elements move and the rest are fixed. Also $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (123)$, $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = (132)$, both are 3-cycles.

This last example gives us motivation for another definition of a cycle :

Definition 2.8. A permutations $\sigma \in S_n$ is a *cycle of length* ℓ where $2 \leq \ell \leq n$, if there exist distinct integers $a_i \in \{1, 2, \ldots, n\}$ for $i = 1, 2, \ldots \ell$ such that

$$\begin{aligned} a_{i+1} &= a_i\sigma \quad \text{for} \quad i = 1, 2, \ldots, \ell - 1 \\ a_1 &= a_\ell \sigma \\ b &= b\sigma \quad \text{for} \quad b \neq a_i, \quad \text{for all} \quad i = 1, 2, \ldots, \ell. \end{aligned}$$

Two questions spring to mind : (i) is the representation in 2.1 unique and (ii) are there other representations? We will come to these shortly. Firstly we need some more properties.

Remark 2.9. Authors usually omit the 1-cycles although you will need to know they could be there.
An ℓ-cycle can be written in ℓ ways (one for each choice of initial element).
Finally some authors use a canonical representation in which the leading number in a cycle σ is the smallest in the cycle and in which the leading elements increase as we go through the cycle.

We have the following Theorem which we will state (you must wait for the next course for its proof!)

Theorem 2.10. *Every permutation $\sigma \in S_n$ can be written as a product of disjoint cycles.*

This cycle decomposition is unique up to the order in which the cycles are written (noting that disjoint cycles commute which needs proving) and up to the choice of starting element for each cycle.

We now look at 2-cycles, transpositions.

Example 2.11. We first see that permutations (eg cycles) can be written as a product of transpositions.

(i) Consider $(\ 1\ 3\ 4) \in S_4$ where $(\ 1\ 3\ 4) = (\ 1\ 3)(\ 1\ 4)$. Multiply the RHS we see that working from left to right

$$\begin{array}{lll} 1 \mapsto 3 & 1 \mapsto 4 & \text{so in the end} \quad 1 \mapsto 3 \\ 2 \mapsto 2 & 2 \mapsto 2 & 2 \mapsto 2 \\ 3 \mapsto 1 & 3 \mapsto 3 & 3 \mapsto 4 \\ 4 \mapsto 4 & 4 \mapsto 1 & 4 \mapsto 1. \end{array}$$

It is clear that this is the LHS. Notice that 2 is fixed under all transpositions on the RHS and on the LHS so we don't need to consider this element. After a time you just do all this in your head.

(ii) Show that $(\ 1\ 2\ 3) = (\ 1\ 3)(\ 2\ 3) \in S_4$. (Now we can see that this representation is not unique.)

Then on the RHS we will have

$$\begin{array}{lllll} 1 \mapsto 3 & 1 \mapsto 1 & \text{so in the end} & 1 \mapsto 2 \\ 2 \mapsto 2 & 2 \mapsto 3 & & 2 \mapsto 3 \\ 3 \mapsto 1 & 3 \mapsto 2 & & 3 \mapsto 1 \end{array}$$

(iii) $(\ 3\ 6\ 2\ 8\ 5) = (\ 3\ 6)(\ 2\ 3)(\ 3\ 8)(\ 5\ 3) \in S_8$. We can see that on both sides 1,4 and 7 are fixed so we omit these. On the RHS we have

$$\begin{array}{llll} 2 \mapsto 2 & 2 \mapsto 3 & 2 \mapsto 2 & 2 \mapsto 2 \\ 3 \mapsto 6 & 3 \mapsto 2 & 3 \mapsto 8 & 3 \mapsto 5 \\ 5 \mapsto 5 & 5 \mapsto 5 & 5 \mapsto 5 & 5 \mapsto 3 \\ 6 \mapsto 3 & 6 \mapsto 6 & 6 \mapsto 6 & 6 \mapsto 6 \\ 8 \mapsto 8 & 8 \mapsto 8 & 8 \mapsto 3 & 8 \mapsto 8. \end{array}$$

Passing through from left to right gives us the LHS.

Exercise : Find a product of transpositions for $(\ 3\ 7\ 5)(\ 4\ 6\ 8)$.

Generally an ℓ-cycle $(\ a_1\ a_2 \ldots\ a_\ell)$ can be written as

$$(\ a_1\ a_2\ \ldots\ a_\ell) = (\ a_1\ a_2)(\ a_1\ a_3)\ \ldots\ (\ a_1\ a_\ell)$$

We will prove this below.

Proposition 2.12. *For $n \geq 2$, every cycle $\sigma \in S_n$ can be written as a product of transpositions.*

We will complete the proof when next we meet but I add this for you to read.

Proof. By theorem 2.10, it suffices to prove this for σ an $\ell-$cycle. Take $\sigma = (\ a_1\ a_2\ \ldots\ a_\ell)$ with $a_i \in \{1, 2, \ldots, n\}$. If $\ell = 1$ then for example, $\sigma = (a_1) = (\ a_1,\ a_2) = \iota = (\ a_1,\ a_2)(\ a_2,\ a_1)$. If $\ell \geq 2$ then we claim that

$$(\ a_1\ a_2\ \ldots\ a_\ell) = (\ a_1\ a_2)(\ a_1\ a_3)\ \ldots\ (\ a_1\ a_\ell)$$

Taking the RHS we can see that a_2 is fixed in all the transpositions except the first. In the first one, a_1 maps to a_2 which agrees with the LHS. Similarly for each i on the right, a_i is fixed everywhere except in $(\ a_1\ a_i)$ when $a_1 \mapsto a_i$. On the left, there is a connected path from a_1 to a_i. $\square$

This representation is not unique but what is unique is whether there are an odd or even number of transpositions in a product.

Definition 2.13. We define the *Parity* of a permutation $\sigma \in S_n$ to be *even* if we can express σ in an even number of transpositions; it is *odd* if we can express σ in an odd number of transpositions.

Lecture 3

3 Parity

We begin with a more formal definition of Parity, in terms of the *sign function* (or *signature function*).

Let $p(x_1, x_2, \dots, x_n)$ be a polynomial and σ be a permutation in S_n. We can always form a new polynomial in terms of p and σ, namely, $p_\sigma(x_1, x_2, \dots, x_n) := p(x_{(1)\sigma}, x_{(2)\sigma}, \dots, x_{(n)\sigma})$.

Now let P be the polynomial $P(x_1, x_2, \dots, x_n) = \prod_{1 \le i < j \le n}(x_i - x_j)$, then :

Definition 3.1. If $\sigma \in S_n$ then *the sign*, $\mathrm{sign}(\sigma)$ of σ is defined to be

$$\mathrm{sign}(\sigma) := \frac{P_\sigma(x_1, x_2, \dots, x_n)}{P(x_1, x_2, \dots, x_n)} = \frac{P(x_{(1)\sigma}, x_{(2)\sigma}, \dots, x_{(n)\sigma})}{P(x_1, x_2, \dots, x_n)}.$$

Equally

$$\mathrm{sign}(\sigma) = \frac{\prod_{1 \le i < j \le n}(x_{(i)\sigma} - x_{(j)\sigma})}{\prod_{1 \le i \le j \le n}(x_i - x_j)}.$$

To see how this definition gives us just a ± 1 we look at an example.

Example 3.2. Let $\sigma = (\ 1\ 2\ 3)$. Compute the sign of σ and check that this agrees with its parity. Well,

$$\mathrm{sign}(\sigma) = \frac{(x_{(1)\sigma} - x_{(2)\sigma})(x_{(1)\sigma} - x_{(3)\sigma})(x_{(2)\sigma} - x_{(3)\sigma})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)}.$$

On the numerator we have $(x_2 - x_3)(x_2 - x_1)(x_3 - x_1)$, and so

$$\mathrm{sign}(\sigma) = \frac{(x_2 - x_3) \times -1(x_1 - x_2) \times -1(x_1 - x_3)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1.$$

Now $(\ 1\ 2\ 3) = (\ 1\ 2)(\ 1\ 3)$ in terms of transpositions, and σ is even and its parity is $+1$.

We shall see that for any permutation, the parity of σ coincides with its sign where $sign(\sigma) = \pm 1$. This needs proving.

Lemma 3.3. *Let $\sigma \in S_n$, then the parity of σ is given by its sign. That is, sign$(\sigma) = +1$ if σ is even, and sign$(\sigma) = -1$ if σ is odd.*

Proof. There are two things to show : (a) that if σ is a transposition, then its sign is -1; (b)sign is a multiplicative function, namely, if σ, π are permutations in S_n then $\text{sign}(\sigma\pi) = \text{sign}(\sigma)\text{sign}(\pi)$.

We begin with (a) : Let $\sigma = (\ k\ \ell)$ be a transposition in S_n with $k < \ell$ where $h, \ell \in \{1, 2, \dots, n\}$.

By definition we need only consider the factors $(x_{(i)\sigma} - x_{(j)\sigma})$ for any $1 \le i \le j \le n$.

We claim that $\text{sign}(\ k\ \ell) = -1$ for $k < \ell$. We decompose $\prod_{1 \le i \le j \le n}(x_{(i)\sigma} - x_{(j)\sigma})$ into 4 disjoint sets of indices :

$$\begin{aligned}
\prod_{1\le i\le j\le n}(x_{(i)\sigma} - x_{(j)\sigma}) &= (x_\ell - x_k)\times \\
&\times \prod_{i<j, i\neq k, j\neq \ell}(x_i - x_j) \\
&\times \prod_{k<j<\ell, j\neq \ell}(x_\ell - x_j) \\
&\times \prod_{k<i<\ell, i\neq k}(x_i - x_k).
\end{aligned}$$

Looking at the number of sign changes on the RHS we arrive at :

$(-1)\times(-1)^{\#\{j:k<j<\ell\}}\times(-1)^{\#\{i:k<i<\ell\}}\prod_{i<j}(x_i - x_j)$ which is $-1\prod_{i<j}(x_i - x_j)$ and so the sign of a transposition is -1.

(b) Finally let σ, π be permutations in S_n.

Then

$$\begin{aligned}
\text{sign}(\sigma\pi) &= \frac{P_{\sigma\pi}(x_1, x_2, \dots, x_n)}{P(x_1, x_2, \dots, x_n)} \\
&= \frac{P(x_{(1)\sigma\pi}, x_{(2)\sigma\pi}, \dots, x_{(n)\sigma\pi})}{P(x_1, x_2, \dots, x_n)} \\
&= \frac{P(x_{(1)\sigma}, x_{(2)\sigma}, \dots, x_{(n)\sigma})\cdot P(x_{(1)\sigma\pi}, x_{(2)\sigma\pi}, \dots, x_{(n)\sigma\pi})}{P(x_1, x_2, \dots, x_n)\cdot P(x_{(1)\sigma}, x_{(2)\sigma}, \dots, x_{(n)\sigma})}.
\end{aligned}$$

Letting $y_i = (x_i)\sigma$ in the numerator and denominator on the right we have $\text{sign}(\sigma\pi) = \ \text{sign}(\sigma)(\pi)$

To conclude the proof we note that if $\sigma_1, \sigma_2, \ldots, \sigma_k$ are transpositions and $\sigma = \sigma_1\sigma_2 \ldots \sigma_k$ then

$$\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma_1\sigma_2 \ldots \sigma_k) = \operatorname{sign}(\sigma_1) \operatorname{sign}(\sigma_2) \ldots \operatorname{sign}(\sigma_k) = (-1)^k.$$

Hence if k is the number of transpositions in σ then if k is odd, $\operatorname{sign}(\sigma) = -1$ and if k is even, $\operatorname{sign}(\sigma) = +1$. □

We have now built sufficient machinery to be able to proof results about Determinants.

4 Determinants

Much of this section is based on Hersteins' *Topics in Algebra*, and Curtis' *Linear Algebra*.

A determinant is a scalar we associate with a square matrix, formally if A is a matrix in $M_{n\times n}(F)$ where F is a field of scalars (eg $\mathbb{R}$), then the *determinant of* A,$det(A)$ (or $|A|$) is $det(A) : F^n \to F$. Determinants play a key role in the solution of a system of linear equations. (We shall see that $det(A) \neq 0 \Leftrightarrow A$ is non-singular - it has an inverse). Determinants also give us some useful geometrical information about the transformation that the matrix represents.

Geometric Motivation for Determinants

We shall see that geometrically the **determinant** is the **volume of a parallelepiped whose edges are the rows of the matrix** (up to a $\pm$ sign).

Consider a parallelogram in the plane whose edges are the vectors $\{a_1, a_2\}$. Consider a function that computes the area $A(a_1, a_2)$ of such a parallelogram. What properties would be desirable?

We would hope this area function A has the following properties :

A1 $A(e_1, e_2) = 1$ where e_i are the usual basis vectors of $\mathbb{R}^2$;

A2 $A(\lambda a_1, a_2) = A(a_1, \lambda a_2) = \lambda A(a_1, a_2)$ for any $a_i \in \mathbb{R}^2$, and $\lambda \in \mathbb{R}$;

A3 $A(a_1 + a_2, a_2) = A(a_1, a_2) = A(a_1, a_1 + a_2)$ for any $a_i \in \mathbb{R}^2$,.

We shall prove that these properties characterise a *Determinant Function* and moreover we have a useful test for the independence of two vectors from the above properties A1 - A3:

$$A(a_1, a_2) \neq 0 \qquad iff \qquad \{a_1, a_2\} \qquad \text{are linearly independent.}$$

Computation of 2×2 and 3×3 matrices

This will be familiar to you.

Definition 4.1. 1. If $A = (a)$ then its determinant, $det(A) = a$

2. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $det(A) = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

So for a diagonal matrix, $A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$ then $det(A) = ad$ which is the product of its diagonal elements . Notice that if we switch a column in A then the sign of the determinant changes. (i.e. $|A'| = \begin{vmatrix} b & a \\ d & c \end{vmatrix} = bc - ad = -|A|$.

We can make some further simple observations.

1. If we form the matrix A' by adding row 2 to row 1, $r_1 \to r_1 + r_2$, we leave the determinant unchanged.

   i.e. $det(A') = \begin{vmatrix} a+c & b+d \\ c & d \end{vmatrix} = (a+c)d - (b+d)c = ad - bc = |A|.$

2. If we form the matrix A'' by multiply a row by a scalar, say $r_1 \to \lambda r_1$, the determinant changes by this factor. i.e.

   i.e. $det(A'') = \begin{vmatrix} \lambda a & \lambda b \\ c & d \end{vmatrix} = \lambda ad - \lambda bc = \lambda(ad - bc) = \lambda |A|.$

Method of Cofactors

Definition 4.2. Let A be a 3×3 matrix where $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

We define the *Minor* of an element a_{ij} to be the 2×2 determinant formed by deleting the row and column from A in which a_{ij} occurs.

For example, the

$$\text{Minor of} \quad a_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} = a_{11}a_{32} - a_{12}a_{31}.$$

Superimposing a chequerboard pattern of signs on A, namely $\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$, we define the *Cofactor* of a_{ij} to be *the signed minor*, denoted by A_{ij}. For example, $A_{23} = -(a_{11}a_{32} - a_{12}a_{31})$.

Definition 4.3. If A is a 3×3 matrix, we define its determinant to be

$$|A| = \sum_{j=1}^{j=3} a_{ij}A_{ij} \quad \text{and} \quad i = 1, 2 \text{ or } 3 \tag{4.1}$$

$$= \sum_{i=1}^{i=3} a_{ij}A_{ij} \quad \text{and} \quad j = 1, 2 \text{ or } 3. \tag{4.2}$$

The two equations in 4.1 and 4.2 are known as the *Laplace expansions of the determinant of* A by the i-th row and j-th column, respectively.

Example 4.4. Expand A by r_1 using the above 4.1.

Solution $|A| = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$ which gives is $|A|$ in terms of 6 products, each of three elements with exactly one from each row and column.

$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$.

Note : a) The subscripts are all of the form $1, (1)\sigma \;\; 2, (2)\sigma \;\; 3, (3)\sigma$. (For clarity we use a comma between the first subscript and its image under σ.)

b) The second product has the sign -1 as this is an *odd* permutation. Here $1 \mapsto 1, 2 \mapsto 3, 3 \mapsto 2$ so as cycles is $(\ 1)(\ 2\ 3)$ and as a product of (one) transposition is $(\ 2\ 3)$.

c) Similarly in the fourth product its subscripts send $1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 1$, which is the cycle $(\ 1\ 2\ 3) = (\ 1\ 2)(\ 1\ 3)$ which is even.

Example 4.5. Evaluate $|A|$ where $A = \begin{pmatrix} 2 & 3 & 1 \\ 5 & 0 & -1 \\ 7 & 2 & 4 \end{pmatrix}$. If we interchange the first and second column, what happens to the determinant?

Solution Expanding by the first row gives us

$$|A| = \sum_{j=1}^{j=3} a_{1j}A_{1j} = +2\begin{vmatrix} 0 & -1 \\ 2 & 4 \end{vmatrix} - 3\begin{vmatrix} 5 & -1 \\ 7 & 4 \end{vmatrix} + 1\begin{vmatrix} 5 & 0 \\ 7 & 2 \end{vmatrix} = -67.$$

We should notice a change in sign if we swop two columns.

Exercise 4.6. State what happens to the value of the determinant if (i) a row or column in the above 3×3 matrix ia zero? (ii) we multiply a row or column by a scalar, eg, λ, (iii) if we add a multiple of λ row 1 to row 3?

It is clear that for larger matrices we need some theory to help us find more efficient expansions of a determinant.

Lecture 4

4.1 Determinants for $n \times n$ Matrices

We now give a formal definition of a determinant. Recall that we saw in the case of a 2×2 matrix that the interchange of two columns produced a sign change in the determinant. This is precisely the sign or parity of a permutation on the number of transpositions of the subscripts (regarded as cycles).

Let A be a matrix in $M_n(\mathbb{R})$. We define the determinant of A to be the algebraic sum of all products of n elements from A which is formed by selecting exactly one element from each row and column of A.

Definition 4.7 (Determinant of Matrix)**.** If a matrix $A = (a_{ij})$ and $A \in M_n(\mathbb{R})$ then the determinant of A is given by

$$det(A) := \sum_{\sigma \in S_n} \text{sign}(\sigma)\ a_{1,(1)\sigma}\ a_{2,(2)\sigma} \dots a_{n,(n)\sigma},$$

where the sum is over all permutations $\sigma \in S_n$. Equally

$$det(A) := \sum_{\sigma \in S_n} (-1)^{\sigma}\ a_{1,(1)\sigma}\ a_{2,(2)\sigma} \cdots\ a_{n,(n)\sigma},$$

where $(-1)^{\sigma}$ denotes the parity of σ which is $+1$ if the permutation σ is even, and -1 is σ is odd.

It is not difficult to see that this agrees with our earlier definition of a determinant for small matrices in terms of cofactors but also generalises the method of cofactors to $n \times n$ matrices. However given $n!$ products in each determinant such a definition does not readily help us compute determinants. The definition tells us that the determinant of a matrix will be zero if any row or column is zero.

We now establish some natural properties of the determinant, regarded as a function of the rows of a matrix. For this let $A \in M_n(\mathbb{R})$ where the ith row of A is r_i, then it is natural to write :

$$Det(A) := D(r_1, \dots, r_i, \dots, r_n).$$

A natural question is what properties does D have? We shall see that it satisfies the Elementary Row Operations and so this will enable us to prove that we can row reduce a matrix to echelon form to determine its determinant.

Theorem 4.8. *The function D on a matrix in $M_n(\mathbb{R})$ is linear in rows, that is,*

D1 $D(r_1, \dots, r_i, \dots, r_n) \ + D(r_1, \dots, r_i', \dots, r_n) \ = D(r_1, \dots, r_i + r_i', \dots, r_n) \ \ \forall i;$

D2 $D(r_1, \dots, r_{i-1}, \lambda r_i, r_{i+1}, \dots, r_n) \ = \lambda D(r_1, \dots, r_i, \dots, r_n),$ *for all* $\lambda \in \mathbb{R}, \ \lambda \neq 0, \ \forall i$ *;*

Also

D3 $D(e_1, e_2, \dots, e_n) \ = 1,$ *for* e_i *the usual* i*th unit vector in* $\mathbb{R}^n$.

Proof. D1 On the right we have

$$\sum_\sigma (-1)^\sigma \ a_{1,(1)\sigma} \ \dots (a_{i,(i)\sigma} + \ a'_{i,(i)\sigma}) \ \dots a_{1(1)\sigma} \dots \ a_{n,(n)\sigma}$$

and as only the ith row is changed, and summation is additive, we have :

$$\sum_{\sigma \in S_n} (-1)^\sigma \ a_{1,(1)\sigma} \ \dots a_{i(i)\sigma} \ \dots\dots \ a_{n,(n)\sigma} \ +$$
$$\sum_\sigma (-1)^\sigma \ a_{1(1)\sigma} \ \dots a'_{i,(i)\sigma} \ \dots\dots \ a_{n,(n)\sigma}.$$

D2 Again only entries in the i-th row change. On the left we have

$$\sum_{\sigma\in S_n} \sigma(-1)^{\sigma} a_{1,(1)\sigma} \ \ldots a_{i-1,(i-1)\sigma} \ \lambda a_{i,(i)\sigma} \ a_{i+1,(i+1)\sigma} \cdots \ a_{n,(n)\sigma} \ =$$
$$\lambda \sum_{\sigma\in S_n} (-1)^{\sigma} \ a_{1,(1)\sigma} \ldots a_{i,(i)\sigma} \ldots a_{n,(n)\sigma},$$

as summation is linear.

D3 $D(e_1, e_2, \ldots, e_n) = \sum_{\sigma} (-1)^{\sigma} e_{1,(1)\sigma} \ \ldots e_{i,(i)\sigma} \ \ldots \ e_{n,(n)\sigma}$ where $e_{i,(j)\sigma} = 1$ if $i = j$, and zero otherwise. Hence the result follows.

Note that D3 tells us that $D(I_n) = 1$, or that D is $1-$ preserving.

□

Proposition 4.9. *If we interchange two rows the resulting determinant is changed by a factor of* -1*, that is,*

$$D(r_1, \ldots, r_i, \ldots, r_j, \ldots, r_n) = -D(r_1, \ldots, r_j, \ldots, r_i, \ldots, r_n)$$

for all $i, j \in \{1, 2, \ldots, n\}, i \neq j$*;* D *is called* Alternating.

Proof. If A is the original matrix, then

$$|A| = \sum_{\sigma\in S_n} \text{sign}(\sigma) \ a_{1,(1)\sigma} a_{2,(2)\sigma} \ldots a_{i,(i)\sigma} \ldots a_{j(j)\sigma} \ldots a_{n,(n)\sigma}.$$

But if A' is the matrix A with the ith and jth row interchanged, then

$$|A'| = \sum_{\sigma\in S_n} \text{sign}(\sigma) \ a_{1,(1)\sigma} a_{2,(2)\sigma} \ldots a_{j,(j)\sigma} \ldots a_{i,(i)\sigma} \ldots a_{n,(n)\sigma}.$$

If we think of the sign(σ) in terms of the definition of the polynomials given in 3.1 we will see that ($(1)\sigma \ (2)\sigma \ldots \ (j)\sigma \ldots \ (i)\sigma \ldots \ (n)\sigma) = -1(\ (1)\sigma \ 2(\sigma) \ldots \ (i)\sigma \ldots \ (j)\sigma \ldots \ (n)\sigma)$ applying Lemma 3.3) as it is a transposition. Thus $|A'| = -|A|$. □

We shall call a function which satisfies $D1, D2$, and $D3$ and is Alternating a **Determinant function.**

Proposition 4.10. *If we replace row r_i in a determinant by row $r_i + \lambda r_j$ the determinant is unchanged, namely,*

$$D(r_1, \ldots, r_i, \ldots, r_n) = D(r_1, \ldots, (r_i + \lambda r_j), \ldots, r_n),$$

for all $\lambda \in \mathbb{R}$, and for all $i, j, i \neq j, 1 \leq i, j \leq n$.

[This will be set as an Exercise in the lecture.]

Proof.

$$\begin{aligned}
& D(r_1, r_2, \ldots, r_i, \ldots, r_j \ldots, r_n) \\
= \; & \frac{1}{\lambda} D(r_1, r_2, \ldots, r_i, \ldots, \lambda r_j \ldots, r_n) \quad \text{by D2,with } \lambda \neq 0 \\
= \; & \frac{1}{\lambda} D(r_1, r_2, \ldots, r_i + \lambda r_j, \ldots, \lambda r_j \ldots, r_n) \quad \text{by} D1' \\
= \; & \frac{\lambda}{\lambda} D(r_1, r_2, \ldots, r_i + \lambda r_j, \ldots, r_j \ldots, r_n) \quad \text{by D2.}
\end{aligned}$$

[To be precise $D1'$ is a modified form of $D1$, where

$$D1' : D(r_1, \ldots, r_i, \ldots, r_n) + D(r_1, \ldots, r_i, \ldots, r_j, \ldots r_n) = D(r_1, \ldots, r_i + r_j, \ldots, r_n)$$

for all $i \neq j$.

which may be proved in a similar manner to $D1$. □

Remark 4.11. Theorem 4.8 and Proposition 4.9 tell us that the Determinant Function D has all the geometric properties that the Area Function had, and in fact they tell us something more important. The Properties $D1, D2, D3$ and Alternating completely characterise the Determinant function : given another function that satisfies the properties $D1, D2, D3$ and is alternating then this agrees with the definition of the determinant,

$$Det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) \; a_{1(1)\sigma} \; a_{2(2)\sigma} \ldots a_{n(n)\sigma}.$$

(We shall prove this in the next lecture.)

Theorem 4.12. *Let D be the determinant function of a matrix $A \in M_n(\mathbb{R})$ let A have rows $\{r_1, r_2, \ldots, r_n\}$ then the following statements hold :*

(a) If A' is the matrix obtained from A by interchanging two rows (r_i and r_j are interchanged where $i \neq j$), then

$$|A'| = -|A|.$$

This amounts to performing an elementary row operation of type I;

(b) If A' is the matrix obtained from A by replacing r_i by λr_i for any i with $\lambda \in \mathbb{R}$, then

$$|A'| = \lambda |A|.$$

This amounts to performing an elementary row operation of type II;

(c) If A' is the matrix obtained from A by replacing r_i by $r_i + \lambda r_j$ for any row $r_j, j \neq i$ and $\lambda \in \mathbb{R}$, then

$$|A'| = |A|.$$

This amounts to performing an elementary row operation of type III;

(d) $D(r_1, r_2, \ldots, r_n) = 0$ if any two rows are equal;

(e) $D(r_1, r_2, \ldots, r_n) = 0$ if $\{r_1, r_2, \ldots, r_n\}$ is linearly dependent.

Proof. Parts (a),(b) and (c) follow immediately from Theorem 4.8, and Propositions 4.9 and 4.10. Case (d) holds since if we subtract the two identical rows then one element of each product of the determinant is zero. Case (e) holds by applying part (c) then (d).

□

4.2 Determinants using the Method of row reduction to echelon form : Consequences of theorem 4.12

Let $A \in M_n(\mathbb{R})$. Theorem 4.12 gives us a method for evaluating a determinant by applying Elementary Row Operations of type I, II and III to a matrix A to obtain a matrix A' which is in **row reduced echelon form** (and so is unique - see Theorem 13.5 from MT). We shall see more precisely below.

[*Reminder- row reduced echelon form* satisfies the following conditions : the zero rows are all after the non-zero rows; non-zero rows have a leading 1; when a column contains a leading 1 the column is a unit vector; for any two consecutive non-zero rows when the first row has a leading 1, the next next row's leading entry is strictly to its right. Last term you proved the following proposition :

Proposition 4.13. *Let $A \in M_n(\mathbb{R})$. TFAE:*

1. *A is invertible;*
2. *$Ax = 0$ has $x = 0$ as a unique solution;*
3. *the row reduced echelon form of A is I_n.*

(from last term, combine prop 15.2 and cor 15.5)].

Once we have A in row reduced echelon form, we will know its row rank so if it has a zero row at the bottom, we can immediately use theorem 4.12 to say that $|A| = 0$. Similarly if we have row reduced A to I_n then Theorem 4.12 shows us how to keep track of the value of the determinant. (See the example below).

[**Some shortcuts** : if A is a diagonal matrix, composed of a leading diagonal of λ_i, then combining D3 and Theorem 4.12 gives us

$$|A| = \lambda_1\lambda_2 \ldots \lambda_n |I_n| = \prod_{i=1}^{i=n} \lambda_i$$

and this in fact holds for a matrix in Upper Triangular or Lower Triangular form (you may say, Triangular form to mean either).

(**Exercise** : Prove that this is true for Triangular matrices).]

Example 4.14. Use the method of row reduction to echelon form to evaluate the determinants :

1. $|A| = \begin{vmatrix} -2 & 5 & 2 \\ 1 & 0 & 3 \\ 0 & -3 & 5 \end{vmatrix}$;

2. $|B| = \begin{vmatrix} 1 & \alpha & \alpha^2 \\ 1 & \beta & \beta^2 \\ 1 & \gamma & \gamma^2 \end{vmatrix}$

3. $C = \begin{vmatrix} 4 & 5 & 1 & 3 \\ 0 & 0 & -4 & 9 \\ 0 & 0 & 6 & 4 \\ 0 & 0 & 0 & 5 \end{vmatrix}$

Solution

1. We state the row operations and keep a record of the scalars we have multiplied by (as these will affect the value of the determinant). Firstly inter-change $r_1 \leftrightarrow r_2$ which will give a negative in the determinant, then $r_2 \to r_2 + 2r_1$ with no change to $|A|$.

   This gives us $|A| = (-1)\begin{vmatrix} 1 & 0 & 3 \\ 0 & 5 & 8 \\ 0 & -3 & 5 \end{vmatrix}$. Next, we have $5r_2$ but want a leading 1 in the a_{22} position, thus

   $|A| = (-1)(5)\begin{vmatrix} 1 & 0 & 3 \\ 0 & 1 & \frac{8}{5} \\ 0 & -3 & 5 \end{vmatrix}$.

   Next $r_3 \to r_3 + 3r_2$ with no change to the determinant, and we have

   $$|A| = (-1)(5)\begin{vmatrix} 1 & 0 & 3 \\ 0 & 1 & \frac{8}{5} \\ 0 & 0 & \frac{49}{5} \end{vmatrix}.$$

   (We could stop here and use the shortcut mentioned above as the matrix is in U Δ form and $|A| = (-1)(5)(\frac{49}{5}) = -49$). Continuing to row reduced echelon form we arrive at $|A| = (-1)(5)(\frac{49}{5})\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}. = (-1)(5)(\frac{49}{5})|I_n| = -49.$

   (2) and (3) can be completed as exercises.

Before we leave Determinants, it remains to prove that there is only one determinant function given by 4.7.

Lecture 5

We know that the determinant of an $n \times n$ matrix given in Definition 4.7 satisfies the properties $D1, D2, D3$ and is alternating. So we know that a determinant function exists. We now show that there is only one determinant function; this is given by Definition 4.7.

5 Uniqueness of Determinants

We adopt the approach given by P.G. Kumpel and J.A. Thorpe in Linear Algebra.

Theorem 5.1. *The determinant function $D : A \in M_n(\mathbb{R}) \to \mathbb{R}$ which is alternating and satisfies the properties $D1, D2, D3$ is unique.*

Proof. We show that given another function Δ that is alternating and satisfies the properties $D1, D2, D3$, then $D = \Delta$. Here D is the determinant function given in definition 4.7. Let $r_1, r_2, \ldots, r_n$ be the rows of A.

As D and Δ satisfy the properties $D1 - D3$ given in Theorem 4.8, we will also have by Theorem 4.12

(i) $\Delta(r_1, \ldots r_n)$ changes sign if rows r_i and r_j are interchanged;

(ii) $\Delta(r_1, \ldots r_n) = 0$ if $r_i = r_j$ where $i \neq j$.

Now $A = (r_1, \ldots r_n) = (a_{ij})\ I_n$ and so if e_j denotes the usual row vector of I_n then,

$$\begin{aligned}
\Delta(A) &= \Delta(r_1, r_2, \ldots, r_n) \\
&= \Delta(\sum_{j=1}^{n} a_{1j} e_j, r_2, \ldots, r_n) \\
&= \sum_{j=1}^{n} \Delta(a_{1j} e_j, r_2, \ldots, r_n) \quad \text{by D1} \\
&= \sum_{j=1}^{n} a_{1j} \Delta(e_j, r_2, \ldots, r_n) \quad \text{by D2.}
\end{aligned}$$

Now call $j = j_1$. Repeating the above process for r_2 gives us

$$\begin{aligned}\Delta(A) &= \sum_{j_1=1}^{n} a_{1j_1}\Delta(e_{j_1}, \sum_{j=1}^{n} a_{2j}e_j, \ldots, r_n), \\ &= \sum_{j_1,j_2=1}^{n} a_{1j_1} \times a_{2j_2}\Delta(e_{j_1}, e_{j_2}, \ldots, r_n),\end{aligned}$$

replacing j by j_2 and applying D1 and D2. Continuing we arrive at

$$\Delta(A) = \sum_{j_1,\ldots,j_n=1}^{n} a_{1j_1} \times a_{2j_2} \times \ldots a_{nj_n}\Delta(e_{j_1}, e_{j_2}, \ldots, e_{j_n}). \tag{5.1}$$

Now by (ii) above, $\Delta(e_{j_1}, e_{j_2}, \ldots, e_{j_n}) \neq 0$ iff $j_1, j_2, \ldots, j_n$ are distinct. Thus the non-zero terms are those for which $(j_1, \ldots, j_n) = ((1)\sigma, (2)\sigma, \ldots, (n)\sigma)$ for some $\sigma \in S_n$. Thus

$$\Delta(A) = \sum_{\sigma \in S_n} a_{1,(1)\sigma} \times a_{2,(2)\sigma} \times \ldots a_{n,(n)\sigma}\Delta(e_{(1)\sigma}, e_{(2)\sigma}, \ldots, e_{(n)\sigma}).$$

Now each permutation is a product of transpositions and so $\sigma = \tau_1 \circ \tau_2, \ldots \tau_k$, say. Then $sign(\sigma) = sign(\sigma^{-1})$. Each transposition swaps two rows in Δ until we have $\Delta(e_1, e_2, \ldots, e_n)$, hence we have the $sign(\sigma^{-1})$ as a factor in Δ. Relabelling σ^{-1} by σ we have the required result.

□

6 Final Properties of Determinants

Recall that the determinant is a sum of products of elements of A each product containing exactly one element from each row and each column. Thus the determinant is unaltered if we compute it for A^t in this manner. Let's prove this formally.

Proposition 6.1. *Let $A \in M_n(\mathbb{R})$, then $Det(A) = Det(A^t)$.*

Proof. Recall that

$$Det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1,(1)\sigma} \times \ldots \times a_{n,(n)\sigma}.$$

If $A^t = (b_{ij})$ where $b_{ij} = a_{ji}$, then

$$\begin{aligned} Det(A^t) &= \sum_{\sigma \in S_n} \text{sign}(\sigma) b_{1,(1)\sigma} \times \ldots \times b_{n,(n)\sigma}, \\ &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{(1)\sigma,1} \times \ldots \times a_{(n)\sigma,n.} \end{aligned}$$

Notice that $\sigma \in S_n$ is a bijection so whenever $(i)\sigma = j$ then $i = (j)(\sigma^{-1})$; also σ and σ^{-1} have the same parity. Thus we may re-write $Det(A^t)$ as

$$\begin{aligned} Det(A^t) &= \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{(1)\sigma,1} \times \ldots \times a_{(n)\sigma,n,} \\ &= \sum_{\sigma^{-1} \in S_n} \text{sign}(\sigma) a_{1,(1)\sigma^{-1}} \times \ldots \times a_{n,(n)\sigma^{-1}} \\ &= Det(A) \end{aligned}$$

□

Thus we have the following:

Proposition 6.2. *Let $A \in M_n(\mathbb{R})$ with rows given by r_i and columns by c_j. Then*

$$D(r_1, r_2, \ldots, r_n) = D(c_1, c_2, \ldots, c_n).$$

Proposition 6.3. *If A and $B \in M_n(\mathbb{R})$, then*

$$Det(AB) = Det(A)Det(B).$$

Proof. Let $C = AB$ and let the k-th column of C be denoted by C_k. Now the entries in the k-th column of C are given by the sums

$$\begin{pmatrix} \sum_{j=1}^n a_{1j} b_{jk} \\ \sum_{j=1}^n a_{2j} b_{jk} \\ \vdots \\ \sum_{j=1}^n a_{nj} b_{jk} \end{pmatrix}.$$

Thus writing B_k for the k-th column of B we have

$$C_k = \sum_{j=1}^{n} a_{kj} B_k.$$

But then

$$\begin{aligned} det(C) &= det(C_1, C_2, \ldots, C_n) \\ &= det(\sum_{j=1}^{n} a_{1j} B_j, \sum_{j=1}^{n} a_{2j} B_j, \ldots, \sum_{j=1}^{n} a_{nj} B_j). \end{aligned}$$

Each column is a sum of n columns so expanding by columns and applying 4.12 part (ii) successively to each column we may expand $det(C)$ as a sum of n^n determinants. This gives us

$$det(C) = \sum det(a_{1j_1} B_{j_1}, a_{2j_2} B_{j_2}, \ldots, a_{nj_n} B_{j_n}) \tag{6.1}$$

where the summation is over which is all possible permutations of $1, 2, \ldots, n$. Writing σ for such a permutation, equation 6.1 becomes

$$det(C) = \sum det(a_{1(1)\sigma} B_{(1)\sigma}, a_{2(2)\sigma} B_{(2)\sigma}, \ldots, a_{n(n)\sigma} B_{(n)\sigma}). \tag{6.2}$$

Pulling out the factors $a_{j(j)\sigma}$ from the j-th column will result in

$$\begin{aligned} det(AB) &= det(C) \\ &= \sum_{\sigma \in S_n} a_{1(1)\sigma} a_{2(2)\sigma} \ldots a_{n(n)\sigma} \; det(B_{(1)\sigma}, B_{(2)\sigma}, \ldots, B_{(2)\sigma}) \\ &= det(A) det(B). \end{aligned}$$

□

Proposition 6.4. *If $A \in M_n(\mathbb{R})$ and A is non-singular, then $Det(A^{-1}) = \frac{1}{Det(A)}$.*

Proof. This follows immediately since

$$1 = Det(I_n) = Det(AA^{-1}) = Det(A) det(A^{-1}).$$

□

Exercise 6.5. Read about Crammer's Rule (very popular with Engineers).

7 Applications of Determinants : Cross Products and Triple Scalar Products

You have met these notions before in Geometry I, but here we see that they are in fact determinants.

Let $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ be non parallel vectors relative to an initial point P. Let $\mathbf{n}$ be a vector normal to the plane containing $\mathbf{a}$ and $\mathbf{b}$. We say $(\mathbf{a},\ \mathbf{b},\ \mathbf{n})$ form a *right-handed set* (in this order). Let θ be smaller angle between $\mathbf{a}$ and $\mathbf{b}$.

Definition 7.1. We define $\mathbf{a}\times\ \mathbf{b}$, the *vector product* or *cross product* between $\mathbf{a}$ and $\mathbf{b}$ to be

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

where $|\mathbf{a}|$ denotes the (Euclidean) norm of $\mathbf{a}$.

Therefore, $\mathbf{a} \times \mathbf{b}$ is a vector $\perp^r$ to both $\mathbf{a}$ and $\mathbf{b}$. **Note** $\mathbf{a} \perp^r \mathbf{b} \Rightarrow \mathbf{a} \times \mathbf{b} = 0$, so for example : $\mathbf{i}\times\mathbf{i} = \ \mathbf{j}\times\mathbf{j} = \ \mathbf{k}\times\mathbf{k} = \ 0$ and,

$$\begin{aligned} \mathbf{i}\times\mathbf{j} &= \mathbf{k}, \mathbf{j}\times\mathbf{k} &= \mathbf{i}, \mathbf{k}\times\mathbf{i} &= \mathbf{j} \\ \mathbf{j}\times\mathbf{i} &= -\mathbf{k}, \mathbf{k}\times\mathbf{j} &= -\mathbf{i}, \mathbf{i}\times\mathbf{k} &= -\mathbf{j} \end{aligned}$$

.

Proposition 7.2. *The vector $\boldsymbol{a}\times\boldsymbol{b}$ is defined to be*

$$\boldsymbol{a}\times\ \boldsymbol{b} \ = \ |\boldsymbol{a}||\boldsymbol{b}|\sin\theta\boldsymbol{n} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

The area of the parallelogram given by the non-parallel sides $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by the norm of this vector product, $|\boldsymbol{a}\times\boldsymbol{b}| = |\boldsymbol{a}||\boldsymbol{b}|\sin\theta$.

Proof. Now

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \quad \text{using the distribute law} \\
&= a_1\mathbf{i} \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) + a_2\mathbf{j} \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) + a_3\mathbf{k} \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \\
&= a_1b_2\mathbf{k} - a_1b_3\mathbf{j} + -a_2b_1\mathbf{k} + a_2b_3\mathbf{i} + a_3b_1\mathbf{j} - a_3b_2\mathbf{i} \\
&= (a_2b_3 - a_3b_2)\mathbf{i} - (-a_3b_1 + a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k}
\end{aligned}$$

which is the determinant, as required.

Clearly $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta$, and a simple geometric argument will give you this as the area of the given parallelogram. □

Immediately we can see that $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$.

7.1 Triple Scalar Product

Consider three non co-planar vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, where for example, $\mathbf{b}, \mathbf{c}$ lie in the $x - y$ plane and $\mathbf{a}$ lies in the $y - z$ plane. Then we have a normal to $\mathbf{b} \times \mathbf{c}$, which we denote by $\mathbf{n}$.

Let θ be smaller angle between $\mathbf{b}$ and $\mathbf{c}$ and ϕ the smaller angle between $\mathbf{a}$ and $\mathbf{n}$.

Definition 7.3. The *triple scalar product* of $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ is defined to be:

$$\begin{aligned}
\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= \mathbf{a} \cdot |\mathbf{b}||\mathbf{c}| \sin\theta\mathbf{n} \\
&= |\mathbf{a}||\mathbf{b}||\mathbf{c}| \sin\theta \cos\phi
\end{aligned}$$

since $\mathbf{a} \cdot\ \mathbf{n} = |\mathbf{a}||\mathbf{n}| \cos\phi = |\mathbf{a}| \cos\phi$.

Thus we have

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = |\mathbf{a}| \cos\phi |\mathbf{b}||\mathbf{c}| \sin\theta = hA$$

which is the volume of a parallelepiped, where A is the area of parallelogram with edges $(\mathbf{b}, \mathbf{c})$ and h is the perpendicular distance of $\mathbf{a}$ to the plane containing $\mathbf{b}, c$.

A simple calculation shows that

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

Here the parenthesis is not needed since $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$ could only be interpreted as $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ as $\underbrace{(\mathbf{a} \cdot \mathbf{b})}_{\text{scalar}} \times \underbrace{\mathbf{c}}_{\text{not defined for a scalar and a vector}}$

In our next lecture we finish this discussion on determinants and more importantly begin the study of *eigenvalues* and *eigenvectors*.

The next two small sections are for you to read.

8 Determinants and Linear Transformation

The above development of determinants is for matrices. Naturally we should know what happens for linear transformations. It is as expected.

Definition 8.1. Let T be a linear transformation on a vector space V with a basis $\{v_i\}$. The determinant function of T, denoted by $Det(T)$ is defined to be equal to the $Det(A)$ where A is the matrix with respect to the basis $\{v_i\}$.

To ensure that the definition is well defined there is one thing to check :

Proposition 8.2. *If A and A' are matrices representing a linear transformation T w.r.t. two bases, then $Det(A) = Det(A')$.*

Proof. This follows from combining Theorem 1.7 which gives us $A' = PAP^{-1}$ and then apply Proposition 6.3. □

Remark 8.3.

1. Two matrices A and B are called *similar* if and only if they represent the same linear Transformation but w.r.t. different bases. i.e. they are matrices for which ∃ invertible matrix P such that

$$B = PAP^{-1}.$$

2. Similar matrices have the same trace.

3. Similar matrices have same the determinant (by Prop 8.2).

9 Polynomials of linear Transformations and Matrices

This section is for you to read and will not be discussed in the lecture.

Let V be a finite dimensional vector space over a field F. We denote by $L(V, V)$ the set of all linear transformations from V to V. If $T, S \in L(V, V)$ then $T + S$ will also be a linear transformation from V to V, as will λT for $\lambda \in F$. This tells us :

Theorem 9.1. *$L(V, V)$ is a vector space.*

Further it can be shown that $L(V, V)$ is isomorphic to $M_n(F)$. (To see this we view $M_n(F)$ as a vector space of $n \times n$-tuples. See Curtis, Theorem 13.3 for proof).

Then $M_{n\times n}(F)$ has dimension n^2 and if $\{A_1, ..., A_{n^2}\}$ is a basis for $M_{n\times n}(F)$ over F, then the linear transformations $T_1, ..., T_{n^2}$ represented by $A_1, ..., A_{n^2}$ w.r.t. basis $\{v_i\}$, form a basis for $L(V, V)$ over F.

Let $F_n[x]$ denote the set of all polynomials in x of degree $\leq n$ with coefficients in the Field F. (Cf. $\mathbb{R}_2[x]$).

If $L(V, V)$ has dim n^2, the set of powers of $T, \{Id_V, T, T^2, ..., T^{n^2}\}$ contains $n^2 + 1$ elements and so is linearly dependent. (In the case that some of the T^i are equal, we may find the smallest integer $r < n$ such that $\{Id_V, T, T^2, ..., T^{r^2}\}$ is linearly dependent).

Therefore $\exists\ a_i \in F$, with not all $a_i = 0$, s.t.

$$a_0 Id_V + a_1 T + \ ... \ + a_{n^2} T^{n^2} = 0,$$

i.e. $\exists$ non-zero polynomial $f \in F_m[T]$, $(m = n^2)$ where

$$f(T) = a_0 Id_V + a_1 T + \ ... \ + a_m T^m \qquad \text{such that} \quad f(T) = 0. \tag{9.1}$$

Equation 9.1 is a *polynomial equation* in T. Similarly we may define a polynomial matrix equation for the matrix A (which represents T).

Definition 9.2. If $A \in M_n(F)$ then we can define a polynomial $f \in F_n[A]$ in A by

$$f(A) = a_n A^n + \ ... \ + a_1 A + a_0 I_n,$$

where $a_i \in F$ and not all a_i are zero.

We say A is a *root* or *zero* of the polynomial f if $f(A) = 0$.

Example 9.3. $A = \begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix}$ and let $f(x) = 2x^2 - 3x + 7$. Compute $f(A)$. Is A a root of this polynomial?

Solution

$$f(A) = 2\begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix} - 3\begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix} + 7\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= 2\begin{pmatrix} 1 & 10 \\ 0 & 16 \end{pmatrix} - \begin{pmatrix} 3 & 6 \\ 0 & 12 \end{pmatrix} + \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 6 & 14 \\ 0 & 27 \end{pmatrix} \neq 0$$

So A is not a root of f.

The aim of the rest of this course is to lay the foundations for the concept of a *minimal polynomial* and determine conditions for when A is a root of a special polynomial, the *characteristic polynomial.* We shall use eigenvalues and eigenvectors to do this.

We state some properties of polynomials in a matrix :

Theorem 9.4. *Let* $f,\ g \in F_n(x)$ *and* $A \in M_n(F)$. *then*

(i) $(f + g)(A) = f(A) + g(A)$

(ii) $(fg)(A) = f(A)g(A)$

(iii) and $(\lambda f)(A) = \lambda f(A)$ *for* $\lambda \in F$.

Property [(ii)] tells us we have product in $F_n[x]$, as well as an additive operation (given by [i]), which you will find out means $F_n[x]$ is a *Ring of Polynomials.* Further since $f(x)g(x) = g(x)f(x)$ (defined pointwise), for any polynomial in a matrix A we have $f(A)g(A) = g(A)f(A)$. This tells us that polynomials in matrix A compute, meaning that $F_n[x]$ is a *commutative ring*).

What happens in the case of a linear transformation $T \in L(V, V)$? If $f(T) = a_nT^n + \ ...\ + a_1T + a_0Id_v$ then we say T is a *root* or *zero* of f if $f(T) = 0$. We note that Theorem 9.4 also holds for linear transformations. Furthermore, if A is a matrix representation of T, then $f(A)$ is a matrix representation of $f(T)$ and not surprisingly $f(T) = 0 \Longleftrightarrow f(A) = 0$.

10 Eigenvalues and Eigenvectors

Let V be a vector space over a field $\mathbb{R}$ and $T : V \to V$. We write $L(V, V)$ for the vector space of all linear transformations from V to V.

In this section we look at the theory of eigenvalues and eigenvectors. This will enable us to investigate when a linear transformation $T \in L(V, V)$ is **diagonalisable,** i.e. when can find the matrix representation of T which is a diagonal matrix.

Firstly, we give some motivation.

10.1 Motivation

Let $T \in L(V, V)$ where V is a finite dimensional vector space. Suppose A is the matrix that represents T w.r.t. a given basis $\{v_1, \dots, v_n\}$ of V.

Firstly we will look at what it means for T to be represented by a diagonal matrix w.r.t. this basis. Suppose A is diagonal, and takes the form $\begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & & \lambda_n \end{pmatrix}$.

What happens to the basis vectors?

By definition of the matrix associated with T we have

$$\begin{array}{ccccccc} T(v_1) & = & \lambda_1 v_1 & +0v_2+ & \dots & +0v_n \\ T(v_2) & = & 0v_1+ & \lambda_2 v_2+ & \dots & +0v_n \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ T(v_n) & = & 0v_1 & +0v_2+ & \dots & +\lambda_n v_n \end{array}$$

ie.

$$T(v_i) = \lambda_i v_i, \qquad \forall i. \tag{10.1}$$

Notice that as $\{v_i\}$ is a basis, then $v_i \neq 0$. The scalars λ_i which satisfy equation 10.1 are called *eigenvalues*; associated vectors are called *eigenvectors*. This leads to the following definition :

Definition 10.1. Let $T : V \to V$ be a linear transformation, then $\lambda \in \mathbb{R}$ (or $\in F$) is called an *eigenvalue of* T if $\exists$ **non-zero vector** $v \in V$ for which $T(v) = \lambda v$.

Every v satisfying this relation is called on eigenvector of T for the eigenvalue λ.

Note that each scalar multiple av of v is an eigenvector associated with the eigenvalue λ since $T(av) = aT(v) = a\lambda v = \lambda(av)$.

The set of all such vectors $\{v : T(v) = \lambda v \quad \text{for some} \quad \lambda\}$ is called the *eigenspace* of λ and is a subspace of V since it contains $0 = 0v$.

Example 10.2. 1. Let $Id_v : V \to V$ be the identity mapping. Then for every $v \in V, Id_v(v) = v = 1v$ hence 1 is an eigenvalue of Id_v and every vector in V is an eigenvector belonging to 1.

2. Let $T \in L(V, V)$ with kernel, ker(T). Every $v \in ker(T)$ satisfies $T(v) = 0$ and so if $v \neq, 0$ it is an eigenvector for the eigenvalue 0.

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which rotates a vector b in $\mathbb{R}^2$ anti-clockwise by $\pi/2$. Under T, no non-zero vector v satisfies $T(v) = \lambda v$. Hence T has no eigenvalues and so no eigenvectors. [Moreover, we will not be able to diagonalise such a T.]

4. Let D be the differential operator on the vector space V of differentiable functions. $D(e^{2t}) = 2e^{2t}$ and so 2 is an eigenvalue of D and e^{2t} is its eigenvector.

Theorem 10.3. *Let T be a linear transformation on a vector space V, over field F. Let $\lambda \in F$. Then λ is an eigenvalue of T if and only if $|T - \lambda Id_v| = 0$ where Id_v is the identity transformation.*

Proof. $\Rightarrow$ Suppose λ is an eigenvalue of T. Then by the definition $\exists\ v \neq 0$ s.t. $T(v) = \lambda v$. Then $(T - \lambda Id_v)v = 0$. Since $v \neq 0$, $T - \lambda Id_v$ must be singular and so its determinant is 0.
$\Leftarrow$ suppose $|T - \lambda Id_v| = 0$. Then $T - \lambda Id_v$ is singular and so it is not an isomorphism, in particular it is not one-to-one. $\therefore$ $\exists$ vectors $v_1,\ v_2, \quad v_1 \neq v_2$ and $(T - \lambda Id_v)(v_1) = (T - \lambda Id_v)(v_2)$. Letting $v = v_1 - v_2 \neq 0$. We have $(T - \lambda Id_v)(v) = 0$ and λ is an eigenvalue of T. □

Remark 10.4. The above theorem is important tells us how to find eigenvalues.

Remark 10.5 (Invariant subspaces). A non-zero vector $v \in V$ is an eigenvector of T iff the 1-dimensional subspace generated by v, $\{\mu v : \mu \in F\}$, is invariant under T. The search for invariant subspaces is important in its own right – and is important in the study of deeper properties of a single linear transformation.

The following definition will be familiar to you :

Definition 10.6. If $A \in M_n(F)$ then λ is an *eigenvalue of* A means that for some non-zero vector $v \in F^n$ satisfying $Av = \lambda v$.

The development of the theory runs in parallel for matrices as for linear transformations. We should check that the notions of eigenvalues in these two settings are compatible.

Proposition 10.7. *Let $T : (V, B) \to (V, B)$ where V is a vector space over a field F and B its basis. Let A be a matrix representation of T w.r.t. B. Then $v \in V$ is a non-zero eigenvector of T for eigenvalue λ if and only if the co-ordinate vector v_B of v w.r.t. B is an eigenvector of A for the same eigenvalue.*

Proof. We need to show that

$$Av_B = \lambda v_B \quad \Leftrightarrow \quad T(v) = \lambda v.$$

This follows immediately from the definition of A. Namely that it acts on a co-ordinate vector in the way that the linear transformation acts on the vector. (Proposition 12.3 in MT notes). □

Remark 10.8.

1. What does this tell us? That the eigenvalues of a matrix representing a linear transformation w.r.t. a certain basis **do not** depend on the basis;
2. We state and prove many of our theorems here in terms of linear transformations but also these work in terms of matrices- replace T by A and Id_v by I_n.

Example 10.9. Find the eigenvalues and eigenvectors associated with the matrix

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

Solution
Required to solve for λ and $x \neq 0$

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

i.e. $(A - \lambda I)(x) = 0$ solve the homogenous system. Recall that we have a non-zero solution for x iff $|A - \lambda I| = 0$ ($\Leftrightarrow$ $(A - \lambda I)$ is singular). (Compare Theorem 10.3). So solving

$$\begin{vmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{vmatrix} = 0,$$

$$(1-\lambda)(2-\lambda) - 6 = 0$$

$$\Rightarrow \lambda^2 - 3\lambda - 4 = 0 \tag{10.2}$$

$(\lambda - 4)(\lambda + 1) = 0$, and so $\lambda = 4$ and $\lambda = -1$. If $\lambda = 4$, then

$$\begin{pmatrix} x + 2y \\ 3x + 2y \end{pmatrix} = \begin{pmatrix} 4x \\ 4y \end{pmatrix}$$

and solving gives us $x, y \neq 0$

$$2y = 3x.$$

Thus $\begin{pmatrix} x \\ \frac{3}{2}x \end{pmatrix}$ or $\begin{pmatrix} 2x \\ 3x \end{pmatrix}$ $x \neq 0$

choose sensible value for x eg. $x = 1$, and the eigenvector is $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

If $\lambda = -1$, we have $\begin{pmatrix} x + 2y \\ 3x + 2y \end{pmatrix} = \begin{pmatrix} -x \\ -y \end{pmatrix}$

and so $x = -y$ which gives us $\begin{pmatrix} x \\ -x \end{pmatrix}$ and the eigenvector is $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Notice these λ_i are distinct and eigenvectors are linearly independent.

Exercise 10.10. In example 10.9, does A satisfy equation 10.2? This is one form of a Theorem called the *Cayley-Hamilton Theorem* which you will prove in Part A.

Exercise 10.11. In example 10.9, form the matrix whose columns are the eigenvectors. Call this matrix P. Find its inverse. Form a diagonal matrix D which has thee eigenvalues on the diagonal. By computation show that $P^{-1}AP = D$. This is what we call a Diagonal representation of A.

11 The Characteristic Polynomial

Let $A \in M_n(F)$ where F is a field.

Definition 11.1. Let $A = (a_{ij})$, then the determinant, $|A - \lambda I_n|$ is called the *Characteristic Polynomial of* A. We denote this by $\chi_A(\lambda)$.

Notice that by Theorem 9.3, the roots of the characteristic equation of a matrix A are its eigenvalues.

Proposition 11.2. *Let $A \in M_n(F)$. There are at most n eigenvalues of A.*

Proof. An eigenvalue of a root of the Characteristic polynomial. This polynomial is of degree n and by the Fundamental Theorem of Algebra, it has at most n roots. □

We look at some illustrative examples before we begin to develop the theory.

Example 11.3.

1. Let $A = \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix}$. Find its characteristic polynomial, all its eigenvalues and eigenvectors, (i) if $F = \mathbb{R}$, (ii) if $F = \mathbb{C}$.
2. Find the form of the characteristic polynomial when $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.
3. Let $B = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix}$.

   Find its characteristic polynomial, its eigenvalues and eigenvectors over $\mathbb{R}$.

Solution

1.
$$\chi_A(\lambda) = \begin{vmatrix} 1-\lambda & -1 \\ 2 & -1-\lambda \end{vmatrix}$$
and solving $\chi_A(\lambda) = 0$ gives us $\lambda^2 + 1 = 0$. (i) Thus there is no real solution for λ and so A has no eigenvalues or eigenvectors over $\mathbb{R}$.

   (ii) As before, we would have $\chi_A(\lambda) = 0$ which gives us $\lambda^2 + 1 = 0$, and $\lambda = \pm\mathrm{i}$. If $\lambda = \mathrm{i}$, we have $(A - \mathrm{i}I_2)x = 0$

which gives us $x - y = \mathrm{i}x$ and $2x - y = \mathrm{i}y$. Subtracting these gives us $x(1 + \mathrm{i}) = iy$ which gives us our only non-zero solution. Thus the eigenvector is $\begin{pmatrix} \frac{\mathrm{i}y}{1+\mathrm{i}} \\ y \end{pmatrix}$, or $\begin{pmatrix} \mathrm{i} \\ 1+\mathrm{i} \end{pmatrix}$.

For $\lambda = -\mathrm{i}$ we have the non-zero eigenvector is $\begin{pmatrix} 1 \\ 1-\mathrm{i} \end{pmatrix}$.

2. A simple computation shows that $\chi_A(\lambda) = \lambda^2 - (a+d)\lambda + (ad - bc)$.

3. The Characteristic polynomial is χ_B which is

$$\chi_B = |B - \lambda I_3| = \begin{vmatrix} 3-\lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 4 & 2 & 3-\lambda \end{vmatrix}.$$

Expanding the determinant by row 1 gives us

$(3-\lambda)(\lambda^2 - 3\lambda - 4) - 2(6 - 2\lambda - 8) + 4(4 + 4\lambda)$ and so

$$\chi_B = \lambda^3 + 6\lambda^2 + 15\lambda + 8.$$

By inspection its roots are $\lambda = -1$ of multiplicity 2, and $\lambda = 8$.

The characteristic polynomial is

$$\chi_B(\lambda) = (\lambda+1)^2(\lambda - 8).$$

If $\lambda = -1$, solving $(B + I_3)x = 0$ will give us its eigenvector(s). The system of equations reduces to :

$$\begin{array}{ccccccc} 4x & + & 2y & + & 4z = & 0 \\ 2x & + & y & + & 2x = & 0 \\ 4x & + & 2y & + & 4z = & 0. \end{array}$$

Equivalently $2x+y+2z = 0$. Thus our eigenvector has the form $\begin{pmatrix} x \\ -2x-2z \\ z \end{pmatrix}$.

We have two parameters and so can find two linearly independent eigenvectors for $\lambda = -1$ (it may not always be so): $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ -2 \\ -1 \end{pmatrix}$. Here the eigenspace of $\lambda = -1$ has dimension 2. If $\lambda = 8$, the eigenvector is of the form $\begin{pmatrix} x \\ 1/2x \\ x \end{pmatrix}$ or $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$. So this is an example where we have 3 eigenvalues, 2 of which are repeated, yet we can 3 linearly independent eigenvectors.

The above example naturally leads us to ask more generally when does this hold.

Theorem 11.4. *Non-zero eigenvectors belonging to distinct eigenvalues are linearly independent.*

Proof. We use induction on n. Consider the predicate $P(n)$ as follows. $P(n)$: the set $\{v_1, ..., v_n\}$ of non-zero eigenvectors belonging to distinct eigenvalues λ_i, is linearly independent. For $n = 1$, $v_1 \neq 0$ and there is nothing to prove.

Assume $P(k)$ is true, that is, $\{v_1, \ldots, v_k\}$ belonging to distinct eigenvalues λ_i, is a set of non-zero linearly independent eigenvectors.

Suppose that

$$a_1 v_1 + \cdots + a_k v_k + a_{k+1} v_{k+1} = 0$$

, (10.1)

where $a_i \in \mathbb{R}$ for $1 \leq i \leq k+1$.

We require that the coefficients a_i are 0. Applying T gives us

$$a_1 \lambda_1 v_1 + \cdots + a_k \lambda_k v_k + a_{k+1} \lambda_{k+1} v_{k+1} = 0$$

. (10.2)

Subtract λ_{k+1} times equation (10.1) from equation (10.2) we get

$$a_1(\lambda_1 - \lambda_{k+1}) v_1 + \cdots + a_k(\lambda_k - \lambda_{k+1}) v_k = 0.$$

By the inductive hypothesis, $v_1, \ldots, v_k$ are linearly independent and so each coefficient must be zero. Then, since $\lambda_i - \lambda_{k+1} \neq 0$ for $1 \leq i \leq k$, we have $a_i = 0$ for $1 \leq i \leq k$. Now look back at equation (10.1) to see that, since $v_{k+1} \neq 0$, also $a_{k+1} = 0$. Thus the vectors $v_1, \ldots, v_k, v_{k+1}$ are linearly independent.

□

Remark 11.5. If a matrix has a repeated eigenvalues, it may still be possible to find independent eigenvectors.

12 Diagonalisation of Linear Transformations

In this section we look at Diagonalisability of a linear transformation. In the next section we look at these notions for matrices.

Let $T \in L(V, V)$ where V is a finite dimensional vector space. Suppose A is the matrix that represents T w.r.t. a given basis $\{v_1, \dots, v_n\}$ of V.

In section 9.1 we showed that if there was a basis $\{v_1, \dots, v_n\}$ with respect to which T is a diagonal matrix, then there exists a basis of $\{v_1, \dots, v_n\}$ of V of *eigenvectors* of T.

The converse also holds. If we have a basis for V of eigenvectors $\{v_1, \dots, v_n\}$ of T so $T(v_i) = \lambda_i v_i$ for $i = 1, \dots, n$, then the matrix of T w.r.t. this basis is a diagonal matrix. Why? The matrix of T is given by $A = (a_{ij})$ where the a_{ij} are given by (assuming a column representation)

$$\begin{array}{ccccccc} T(v_1) & = & a_{11}v_1 & +a_{21}v_2+ & \dots & +a_{n1}v_n \\ T(v_2) & = & a_{12}v_1+ & a_{22}v_2+ & \dots & +a_{n2}v_n \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ T(v_n) & = & a_{1n}v_1 & +a_{2n}v_2+ & \dots & +a_{nn}v_n. \end{array}$$

But

$$T(v_i) = \lambda_i v_i. \tag{12.1}$$

and so it follows that $a_{11} = \lambda_1, a_{i1} = 0$, for $i = 2, \dots, n$, similarly that $a_{ii} = \lambda_i, a_{ji} = 0$ for $j \neq i$. That is, T is a diagonal matrix where the diagonal elements are λ_i.

In fact we have then showed that the representation of T as a diagonal matrix is equivalent to finding a basis of V of eigenvectors of the linear transformation T.

This leads us to the following definition :

Definition 12.1. Let $T \in L(V, V)$.We say that T is *diagonalisable* if there exists a basis of V with respect to which T may be represented by a diagonal matrix.

Combining the above results and this definition we have a theorem :

Theorem 12.2. *Let $T : V \to V$ be a linear mapping where V is a finite dimensional vector space. T is diagonalisable if and only if V has a basis consisting of eigenvectors of T.*

Applying Theorem 10.4 gives us :

Theorem 12.3. *Let $T \in L(V, V)$ where V is an $n-$dimensional vector space over F. If the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct, then T is diagonalisable.*

Proof. Let $v_1, v_2, \ldots, v_n$ be the eigenvectors corresponding to the λ_i. Thus

$$T(v_i) = \lambda_i v_i, \quad \text{for} \quad i = 1, \ldots, n, v_i \neq 0.$$

Since $v_1, v_2, \ldots, v_n$ consist of n vectors, these form a basis for V if and only if they are linearly independent. By Theorem 10.4, as the λ_i are distinct, the set $v_1, v_2, \ldots, v_n$ is linearly independent; it is a basis as its cardinality in n.

□

Clearly the converse of 11.3 is false (example 10.3 part 3). A natural question to ask is what happens when the eigenvalues are repeated? Can we still diagonalise? This in effect means that Theorem 11.3 is sufficient, but it is not necessary (see earlier example). In fact we need to consider a special polynomial, the minimal polynomial, to answer this question. (You may wish to read about this - or wait till MT 2008!)

Clearly if the dimension of the eigenspace equals the multiplicity of each eigenvector, then we can diagonalise.

A related result :

Theorem 12.4. *Let $T : V \to V$ be a linear transformation where V is a vector space over F. Then, $\lambda \in F$ is an eigenvalue of T if and only if $T - \lambda Id_v$ is a singular transformation. The eigenspace of λ is then the kernel of $(T - \lambda Id_v)$.*

Proof. λ is an eigenvalue of T if and only if $\exists$ non-zero vector v such that $T(v) = \lambda v$ or $T(v) - \lambda Id_v(v) = 0$
or

$$(T - \lambda Id_v)(v) = 0, \tag{12.2}$$

But then $T - \lambda Id_v$ is singular. Therefore v is in the eigenspace of λ iff 12.2 holds, in which case $v \in Ker(T - \lambda Id_v)$. □

Proposition 12.5. *Show that $\lambda = 0$ is an eigenvalue of T if and only if T is singular.*

Proof. Well 0 is eigenvalue iff $\exists\ v \neq 0$ st. $T(v) = 0v = 0$ iff T is singular. □

Example Let T be an orthogonal projection of $\mathbb{R}^3$ onto the $x-y$ plane. Geometrically determine the eigenvalues and eigenvectors of T.

Solution $T(x,y,z) = (x,y,0)$ and so each vector x of the $x-y$ plane is mapped to itself, $T : x \mapsto x$. Thus $\lambda = 1$ is the eigenvalue with eigenvector $\begin{pmatrix} s \\ t \\ 0 \end{pmatrix}$ where $s, t \in \mathbb{R}$. For each x on the $z-$axis, $T : x \mapsto 0$ and so $\lambda = 0$. The eigenvector is $\begin{pmatrix} 0 \\ 0 \\ u \end{pmatrix}$ where $u \in \mathbb{R}$.

No other vector is mapped onto a scalar multiple of itself so these are the only e-values and e-vectors of T.

13 Diagonalisation of Matrices

In our definition of a linear transformation being diagonalisable we have a parallel definition for a matrix.

Definition 13.1. Let $A \in M_{n\times n}(\mathbb{R})$ be a matrix. We say that A is *diagonalisable* if it is similar to a diagonal matrix, D, that is, there is an invertible matrix P such that $P^{-1}\ A\ P = D$.

How do we find such a matrix P? Observe that $P^{-1}\ A\ P = D$ may be written as $AP = PD$. Set p_j to be the columns of P, and $D = (a_{ii}) = (\lambda_i)$ the matrix of eigenvalues of A. Compare the j-th columns on both sides of $AP = PD$. We have $Ap_j = \lambda_j\ p_j$. That is, the j-th column of P is an eigenvector of A corresponding to the eigenvalue λ_j.

To be precise, we form an invertible matrix P whose columns are the eigenvectors of A, and a diagonal matrix D of eigenvalues of A such that

$$D = P^{-1}AP.$$

This gives us an important theorem :

Theorem 13.2. *$A \in M_n(F)$ is similar to a diagonal matrix if and only if A has n linearly independent eigenvectors.*

Proof. This follows as per Theorem 11.2. □

In Remark 9.8 we noted that the eigenvalues of a linear matrix do not depend on the basis of the linear transformation. We now prove this by showing that

Lemma 13.3. *Suppose $A \in M_n(F)$ is similar to a matrix $B = P^{-1}AP$, then A and B have the same characteristic polynomial.*

Proof. Let $|A - \lambda I_n|$ be the characteristic polynomial of A and let P be as stated in the lemma. Now $I_n = P^{-1}P$ and $I_n = P^{-1}I_nP$. Thus

$$
\begin{aligned}
&|B - \lambda I_n| \\
&\quad = |P^{-1}AP - \lambda P^{-1}I_nP| \\
&\quad = |P^{-1}AP - P^{-1}\lambda I_nP| \\
&\quad = |P^{-1}(AP - \lambda I_nP)| \\
&\quad = |P^{-1}(A - \lambda I_n)P| \\
&\quad = |P^{-1}||A - \lambda I_n||P| \quad \text{as } |AB| = |A||B| \text{ by Proposition 7.1,} \\
&\quad = |A - \lambda I_n| \quad \text{as } |P^{-1}||P| = 1
\end{aligned}
$$

□

You will have proved the following on Problem sheet 4.

Proposition 13.4. *Suppose λ is an eigenvalue of a non-singular matrix A. Then λ^{-1} is an eigenvalue of A^{-1}. The respective eigenvectors remain unchanged.*

Proposition 13.5. *If A and $B \in M_n(\mathbb{R})$ then AB and BA have the same eigenvalues.*

Proof. **case 1** $\lambda = 0$ Suppose λ is an eigenvalue. By Proposition 12.5 and the fact that the product of two non-singular matrices is non-singular, the following statements are equivalent

1. 0 is an eigenvalue of AB;
2. AB is singular;
3. A or B is singular;
4. BA is singular;

5. 0 is an eigenvalue of BA.

(Why? The first two are equivalent by proposition 12.5. Recall that $(AB)^{-1} = B^{-1}A^{-1}$ if it exists and so AB is singular if either A or B is singular. Similarly $(BA)^{-1} = A^{-1}B^{-1}$ if it exists, and BA is singular if either A or B is. Thus the rest follows).

case 2 $\lambda \neq 0$ Suppose λ is an eigenvalue of AB. Then $\exists\ v \neq 0$ such that $ABv = \lambda v$. Write $w = Bv$. Then as $\lambda \neq 0$ and $v \neq 0$ then

$$Aw = ABv = \lambda v \neq 0$$

then $w \neq 0$. But

$$\begin{aligned} BAw &= BABv \\ &= \lambda Bv \\ &= \lambda w \end{aligned}$$

and so w is an eigenvector of BA. Also λ is an eigenvalue of BA. Conversely, any non-zero eigenvalue of BA is also an eigenvalue of AB, and so AB and BA have the same eigenvalues.

□

We have some useful results for finding the eigenvalues of a triangular or diagonal matrix.

Theorem 13.6. *The eigenvalues of a triangular matrix are its diagonal entries. Hence the eigenvalues of a diagonal matrix are the diagonal elements.*

Proof. Let $A \in M_n(F)$ and suppose A is triangular. W.L.O.G. suppose A is upper triangular, that is $A = (a_{ij})$ where $a_{ij} = 0$ for $j > i$.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots 0 & a_{nn} \end{pmatrix}.$$

Let $\chi_A(\lambda) = |A - \lambda I_n|$. We shall prove by induction on n that if $A \in M_n(F)$ and A is upper triangular, then its eigenvalues are its n diagonal elements.

It is clear that when $n = 2$ we have

$$\begin{vmatrix} a_{11} - \lambda & 0 \\ 0 & a_{22} - \lambda \end{vmatrix} = 0 = (a_{11} - \lambda)(a_{22} - \lambda),$$

and its diagonal elements are the eigenvalues of A. Assume our hypothesis is true for $n-1$ and consider $\chi_A(\lambda) = |A - \lambda I_n|$ where $A \in M_n(F)$. Expand the determinant by row R_1 and we have

$$\chi_A(\lambda) = (a_{11} - \lambda) \begin{vmatrix} a_{22} - \lambda & \dots & a_{2n} \\ 0 & \vdots & \vdots \\ 0 & \dots & a_{nn} - \lambda \end{vmatrix} = (a_{11} - \lambda) \prod_{i=2}^{n} (a_{ii} - \lambda)$$

by the case for $n-1$. Hence the roots of $\chi_A(\lambda) = 0$ are $\lambda = a_{ii}$ for all $i = 1, 2, \dots, n$, which are the eigenvalues of A. Now we know that $\chi_A(\lambda) = |A - \lambda I_n|$ is a polynomial in λ of degree n with leading term $(-1)^n \lambda^n$ and constant term $|A|$.

(Check for $A \in M_{2\times2}(F)$ and $A \in M_{3\times3}(F)$. Recall that the trace of $A, tr(A)$ is the sum of its diagonal elements. Thus for$A \in M_{2\times2}(F)$,

$$\begin{aligned} \chi_A(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda) \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} \\ &= \lambda^2 - tr(A)\lambda + |A|. \end{aligned}$$

Similarly for $A \in M_{3\times3}(F)$ we have

$$\begin{aligned} \chi_A(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) \\ &= (-1)^3\lambda^3 + tr(A)\lambda^2 - (a_{11}a_{22} + a_{11}a_{33} + a_{22}a_{33})\lambda + |A|. \end{aligned}$$

In general we can show that for $A \in M_n(F)$ and A a triangular matrix, the coefficient of $\lambda^{n-1} = tr(A)$ and the constant term is $|A|$. Thus if $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of A and hence roots of its characteristic function, then its constant term $|A| = \prod_{i=1}^{n} \lambda_i$ and $tr(A) = \sum \lambda_i$.

$\square$

Remark 13.7. The question as to when a linear transformation may be represented by a triangular matrix (or has a so-called *triangular form*) is answered in your Part A course - in this case the characteristic function has a simple form (a product of linear terms).

14 Applications of Eigenvectors to linear systems of first order Ordinary Differential Equations

Consider the linear system of 1st order ODE's with constant coefficients given by

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ \vdots &= \vdots + \vdots \ldots \vdots \\ \frac{\mathrm{d}x_n}{\mathrm{d}t} &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n, \end{aligned}$$

where $\{x_i\}_1^n$ is a set of differentiable functions in t and $a_{ij} \in \mathbb{R} \, \forall \, | \leq i, j \leq n$.

1.

$$\Leftrightarrow \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} A = (a_{ij}).$$

Case 1 A diagonal matrix; $A = (\lambda_1, \lambda_2 \ldots \lambda_n)$
Then (1) reduces to

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda_1 x_1 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= \lambda_2 x_2 \\ \frac{\mathrm{d}x_n}{\mathrm{d}t} &= \lambda_n x_n \end{aligned}$$

Which has general solution

$$\mathbf{x} = \begin{pmatrix} c_1 e^{\lambda_1 t} \\ c_n e^{\lambda_n t} \end{pmatrix}$$

Case 2 A is not diagonal, but it diagnalisable. Suppose $\exists$ matrix $p \in m_n(\mathbb{R},$ $P^{-1}AP = D$ (or $A = PDP^{-1}$)

(1) Becomes $\frac{\mathrm{d}x}{\mathrm{d}t} = PDP^{-1}\mathbf{x}$
$\Leftrightarrow P^{-1}\frac{\mathrm{d}x}{\mathrm{d}t} = DP^{-1}\mathbf{x}$

Set $\mathbf{u} = P^{-1}\, so\, \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = P^{-1}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$

so the system (3) becomes

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = D\mathbf{u}$$

using (2) we know the general solution for $\mathbf{u}$ is

$$\mathbf{u} = \begin{pmatrix} c_1 e^{\lambda_1 t} \\ c_n e^{\lambda_n t} \end{pmatrix}$$

λ_i are diagonal entries in D i.e. λ_i are eigenvalues of A.

But $\mathbf{x} = P\mathbf{u}$ so

$$\mathbf{x} = P\begin{pmatrix} c_1 e^{\lambda_1 t} \\ c_n e^{\lambda_n t} \end{pmatrix}$$

Where P is matrix of eigenvectors $\mathbf{V_i}$ of A. (Corresponds to eigenvalues).

Write

$$P = \begin{pmatrix} V_1 & V_n \\ \Downarrow & \Downarrow \end{pmatrix}$$

and eigenvectors form basis for $\mathbb{R}^n$

$$\mathbf{x} = \begin{pmatrix} v_1 & v_n \\ \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_1 t} \end{pmatrix} = \sum_{i=1}^{n} c_i e^{\lambda i} \mathbf{v_i}$$

Theorem 14.1. *Let*$A \in M_n(\mathbb{R})$ *the general solution of the matrix d.e.*

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x} \textit{ is } \mathbf{x} = e^{tA}\mathbf{c}$$

where $\mathbf{c} \in \mathbb{R}^n$ *is arbitrary.*

Proof.

$$\begin{aligned}
\mathrm{d}\frac{\mathbf{x}}{\mathrm{d}t} = A\mathbf{x} \quad &\leftrightarrow \quad \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - A\mathbf{x} = \mathbf{0} \\
&\leftrightarrow \quad e^{-tA}(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - A\mathbf{x}) = \mathbf{0} \\
&\leftrightarrow \quad e^{-tA}\underbrace{\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - A}_{exact} e^{-tA}\mathbf{x} = \mathbf{0} \\
&\leftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}(e^{-tA}\mathbf{x}) = \mathbf{0} \\
&\leftrightarrow \quad e^{-tA}\mathbf{x} = \mathbf{c} \\
&\leftrightarrow \quad \mathbf{x} = e^{-tA}\mathbf{c}
\end{aligned}$$

$$\text{some } \mathbf{c} \in \mathbb{R}^n$$

□

So we have proved

Theorem 14.2. *Suppose* $A \in m_n(\mathbb{R})$ *has n linear independent e-vectors corresponding to eigenvalues* $\lambda_1 \dots \lambda_n$. *Then the general solution of the linear system*

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A\mathbf{x}$$

is

$$\mathbf{x} = \sum_{i=1}^{n} c_i e^{\lambda_i t}\mathbf{v_i}$$

where $\mathbf{c_i} \dots \mathbf{c_n}$ ***are arbitrary constants.***

Case 3 Suppose A is not a diagonalisable matrix. We can devise a solution using the exponential matrix e^A, where

$\mathrm{e}^A = I + A + \frac{1}{1}A^2 + \frac{1}{3^1}A^3 + \ldots$. Considering S_n, sequence of partial sums of e^A, it can be shown that $\underbrace{lin}_{n\to\infty} s_{ij}, n$ exists $for alli, j$, where s_{ij}, n is i, j entry of s_n. Define xxx to be sum of $p.s.$

we need the following Lemma.

Lemma 14.3. *Let* $A \in M_n(\mathbb{R})$

$$\frac{\mathrm{d}e^{tA}}{\mathrm{d}t} = Ae^{tA}; \tag{14.1}$$
$$e^{-tA} = (e^{tA})^{-1} \tag{14.2}$$

Proof P263 Kumpel and Thorpe. Linear Algebra