Linear Algebra and Groups MATH40003

Dr. C. Kestner and Prof. D. Evans

Typed by Benjamin Bateman

October 2019 - March 2020

Contents

Chapter 1

Linear Algebra

1 Introduction

Often the first step to tackling a problem is to "Linearise" it, so to put it into the form of a system of linear equations, like a Taylor Series:

$$f(x) \approx f(a) + (x-a)f'(a) + \cdots, |x-a| \leq 1$$

A function or 'Transformation' of L is linear iff $L(af_1 + bf_2) = aLf_1 + bLf_2$ This makes linear transformations easier to handle than non-linear ones. In the linear algebra part of this course we will look at the maths developed to deal with linear transformations.

2 Systems of Linear Equations and Matrices

2.1 Introduction

A system of linear equations is a set of equations in the same variables. For example:

$$\begin{aligned} -x + y + 2z &= 2 \\ 3x - y + z &= 6 \\ -x + 3y + 4z &= 4 \end{aligned}$$

In general, a system of m linear equations in n unknowns will have the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad &\quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Definition 2.1.1. Given a system of m linear equations in n unknowns, we can write this in matrix form $Ax = b$ where:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \; b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \; A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix}$$

We can use an *Augmented Matrix* to represent the system of linear equations:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} & b_m \end{array}\right) (A|b)$$

Example 2.1.2.

$$\begin{aligned} w - x + y + 2z &= 2 \\ w + 3x - y + z &= 6 \end{aligned}$$

As an augmented matrix:

$$\left(\begin{array}{rrrr|r} 1 & -1 & 1 & 2 & 2 \\ 1 & 3 & -1 & 1 & 6 \end{array}\right)$$

Remark 2.1.3. Matrix Multiplication is defined precisely so that the above equations work out.

2.2 Matrix Algebra

Here's a quick recap/introduction (N.B, we will mainly be working in $\mathbb{R}$ but any field F will do).

If we want to add matrices, they need to have the same size and shape.

Definition 2.2.1. Given

$$\begin{aligned} A &= [a_{ij}]_{m\times n} \in M_{m\times n}(\mathbb{R}) \\ B &= [b_{ij}]_{m\times n} \in M_{m\times n}(\mathbb{R}) \end{aligned}$$

Then define

$$A + B = [a_{ij} + b_{ij}]_{m\times n}$$

We can also multiply by a scalar product.

Definition 2.2.2. Let $A = [a_{ij}]_{m\times n} \in M_{m\times n}(\mathbb{R})$ and $\lambda \in \mathbb{R}$. Then the *Scalar multiple of A by λ* denoted by λA is the matrix

$$\lambda A = [\lambda a_{ij}]_{m\times n}$$

.

We can also multiply matrices together.

Definition 2.2.3. Let $A = [a_{ij}]_{p\times q} \in M_{p\times q}(\mathbb{R})$ and $B = [b_{ij}]_{q\times r} \in M_{q\times r}(\mathbb{R})$. Then the *matrix product* AB is the matrix C where:

$$C = [c_{ij}]_{p\times r}, c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj}$$

Theorem 2.2.4. *Matrix multiplication is associative. I.e. for matrices A, B, C, we have $(AB)C = A(BC)$*

Proof: for $A(BC)$ to be defined, we require the respective orders to be $m \times n, n \times p, p \times q$, in which case the product $A(BC)$ is also defined (and vice versa). Consider the ij^{th} element of $A(BC)$.

$$\begin{aligned}
[A(BC)]_{ij} &= \sum_{k=1}^{n} a_{ik}[BC]_{kj} \\
&= \sum_{k=1}^{n}\sum_{t=1}^{p} a_{ik}b_{kt}c_{tj} \\
&= \sum_{t=1}^{p}\left(\sum_{k=1}^{n} a_{ik}b_{kt}\right)c_{tj} \\
&= \sum_{t=1}^{p}[AB]_{it}c_{tj} \\
&= [(AB)C]_{ij}
\end{aligned}$$

So $A(BC) = (AB)C$ as every ij^{th} element is equal.

2.3 Row Operations

Example 2.3.1. The augmented matrix for the following system of linear equations

$$\begin{aligned}
-x + y + 2z &= 2 \\
3x - y + z &= 6 \\
-x + 3y + 4z &= 4
\end{aligned}$$

is

$$\left(\begin{array}{rrr|r} -1 & 1 & 2 & 2 \\ 3 & -1 & 1 & 6 \\ -1 & 3 & 4 & 4 \end{array}\right)$$

You will already know how to solve system of linear equations (simultaneous equations). There are three different operations:

- Multiply an equation by a non-zero factor;

- Add a multiple of one equation to another;
- Swap equations around.

In the augmented matrix we can do these operations more efficiently.

Definition 2.3.2. *Elementary row operations* are performed on an augmented matrix. The allowed options for these operations are detailed in the list above.

Remark 2.3.3. Performing row operations preserves the solutions of a linear system, and every row operation has an inverse.

Example 2.3.4. Consider the following system of linear equations in augmented matrix form. The left column represents the simultaneous equations method of solving the equations, and the right represents using row operations and augmented matrices:

$$
\begin{aligned}
3x - 2y + z &= -6 \quad (1.1)\\
4x + 6y - 3z &= 5 \quad (1.2)\\
-4x + 4y &= 12 \quad (1.3)
\end{aligned}
$$

Multiply (1.3) by $\frac{1}{4}$

$$-x + y = 3 \qquad (1.4)$$

Add $3 \times (1.4)$ to (1.1) and $4 \times (1.4)$ to (1.2)

$$
\begin{aligned}
y + z &= 3 \quad (1.5)\\
10y - 3z &= 17 \quad (1.6)
\end{aligned}
$$

Take $10 \times (1.5)$ from (1.6)

$$-13z = -13$$

So $z = 1$. Plug into (1.5) to get

$$y + 1 = 3$$

So $y = 2$ Plug into (1.4)

$$-x + 2 = 3 \qquad x = -1$$

$$
\left(\begin{array}{ccc|c} 3 & -2 & 1 & 6\\ 4 & 6 & -3 & 5\\ -4 & 4 & 0 & 12 \end{array}\right)
\xrightarrow{R_3 \to R_3 \times \frac{1}{4}}
\left(\begin{array}{ccc|c} 3 & -2 & 1 & 6\\ 4 & 6 & -3 & 5\\ -1 & 1 & 0 & 3 \end{array}\right)
\xrightarrow[R_2 \to R_2 + 4R_3]{R_1 \to R_1 + 3R_3}
\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 10 & -3 & 17\\ -1 & 1 & 0 & 3 \end{array}\right)
$$

$$
\xrightarrow{R_2 \to R_2 - 10R_1}
\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 0 & -13 & -13\\ -1 & 1 & 0 & 3 \end{array}\right)
\xrightarrow{R_2 \to -\frac{1}{13} R_2}
\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 0 & 1 & 1\\ -1 & 1 & 0 & 3 \end{array}\right)
\xrightarrow{R_1 \to R_1 - R_2}
\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2\\ 0 & 0 & 1 & 1\\ -1 & 1 & 0 & 3 \end{array}\right)
$$

$$
\xrightarrow{R_3 \to R_3 - R_1}
\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 0 & 1 & 1\\ -1 & 0 & 0 & 1 \end{array}\right)
\xrightarrow{R_3 \to -R_3}
\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 0 & 1 & 1\\ 1 & 0 & 0 & -1 \end{array}\right)
\xrightarrow{R_1 \to R_1 - R_2}
\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2\\ 0 & 0 & 1 & 1\\ 1 & 0 & 0 & -1 \end{array}\right)
$$

Definition 2.3.5. Two systems of linear equations are *equivalent* iff either:

- They are both inconsistent.
- The augmented matrix of the first system can be obtained using row op-

erations from the augmented matrix of the second system.

Remark 2.3.6. Equivalently, by Remark 2.3.3, two systems of linear equations are equivalent if and only if they have the same solutions.

If a row consists of mainly 0s and 1s then it is easier to read off solutions.

Example 2.3.7.

$$\left(\begin{array}{ccc|c} -2 & 1 & 2 & 2 \\ 3 & -3 & 1 & 5 \end{array}\right) \qquad \left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 5 \end{array}\right)$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$\begin{aligned} -2x + y + 2z &= 2 \\ 3x - 3y + z &= 5 \end{aligned} \qquad \begin{aligned} y &= 2 \\ z &= 5 \end{aligned}$$

Definition 2.3.8. We say a matrix is in *Echelon form (e.f.)* if it satisfies the following:

- All the zeroes are at the bottom.
- The first non-zero entry in each row is 1
- The first non-zero entry in each row i is strictly to the left of the first non-zero entry in row $i + 1$.

We say a matrix is in *row reduced echelon form* if it is in echelon form and:

- If the first non-zero entry in row i also appears in column j then every other entry in column j is zero.

Example 2.3.9. A matrix in echelon form looks like:

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 2 \\ 0 & 1 & 7 & 12 \\ 0 & 0 & 1 & -10 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

And a matrix in row reduced echelon form looks like:

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

2.4 Elementary Matrices

Definition 2.4.1. Any matrix that can be obtained from an identity matrix by means of one elementary row operation is in *elementary matrix*.

We have three types:

- Multiplies a row by a non-zero number:

$$E_r(\alpha) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \alpha & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

- Adds a multiple of row r to row s:

$$E_{rs}(\alpha) = \begin{array}{c} \\ \\ \\ s \\ r \end{array} \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & \alpha \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

- Swapping row s with row r

$$E_{rs} = \begin{array}{c} \\ s \\ r \\ \\ \\ \end{array} \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & 1 & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

Theorem 2.4.2. *Let $A \in M_{m\times n}(\mathbb{R})$, E an elementary $m \times m$ matrix. The matrix multiplication EA applies the same elementary row operation to A that was performed on I_m to obtain E*

Proof: From the definition of matrix multiplication, we have

$$ea_{ij} = \sum_{k=1}^{n} e_{ik}a_{kj}$$

Now we have E is the identity matrix (so will preserve the row i of A) apart from one (or two, in the third case) critical row(s), which we will call r, and the other involved row (if any) s. Now

$$ea_{rj} = \sum_{k=1}^{n} e_{rk}a_{kj}$$

Now for the different types of elementary matrices:

1.
$$e_{rr} = \alpha \implies ea_{rj} = \alpha a_{rj}$$

2.
$$e_{rs} = \alpha \implies ea_{rj} = a_{rj} + \alpha a_{sj}$$

3.
$$e_{rs} = e_{sr} = 1, e_{rr} = e_{ss} = 0 \implies ea_{rj} = a_{sj}, ea_{sj} = ea_{rj}$$

Which is the desired result. $\square$

2.5 More Matrices

Definition 2.5.1. We say a matrix is *square* if it has the same number of columns as it does rows.

Definition 2.5.2. A square matrix $A = (a_{ij}) \in M_{n\times n}(\mathbb{F})$ is:

- *Upper triangular* if $a_{ij} = 0$ whenever $i > j$ (zeroes below the diagonal)
- *Lower triangular* if $a_{ij} = 0$ whenever $i < j$ (zeroes above the diagonal)
- *Diagonal* if $a_{ij} = 0$ whenever $i \neq j$.

Example 2.5.3.

$$\text{U.T}: \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix} \text{L.T}: \begin{pmatrix} 3 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{Diagonal} \begin{pmatrix} 3 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Definition 2.5.4. The $n \times n$ *identity matrix*, denoted I_n has all its diagonal entries equal to one, all other entries equal to zero. It is called the identity matrix because it is the *multiplicative identity* matrix for $M_{n\times n}(\mathbb{F})$. I.e.

$$\forall A \in M_{n\times n}(\mathbb{F}) : AI_n = I_n A = A$$

Definition 2.5.5. If, for $B \in M_{n\times n}(\mathbb{F}), \exists A \in M_{n\times n}(\mathbb{F}) : AB = BA = I_n$, then we say B is *invertible*, with *inverse* A. We write $A = B^{-1}$

Note: not all $n \times n$ matrices are invertible, e.g. $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Definition 2.5.6. A matrix without an inverse is called *singular*.

Example 2.5.7.

$$\text{Let} A = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix} \text{Verify} A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix} AA^{-1} = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Theorem 2.5.8. *The inverse of a given matrix is unique. I.e., suppose $A, B, C \in M_{n\times n}(F)$ such that $AB = BA = I_n$ and $AC = CA = I_n$. Then $B = C$.*

Proof: suppose

$$\begin{aligned} AB = BA = I_n &\text{ and } AC = CA = I_n \\ B &= BI_n \\ b &= BI_n \\ &= B(AC) \\ &= (BA)C \\ &= I_nC \\ &= C \end{aligned}$$

□ 2.5.8 allows us to talk about the **only** inverse of a matrix.

Definition 2.5.9. If $A = (a_{ij})_{m\times n}$ then the *transpose of* A, $A^T = (a_{ji})_{n\times m}$

Example 2.5.10. If

$$A = \begin{pmatrix} 1 & 0 & 5 \\ 4 & 2 & 1 \end{pmatrix} \text{ then } A^T = \begin{pmatrix} 1 & 4 \\ 0 & 2 \\ 5 & 1 \end{pmatrix}$$

Corollary. *Let $A \in M_{n\times n}(\mathbb{R})$ be invertible, then A^T is also invertible. And $(A^T)^{-1} = (A^{-1})^T$*

Proof:

$$\begin{aligned} AA^{-1} &= I_n \\ (AA^{-1})^T &= I_n^T = I_n \\ (A^{-1})^T A^T &= I_n \end{aligned}$$

□

Lemma 2.5.11. *Let $A \in M_{n\times m}(\mathbb{R}), B \in M_{m\times p}(\mathbb{R})$. Then*

$$(AB)^T = B^T A^T$$

Proof: First note that $(AB)^T \in M_{p\times n}(\mathbb{R})$. $B^T A^T$ is defined and has order $p \times n$.
Let $A = (a_{ij})$ and $B = (b_{ij})$. Then:

- the ji^{th} entry of $(AB)^T$ is the ij^{th} entry of (AB) which is $\sum_{k=1}^{m} a_{ik}b_{kj}$
- the ji^{th} entry of $B^T A^T$ is

$$\begin{aligned} &\sum_{k=1}^{m} (b^T)_{jk}(a^T)_{ki} \\ &= \sum_{k=1}^{m} (b^T)_{jk}(a^T)_{ki} = \sum_{k=1}^{M} a_{ik}b_{kj} \end{aligned}$$

□

2.6 Inverses using row operations

We can use an e.r.o to find inverses (if they exist)

Theorem 2.6.1. *Every elementary matrix is invertible, and its inverse is an elementary matrix.*

Proof: check.

- $E_r(\alpha)E_r(\alpha^{-1}) = I_n$ and vice versa
- $E_{rs}(\alpha)E_{rs}(-\alpha) = I_n$ and vice versa.
- $E_{rs}E_{rs} = I_n$

So we have $(E_r(\alpha))^{-1} = E_r(\alpha^{-1}), (E_{rs}(\alpha))^{-1} = E_{rs}(-\alpha)$ and $E_{rs}^{-1} = E_{rs}$ $\square$

Theorem 2.6.2. *If $A \in M_{n\times n}(\mathbb{R})$ can be reduced to I_n by a sequence of elementary row operations, then A is invertible, and its inverse is found by applying the same sequence to I_n*

Proof: Let $E_1, E_2, \cdots, E_k$ be the elementary matrices corresponding to the elementary row ops. So

$$E_k \cdots E_3 E_2 E_1 A = I_n$$

The previous theorem states that the E_i are invertible. Recall also that $(AB)^{-1} = B^{-1}A^{-1}$. So:

$$A = E_1^{-1}E_2^{-1}\cdots E_k^{-1}$$

This shows that A is a product of invertible matrices, so A^{-1} exists.

$$\begin{aligned} A^{-1} &= (E_1^{-1}\cdots E_k^{-1})^{-1} \\ &= E_k \cdots E_1 \\ &= E_k \cdots E_1 I_n \end{aligned}$$

which is what you get from applying the row operations to I_n $\square$

Example 2.6.3. Let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ 3 & 0 & 4 \end{pmatrix}$$

First, construct an augmented matrix with the identity matrix.

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 4 & 0 & 0 & 1 \end{array}\right)$$

This gives us a good way of keeping track of the e.r.o.'s. We're looking to get the LHS equal to I_3 and the RHS will give us A_{-1}

$$\xrightarrow{R_3 \to R_3 - 3R_1}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to R_2 - R_1}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to R_2 + R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -4 & 1 & 1 \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_1 \to R_1 - R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & 0 & -1 \\ 0 & 2 & 0 & -4 & 1 & 1 \\ 0 & 1 & 1 & -3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to \frac{R_2}{2}}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & 0 & -1 \\ 0 & 1 & 0 & -2 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & -3 & 0 & 1 \end{array}\right)$$

2.7 Geometric Interpretation

Ben's remark: From this point, we shall write scalar quantities in italicised form: $x, y, \alpha, \lambda, \mu$ and vectors in bold: $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Matrices will stay italicised until part 3.8, where they will be in bold font to distinguish from subspaces (and later, groups).

As you have seen in the introductory module, vectors are $n \times 1$ matrices and vectors in $\mathbb{R}^2/\mathbb{R}^3$ can be represented as points in 2 or 3 (respectively) dimensional space. In this section we will look at the geometric interpretations of some of the things we have seen so far.

A system of linear equations in n unknowns specifies a set in n-space.

Example 2.7.1. Consider

$$\begin{aligned} x_1 + x_2 + x_3 &= -1 \\ 2x_1 + x_3 &= 1 \\ 3x_1 + x_2 &= -4 \end{aligned}$$

Using row reduction we get

$$
\begin{aligned}
x_1 &= -0.5 \\
x_2 &= -2.5 \\
x_3 &= 2
\end{aligned}
$$

This specifies a point. Whereas

$$
\begin{aligned}
x_1 + x_2 + x_3 &= -1 \\
2x_1 + x_3 &= 1
\end{aligned}
$$

Using row reduction we get

$$
\begin{aligned}
x_1 &= -2.5 - 0.5x_3 \\
x_2 &= 1.5 - 0.5x_3
\end{aligned}
$$

Giving us the line

$$\begin{pmatrix} -2.5 \\ 1.5 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -0.5 \\ -0.5 \\ 1 \end{pmatrix}$$

for $\lambda \in \mathbb{R}$. Just taking the first equation

$$x_1 + x_2 + x_3 = -1$$

gives us a plane with normal $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

We have seen that we can apply matrices to vectors via matrix multiplication, so we can see a matrix $A \in M_{m\times n}(\mathbb{R})$ as a map.

$$
\begin{aligned}
A : &\mathbb{R}^n \to \mathbb{R}^m \\
&\mathbf{v} \to A\mathbf{v}
\end{aligned}
$$

We can use matrices to represent many different operations.

Example 2.7.2. Consider $A = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$. Then

$$A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5x_1 \\ 5x_2 \end{pmatrix}$$

so A is a stretch by a factor of 5.

Definition 2.7.3. Let T be a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, we say T is a *linear transformation* if for every $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ and every $\alpha, \beta \in \mathbb{R}$, we have

$$T(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2)$$

Theorem 2.7.4. *Let $A \in M_{m\times n}(\mathbb{R})$ be seen as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, then A is a linear transformation.*

Proof: let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n, \alpha, \beta \in \mathbb{R}$. Then

$$\begin{aligned} A(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) &= A(\alpha\mathbf{v}_1) + A(\beta\mathbf{v}_2) && \text{By Distributivity} \\ &= \alpha(A\mathbf{v}_1) + \beta(A\mathbf{v}_2) \end{aligned}$$

$\square$

Lemma 2.7.5. *Let $A \in M_{n\times n}(\mathbb{R})$. The following are equivalent:*

(1) A is invertible with inverse $A^{-1} = A^T$.

(2) $AA^T = I_n = A^TA$

(3) A preserves inner products (i.e. dot products) i.e. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, (A\mathbf{x}) \cdot (A\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

Proof: (1) $\iff$ (2) by definition.
(2) $\iff$ (3): First note for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$ as per intro to maths. So A preserves inner products if and only if:

$$\begin{aligned} &(A\mathbf{x}) \cdot (A\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \\ \iff &(A\mathbf{x})^T(A\mathbf{y}) = \mathbf{x}^T\mathbf{y} \\ \iff &(A\mathbf{x})^T(A\mathbf{y}) = \mathbf{x}^T I_n\mathbf{y} \\ \iff &\mathbf{x}^T(A^TA)\mathbf{y} = \mathbf{x}^T I_n\mathbf{y} \\ \iff &\mathbf{x}^T(A^TA - I_n)\mathbf{y} = 0 \end{aligned}$$

If (2) then $A^TA = I_n$ so $\mathbf{x}^T(A^TA - I_n)\mathbf{y} = 0$ so we can conclude (3)
then if (3), let

$$\mathbf{x}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

where the 1 is on the i^{th} row. So for each $\mathbf{x}_i$:

$$\mathbf{x}_i^T(A^TA - I_n)\mathbf{y} = 0$$

The LHS of which is the i^{th} row of the column vector. So

$$(A^TA - I_n)\mathbf{y} = \mathbf{0}_n \in \mathbb{R}^n$$

where $\mathbf{0}_n$ is the zero vector for $\mathbb{R}^n$. Now do the same thing choosing $\mathbf{y}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ where the 1 is in the j^{th} row. This gives us

$$(A^T A - I_n) = 0 \in M_{n \times n}(\mathbb{R})$$

So $A^T A = I_n$ $\square$

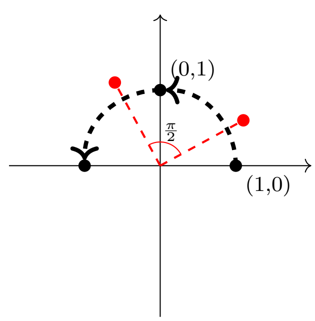
Definition 2.7.6. $A \in M_{n\times n}(\mathbb{R})$ is called *orthogonal* if it is such that $A^{-1} = A^T$.

Here are some examples of useful orthogonal matrices in $\mathbb{R}^2$, and a very useful result proved in a lemma at the end.

Example 2.7.7. ① Consider the matrix

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

We have $A^{-1} = A^T$ (easy to check!) and it is therefore orthogonal. Now the behaviour of this matrix as a transformation $A : \mathbb{R}^2 \to \mathbb{R}^2$ is:

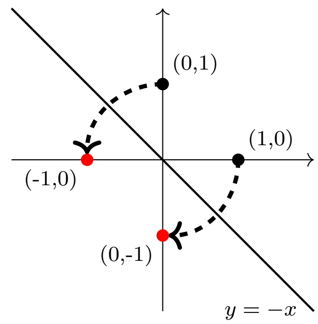


As we can see, the matrix takes a value $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and transforms it to $\begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$, which is a rotation of $\pi/2$ about the origin, anticlockwise.

② The matrix

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

Has the same property as A in that it is orthogonal (Again, easy to check!) Now, if we transform the same coordinate pair $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ we get $\begin{pmatrix} -x_2 \\ -x_1 \end{pmatrix}$ Which is a reflection in the line $y = -x$:



Lemma 2.7.8. *The rotation about the origin of angle θ is given by the matrix*

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

Proof: Consider the unit square. We can write the transformation R_θ depending on what happens to the square when we transform it, and using trigonometry, we know we want the "Bottom right" corner to go to the position $(\cos\theta, \sin\theta)$, and the "top left" to $(-\sin\theta, \cos\theta)$. Now we can simply combine these into the transformed unit square, using the fact that the coordinates of the unit square form the Identity matrix. This gives us our result. $\square$

In general, when working in $\mathbb{R}^2$, we can use the trick:

$$A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$$
$$A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$
$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Finally, we finish this section with an interesting theorem:

Theorem 2.7.9. *If $A \in M_{n\times n}(\mathbb{R})$ is orthogonal, then it is either a rotation or a reflection.*

First we make a claim and prove it to add some simplicity (and clarity) to why we call matrices orthogonal. Recall we say that two vectors are orthonormal if and only if their inner (dot) product is zero, and each vector is such that $\|\mathbf{x}\| = 1$. We claim that an equivalent definition for orthogonal matrices is:

Definition 2.7.10. A matrix A is orthogonal if an only if the columns are orthonormal.

Proof of claim: Consider $C = AA^T$, using the definition of matrix multiplication, we obtain:

$$c_{ij} = \sum_{k}^{n} a_{ik}a_{kj}^T = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

because we have $a^T_{ij} = a_{ji}$. This gives us $A^T A = I_n$ as required.

Proof of 2.7.9: Any point of unit norm in $\mathbb{R}^2$ can be written as $(\cos\theta, \sin\theta)$ and its orthogonal twin will either be $(-\sin\theta, \cos\theta)$ or $(\sin\theta, -\cos\theta)$. So if we construct a matrix from these two vectors, we will *always* end up with a rotation or reflection. □

We discuss orthonormal and orthogonal vectors more in section 6.4.

2.8 Fields

So far we have seen matrices and systems of linear equations in $\mathbb{R}$. We could have used ANY field.

Every field has distinguished elements 0 (additive identity) and 1 (multiplicative identity). As a result, over any field F we can define

1. The null matrix (i.e. the additive identity matrix):

$$\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

2. The identity matrix (the matrix multiplicative identity):

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

Remark 2.8.1. It is important to know what field we are working in, especially for scalar multiplication. e.g. if we take $M_{n\times m}(\mathbb{Q})$ this is not closed under scalar multiplication from $\mathbb{R}$ for example, multiplication by $\sqrt{2}$. You have seen $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ and there are also finite fields.

Theorem 2.8.2. *Let $\mathbb{F}_p = \{0, 1, \cdots, p-1\}$ for some prime number p. Define addition as $a + b \equiv (a + b) \mod p$. Define multiplication as $ab \equiv (ab) \mod p$. Then $(\mathbb{F}_p, +_{\mod p}, \times_{\mod p}, 0, 1)$ is a field.*

Proof: A1-4 are clear from the relevant properties in $\mathbb{Z}$. Similarly, M1-3 are clear from the same properties. But M4 about the multiplicative inverse is less clear.

For $x \in \mathbb{F}_p\{0\}$ we have $\gcd(x, p) = 1$. By the intro module there are $s, t \in \mathbb{Z}$ such that

$$1 = sx + tp$$

i.e. $sx \equiv 1 \mod p$. Take $s \mod p \in \mathbb{F}_p$. This is then the multiplicative inverse of x in $\mathbb{F}_p$. D1 gets its properties similarly from $\mathbb{Z}$. □

[0.5cm] (For a reminder of the Fields axioms, see the material on the introduction to university mathematics course MATH40001)

3 Vector Spaces

3.1 Introduction to vector spaces

Definition 3.1.1. Let F be a field. A *vector space over* F is a non empty set V together with the following maps:

1. Addition of vectors:

$$\begin{aligned} + : V \times V &\to V \\ (\mathbf{v}_1, \mathbf{v}_2) &\to \mathbf{v}_1 + \mathbf{v}_2 \end{aligned}$$

2. Scalar multiplication:

$$\begin{aligned} \cdot : F \times V &\to V \\ (f, \mathbf{v}) &\to f \cdot \mathbf{v} \end{aligned}$$

These satisfy the following axioms:

For vector addition we have:

A1. Associativity:

$$\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V : \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$

A2. Commutativity:

$$\forall \mathbf{u}, \mathbf{v} \in V : \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

A3. Additive identity element:

$$\exists \mathbf{0} \in V, \forall \mathbf{v} \in V : \mathbf{0} + \mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v}$$

$\mathbf{0}$ is called the zero vector.

A4. Additive inverses:

$$\forall \mathbf{v} \in V, \exists \mathbf{u} \in V : \mathbf{v} + \mathbf{u} = \mathbf{u} + \mathbf{v} = \mathbf{0}$$

Then for multiplication by a scalar:

A5. Distributive law 1:

$$\forall r \in F, \forall \mathbf{u}, \mathbf{v} \in V : r \cdot (\mathbf{u} + \mathbf{v}) = r \cdot \mathbf{u} + r \cdot \mathbf{v}$$

A6. Distributive law 2:

$$\forall r, s \in F, \forall \mathbf{u} \in V : (r + s) \cdot \mathbf{u} = r \cdot \mathbf{u} + s \cdot \mathbf{u}$$

A7. Associativity:

$$\forall r, s \in F, \mathbf{u} \in V : (rs) \cdot \mathbf{u} = r \cdot (s \cdot \mathbf{u})$$

A8. Identity scalar:

$$1 \cdot \mathbf{v} = \mathbf{v}$$

Definition 3.1.2. Let V be a vector space over F.

- Elements of V are called *vectors*
- Elements of F are called *scalars*
- We call V an F-Vector space (sometimes, and a lot in this course)

Example 3.1.3. The following are real vector spaces:

1. The canonical example is $\mathbb{R}^n$ where $+$ is normal matrix addition and $\cdot$ is scalar multiplication.
2. $M_{m \times n}(\mathbb{R})$ with $+$ matrix addition and $\cdot$ normal scalar multiplication and

$$0 = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

3. Define $V = \mathbb{R}^X$ to be the set of *real-valued functions on X*.

$$\mathbb{R}^X = \left\{ \begin{array}{cc} f: & f \text{ is a function} \\ f: & X \to \mathbb{R} \end{array} \right\}$$

Then for $f, g \in \mathbb{R}^X, \alpha \in \mathbb{R}$, define

$$f + g : X \to \mathbb{R}$$

$$\forall x \in X, (f + g)(x) = f(x) + g(x)$$

and define:

$$\alpha \cdot f : X \to \mathbb{R}$$

$$\forall x \in X : (\alpha \cdot f)(x) = \alpha(f(x))$$

We'll now present some non-examples of vector spaces and examine why they aren't vector spaces in $\mathbb{R}$.

Example 3.1.4. 1. The set of vectors $V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : a, b \in \mathbb{Z} \right\}$ is NOT an $\mathbb{R}$-vector space. This is because we can't define scalar multiplication. Take the following counterexample:

$$\sqrt{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin V$$

2. $V = \left\{ \begin{pmatrix} a+1 \\ 2 \end{pmatrix} : a \in \mathbb{R} \right\}$ with standard multiplication and addition. This isn't a vector space because it doesn't contain the identity element, $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

3. Consider the following addition and scalar multiplication operations.

$$\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix} \tag{1.1}$$

$$r \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ ry \end{pmatrix} \tag{1.2}$$

$$r \in \mathbb{R}$$

this isn't an $\mathbb{R}$-vector space because it doesn't satisfy axiom 8.

$$1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

3.2 Subspaces

Definition 3.2.1. A subset W of a vector space V (over F) is a *subspace* if:

S1. W is not empty.

S2. Let $\mathbf{v}, \mathbf{w} \in W$, then $\mathbf{v} + \mathbf{w} \in W$ i.e., the set is *closed under vector addition.*

S3. Let $\mathbf{v} \in W, \alpha \in F$, then $\alpha \cdot \mathbf{v} \in W$ i.e. the set is *closed under scalar multiplication.*

We use the notation $W \leq V$ to mean W is a subspace of V.

Remark 3.2.2. Note that V and the *zero subspace* $\{\mathbf{0}\}$ are always subspaces of V. Any other subspace is called a *proper subspace*. In these notes we differentiate between the scalar 0 and the vector $\mathbf{0}$ using bold face. The difference may also sometimes be denoted by 0_v being the zero vector.

Proposition 3.2.3. Every subspace must contain the zero vector.

Proof: First, we claim that for an F-vector space V with $0 \in F$ [the field additive identity] we have

$$0 \cdot \mathbf{v} = \mathbf{0}, \forall \mathbf{v} \in V$$

We must simply show that $0 \cdot \mathbf{v}$ works as an additive identity since the additive identity is always unique. Let $\mathbf{v}$ be any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ so:

$$0 \cdot \mathbf{v} + \mathbf{v} = (0+1) \cdot \mathbf{v} = 1 \cdot \mathbf{v}$$

and so the claim is proven.

Let $W \leq V, \mathbf{v} \in W$(S1) then $-1 \cdot \mathbf{v} \in W$(S3)

$$\begin{aligned}
\mathbf{0} &= 0 \cdot \mathbf{v} && \text{By our claim} \\
&= (1-1) \cdot \mathbf{v} && \text{field axioms} \\
&= 1 \cdot \mathbf{v} + -1 \cdot \mathbf{v} && \text{(A6)} \\
&= \mathbf{v} + -1 \cdot \mathbf{v} \in W && \text{(S2)}
\end{aligned}$$

□ This can also be proved kinda easier: Once we know that $0 \cdot \mathbf{v} = \mathbf{0}$ then we have $\mathbf{v} \in W$ and $0 \in F$ so

$$0 \cdot \mathbf{v} = \mathbf{0} \in W \text{ by (S3)}$$

Example 3.2.4. Show that the set $X = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^2$

Proof:

S1: $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \in X$ so $X \neq \emptyset$

S2: Let $\mathbf{v}, \mathbf{w} \in X$ then

$$\mathbf{v} = \begin{pmatrix} a \\ 0 \end{pmatrix},\ \mathbf{w} = \begin{pmatrix} b \\ 0 \end{pmatrix}, a, b \in \mathbb{R}$$

$$\begin{aligned}
\mathbf{v} + \mathbf{w} &= \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} b \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} a+b \\ 0 \end{pmatrix}
\end{aligned}$$

and since $a + b \in \mathbb{R}, \mathbf{v} + \mathbf{w} \in X$

S3: let $\mathbf{v} \in X$ and $r \in \mathbb{R}$. Then $\mathbf{v} = \begin{pmatrix} a \\ 0 \end{pmatrix}$ for some $a \in \mathbb{R}$.

$$r \cdot \mathbf{v} = r \begin{pmatrix} a \\ 0 \end{pmatrix} = \begin{pmatrix} ra \\ 0 \end{pmatrix}$$

and as $ra \in \mathbb{R}, r \cdot \mathbf{v} \in X$ □

Theorem 3.2.5. *let $U \leq V, W \leq V, V$ an F-vector space. Then $U \cap W \leq V$. In general, the intersection of any set of subspaces is a subspace.*

Proof: again, we'll go through the subspace axioms and check!

S1: $\mathbf{0} \in U, \mathbf{0} \in W$ by 3.2.3. So $\mathbf{0} \in U \cap W$ i.e. $U \cap W \neq \emptyset$

S2: Suppose $\mathbf{v}_1, \mathbf{v}_2 \in U \cap W$. Then $\mathbf{v}_1, \mathbf{v}_2 \in U$ and by $U \leq V$(S2) we have $\mathbf{v}_1 + \mathbf{v}_2 \in U$. The same holds with W in place of U.

S3: Suppose $\mathbf{v} \in U \cap W, \alpha \in F$. Then $\mathbf{v} \in U \implies \alpha\mathbf{v} \in U$ (since $U \leq V$ by S3) Similarly, $\mathbf{v} \in W \implies \alpha\mathbf{v} \in W$ (since $W \leq V$ by S3). So then $\alpha\mathbf{v} \in U \cap W$.

As $U \cap W \subset V$ we get $U \cap W \leq V$. □

Proposition 3.2.6. Let V be an F-Vector space $W \leq V$. Then W is an F-vector space too.

Proof: All we need to do is show that every vector in the subspace W is also in V. In more mathematical terms, $W \cup V$ is just V. We already have that in order for W to be a subspace, it must also be a subset of V. By the definition of subsets then, every member of W is a member of V, proving the proposition. □

Example 3.2.7. In general if $U \leq V$ and $W \leq V$ for V an F-vector space, we don't get $U \cup W$ being a subspace. For example, let

$$\begin{aligned} U &= \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^2 : x \in \mathbb{R} \right\} && \text{x-axis} \\ W &= \left\{ \begin{pmatrix} 0 \\ y \end{pmatrix} \in \mathbb{R}^2 : y \in \mathbb{R} \right\} && \text{y-axis} \\ V &= \mathbb{R}^2 \end{aligned}$$

Then

$$U \leq V,\ W \leq V$$

Then

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in U \cup W$$

But

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin U \cup W$$

so $U \cup W$ is not generally closed under vector addition.

3.3 Spanning sets

Definition 3.3.1. Let V be an F-vector space, and let $\mathbf{u}_1, \cdots, \mathbf{u}_m \in V$.

- A *linear combination* of $\mathbf{u}_1 \cdots, \mathbf{u}_m$ is a vector of the form $\alpha_1\mathbf{u}_1 + \cdots + \alpha_m\mathbf{u}_m = \sum_{i=1}^m \alpha_i\mathbf{u}_i$ for scalars $\alpha_1, \cdots, \alpha_m \in F$.
- The *Span* of $\mathbf{u}_1, \cdots, \mathbf{u}_m$ is the set of linear combinations of $\mathbf{u}_1, \cdots, \mathbf{u}_m$. i.e. $\text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \{\alpha_1\mathbf{u}_1 + \cdots + \alpha_1\mathbf{u}_1 \in V : \alpha_1, \cdots, \alpha_m \in F\ \}$

Lemma 3.3.2. *Let V be an F-vector space and $\mathbf{u}_1 \cdots \mathbf{u}_m \in V$. Then Span$(\mathbf{u}_1, \cdots, \mathbf{u}_m)$ is a subspace of V.*

Proof: It's quite clear that it's a subspace, but we do the "test" (i.e. check the axioms again) to make sure.

S1: $\mathbf{u}_1 \in \text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m)$ so $\text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) \neq \emptyset$

S2: Suppose $\mathbf{v}, \mathbf{w} \in \text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m)$, so

$$\mathbf{v} = \sum_{i=1}^m \alpha_i\mathbf{u}_i \qquad \alpha_i \in F$$

and

$$\begin{aligned}
\mathbf{w} &= \sum_{i=1}^m \beta_i\mathbf{u}_i && \beta_i \in F \\
\mathbf{v} + \mathbf{w} &= \sum_{i=1}^m \alpha_i\mathbf{u}_i + \sum_{i=1}^m \beta_i\mathbf{u}_i \\
&= \sum_{i=1}^m (\alpha_i + \beta_i)\mathbf{u}_i && \text{(A6)} \\
\text{As } &\alpha_i + \beta_i \in F \\
\mathbf{v} + \mathbf{w} &\in \text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m)
\end{aligned}$$

S3: Suppose $\mathbf{v} \in \text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m)$ and $\lambda \in F$, then

$$\mathbf{v} = \sum_{i=1}^m \alpha_i\mathbf{u}_i \qquad \alpha_i \in F$$

so

$$\lambda v = \lambda\left(\sum_{i=1}^m \alpha_i\mathbf{u}_i\right) = \sum_{i=1}^m (\lambda\alpha_i)\mathbf{u}_i \qquad \text{(A5)}$$

And as $\lambda\alpha_i \in F$ we can have $\lambda\mathbf{v} \in \text{Span}(\mathbf{u}_1 \cdots, \mathbf{u}_m)$.

Remark 3.3.3. By convention we take the empty sum to be $\mathbf{0}$. So $\text{Span}(\emptyset) = \{\mathbf{0}\}$. For an infinite set S we still take finite sums for $\text{Span}(S)$. I.e.

$$\text{Span}(S) = \left\{ \sum_{\mathbf{s}_i \in S'} \alpha_i \mathbf{s}_i : S' \text{ a finite subset of } S, \alpha_i \in F \right\}$$

Proposition 3.3.4. For an infinite subset S of an F-vector space V, $\text{Span}(S)$ is a subspace.

Proof: Left as an exercise (consider the union of 2 finite subspaces which you can make from our infinite set S, and vectors $\mathbf{v}$ and $\mathbf{u}$ which are in the span of S and can be made from elements of the finite subspaces. All that's left is to also prove that these finite subsets are disjoint.) $\square$

Definition 3.3.5. Let V be an F-vector space, suppose $s \leq V$ such that $\text{Span}(s) = V$. Then we say S is a *spanning set* for V, or equivalently S *spans* V.

Example 3.3.6. Which of the following sets spans $\mathbb{R}^3$?

1. $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$: yes this is a spanning set! You can make $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ by taking the second vector away from the fourth.

2. $\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$: Nope, this guy is not :(

3. $\left\{ \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} \right\}$ Yep, this set spans $\mathbb{R}^3$

4. $\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$: No again :c

The basic takeaway from this is the fact to get a spanning set over $\mathbb{R}^n$ you have to be able to make all the e_i unit vectors using addition and scalar multiplication, like what was taught during the introduction module in regards to bases. Covered in more detail in just a bit.

In part 1 of that previous exercise, we get a 'redundant' vector. If as well as being a spanning set, it's also *linearly independent*, then this won't happen.

3.4 Linear Independence

Definition 3.4.1. Let V be an F-vector space. We say $\mathbf{u}_1, \cdots, \mathbf{u}_m \in V$ are *linearly independent* if whenever

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0} \qquad \alpha_i \in F$$

then

$$\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0 \in F$$

we say $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ is a *linearly independent set.*

Alternatively, a set $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ is *linearly dependent* if $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = 0_v$ where at least one of the $\alpha_i \neq 0 \in F$. A set is linearly independent if it is NOT linearly dependent (duh).

Example 3.4.2. The set $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a linearly independent subset of $\mathbb{R}^3$.

[0.5cm] *Proof:* Suppose

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Then

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

so

$$\alpha_1 = \alpha_2 = \alpha_3 = 0 \in F$$

$\square$

Example 3.4.3.

- Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions and suppose $f(x) = x, g(x) = x^2$. The set $\{f, g\}$ is a linearly independent subset of $\mathbb{R}^{\mathbb{R}}$, i.e. the set of functions from $\mathbb{R} \to \mathbb{R}$

  Proof: Assume $\alpha, \beta \in \mathbb{R}$ such that $(\alpha f + \beta g) = 0_v$. Our aim is to prove $\alpha = \beta = 0$. Two functions are equal iff they are equal on all elements of the domain. Now $1, 2 \in \mathbb{R}$:

$$\begin{aligned} 0_v(1) &= (\alpha f + \beta g)(1) \\ 0 &= \alpha f(1) + \beta g(1) \\ &= \alpha(1) + \beta(1) \\ &= \alpha + \beta \\ \alpha &= -\beta \end{aligned}$$

also,

$$\begin{aligned}
0_v(2) &= (\alpha f + \beta g)(2) \\
0 &= \alpha f(2) + \beta g(2) \\
&= \alpha 2 + \beta 4 \\
\alpha &= -2\beta
\end{aligned}$$

Therefore $\alpha = \beta = 0$

- The set $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a linearly dependent subset of $\mathbb{R}^3$.

  Proof: Note

  $$1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-1) \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

  So there is a combination of these vectors where the α_i don't need to all be zero to obtain the zero vector. This can also be shown another way:

  $$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

  Thereby showing that one member of the set can be made using the two others, which is sufficient for our definition of linear dependence.

- V an F-vector space then $\{\mathbf{0}\}$ is linearly dependent, by the definition (since the α_i can be anything and we get $\mathbf{0}$ anyway)

- V an F-vector space, and for $\mathbf{v} \in V, \{\mathbf{v}\}$ is linearly independent iff $\mathbf{v} \neq \mathbf{0}$.

Lemma 3.4.4. *Let $\mathbf{v}_1, ..., \mathbf{v}_m$ be linearly independent in an F-Vector space V. Let $\mathbf{v}_{m+1}$ be such that $\mathbf{v}_{m+1} \notin \text{Span}(\mathbf{v}_1, ..., \mathbf{v}_m)$. Then $\{\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{v}_{m+1}$ is linearly independent.*

Proof: suppose $\alpha_1, ..., \alpha_m, \alpha_{m+1} \in F$ such that

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m + \alpha_{m+1} \mathbf{v}_{m+1} = \mathbf{0}$$

Our aim is to show that $\alpha_1 = \cdots = \alpha_m = \alpha_{m+1} = 0$. Suppose $\alpha_{m+1} \neq 0$, then

$$\begin{aligned}
\mathbf{v}_{m+1} &= \frac{-1}{\alpha_{m+1}} (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m) \\
&\in \text{Span}(\mathbf{v}_1, ... \mathbf{v}_m)
\end{aligned}$$

so $\alpha_{m+1} = 0$. So:

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m + \mathbf{0} = \mathbf{0}$$

i.e.

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_m\mathbf{v}_m = \mathbf{0}$$

but $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is linearly independent, so

$$\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$$

$\square$

3.5 Bases

Definition 3.5.1.

- Let V be an F-vector space. A *basis* of V is a linearly independent spanning set.
- If V has a finite basis, then we say V is *finite dimensional*.

Example 3.5.2.

1. The set $B = \left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

   Proof:

   - Show B is Linearly independent. Suppose $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ such that

     $$\alpha_1 \begin{pmatrix}1\\0\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\\0\end{pmatrix} + \alpha_3 \begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

     Then $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

   - Show B spans $\mathbb{R}^3$. Let $\mathbf{v} \in \mathbb{R}^3$, then $\mathbf{v} = \begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}$ for $v_1, v_2, v_3 \in \mathbb{R}$. Then

     $$\mathbf{v} = \begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix} = v_1 \begin{pmatrix}1\\0\\0\end{pmatrix} + v_2 \begin{pmatrix}0\\1\\0\end{pmatrix} + v_3 \begin{pmatrix}0\\0\\1\end{pmatrix} \in \text{span}(B)$$

   $\square$

2. Let F be a field, then in F^n, let $\hat{\mathbf{e}}_i$ be the column vectors with zeroes everywhere except the row i, where the entry is 1. Then $\{\hat{\mathbf{e}}_i, ..., \hat{\mathbf{e}}_n\}$ forms a basis for F^n. The proof is very similar to the last one, just check that it's linearly dependent and that it spans F! $\square$

Remark 3.5.3. We can see from the following example that not all vector spaces are finite dimensional. Take the vector space $\mathbb{R}[x] :=$ polynomials with variable x, which is a problem in problem sheet 3 to prove that it's a vector space. This vector space has basis $\{1, x, x^2, x^3 ...\}$ which is evidently not finite dimensional.

Proposition 3.5.4. Let V be an F-vector space, $S = \{\mathbf{u}_1 ... \mathbf{u}_m\} \subset V$. Then S is a basis if and only if every vector in V has a *unique* expression as a linear combination of elements of S.

Proof: " $\implies$ " Suppose S is a basis, and take $\mathbf{v} \in V$. [We want there to be unique $\alpha_1 ... \alpha_m \in F$ such that $v = \sum_{i=1}^m \alpha_i \mathbf{u}_i$]
EXISTENCE: Since V is spanned by S, we have $\alpha_1, ..., \alpha_m \in F$ such that

$$\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m$$

UNIQUENESS: Suppose for contradiction that we also have $\beta_1, ..., \beta_m \in F$ such that $\mathbf{v} = \beta_1 \mathbf{u}_1 + \cdots + \beta_m \mathbf{u}_m = \sum_{i+1}^m \beta_i \mathbf{u}_i$. Then

$$\sum_{i=1}^m \alpha_i \mathbf{u}_i = \sum_{i=1}^m \beta_i \mathbf{u}_i$$

i.e.

$$\left(\sum_{i=1}^m \alpha_i \mathbf{u}_i \right) - \left(\sum_{i=1}^m \beta_i \mathbf{u}_i \right) = \mathbf{0}$$
$$\sum_{i=1}^m (\alpha_i - \beta_i) \mathbf{u}_i = \mathbf{0}$$

but $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ is linearly independent. So $\alpha_i - \beta_i = 0, \forall i \in \{1, 2, ..., m\}$ i.e. $\alpha_i = \beta_i$ which is our contradiction. So the α_i are unique.
" $\impliedby$ ": Suppose conversely for every $\mathbf{v} \in V$ there are unique $\alpha_1, ..., \alpha_m \in F$ such that $\mathbf{v} = \sum_{i=1}^m \alpha_i \mathbf{u}_i$. [Our aim is to show that the set $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ must be spanning, and linearly independent.] Suppose for every $\mathbf{v} \in V$ there are unique $\alpha_1, ..., \alpha_n \in F$ such that $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{v}$

1. Spanning: Let $\mathbf{v} \in V$ then there exist $\alpha_1, ..., \alpha_m \in F$ such that $\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m \in \mathrm{Span}(\mathbf{u}_1, ..., \mathbf{u}_m)$

2. Linear Independence: suppose $\alpha_1, ..., \alpha_m \in F$ with $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m = \mathbf{0}$. Note that $0\mathbf{u}_1 + \cdots + 0\mathbf{u}_m = \mathbf{0}$ so by uniqueness,

$$\alpha_1 = \cdots = \alpha_m = 0$$

□

Remark 3.5.5. Let $B = \{\mathbf{u}_1, ..., \mathbf{u}_m\}$ be a basis for an F-vector space. By proposition 3.5.4 we have a bijective map

$$V \longrightarrow F^m$$
$$v = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_m \mathbf{u}_m \to (\alpha_1, \alpha_2, ..., \alpha_m)$$

we call $(\alpha_1, ..., \alpha_m)$ the *coordinates of* V (with respect to B)

Proposition 3.5.6. Let V be a non-trivial (i.e. not $\{\mathbf{0}\}$ F-vector space, and suppose V has a finite spanning set S. Then S contains an LI spanning set (i.e. a basis).

Proof: Consider T, such that

- T is LI
- $T \subset S$
- T is the largest such subset (maximal)

We have such a T because $V \neq \{\mathbf{0}\}$, so there is $\mathbf{v} \in V$, thus for S to be a spanning set, $S \neq \{\mathbf{0}\}$. Take $\mathbf{v'} \in V, \{\mathbf{v'}\}$ is LI.
Claim: T is spanning. The proof of this is as follows: assume on the contrary that $\mathbf{v} \in V(= \mathrm{Span}(S))$ such that $\mathbf{v} \notin \mathrm{Span}(T)$. So $\mathbf{v} \in \mathrm{Span}(s) \backslash \mathrm{Span}(T)$. By Lemma 3.4.4, $\{\mathbf{v}\} \cup T$ is LI. We may assume that $\mathbf{v} \in S$ because (assuming T is maximal but not spanning), $\mathbf{v}$ must be in S to give us $\mathrm{Span}(T) \neq \mathrm{Span}(S)$, since if $\mathbf{v} \notin S$, we get $\mathrm{Span}(T) = \mathrm{Span}(S)$, which contradicts T not spanning. Therefore $|\{\mathbf{v}\} \cup T| > |T|$, which is a contradiction. □

This last proof is quite difficult to follow, but the intuition is hopefully easy to understand. If you have a spanning set, then you can find the vectors who are redundant (linearly dependent) and remove them to get a basis. This intuition is critical to the following extremely important Lemma.

3.6 Dimension

Lemma 3.6.1. (Steinitz Exchange Lemma) *Let V be a vector space over F. Take $X \leq V$ and suppose $u \in \mathrm{Span}(X)$. But $u \notin \mathrm{Span}(X \backslash \{\mathbf{v}\})$ for some $\mathbf{v} \in X$. Now let $Y = (X \backslash \{\mathbf{v}\}) \cup \mathbf{u}$, "exchange $\mathbf{v}$ for $\mathbf{u}$". Then $\mathrm{Span}(X) = \mathrm{Span}(Y)$.*

Proof: Since $\mathbf{u} \in \mathrm{Span}(X)$ we have $\alpha_1, ..., \alpha_n \in F$ and $\mathbf{v}_1, ..., \mathbf{v}_n \in X$ such that

$$\mathbf{u} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$$

As $\mathbf{u} \notin \mathrm{Span}(X \backslash \{\mathbf{v}\})$ we may assume $\mathbf{v} = \mathbf{v}_n$ and $\alpha_n \neq 0$. So $\mathbf{v} = \mathbf{v}_n = (\alpha_n)^{-1}(\mathbf{u} - (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_{n-1} \mathbf{u}_{n-1}))$. Then $\mathrm{Span}(Y) \leq \mathrm{Span}(X)$.

Take $\mathbf{w} \in \mathrm{Span}(Y)$. There are $\beta_1, ..., \beta_m \in F$ and $\mathbf{x}_1, ..., \mathbf{x}_m \in Y = (X \backslash \{\mathbf{v}\}) \cup \{\mathbf{u}\}$ such that

$$\mathbf{w} = \beta_1 \mathbf{x}_1 + \cdots + \beta_m \mathbf{x}_m$$

We may assume that $\mathbf{x}_1 = \mathbf{u}$ (and if it doesn't appear, then set $\beta_1 = 0$ and try the next, and so forth). Then

$$\mathbf{w} = \beta_1 \mathbf{u} + \sum_{i=2}^{m} b_i \mathbf{x}_i$$

$$\mathbf{w} = \beta_1 (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) + \sum_{i=2}^{m} \beta_i \mathbf{x}_i$$

so $\mathbf{w} \in \text{Span}(X)$ Similarly, using the case if $\mathbf{w} \in \text{Span}(X)$ then $\mathbf{w} \in \text{Span}(Y)$ i.e.

$$\text{Span}(X) \subset \text{Span}(Y)$$

Thus

$$\text{Span}(X) = \text{Span}(Y)$$

Remark 3.6.2. We need this lemma to be able to define dimensions... it relied on taking inverses in F.

Theorem 3.6.3. *Let V be a vector space. Then let S, T be finite subsets of V. Suppose that*

- *S is an LI set.*
- *T spans V*

Then $|S| \leq |T|$

Aside: Read this as "LI sets are smaller than or equal to spanning sets"
Proof:

$$\begin{aligned} S &= \{\mathbf{s}_1, ..., \mathbf{s}_m\} && \text{LI} \\ T &= \{\mathbf{t}_1, ..., \mathbf{t}_n\} && \text{spans} \end{aligned}$$

IDEA: let's use S.E.L. and swap elements of T for elements of S, retaining that the set spans V. We cannot run out of space in T, as this would mean the remaining elements of S were in the span of the ones already placed in T.

Assume S is LI, T spans V.

$$\begin{aligned} S &= \{\mathbf{s}_1, ..., \mathbf{s}_m\} \\ T &= \{\mathbf{t}_1, ..., \mathbf{t}_n\} \end{aligned}$$

Let $T = T_0$, since $\text{span}(T_0) = V$, there is some i such that

$$\begin{aligned} &\mathbf{s}_1 \in \text{Span}(\mathbf{t}_1, ..., \mathbf{t}_i) \\ &\mathbf{s}_1 \notin \text{Span}(\mathbf{s}_1, \mathbf{t}_1, ..., \mathbf{t}_{i-1}, \mathbf{t}_{i+1}, ..., \mathbf{t}_n) \end{aligned}$$

Let $T_1 = \{\mathbf{s}_1, \mathbf{t}_1, ..., \mathbf{t}_{i-1}, \mathbf{t}_{i+1}, ..., \mathbf{t}_n\}$, which is everything in T except for $\mathbf{t}_i$. By the S.E.L we get $V = \text{Span}(T_0) = \text{Span}(T_1)$. We continue inductively.

Suppose for some j with $1 \leq j \leq m$ we have

$$T_j = \{\mathbf{s}_1, ..., \mathbf{s}_j, \mathbf{t}_{i1}, ..., \mathbf{t}_{i(n-j)}\}$$

with $\text{Span}(T_j) = \text{Span}(T) = V, \mathbf{t}_{ij} \in T$ Now

$$\begin{aligned} &\mathbf{s}_{j+1} \in \text{Span}(\mathbf{t}_j) \\ &\mathbf{s}_{j+1} \notin \text{Span}(\mathbf{s}_1, ..., \mathbf{s}_j) \end{aligned}$$

S is LI. So there is a k such that

$$\mathbf{s}_{j+1} \in \text{Span}\{\mathbf{s}_1, ..., \mathbf{s}_j, \mathbf{t}_{i_1}, ..., \mathbf{t}_{i_k}\}$$
$$\mathbf{s}_{j+1} \notin \{\mathbf{s}_1, ..., \mathbf{s}_{j+1}, \mathbf{t}_{i_1}, ..., \mathbf{t}_{i_{k+1}}\}$$

Then let $T_{j+1} = \{\mathbf{s}_1, ..., \mathbf{s}_{j+1}, \mathbf{t}_{i_1}, ..., \mathbf{t}_{i_{k-1}}, \mathbf{t}_{i_{k+1}}, ..., \mathbf{t}_{i_{n-j}}$ which is T_j with $\mathbf{t}_{i_k}$ removed. Then $\text{Span}(T_{j+1}) = \text{Span}(T_j) = V$ by relabelling we have a set of the form

$$T_{j+1} = \{\mathbf{s}_1, ..., \mathbf{s}_{j+1}, \mathbf{t}_{i_1}, ..., \mathbf{t}_{i_{n-(j+1)}}\}$$

After j steps, we have replaced j elements of T with j elements of S. We cannot run out of elements of T before we run out of elements of S, as otherwise the remaining elements of S would be in the span of the elements of S that have already been swapped, which contradicts that S is linearly independent. □

Ben's remark: Wow this proof is hard to follow. It's basically a how-to of the SEL, I guess. Hopefully the LaTeX formatting will make it easier for you guys. See Kestner's notes here for some examples with diagrams.

Corollary. *Let V be a finite dimensional vector space. Let T, S be bases of V. Then T and S are both finite and $|S| = |T|$.*

Proof: Since V is finite dimensional, it has a finite basis B. Suppose $|B| = n$. By 3.6.3, any linearly independent set has size $\leq n$. i.e. $|S| \leq n, |T| \leq n$. But S is spanning, and T is LI. So by 3.6.5 $|T| \leq |S|$. And T is spanning, and S is LI. So $|S| \leq |T|$. Therefore $|S| = |T|$. □

Definition 3.6.4. Let V be a finite dimensional vector space. The *dimension of V*, written $\text{dim}V$, is the size of any basis of V.

We **need** the previous corollary and the SEL to know that the dimension of V is unique.

Example 3.6.5. Describe the subspaces of $\mathbb{R}$. Problem sheet 3, Q4.

By 3.6.3 these have dim ≤ 3.

$$\begin{aligned}
&\text{dim3} : \mathbb{R}^3 \\
&\text{dim2} : \text{Planes through the origin.} \cong \mathbb{R}^2 \\
&\text{dim1} : \text{lines through the origin.} \\
&\text{dim0} : \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}
\end{aligned}$$

3.7 More subspaces

Definition 3.7.1. Let V be a vector space, U, W subspaces.

- The *intersection of U and W* is $U \cap W = \{\mathbf{v} \in V : \mathbf{v} \in U, \mathbf{v} \in W\}$

- The *sum of U and W* is $U + W := \{\mathbf{u} + \mathbf{w} : \mathbf{u} \in U, \mathbf{w} \in W\}$

Remark 3.7.2. $U \leq U + W, W \leq U + W$ as $0 \in U$, and $\forall \mathbf{w} \in W, \mathbf{w} = 0 + \mathbf{w} \in U + W$

Example 3.7.3. Let $V = \mathbb{R}^2, U = \text{Span}\{(1,0)\}, W = \text{Span}\{(0,1)\}$. Then $U + W = V$ *Proof:* $U + W \leq V$. Now let $\mathbf{v} \in V$ then $\mathbf{v} = (\alpha, \beta), \alpha, \beta \in \mathbb{R}$. Also:

$$V = \alpha(1,0) + \beta(0,1) \in U + W$$

so $U + W \geq V \therefore U + W = V$.

Example 3.7.4. Let U and W be subspaces of an F-vector space V. Then $U + W$ and $U \cap W$ are subspaces of V.

Proof:

- $U \cap W$ is 3.2.6
- $U + W$ we check with the subspace test.

S1. $0 \in U, 0 \in W$, so $0 + 0 \in U + W$

S2. Suppose $\mathbf{v}_1, \mathbf{v}_2 \in U + W$. Then

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 + \mathbf{w}_1 && \text{for some } \mathbf{u}_1 \in U, \mathbf{w}_1 \in W \\ \mathbf{v}_2 &= \mathbf{u}_2 + \mathbf{w}_2 && \text{for some } \mathbf{u}_2 \in U, \mathbf{w}_2 \in W \end{aligned}$$

So

$$\begin{aligned} \mathbf{v}_1 + \mathbf{v}_2 &= (\mathbf{u} + 1 + \mathbf{w}_1) + (\mathbf{u}_2 + \mathbf{w}_2) \\ &= (\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{w}_1 + \mathbf{w}_2) \end{aligned}$$

so $\mathbf{v}_1 + \mathbf{v}_2 \in U + W$

S3. Suppose $\lambda \in F, \mathbf{v} \in U + W$. Then

$$\begin{aligned} \mathbf{v} &= \mathbf{u} + \mathbf{w} && \text{for some} \mathbf{u} \in U, \mathbf{w} \in W \\ \lambda\mathbf{v} &= \lambda(\mathbf{u} + \mathbf{w}) \\ &= \lambda\mathbf{u} + \lambda\mathbf{w} \end{aligned}$$

so $\lambda\mathbf{v} \in U + W$.

□

Proposition 3.7.5. Let V be a vector space over F. $U, W \leq V$. Suppose

- $U = \text{Span}\{\mathbf{u}_1, ..., \mathbf{u}_s\}$
- $W = \text{Span}\{\mathbf{w}_1, ..., \mathbf{w}_r\}$

Then $U + W = \text{Span}\{\mathbf{u}_1, ..., \mathbf{u}_2, \mathbf{w}_1, ..., \mathbf{w}_r\} = S$

Proof: We want $U + W \subset S$. Let $\mathbf{v} \in U + W, \mathbf{v} = \mathbf{u} + \mathbf{w}, \mathbf{u} \in U, \mathbf{w} \in W$ So

$$\begin{aligned}\mathbf{u} &= \alpha_1 \mathbf{u}_1 + \cdots + \alpha_s \mathbf{u}_s \\ \mathbf{w} &= \beta_1 \mathbf{w}_1 + \cdots + \beta_r \mathbf{w}_r\end{aligned}$$

Thus

$$\mathbf{u} + \mathbf{w} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_s \mathbf{u}_s + \beta_1 \mathbf{w}_1 + \cdots + \beta_r \mathbf{w}_r \in S$$

Now we want $S \subset U + W$. Suppose $\mathbf{v} \in S$, then

$\mathbf{v} = \lambda_1 \mathbf{u}_1 + \cdots + \lambda_s \mathbf{u}_s + \mu_1 \mathbf{w}_1 + \cdots + \mu_r \mathbf{w}_r \text{Span}\{\mathbf{u}_1, ..., \mathbf{u}_s\} + \text{Span}\{\mathbf{w}_1, ..., \mathbf{w}_r\}$

So $\mathbf{v} \in U + W$

Example 3.7.6. Let

$$\begin{aligned}\mathbf{v} &= \mathbb{R}^2 \\ U &= \text{Span}\{(1, 0)\} \\ W &= \text{Span}\{(0, 1)\} \\ U + W &= \text{Span}\{(0, 1), (, 1, 0)\} \\ &= \mathbb{R}^2\end{aligned}$$

Example 3.7.7. Let $V = \mathbb{R}^3$.

$$\begin{aligned}U &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\} \\ W &= \{(x_1, x_2, x_3 \in \mathbb{R}^3 : -x_1 + 2x_2 + x_3 = 0\}\end{aligned}$$

Find bases for $U, W, U \cap W, U + W$.

- U: a general vector in U is of the form

  $$\begin{aligned}\mathbf{u} &= (a, b, -a - b) && \text{for } a, b \in \mathbb{R} \\ \mathbf{u} &= a(1, 0, -1) + b(0, 1, -1)\end{aligned}$$

  so $\{(1, 0, -1.0, 1, -1)\}$ is a spanning set for U. Clearly it is also linearly independent, as the equation

  $$\alpha_1(1, 0, -1) + \alpha_2(0, 1, -1) = \begin{pmatrix}\alpha_1 \\ 0 \\ \alpha_1\end{pmatrix} + \begin{pmatrix}0 \\ \alpha_2 \\ -\alpha_2\end{pmatrix} = 0$$

  which gives $\alpha_1 = \alpha_2 = 0$

- W: Using similar methods, $\{(2, 1, 0), (1, 0, 1)\}$ is a basis for W, as any $\mathbf{w} \in W$ is of the form $\mathbf{w} = (2a + b, a, b)$

- $U+W$: By 3.7.6, $\{(1,0,-1),(0,1,-1),(2,1,0),(1,0,1)\}$ spans $U+W$. Clearly it's not linearly independent, so we can row reduce to get one. By row reductions we get

$$U+W = \text{Span}\{(1,0,0),(0,1,0),(0,0,1)\} = \mathbb{R}^3$$

- $U \cap W$: Let $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$.

$$\begin{aligned} \mathbf{x} \in U &\iff x_1 + x_2 + x_3 = 0 \\ \mathbf{x} \in Q &\iff -x_1 + 2x_2 + x_3 = 0 \end{aligned}$$

  so $\mathbf{x} \in U \cap W$ iff

$$x + 1 + x_2 + x_3 = -x_1 + 2x_2 + x_3 = 0$$

  so $U \cap W = \{(x_1, x_2, x_3 \in \mathbb{R}^3 : x_1 + x_2 + x_3 = -x - 1 + 2x_2 + x_3 = 0\}$. From this equation we get $\mathbf{x} = (x_1, 2x_1, -3x_1)$ so a spanning set is of the form $\{(1,2,-3)\}$ which is LI, so a basis for $U \cap W$.

Remark 3.7.8. A neater way of finding a basis for $U+W$ would be to find a basis for $U \cap W$. Since $U \cap W \leq W$ we could extend this basis to one for W. Similarly, we could extend to a basis for U. The union of these bases will be a basis for $U+W$. In 3.7.7,

$$\begin{aligned} B_{U\cap W} &= \{(1,2,-3)\} \\ B_U &= \{(1,2,-3),(1,0,-1)\} \\ B_W &= \{(1,2,-3),(1,0,1\} \\ B_{U+W} &= \{(1,2,-3),(1,0,-1),(1,0,1)\} \end{aligned}$$

Theorem 3.7.9. *Le V be a vector space over $F, U, W \leq V$. Then* $\dim(U+W) = \dim(U) + \dim(W) - \dim(U \cap W)$

Proof: Suppose $\dim(U) = r, \dim(W) = s, \dim(U \cap W) = m$. Now we have a basis of $U \cap W$:

$$B_{U\cap W} = \{\mathbf{v}_1, ..., \mathbf{v}_m\}$$

Now $U \cap W \leq U$ and $B_{U\cap W}$ is linearly independent, so it is contained in some basis of U.

$$B_U = \{\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{u}_{m+1}, ..., \mathbf{u}_r\}$$

Similarly, we have a basis for W:

$$B_W = \{\mathbf{v}_1, ..., \mathbf{v}_m, w_{m+1}, ..., \mathbf{w}_s\}$$

Claim:

$$B_U \cup B_W = \{\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{u}_{m+1}, ..., \mathbf{u}_r, \mathbf{w}_{m+1}, ..., \mathbf{w}_s\}$$

is a basis for $U+W$.

Proof of claim:

- By prop 3.7.5, $B_U \cup B_W$ is a **spanning set**.

LI: Suppose

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m + \alpha_{m+1}\mathbf{v}_{m+1} + \cdots + \alpha_r \mathbf{u}_r + \beta_{m+1}\mathbf{w}_{m+1} + \cdots + \beta_s \mathbf{w}_s = \mathbf{0}$$

i.e.

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i + \sum_{i=m+1}^{r} \alpha_i \mathbf{u}_i + \sum_{i=m+1}^{s} \beta_i \mathbf{w}_i = \mathbf{0}$$

We want $\lambda_i = \alpha_j = \beta_k = 0 \forall i, l, k \in F$. Since the first 2 sums are in U, and the second is in W, we have

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i + \sum_{i=m+1}^{r} \alpha_i \mathbf{u}_i = -\sum_{i=m+1}^{s} \beta_i \mathbf{w}_i$$

Thus

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i + \sum_{i=m+1}^{r} \alpha_i \mathbf{u}_i \in U \cap W$$

So it's in Span($\{\mathbf{v}_1, ..., \mathbf{v}_m\}$)

$$\sum_{i=1}^{m} \lambda_i \mathbf{v}_i + \sum_{i=m+1}^{r} \alpha_i \mathbf{u}_i = \sum_{i=1}^{m} \mu_i \mathbf{v}_i$$

Therefore

$$\sum_{i=1}^{m} (\lambda_i - \mu_i)\mathbf{v}_i + \sum_{i=m+1}^{r} \alpha_i \mathbf{u}_i = \mathbf{0}$$

But $\{\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{u}_{m+1}, ..., \mathbf{u}_r\}$ is LI. so

$$\begin{aligned} \lambda_i - \mu_i &= 0 && \text{for} i \in \{1, ..., m\} \\ \alpha_j &= 0 && \text{for} j \in \{m+1, ..., r\} \end{aligned}$$

But $\{\mathbf{v}_1, ..., \mathbf{u}_m, \mathbf{w}_{m+1}, ..., \mathbf{w}_s\}$ is LI. Therefore

$$\begin{aligned} \lambda_i &= 0 && \text{for } i \in \{1, ..., m\} \\ \beta_j &= 0 && \text{for } j \in \{m+1, ..., s\} \end{aligned}$$

Which completes the proof of the claim. □

So $B_U \cup B_W$ is a basis for $U + W$

$$\begin{aligned} |B_U| \cup |B_W| &= |B_U| + |B_W| - |B_U \cap B_W| \\ &= |B_U| + |B_W| - |B_{U \cap W}| \\ &= r + s - m \end{aligned}$$

□

3.8 Rank of matrix

Definition 3.8.1. Let $\mathbf{A} \in M_{m\times n}(F)$. Define

- The row space of $\mathbf{A}$ (write $\mathrm{RSp}(\mathbf{A})$) as the span of the rows of $\mathbf{A}$. (This is a subspace of F^n)
- The column space of $\mathbf{A}$ (write $\mathrm{CSp}(\mathbf{A})$) as the span of the columns of $\mathbf{A}$. (This is a subspace of F^m
- The row rank of $\mathbf{A}$ is $\dim(\mathrm{RSp}(\mathbf{A}))$
- The column rank of $\mathbf{A}$ is $\dim(\mathrm{CSp}(\mathbf{A}))$.

Example 3.8.2.

$$F = \mathbb{R}, \mathbf{A} = \begin{pmatrix} 3 & 1 & 2 \\ 0 & -1 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} \mathrm{RSp}(\mathbf{A}) &= \mathrm{Span}\{(3,1,2), 0, -1, 1\} \\ \mathrm{CSp}(\mathbf{A}) &= \mathrm{Span}\left\{\begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\} \end{aligned}$$

Here's a procedure to calculate the row rank of a matrix $\mathbf{A}$.

Example 3.8.3. 1. Reduce $\mathbf{A}$ to row echelon form (using row ops)

$$\mathbf{A}_{ech} = \begin{pmatrix} 1 & x & x & x & \cdots & \cdots \\ 0 & 0 & 1 & x & x & \cdots \\ 0 & 0 & 0 & 1 & x & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

Actually it doesn't matter whether the leading entries are 1; they just need to be non-zero.

2. The row rank of $\mathbf{A}$ is the number of non-zero rows in $\mathbf{A}_{ech}$. In fact the non-zero rows of $\mathbf{A}_{ech}$ form a basis for $\mathrm{RSp}(\mathbf{A})$.

Justification: It is enough to show that

1. $\mathrm{RSp}(\mathbf{A}) = \mathrm{RSp}(\mathbf{A}_{ech})$
2. The rows of $\mathbf{A}_{ech}$ are LI.

1) Note that to obtain $\mathbf{A}_{ech}$ from $\mathbf{A}$ we use row operations:

$$\begin{aligned} &\mathbf{r}_i \to \mathbf{r}_i + \lambda\mathbf{r}_j && \lambda \in F, i \neq j \\ &\mathbf{r}_i \to \lambda\mathbf{r}_i && \lambda \in F \\ &\mathbf{r}_i \to \mathbf{r}_j \end{aligned}$$

Let $\mathbf{A}'$ be obtained from $\mathbf{A}$ by using one row operation, then clearly every row of $\mathbf{A}'$ lies in $\mathrm{RSp}(\mathbf{A})$. So

$$\mathrm{RSp}(\mathbf{A}') \subset \mathrm{RSp}(\mathbf{A})$$

As every row has an inverse which is also a row op, we get

$$\mathrm{RSp}(\mathbf{A}) \subset \mathrm{RSp}(\mathbf{A}')$$

i.e. their row spans are equal. Therefore doing row operations maintenance the row spaces. So

$$\mathrm{RSp}(\mathbf{A}) = \mathrm{RSp}(\mathbf{A}_{ech})$$

2) Let $i_1, ..., i_k$ be the numbers of the columns of $\mathbf{A}_{ech}$ containing the leading entries.

$$\begin{array}{c} \begin{array}{cccccc} i_1 = 1 & & i_2 = 3 & i_3 = 4 & & i_4 = 6 \end{array} \\ \begin{pmatrix} 1 & x & x & x & x & x \\ 0 & 0 & 1 & x & x & x \\ 0 & 0 & 0 & 1 & x & x \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \end{array}$$

Let $\mathbf{r}_1, ..., \mathbf{r}_k$ be the non-zero rows of $\mathbf{A}_{ech}$ Suppose also that

$$\lambda_1 \mathbf{r}_1 + \cdots + \lambda_k \mathbf{r}_k = \mathbf{0} \text{ for some } \lambda_i \in F$$

Consider the i_1^{th} entry of $\lambda_1 \mathbf{r}_1 + \cdots + \lambda_k \mathbf{r}_k$, this will be equal to the i_1^{th} entry of $\lambda_1 \mathbf{r}_1$ which is $\lambda_1 \cdot 1 = \lambda_1$. As

$$\lambda_1 \mathbf{r}_1 + \cdots + \lambda_k \mathbf{r}_k = 0$$

We must have $\lambda_1 = 0$. Now consider the i_2^{th} entry similarly, and conclude that $\lambda_2 = 0$. By continuing similarly we get $\lambda_1 = \cdots = \lambda_k = 0$ as required.

Example 3.8.4. Find the row rank of

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 0 \\ -1 & 4 & 15 \end{pmatrix}$$

Using row reductions, we get

$$\begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 0 \\ -1 & 4 & 15 \end{pmatrix}$$
$$\xrightarrow[R_2 \to R_2 - 2R_1]{R_3 \to R_3 + R_1}$$
$$\begin{pmatrix} 1 & 2 & 5 \\ 0 & -3 & -10 \\ 0 & 6 & 20 \end{pmatrix}$$
$$\xrightarrow[R_2 \to -\frac{R_2}{3}]{R_3 \to R_3 + 2R_2}$$
$$\begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & \frac{10}{3} \\ 0 & 0 & 0 \end{pmatrix}$$

So $\mathbf{A}_{ech}$ has 2 non-zero rows, so the row rank is 2.

Example 3.8.5. Find the dimension of

$$W = \text{Span}\{\begin{pmatrix} -1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 & 3 \end{pmatrix}\}$$

W is the row span of

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$
$$\xrightarrow{R_2 \to 2R_1} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$
$$\xrightarrow[R_3 \to R_2]{R_2 \to R_3} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 5 & 1 & 2 \end{pmatrix}$$
$$\xrightarrow{R_3 \to R_3 - 5R_2} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -9 & -13 \end{pmatrix}$$
$$\xrightarrow[R_1 \to -R_1]{R_3 \to -\frac{R_3}{9}} \begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & \frac{13}{9} \end{pmatrix}$$

Which is $\mathbf{A}_{ech}$.

We can find the column rank of a matrix in a similar way. One way is to simply use $\mathbf{A}^T$ and find the row rank of that matrix, or alternatively to use column operations.

Theorem 3.8.6. *For any matrix $\mathbf{A} \in M_{n\times m}(F)$, the row rank of $\mathbf{A}$ equals the column rank of $\mathbf{A}$.*

Proof:

$$\begin{array}{cc} & \mathbf{c_j} \\ \mathbf{A} = \ \mathbf{r_i} & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{ij} & \cdots & a_{1m} \\ \vdots & & & \vdots & & \vdots \\ a_{i1} & \cdots & \cdots & a_{ij} & \cdots & a_{im} \\ \vdots & & & \vdots & & \vdots \\ a_{n1} & \cdots & \cdots & a_{nj} & \cdots & a_{nm} \end{pmatrix} \end{array}$$

Let $\mathbf{A} = (a_{ij})_{n\times m}$. Let the rows be $\mathbf{r}_1, ..., \mathbf{r}_n$. with

$$\mathbf{r}_i = (a_{i1}, ..., a_{im})$$

Let the columns be $\mathbf{c}_1, ..., \mathbf{c}_m$ with

$$\mathbf{c}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}$$

Let k be the row rank of $\mathbf{A}$, the RSp($\mathbf{A}$) has a basis $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$. Every $\mathbf{r}_i$ is a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_k$ i.e.

$$\mathbf{r}_i = \lambda_{i1}\mathbf{v}_1 + \cdots + \lambda_{ik}\mathbf{v}_k$$

Suppose now that $\mathbf{v}_i = (b_{i1}, ..., b_{ij}, ..., b_{im})^T$. Consider the j^{th} element of

$$a_{ij} = \lambda_{i1}b_{1j} + \cdots + \lambda_{ik}b_{kj}$$

Then

$$\mathbf{c}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = \begin{pmatrix} \lambda_{11}b_{1j} + \cdots + \lambda_{1k}b_{kj} \\ \vdots \\ \lambda_{m1}b_{1j} + \cdots + \lambda_{mk}b_{kj} \end{pmatrix}$$

so

$$\mathbf{c}_j = \begin{pmatrix} \lambda_{11} \\ \vdots \\ \lambda_{m1} \end{pmatrix} b_{1j} + \cdots + \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{mk} \end{pmatrix} b_{kj}$$

$$\mathbf{c}_j \in \text{Span}\left\{ \begin{pmatrix} \lambda_{11} \\ \vdots \\ \lambda_{m1} \end{pmatrix}, ..., \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{mk} \end{pmatrix} \right\}$$

So the column rank of $\mathbf{A}$ is k. $\square$

Example 3.8.7. Let
$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 3 & -1 & 1 \end{pmatrix}$$
Note that $\mathbf{r}_3 = \mathbf{r}_1 + \mathbf{r}_2$ so $\{\mathbf{r}_1, \mathbf{r}_2\}$ is LI. So a basis for $\mathrm{RSp}(\mathbf{A})$ is $\{\begin{pmatrix}1 & 2 & -1 & 0\end{pmatrix}, \begin{pmatrix}-1 & 1 & 0 & 1\end{pmatrix}\}$. Next we write the rows as a linear combination of the above basis (the first being $\mathbf{v}_1$ and the second $\mathbf{v}_2$.)

$$\begin{aligned} \mathbf{r}_1 &= 1\mathbf{v}_1 + 0\mathbf{v}_2 \\ \mathbf{r}_2 &= 0\mathbf{v}_1 + 1\mathbf{v}_2 \\ \mathbf{r}_3 &= 1\mathbf{v}_1 + 1\mathbf{v}_2 \end{aligned}$$

So our λ_{ij} are

$$\begin{matrix} \lambda_{11} = 1 & \lambda_{12} = 0 \\ \lambda_{21} = 0 & \lambda_{22} = 1 \\ \lambda_{31} = 1 & \lambda_{32} = 1 \end{matrix}$$

According to the proof,
$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$
Is a spanning set for $\mathrm{CSp}(\mathbf{A})$.

Definition 3.8.8. Let $\mathbf{A}$ be a matrix. The *Rank of* $\mathbf{A}$ written $\mathrm{rank}(\mathbf{A})$ or $\mathrm{rk}(\mathbf{A})$ is the row rank (equivalently, the column rank) of $\mathbf{A}$.

Proposition 3.8.9. Let $\mathbf{A} \in M_{n\times n}(F)$. The following are equivalent.

1. $\mathrm{rk}(\mathbf{A}) = n$
2. The rows of $\mathbf{A}$ form a basis for F^n.
3. The columns of $\mathbf{A}$ form a basis for F^n.
4. $\mathbf{A}$ is invertible.

Proof:

- 1. $\iff$ 2.

$$\begin{aligned} \mathrm{rk}(\mathbf{A}) = n &\iff \dim(\mathrm{RSp}(\mathbf{A})) = n \\ &\iff \mathrm{RSp}(\mathbf{A}) = F^n \\ &\iff \text{The } \lambda \text{ rows of } \mathbf{A} \text{ Span } F^n \\ &\iff \text{The} n \text{ rows of } \mathbf{A} \text{ form a basis for } F^n \end{aligned}$$

- 1. $\iff$ 3.: Do the same as above, but replace RSp with CSp.

- 1. $\implies$ 4.:

$$\text{rk}(\mathbf{A}) = n \iff \mathbf{A}_{ech} = \begin{pmatrix} 1 & * & * & \cdots & \cdots \\ 0 & 1 & * & \cdots & \cdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

  We can eliminate the $*$ entries with row ops. So $\mathbf{A}$ is reducible to $\mathbf{I}_n$ and therefore invertible.

- 4 $\implies$ 1: If $\mathbf{A}$ is invertible, then $\mathbf{A}$ is reducible to $\mathbf{I}_n$ and by definition, the rank of $\mathbf{A}$ is n.

4 Linear transformations

4.1 Introduction

Definition 4.1.1. Suppose V, W are vector spaces over F, and $T : V \to W$ a function. We say

- T *preserves addition* if

$$\forall \mathbf{v}_1, \mathbf{v}_2 \in V, T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$$

- T *preserves scalar multiplication* if

$$\forall \mathbf{v} \in V, \lambda \in F, T(\lambda \mathbf{v}) = \lambda T(\mathbf{v})$$

- T is a linear transformation if it does both of these things.

There are several different names for linear transformations, e.g. Linear maps, linear operators and just operators.

Example 4.1.2. Here are a few examples of different linear transformations (or non-examples). With proof of their linearity (or non-linearity).

1. The identity map of any VS. This is obviously linear.

2.

$$\begin{aligned} &T : \mathbb{R}^2 \to \mathbb{R} \\ &T(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y} \end{aligned}$$

Preserves addition: let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2, \mathbf{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Then

$$\begin{aligned}
T(\mathbf{v}_1 + \mathbf{v}_2) &= T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) \\
&= T\left(\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}\right) \\
&= (x_1 + x_2) + (y_1 + y_2) \\
&= (x_1 + y_1) + (x_2 + y_2) \\
&= T(\mathbf{v}_1) + T(\mathbf{v}_2)
\end{aligned}$$

Preserves scalar multiplication: Let $\mathbf{v} \in \mathbb{R}^2, \lambda \in \mathbb{R}, \mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$

$$\begin{aligned}
T(\lambda \mathbf{v}) &= T\left(\lambda \begin{pmatrix} x \\ y \end{pmatrix}\right) \\
&= T\left(\begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}\right) \\
&= \lambda x + \lambda y \\
&= \lambda(x + y) = \lambda T(\mathbf{v})
\end{aligned}$$

3.

$$\begin{aligned}
V = \mathbb{R}[x], T : \mathbb{R}[x] &\to \mathbb{R}[x] \\
T(f(x)) &= \frac{\mathrm{d}}{\mathrm{d}x} f(x)
\end{aligned}$$

Preserves addition: let $f(x), g(x) \in \mathbb{R}[x]$. Then

$$\begin{aligned}
T(f(x) + g(x)) &= \frac{\mathrm{d}}{\mathrm{d}x}(f(x) + g(x)) \\
&= \frac{\mathrm{d}}{\mathrm{d}x} f(x) + \frac{\mathrm{d}}{\mathrm{d}x} g(x) \\
&= T(f(x)) + T(g(x))
\end{aligned}$$

Preserves multiplication: let $f(x) \in \mathbb{R}[x], \lambda \in \mathbb{R}$. Then

$$\begin{aligned}
T(\lambda f(x)) &= \frac{\mathrm{d}}{\mathrm{d}x}(\lambda f(x)) \\
&= \lambda\left(\frac{\mathrm{d}}{\mathrm{d}x} f(x)\right) \\
&= \lambda T(f(x))
\end{aligned}$$

So differentiation is linear.

4. $\mathbb{C}$ as a 1-dim vector space over $\mathbb{C}, T(\mathbf{z}) = \overline{\mathbf{z}}$ *Counterexample:* Let $\mathbf{z}, \boldsymbol{\lambda} \in \mathbb{C}$. Then

$$\begin{aligned} T(\boldsymbol{\lambda}\mathbf{z}) &= \overline{\boldsymbol{\lambda}\mathbf{z}} \\ &= \bar{\boldsymbol{\lambda}}\bar{\mathbf{z}} \\ &\neq \boldsymbol{\lambda}\bar{\mathbf{z}} \text{ if } \boldsymbol{\lambda} \notin \mathbb{R} \end{aligned}$$

Proposition 4.1.3. Let $\mathbf{A} \in M_{m \times n}(F)$. Define

$$\begin{aligned} T &: F^n \to F^m \\ T(\mathbf{v}) &= \mathbf{A}\mathbf{v} \end{aligned}$$

Then T is a linear transformation.

Proof: See 2.7.4

Proposition 4.1.4.

(a) Define

$$T : \mathbb{R}^3 \to \mathbb{R}^2$$
$$T\left(\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}\right) = \begin{pmatrix} a_1 - 3a_2 + a_3 \\ a_1 + a_2 - 2a_3 \end{pmatrix}$$

Then T is linear by 4.1.3 as

$$T\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 & -3 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

(b) define $\rho_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ to be the anticlockwise rotation through angle θ about the origin. Then

$$\rho_\theta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

So ρ_θ is linear.

Proposition 4.1.5. Let $T : V \to W$ be a linear transformation. V, W vector spaces over F, $\mathbf{0}_V, \mathbf{0}_W$ zeros of V and W respectively. Then

1. $T(\mathbf{0}_V) = \mathbf{0}_W$
2. If $\mathbf{v} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$, then $T(\mathbf{v}) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_k T(\mathbf{v}_k)$

Proof:

1. T preserves scalar multiplication. So

$$\begin{aligned} T(0 \cdot \mathbf{0}_V) &= 0 \cdot T(\mathbf{0}_V)(\in W) && 0 \in F \\ &= 0 \cdot T(\mathbf{0}_W) \end{aligned}$$

   But

$$\begin{aligned} 0 \cdot \mathbf{0}_V &= \mathbf{0}_V \\ 0 \cdot \mathbf{w} &= \mathbf{0}_W && \forall \mathbf{w} \in W \end{aligned}$$

   so

$$T(\mathbf{0}_V) = \mathbf{0}_W$$

2. Induction on k:
   Base case where k=1:

$$T(\lambda_1 \mathbf{v}_1) = \lambda_1 T(\mathbf{v}_1)$$

   As T preserves scalar multiplication. Now for the inductive step, suppose we know

$$T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k) = \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_k T(\mathbf{v}_k)$$

   for any $\lambda_1, ..., \lambda_k \in F, \mathbf{v}_1, ..., \mathbf{v}_k \in V$. Now consider

$$T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_{k+1} \mathbf{v}_{k+1}) = T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k) + T(\lambda_{k+1} \mathbf{v}_{k+1})$$

   And by our inductive hypothesis

$$= \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_k T(\mathbf{v}_k) + \lambda_{k+1} T(\mathbf{v}_{k+1})$$

   So by induction part 2 is true

$\square$

Example 4.1.6. Find a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that

$$T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

and

$$T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$$

Since $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ forms a basis of $\mathbb{R}^2$ and

$$\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

we should define

$$
\begin{aligned}
T\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) &= aT(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= a \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} a \\ -a+b \\ 2a+3b \end{pmatrix}
\end{aligned}
$$

Then this is a linear transformation as it can be expressed as a matrix

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

Proposition 4.1.7. Let V, W be vector spaces over F. Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be a basis for V and let $\mathbf{w}_1, ..., \mathbf{w}_n$ be any vectors in W. Then there exists a *unique* linear transformation $T : V \to W$ with $T(\mathbf{v}_i) = \mathbf{w}_i, i = 1, ..., n.$

Proof: First define T: Let $\mathbf{v} \in V$. As $\mathbf{v}_1, ..., \mathbf{v}_n$ is a basis for V, there exist unique $\lambda_1, ..., \lambda_n \in F$ with $\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$. Define

$$
T(\mathbf{v}) = \lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n
$$

So

$$
T(\mathbf{v}_i) = \mathbf{w}_i
$$

As required. Now we need to show T is linear. Let $\mathbf{u}, \mathbf{v} \in V$. Write

$$
\begin{aligned}
\mathbf{v} &= \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n & \lambda_i \in F, \forall i = 1, ..., n \\
\mathbf{u} &= \mu_1 \mathbf{v}_1 + \cdots + \mu_n \mathbf{v}_n & \mu_i \in F, \forall i = 1, ..., n
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathbf{u} + \mathbf{v} &= (\mu_1 + \lambda_1)\mathbf{v}_1 + \cdots + (\mu_n + \lambda_n)\mathbf{v}_n \\
T(\mathbf{u} + \mathbf{v}) &= (\mu_1 + \lambda_1)\mathbf{w}_1 + \cdots + (\mu_n + \lambda_n)\mathbf{w}_n \\
&= (\lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n) + (\mu_1 \mathbf{w}_1 + \cdots + \mu_n \mathbf{w}_n) \\
&= T(\mathbf{u}) + T(\mathbf{v})
\end{aligned}
$$

Similarly, for $\alpha \in F, \mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$,

$$
\begin{aligned}
\alpha \mathbf{v} &= \alpha \cdot \lambda_1 \mathbf{v}_1 + \cdots + \alpha \cdot \lambda_n \mathbf{v}_n \\
T(\alpha \mathbf{v}) &= \alpha \cdot \lambda_1 \mathbf{w}_1 + \cdots + \alpha \cdot \lambda_n \mathbf{w}_n \\
&= \alpha(\lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n) \\
&= \alpha T(\mathbf{v})
\end{aligned}
$$

So T is linear. Now to prove uniqueness, suppose instead there exists $S : V \to W$ a linear transformation with $S(\mathbf{v}_i) = \mathbf{w}_i, \forall i = 1, ..., n$. Let

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n \in V \qquad \lambda_i \in F, \forall i = 1, ..., n$$

Then

$$S(\mathbf{v}) = \lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n = T(\mathbf{v})$$

□

Remark 4.1.8. This also shows that linear transformations are determined by what they do to a basis.

Example 4.1.9. Let V be $\mathbb{R}[x]$ with degree less than or equal to 2. A basis for this is then $\{1, x, x^2\}$. Consider

$$\begin{aligned} \mathbf{w}_1 &= 1 + x \\ \mathbf{w}_2 &= x - x^2 \\ \mathbf{w}_3 &= 1 + x^2 \end{aligned}$$

Then by 4.1.7 there is a unique linear transformation

$$T : V \to V$$

With

$$\begin{aligned} T(1) &= 1 + x \\ T(x) &= x - x^2 \\ T(x^2) &= 1 + x^2 \end{aligned}$$

If $v = a + bx + cx^2$ then

$$\begin{aligned} T(\mathbf{v}) &= aT(1) + bT(x) + cT(x^2) \\ &= (a + c) + (a + b)x + (-b + c)x^2 \end{aligned}$$

4.2 Image and Kernel

Definition 4.2.1. Suppose $T : V \to W$ is a linear transformation. Then

- The *Image* of T is the set

$$\mathrm{Im}(T) = \{T(\mathbf{v}) : \mathbf{v} \in V\} \subset W$$

- the *Kernel* of T is the set

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = 0\} \subset V$$

Example 4.2.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 2 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

So the image of T is

$$\begin{aligned} \operatorname{Im}(T) &= \left\{ \begin{pmatrix} 3x_1 + x_2 + 2x_3 \\ -x_1 + x_3 \end{pmatrix} : x_1, x_2, x_3 \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 3 \\ -1 \end{pmatrix} x_1 + \begin{pmatrix} 1 \\ 0 \end{pmatrix} x_2 + \begin{pmatrix} 2 \\ 1 \end{pmatrix} x_2 : x_1, x_2, x_3 \in \mathbb{R} \right\} \\ &= \operatorname{CSp}(A) = \mathbb{R} \end{aligned}$$

And the kernel of T is

$$\begin{aligned} \operatorname{Ker}(T) &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} 3x_1 + x_2 + 2x_3 \\ -x_1 + x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \\ &= \operatorname{Span} \begin{pmatrix} 1 \\ -5 \\ 1 \end{pmatrix} \end{aligned}$$

Proposition 4.2.3. Let $T : V \to W$ be a linear transformation. Then

1. $\operatorname{Im}(T)$ is a subspace of W
2. $\operatorname{Ker}(T)$ is a subspace of V

Proof: 1: As $\mathbf{0}_V \in V, T(\mathbf{0}_V) \in \operatorname{Im}(T)$. So $\operatorname{Im}(T)$ isn't empty. Let $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{Im}(T)$ so there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ with $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. So

$$\begin{aligned} T(\mathbf{v}_1 + \mathbf{v}_2) &= T(\mathbf{v}_1)_T(\mathbf{v}_2) \\ &= \mathbf{w}_1 + \mathbf{w}_2 \\ &\therefore \mathbf{w}_1 + \mathbf{w}_2 \in \operatorname{Im}(T) \end{aligned}$$

Likewise if $\alpha \in F$

$$\begin{aligned} T(\alpha \mathbf{v}_1) &= \alpha T(\mathbf{v}_1) \\ &= \alpha \mathbf{w}_1 \\ &\therefore \alpha \mathbf{w}_1 \in \operatorname{Im}(T) \end{aligned}$$

Hence $\operatorname{Im}(T)$ is a subspace of W.

2: Left as an exercise (Very similar to part 1, do the subspace test!)

Example 4.2.4. Let V_n be the vector space of polynomials in x over $\mathbb{R}$ of degree less than n. We have

$$V_0 \leq V_1 \leq \cdots \leq V_n$$

Define

$$\begin{aligned} T : V_n &\to V_{n-1} \\ T(f(x)) &= f'(x) \end{aligned}$$

And note T is linear.

$$\begin{aligned} \mathrm{Ker}(T) &= \{f(x) : f'(x) = 0\} \\ &= \text{constant polynomials} \\ &= V_0 \end{aligned}$$

Suppose $g(x) \in V_{n-1}$, then by integrating we can find $f(x) \in V_n$ such that $f'(x) = g(x)$. So

$$\begin{aligned} T(f(x)) &= g(x) \in \mathrm{Im}(T) \\ \mathrm{Im}(T) &= V_{n-1} \end{aligned}$$

Note: for $c \in V_0, T(f(x) + c) = g(x)$. In fact, the set

$$\{h(x) : h'(x) = g(x)\} = \{f(x) + s(x) : s(x) \in \mathrm{Ker}(T)\}$$

Proposition 4.2.5. Let $T : V \to W$ be a linear transformation. Let $\mathbf{v}_1, \mathbf{v}_2 \in V$, then $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ iff $\mathbf{v}_1 - \mathbf{v}_2 \in \mathrm{Ker}(T)$.

Proof: $T(\mathbf{v}_1) = T(\mathbf{v}_2) \iff T(\mathbf{v}_1) - T(\mathbf{v}_2) = \mathbf{0}_W \iff T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}_W \iff \mathbf{v}_1 - \mathbf{v}_2 \in \mathrm{Ker}(T)$ □

Proposition 4.2.6. Let $T : V \to W$ be a linear transformation. Suppose $\mathbf{v}_1, ..., \mathbf{v}_n$ is a basis for V. Then $\mathrm{Im}(T) = \mathrm{Span}\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$.

Proof: It's clear that $\mathrm{Span}\{T(v_1), ..., T(v_n)\} \subset \mathrm{Im}(T)$. Now let $\mathbf{w} \in \mathrm{Im}(T)$. Then there is $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. Now, as $\mathbf{v} \in V$ there are $\lambda_1, ..., \lambda_n \in F$ such that

$$\begin{aligned} \mathbf{v} &= \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n \\ T(\mathbf{v}) &= T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n) \\ &= \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n) \\ \mathbf{w} &= \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_n T(\mathbf{v}_n) \in \mathrm{Span}\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\} \end{aligned}$$

□

Proposition 4.2.7. Let $\mathbf{A} \in M_{m \times n}(F)$. Let

$$\begin{aligned} T : F^n &\to F^m \\ T(\mathbf{v}) &= \mathbf{A}\mathbf{v} \end{aligned}$$

Then

1. $\mathrm{Ker}T$ is the solution space for $\mathbf{Av} = 0$.
2. $\mathrm{Im}T$ is the column space of $\mathbf{A}$.
3. $\dim(\mathrm{Im}(T)) = \mathrm{rank}\mathbf{A}$

Proof:

1. Immediate from definitions.
2. Take the standard basis for F^n i.e.:

$$\hat{\mathbf{e}}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

with the 1 in the i^{th} row. Then by 4.2.6 we have $\mathrm{Im}(T) = \mathrm{Span}\{T(\hat{\mathbf{e}}_1), ..., T(\hat{\mathbf{e}}_n)\}$

$$T\hat{\mathbf{e}}_i = \mathbf{A}\hat{\mathbf{e}}_i = \mathbf{c}_i$$

where $\mathbf{c}_i$ is the i^{th} column of $\mathbf{A}$. So

$$\begin{aligned} \mathrm{Im}(T) &= \mathrm{Span}\{\mathbf{c}_1, ..., \mathbf{c}_n\} \\ &= \mathrm{CSp}(\mathbf{A}) \end{aligned}$$

3.

$$\begin{aligned} \mathrm{Im}(T) &= \dim(\mathrm{CSp}(\mathbf{A})) \\ &= \mathrm{rank}(\mathbf{A}) \end{aligned}$$

□

Theorem 4.2.8. Rank-Nullity Theorem*: Let $T : V \to W$ be a linear transformation. Then*

$$\dim(V) = \dim(\mathrm{Im}T) + \dim(\mathrm{Ker}T)$$

Proof: Let $\{\mathbf{u}_1, ..., \mathbf{u}_s\}$ be a basis for $\mathrm{Ker}T$. And let $\{\mathbf{w}_1, ..., \mathbf{w}_r\}$ be a basis for $\mathrm{Im}T$. Then for each $\mathbf{w}_i \in \mathrm{Im}T$, there is a $\mathbf{v}_i \in V$ with $T(\mathbf{v}_i) = \mathbf{w}_i$. We first claim that $B = \{\mathbf{u}_1, ..., \mathbf{u}_s\} \cup \{\mathbf{v}_1, ..., \mathbf{v}_r\}$ is a basis for V.

Proof of claim:

1. Spanning set: Let $\mathbf{v} \in V$, since $T(v) \in \mathrm{Im}(T)$ we have

$$T(\mathbf{v}) = \lambda_1\mathbf{w}_1 + \cdots + \lambda_r\mathbf{w}_r$$

For some $\lambda_1, ..., \lambda_r \in F$

$$\begin{aligned} &= \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_r T(\mathbf{v}_r) \\ &= T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r) \end{aligned}$$

as T is a linear transformation. Then by 4.2.5

$$\mathbf{v} - (\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r) \in \mathrm{Ker}(T)$$

so

$$\mathbf{v} - (\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r) = \mu_1 \mathbf{u}_1 + \cdots + \mu_s \mathbf{u}_s$$

for some $\mu_1, ..., \mu_s \in F$. So

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r + \mu_1 \mathbf{u}_1 + \cdots + \mu_s \mathbf{u}_s \in \mathrm{Span}(B)$$

2. Linear Independence: Suppose

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r + \mu_1 \mathbf{u}_1 + \cdots + \mu_s \mathbf{u}_s = \mathbf{0}_V$$

Apply T to this equation:

$$\begin{aligned} T(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_r \mathbf{v}_r + \mu_1 \mathbf{u}_1 + \cdots + \mu_s \mathbf{u}_s) &= T(\mathbf{0}_V) \\ \lambda_1 T(\mathbf{v}_1) + \cdots + \lambda_r T(\mathbf{v}_r) + \mu_1 T(\mathbf{u}_1) + \cdots + \mu_s T(\mathbf{u}_s) &= \mathbf{0}_W \end{aligned}$$

And since $T(\mathbf{v}_i) = \mathbf{w}_i, T(\mathbf{u}_i) = \mathbf{0}_W, \forall i,$

$$\lambda_1 \mathbf{w}_1 + \cdots + \lambda_r \mathbf{w}_r = \mathbf{0}_W$$

As $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is a basis for $\mathrm{Im}(T)$,

$$\begin{aligned} \lambda_1 = \lambda_2 = \cdots = \lambda_r &= 0 \\ \mu_1 \mathbf{u}_1 + \cdots + \mu_s \mathbf{u}_s &= \mathbf{0}_V \end{aligned}$$

Since $\{\mathbf{u}_1, ..., \mathbf{u}_s\}$ is a basis for $\mathrm{Ker}(T)$,

$$\mu_1 = \cdots = \mu_s = 0$$

So B is linearly independent.

$\square$

Corollary. *A system of linear equations in n unknowns with coefficients in F is called* homogenous *if all equations are equal to zero. We can represent this as $\mathbf{Ax} = \mathbf{0}_{F^m}$. We know we will always get at least a trivial solution i.e. $\mathbf{x} = \mathbf{0}_{F^n}$. We saw in the mid-module tests that the solution space is a subspace... but of what dimension?*

We can use the rank-nullity theorem: using $\mathbf{T} : F^n \to F^m$, represented by the matrix $\mathbf{A}$. By 4.2.7 the solution space is $\mathrm{Ker}(\mathbf{A})$ and by the rank-nullity theorem

$$\dim(\mathrm{Ker}(\mathbf{A})) = \dim(F^n) - \dim(\mathrm{Im}(\mathbf{A}))$$

And if $\mathrm{rank}(\mathbf{A}) = n$ then we get one solution (the trivial one). If $\mathrm{rank}(\mathbf{A}) < n$, then the solution space has $\dim \mathbf{A} \geq 1$. If F is infinite, then you get infinitely many solutions.

4.3 Representing vectors and transformations with respect to a basis

Let V be n-dimensional F-vector space and denote the basis of F by $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$.

Definition 4.3.1. For $\mathbf{v} \in V$ such that $\mathbf{v} = \lambda_1 zvv_1 + \ldots + \lambda_n \mathbf{v}_n$, the vector of $\mathbf{v}$ with respect to B is

$$[\mathbf{v}]_B = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

Observe simply that this is well defined by the linear independence of basis B.

Example 4.3.2. $V = \mathbb{R}^3$, $B = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$. Then:

$$\left[\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

Example 4.3.3. Let V be a vector space of polynomials with degree less than 2. Let $B = \{1, x, x^2\}$. Then:

$$[a + bx + cx^2]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

If $B = \{x^2, x, 1\}$ then

$$[a + bx + cx^2]_B = \begin{pmatrix} c \\ b \\ a \end{pmatrix}$$

Proposition 4.3.4. Let V be an n-dimensional F-vector space with basis B. Then the map $T : V \to F^n$ such that $T(\mathbf{v}) = [\mathbf{v}]_B$ is bijective and linear.

Proof: WLOG, denote $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$.

1. Linear transformation

- Preserves addition

Let $\mathbf{u}, \mathbf{v} \in V$ and represent $\mathbf{u}, \mathbf{v}$ as such:

$$u = \lambda_1 \mathbf{b}_1 + \ldots + \lambda_n \mathbf{b}_n$$

$$\mathbf{v} = \mu_1 \mathbf{b}_1 + \ldots + \mu_n \mathbf{b}_n.$$

$$[\mathbf{u}]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, [\mathbf{v}]_B = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, [\mathbf{u} + \mathbf{v}]_B = \begin{pmatrix} \lambda_1 + \mu_1 \\ \vdots \\ \lambda_n + \mu_n \end{pmatrix}$$

$$[\mathbf{u}+\mathbf{v}]_B = [\mathbf{u}]_B + [\mathbf{v}]_B.$$

i.e.

$$T(\mathbf{u}+\mathbf{v}) = [\mathbf{u}+\mathbf{v}]_B = [\mathbf{u}]_B + [\mathbf{v}]_B = T(\mathbf{u}) + T(\mathbf{v}).$$

- Preserves scalar multiplication

Similar to checking that addition is well defined: let $\mathbf{v} = \lambda_1\mathbf{b}_1 + ... + \lambda_n\mathbf{b}_n$. Then we have

$$\alpha[\mathbf{v}]_B = \begin{pmatrix} \alpha\lambda_1 \\ \vdots \\ \alpha\lambda_n \end{pmatrix} = [\alpha\mathbf{v}]_B$$

giving us

$$T(\alpha\mathbf{v}) = [\alpha\mathbf{v}]_b = \alpha[\mathbf{v}]_B = \alpha T(\mathbf{v})$$

2. Bijective

- Injective

Let $\mathbf{u}, \mathbf{v}$ such that $T(\mathbf{u}) = T(\mathbf{v})$, implying $T(\mathbf{u}-\mathbf{v}) = 0$. so

$$[\mathbf{u}-\mathbf{v}]_B = 0_{F^n} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{aligned} \mathbf{u}-\mathbf{v} &= 0b_1 + \ldots + 0b_n \\ &= 0 \end{aligned}$$

so $u = v$ and T is injective.

- Surjective

Let $\mathbf{v} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in F^n$. Then clearly $[\alpha_1 b_1 + \ldots + \alpha_n b_n]_B = \mathbf{v}$ so $T(\alpha_1 b_1 + \ldots + \alpha_n b_n) = \mathbf{v}$ so T is surjective. $\square$

Construction

Let V, W be finite-dimension F-vector spaces with bases B and C respectively. Let $T : V \to W$ be a linear transformation. We want to construct a map $\varphi : F^n \mapsto F^m$ to give rise to the following commutative diagram:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, leftrightarrow] & W \arrow[d, leftrightarrow] \\ F^n \arrow[r, "\varphi"] & F^m \end{tikzcd}$$

φ is a linear transformation since the composition of two linear transformations is a linear transformation itself. By our hand in, $\varphi : F^n \mapsto F^m$ is a matrix transformation (coursework 1). Let $\mathbf{A}$ be this matrix. Then

$$\mathbf{A}[\mathbf{v}]_B = [T\mathbf{v}]_C$$

We calculate $\mathbf{A}$ by figuring out its columns $\gamma_1, \ldots, \gamma_n$. To calculate γ_1 we work out $T(\mathbf{b}_i) = a_{1i}\mathbf{c}_1 + \ldots + a_{mi}\mathbf{c}_m$ so

$$\gamma_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}$$

Definition 4.3.5. The matrix constructed above is the *matrix of* T *with respect to* B *and* C. We write ${}_C[T]_B$. So

$${}_C[T]_B[\mathbf{v}]_B = [T(\mathbf{v})]_C$$

Proposition 4.3.6. If $T : V \to V$ and B is a basis for V then for all $\mathbf{v} \in V, [T(\mathbf{v})]_B = [T]_B[\mathbf{v}]_B$.

Example 4.3.7.

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$
$$T(x_1, x_2) = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{pmatrix}$$

- Take $E = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. Find $[T]_E$.

$$T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -1\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$[T]_E = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$$

- Let $B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. Find $[T]_B$.

$$T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 3\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$[T]_B = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$$

- Find ${}_B[T]_E$

$$T\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix} = 2\begin{pmatrix}1\\1\end{pmatrix} - 1\begin{pmatrix}0\\1\end{pmatrix}$$
$$T\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-1\\2\end{pmatrix} = -1\begin{pmatrix}1\\1\end{pmatrix} + 3\begin{pmatrix}0\\1\end{pmatrix}$$
$${}_B[T]_E = \begin{pmatrix}2 & -1\\-1 & 3\end{pmatrix}$$

Proposition 4.3.8. Let V be a vector space over F.

$$\left.\begin{matrix}B = \{\mathbf{v}_1, ..., \mathbf{v}_n\}\\ C = \{\mathbf{w}_1, ..., \mathbf{w}_n\}\end{matrix}\right\}\text{bases for } V$$

Then for $j \in [1, ..., n]$,

$$\mathbf{v}_j = \lambda_{1j}\mathbf{w}_1 + \cdots \lambda_{nj}\mathbf{w}_n$$

Let $\mathbf{P}$ be the matrix $[\lambda_{ij}]_{n\times n}$ so the j^{th} column is $[\mathbf{v}_j]_C$. Then

- $\mathbf{P} = [X]_C$ where $X : V \to V$ is the unique linear transformation such that $X(\mathbf{w}_j) = \mathbf{v}_j$ for all j.
- For all $\mathbf{v} \in V, \mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C$.
- $\mathbf{P} =_C [Id]_B$ where $Id : V \to V$ is the identity transformation.

Proof:

- the j^{th} column of $[X]_C$ is the image of $X(\mathbf{w}_j)$ written as a vector in C. Now $X(\mathbf{w}_j) = \mathbf{v}_j$, so the j^{th} column is $[\mathbf{v}_j]_C$ which is the j^{th} column of $\mathbf{P}$. Thus $[X]_C = \mathbf{P}$.
- For a basis vector $\mathbf{v}_j \in B$, we have

$$\begin{aligned}\mathbf{P}[\mathbf{v}_j]_B &= \mathbf{P}_{ej}\\ &= j^{th} \text{ column of } \mathbf{P}\\ &= [\mathbf{v}_j]_C\end{aligned}$$

so this is true for elements of the basis B - hence is true for all $\mathbf{v} \in V$ $\square$

Definition 4.3.9. $\mathbf{P}$ is the *change of basis matrix* from B to C Warning: This is confusing because of Part 1 of Proposition 4.3.8 i.e $= [X]_C$ where $X(\mathbf{w}_j) = \mathbf{v}_j$ with $\mathbf{w}_j \in C$ and $\mathbf{v}_j \in B$. Be careful when reading about this!

Proposition 4.3.10. Let V, B, C and $\mathbf{P}$ be as above.

- $\mathbf{P}$ is invertible and its inverse is the change of basis matrix from C to B
- Let $T : V \to V$ be a linear transformation, then $[T]_C = P[T]_B\mathbf{P}^{-1}$

Proof:

- Let $\mathbf{Q}$ be the change of basis matrix from C to B.

$$\mathbf{Q}[\mathbf{v}]_C = [\mathbf{v}]_B \qquad \forall\, \mathbf{v} \in V$$

$$\mathbf{P}[\mathbf{v}]_B = [\mathbf{v}]_C \qquad \forall\, \mathbf{v} \in V$$

  Hence:

$$\mathbf{QP}[\mathbf{v}]_B = Q[\mathbf{v}]_C = [\mathbf{v}]_B$$

  As $\mathbf{v}$ ranges over V, we have that $[\mathbf{v}]_B$ ranges over F^n
  Hence we get $\mathbf{QPx} = \mathbf{x}\ \forall \mathbf{x} \in F^n$ and similarly, $\mathbf{PQx} = \mathbf{x}\ \forall \mathbf{x} \in F^n$.
  Therefore $\mathbf{QP} = \mathbf{PQ} = \mathbf{I}_n$ and we get $\mathbf{Q} = \mathbf{P}^{-1}$

- Take a vector $[\mathbf{v}]_C \in F^n$:

$$\begin{aligned}
[T]_C[\mathbf{v}]_C &= [T(\mathbf{v})]_C \\
(P[T]_B\mathbf{P}^{-1})[\mathbf{v}]_C &= (\mathbf{P}[T]_B\mathbf{P}^{-1})(\mathbf{P}[\mathbf{v}]_B) \\
&= (\mathbf{P}[T]_B)(\mathbf{P}^{-1}\mathbf{P})([\mathbf{v}]_B) \\
&= \mathbf{P}[T]_B[\mathbf{v}]_B \\
&= \mathbf{P}[T(\mathbf{v})]_B \\
&= [T(\mathbf{v})]_C
\end{aligned}$$

  As this holds $\forall\, \mathbf{v} \in V$, we get that $[T]_C = \mathbf{P}[T]_B\mathbf{P}^{-1}$

Example 4.3.11.

$$V = \mathbb{R}^2 \qquad T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -2x_1 + 3x_2 \end{pmatrix}$$

$$B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\} \qquad E = \{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2\}$$

Calculate $[T]_B$ and $\mathbf{P}$, the change of basis matrix from E to B and verify that:

$$[T]_E = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$$

For $[T]_B$, we have:

$$\begin{aligned}
T\begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0\begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
T\begin{pmatrix} 1 \\ 2 \end{pmatrix} &= \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 0\begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
[T]_B &= \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}
\end{aligned}$$

For $\mathbf{P}$, we want to take a vector from B and end with a vector in E by multiplication by $\mathbf{P}$. We can make the task easier by instead going from E to B through $\mathbf{P}^{-1}$ in the following way:

$$\begin{aligned}
P^{-1}\begin{pmatrix}1\\0\end{pmatrix} &= \begin{pmatrix}1\\1\end{pmatrix} = 1\begin{pmatrix}1\\0\end{pmatrix} + 1\begin{pmatrix}0\\1\end{pmatrix}\\
P^{-1}\begin{pmatrix}0\\1\end{pmatrix} &= \begin{pmatrix}1\\2\end{pmatrix} = 1\begin{pmatrix}1\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\end{pmatrix}\\
P^{-1} &= \begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}
\end{aligned}$$

We can then use this to calculate $\mathbf{P}$

Remark 4.3.12. In fact, if $\mathbf{P}$, the change in basis matrix from B to C is ${}_C[Id]_B$ and $\mathbf{Q}$ is the change of basis matrix from C to D (where D is also a basis for V) then:

$$\begin{aligned}
\mathbf{QP} &=_D [Id]_{CC}[Id]_B\\
&=_D [Id]_B
\end{aligned}$$

And $\mathbf{QP}$ is the change of basis matrix from B to D.

It is easier to find for a given basis B the matrix ${}_E[Id]_B$ where E is the standard basis. So an easy way to calculate ${}_C[Id]_B$ is to do this:

$$\begin{aligned}
{}_C[Id]_B &=_C [Id]_{EE}[Id]_B\\
&= ({}_E[Id]_C)_E^{-1}[Id]_B
\end{aligned}$$

This gives us a quick method to calculate change of bases matrices.

5 Determinants

5.1 Definitions and some properties

Definition 5.1.1.

- Notation: F a field (e.g. $\mathbb{R}, \mathbb{C}, \mathbb{Q}$, etc).
- $n \in \mathbb{N} = \{1, 2, ...\}$
- $M_n(F)$ is the set of $n \times n$ matrices in F.
- $\mathbf{A} \in M_n(F)$ write entries as $\mathbf{A} = (a_{ij})$.

Definition 5.1.2. If $\mathbf{A} \in M_n(F), 1 \leq i, j \leq n$, let $\mathbf{A}_{ij}$ denote the $(n-1)\times(n-1)$ matrix obtained by deleting row i and column j from $\mathbf{A}$. This is the *ij-minor of* $\mathbf{A}$

E.g.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

$$\mathbf{A}_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}$$

Definition 5.1.3. Let $\mathbf{A} = (a_{ij}) \in M_n(F)$. Define $\det(\mathbf{A})$ *the determinant* of $\mathbf{A}$ inductively on n.

i $n = 1 : \det(\mathbf{A}) = a_{11}$

ii $n = 2 :$

$$\begin{aligned} &\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\ &= a_{11}a_{22} - a_{12}a_{21} \\ &= a_{11} \det(\mathbf{A}_{11}) - a_{12} \det(\mathbf{A}_{12}) \end{aligned}$$

iii $n = 3 :$

$$\det(\mathbf{A}) = a_{11} \det(\mathbf{A}_{11}) - a_{12} \det(\mathbf{A}_{12}) + a_{13} \det(\mathbf{A}_{13})$$

iv General n. Suppose we have defined the determinant of $(n-1) \times (n-1)$ matrices of $\mathbf{A} = (a_{ij}) \in M_n(F)$.

$$\begin{aligned} \det(\mathbf{A}) &= a_{11} \det(\mathbf{A}_{11}) - a_{12} \det(\mathbf{A}_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(\mathbf{A}_{1n}) \\ &= \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(\mathbf{A}_{1j}) \end{aligned}$$

We can also write $\det(\mathbf{A}) = |\mathbf{A}|$.

E.g.:

$$\begin{aligned} \det(\mathbf{A}) &= \begin{vmatrix} 1 & 2 & 0 & 1 \\ 2 & 0 & -1 & 1 \\ -1 & 2 & 1 & 0 \\ 1 & 0 & -2 & 1 \end{vmatrix} \\ &= \begin{vmatrix} 0 & -1 & 1 \\ 2 & 1 & 0 \\ 0 & -2 & 1 \end{vmatrix} - 2 \begin{vmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{vmatrix} + 0 - 1 \begin{vmatrix} 2 & 0 & -1 \\ -1 & 2 & 1 \\ 1 & 0 & -2 \end{vmatrix} \end{aligned}$$

Theorem 5.1.4. *Let $\mathbf{A} \in M_n(F)$ and $\alpha \in F$. Let $\leq l \leq n$ and let $\mathbf{B}$ be the matrix obtained by multiplying row l of $\mathbf{A}$ by α. Then*

$$\det(\mathbf{B}) = \alpha \det(\mathbf{A})$$

Proof: Case $l = 1$: The ij entry of $\mathbf{B}$ is α_{ij} of

$$\mathbf{A}_{ij} = \mathbf{B}_{ij}$$

So by definition,

$$\begin{aligned}\det(\mathbf{B}) &= \sum_{j=1}^{n} (-1)^{j+1} \alpha a_{1j} \det(\mathbf{A}_{1j}) \\ &= \alpha \det(\mathbf{A})\end{aligned}$$

Case $l > 1$: The $1j$ minor $\mathbf{B}_{1j}$ has $l-1$ rows equal to α times the $l-1^{th}$ row of $\mathbf{A}_{1j}$. So by induction,

$$\det(\mathbf{B}_{1j}) = \alpha \det(\mathbf{A}_{1j})$$

and as $b_{1j} = a_{1j}$ we obtain by definition

$$\det(\mathbf{B}) = \alpha \det(\mathbf{A})$$

□

Theorem 5.1.5. *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in M_n(F)$ *and* $1 \le l \le n$. *Suppose* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *are the same except in row* l*, where the* l^{th} *row of* $\mathbf{C}$ *is the sum of the* l^{th} *row of* $\mathbf{A} = \mathbf{B}$. *Then*

$$\det(\mathbf{C}) = \det(\mathbf{A}) + \det(\mathbf{B})$$

Proof: exactly like in 5.1.4.

Theorem 5.1.6. *Let* $\mathbf{A} \in M_n(F)$ *and* $1 \le l < n$. *Suppose rows* l *and* $l+1$ *of* $\mathbf{A}$ *are equal. Then* $\det(\mathbf{A}) = 0$.

Proof: by induction on n. If $l \ge 2$, this is like the previous result, it follows by an easy induction. So instead suppose that rows 1 and 2 of $\mathbf{A}$ are equal, $a_{1j} = a_{2j}$.

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(\mathbf{A}_{1j})$$

$$\mathbf{A}_{1j} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1n} \\ a_{31} & a_{32} & \cdots & a_{3(j-1)} & a_{3(j+1)} & \cdots & a_{3n} \\ \vdots & & & & & & \end{pmatrix}$$

Let $\mathbf{A}_{1j,k}$ be obtained from $\mathbf{A}$ by deleting rows $1, 2$ and columns j and k

$$\det(\mathbf{A}_{1j}) = \sum_{k<j} (-1)^{1+k} a_{1k} \det(A_{1j,k}) - \sum_{k>j} (-1)^{1+k} a_{1k} \det(\mathbf{A}_{1j,k})$$

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{j=1}^{n} \sum_{k<j} (-1)^{1+j} (-1)^{1+k} a_{1j} a_{1k} \det(\mathbf{A}_{1j,k}) \\ &- \sum_{j=1}^{n} \sum_{j<k} (-1)^{1+j} (-1)^{1+k} a_{1j} a_{1k} \det(\mathbf{A}_{1j,k}) \\ &= 0\end{aligned}$$

If it is not clear why it is 0, look at the matrix $\mathbf{A}$ again and just consider the case of the first two rows. If we take the part of the determinant corresponding to a_{11} and a_{22}, this is $a_{11} \cdot a_{22} \cdot \det(\mathbf{A}_{11,2})$. Now look at the part of the determinant corresponding to a_{12} and a_{21}: this is $a_{12} \cdot a_{21} \cdot \det(\mathbf{A}_{11,2}) = a_{22} \cdot a_{11} \cdot \det(\mathbf{A}_{11,2})$, they are exactly the same and they cancel out! Since we have to then check for a_{23} and so on, it turns out that you can use the same argument by looking at a_{13} since the part of the determinant on a_{13} will also have a part with the first and third row deleted. And then it is quite obvious to see why the determinant is 0. $\square$

Define the determinant of A to be a map:

$$\begin{aligned} A &\to \det A \\ M_n(F) &\to F \end{aligned}$$

Rules:

D1. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by multiplying a row of $\mathbf{A}$ by a scalar $\alpha \in F$, then

$$\det(\mathbf{B}) = \alpha \det(\mathbf{A})$$

D2. the Det map is a linear function of the rows of $\mathbf{A}$:

$$\det \begin{pmatrix} R_1 \\ \vdots \\ R_i + R_i' \\ \vdots \\ R_n \end{pmatrix} = \det \begin{pmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_n \end{pmatrix} + \det \begin{pmatrix} R_1 \\ \vdots \\ R_i' \\ \vdots \\ R_n \end{pmatrix}$$

D3. If two consecutive rows of $\mathbf{A}$ are equal then $\det(\mathbf{A}) = 0$

D4. $\det \mathbf{I}_n = 1$ (next theorem)

Theorem 5.1.7. $\det(\mathbf{I}_n) = 0$

Proof: by induction on n. The result is obvious for $n = 1$, and by definition of determinant,

$$\begin{aligned} \det I_n &= 1 \cdot \det(I_{n-1}) \\ &= 1 && \text{by inductive hypothesis} \end{aligned}$$

$\square$ It's important to have efficient methods to compute determinants. We work out the effect of elementary row operations on the determinant:

Theorem 5.1.8. *Let* $\mathbf{A}, \mathbf{B} \in M_n(F)$.

i Suppose $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by swapping rows* i *and* $i+1$. *Then* $\det(\mathbf{B}) = -\det(\mathbf{A})$.

ii Suppose $\mathbf{A}$ *has two equal rows, then* $\det(\mathbf{A}) = 0$

iii Suppose $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by swapping two rows. Then* $\det(\mathbf{B}) = -\det(\mathbf{A})$.

iv Suppose $i \neq j$ *and* $\mathbf{B}$ *is obtained from* $\mathbf{A}$ *by adding* $\alpha \cdot R_i$ *to* R_j*:*

$$\mathbf{B} = \begin{pmatrix} R_1 \\ \vdots \\ R_j + \alpha R_i \\ \vdots \\ R_n \end{pmatrix}$$

Then $\det(\mathbf{B}) = \det(\mathbf{A})$. *This makes determinants easier to compute using Gaussian elimination, e.g.*

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & 0 & 1 \\ 2 & 0 & -1 & 1 \\ -1 & 2 & 1 & 0 \\ 1 & 0 & -2 & 1 \end{vmatrix} &\overset{iv}{=} \begin{vmatrix} 1 & 2 & 0 & 1 \\ 0 & -4 & -1 & -1 \\ 0 & 4 & 1 & 1 \\ 0 & -2 & 02 & 0 \end{vmatrix} \\
&\overset{D1}{=} - \begin{vmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & 1 & 1 \\ 0 & 4 & 1 & 1 \\ 0 & -2 & -2 & 0 \end{vmatrix} \\
&= 0
\end{aligned}$$

Proof:

i) Just display rows $i, i+1$

$$\begin{aligned}
0 &= \det \begin{pmatrix} \vdots \\ R_i + R_{i+1} \\ R_i + R_{i+1} \\ \vdots \end{pmatrix} \\
&= \det \begin{pmatrix} \vdots \\ R_i \\ R_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ R_i \\ R_{i+1} \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ R_{i+1} \\ R_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ R_{i+1} \\ R_{i+1} \\ \vdots \end{pmatrix} \\
&= 0 + \det(\mathbf{A}) + \det(\mathbf{B}) = 0 \\
&\therefore \det(\mathbf{B}) = -\det(\mathbf{A})
\end{aligned}$$

ii) Suppose $\mathbf{A}$ has two equal rows. Repeatedly swap consecutive rows of $\mathbf{A}$ to get a matrix $\mathbf{B}$ with two consecutive rows equal. Then $\det(\mathbf{B}) = 0$ by part i.

iii) Same proof for i), but with D3 replaced by ii

iv) Just display rows $i \neq j$

$$\det(\mathbf{B}) = \det \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_j + \alpha R_i \\ \vdots \end{pmatrix}$$

$$\stackrel{\text{D1,D2}}{=} \det \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \end{pmatrix} + \alpha \det \begin{pmatrix} \vdots \\ R_i \\ \vdots \\ R_i + \alpha R_i \\ \vdots \end{pmatrix}$$

$$\stackrel{\text{ii}}{=} \det(\mathbf{A}) + 0$$

□

Corollary. *If* $\mathbf{A}, \mathbf{B} \in M_n(F)$ *are row-equivalent, then* $\exists$ *a non-zero scalar* $\beta \in F$ *such that* $\det(\mathbf{B}) = \beta \det(\mathbf{A})$. *Hence*

$$\det(\mathbf{A}) = 0 \iff \det(\mathbf{B}) = 0$$

Definition 5.1.9. Say $\mathbf{A} \in M_n(F)$ is singular if $\exists$ a nonzero vector $\mathbf{v} \in F^n$ such that $\mathbf{A}\mathbf{v} = 0$. Otherwise $\mathbf{A}$ is non-singular.

Theorem 5.1.10. *Let* $\mathbf{A} \in M_n(F)$. *The following are equivalent:*

1. $\mathbf{A}$ *is invertible.*
2. $\mathbf{A}$ *is non-singular.*
3. *The rows of* $\mathbf{A}$ *are linearly independent*
4. $\mathbf{A}$ *is row-equivalent to* $\mathbf{I}_n$.
5. $\det(\mathbf{A}) \neq 0$

Proof: (1)-(4) are equivalent by theorems proved last term.
$4 \implies 5$: since $\det(\mathbf{I}_n) = 1 \neq 0$, this follows from 5.1.9
$5 \implies 4$: We prove the contrapositive. Suppose $\mathbf{A}$ is not row equivalent to $\mathbf{I}_n$. By Gaussian Elimination, $\mathbf{A}$ is row-equivalent to a matrix $\mathbf{B}$ with a row of zeros. Then $\det(\mathbf{B}) = 0$ and hence $\det(\mathbf{A}) = 0$ by 5.1.9 □

Theorem 5.1.11. *Let $\mathbf{A} \in M_n(F)$. Then*

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i+1} a_{ij} \det(\mathbf{A}_{ij})$$

e.g.

Proof: Can assume $i > 1$. Let

$$\mathbf{A} = \begin{pmatrix} R_1 \\ \vdots \\ R_n \end{pmatrix}$$

By doing $i-1$ row swaps, we obtain:

$$\mathbf{B} = \begin{pmatrix} R_i \\ R_1 \\ \vdots \\ R_{i-1} \\ R_{i+1} \\ \vdots \\ R_n \end{pmatrix}$$

(Swap R_i with the row above it $i-1$ times.) Then $\det(\mathbf{B}) = (-1)^{i-1} \det(\mathbf{A})$. Also $\mathbf{A}_{ij} = \mathbf{B}_{1j}$ for each j. So

$$\begin{aligned}
\det(\mathbf{A}) &= (-1)^{i-1} \det(\mathbf{B}) \\
&= (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j+1} b_{1j} \det(\mathbf{B}_{1j}) \\
&= (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j+1} a_{ij} \det(\mathbf{A}_{ij}) \\
&= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij})
\end{aligned}$$

Corollary. *Suppose $\mathbf{A} \in M_n(F)$ is upper triangular.*

$$\mathbf{A} = \begin{pmatrix} a_{11} & * & * & * & \\ 0 & a_{22} & \cdots & \cdots & \vdots \\ & \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & a_{nn} \end{pmatrix}$$

Then $\det(\mathbf{A}) = a_{11} a_{22} \cdots a_{nn}$

Proof: By induction on n. True for $n = 1$. Assume true for $(n-1) \times (n-1)$ upper triangular matrices and let $\mathbf{A} \in M_n(F)$. Expand $\det(\mathbf{A})$ along nth row:

$$\det(\mathbf{A}) = a_{nn} \det(\mathbf{A}_{nn})$$

Then by inductive hyothesis

$$\det(\mathbf{A}_n n) = a_{11} \cdots a_{n-1,n-1}$$

Hence $\det(\mathbf{A}) = a_{11} \cdots a_{nn}$ and result follows by induction. $\square$

5.2 Further properties of matrices

STAGGERING RESULT!!!! If $\mathbf{A}, \mathbf{B} \in M_n(F)$, then

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$$

i.e. $\det : M_n(F) \to F$ is a multiplicative function.

We approach the proof of this theorem via elementary matrices: recap from section 2.4:

1. Elementary matrices ate obtained from $\mathbf{I}_n$ by doing a single row operation.They are:
    - Multiplies a row by a non-zero number:

$$\mathbf{E}_r(\alpha) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \alpha & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

    - Adds a multiple of row r to row s:

$$\mathbf{E}_{rs}(\alpha) = \begin{array}{r} \\ \\ \\ s \\ r \end{array} \left(\begin{array}{ccccc} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & \alpha \\ 0 & \cdots & \cdots & \cdots & 1 \end{array}\right)$$

    - Swapping row s with row r

$$\mathbf{E}_{rs} = \begin{array}{r} \\ s \\ r \\ \\ \\ \end{array} \left(\begin{array}{ccccc} 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & 1 & & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{array}\right)$$

2. If $\mathbf{E}$ is an elementary matrix and $\mathbf{A} \in M_n(F)$ then $\mathbf{EA}$ is the matrix obtained by doing the row op to $\mathbf{A}$.

$$\mathbf{E}_{rs}(\alpha)\mathbf{A} = \begin{pmatrix} R_1 \\ \vdots \\ R_r + \alpha R_s \\ \vdots \\ R_n \end{pmatrix}$$

Lemma 5.2.0. *If $\mathbf{A}$ is nonsingular, then $\exists$ elementary matrices such that*

$$\mathbf{A} = \mathbf{E}_1 \cdots \mathbf{E}_r$$

Proof of 5.2.1: By 5.1.10, $\exists$ a sequence of row ops reducing $\mathbf{A}$ to $\mathbf{I}_n$. Hence $\exists$ elementary matrices

$$\begin{gathered} \mathbf{E}'_1, ..., \mathbf{E}'_r \text{ such that} \\ \mathbf{E}'_r \cdots \mathbf{E}'_2 \cdot \mathbf{E}'_1 = \mathbf{I}_n \end{gathered}$$

Hence

$$\begin{aligned} \mathbf{A} &= (\mathbf{E}'_1)^{-1} \cdots (\mathbf{E}'_r)^{-1} \\ &= \mathbf{E}_r \cdots \mathbf{E}_1 \end{aligned}$$

We'll be right back to the proof after this lemma:

Lemma 5.2.1. *If $\mathbf{A}, \mathbf{E} \in M_n(F)$ with $\mathbf{E}$ elementary, then*

$$\det(\mathbf{EA}) = (\det(\mathbf{E}))(\det(\mathbf{A}))$$

Proof:

1. If $\mathbf{E} = \mathbf{E}_{rs}(\alpha)$ then $\det(\mathbf{E}) = 1$ and

$$\det(\mathbf{EA}) = \det \begin{pmatrix} R_1 \\ \vdots \\ R_i + \alpha R_j \\ \vdots \\ R_n \end{pmatrix} = \det(\mathbf{A})$$

   By 5.1.8. So $\det(\mathbf{A}) = \det(\mathbf{E}) \cdot \det(\mathbf{A})$

2. If $\mathbf{E} = \mathbf{E}_{rs}$ then $\det(\mathbf{E}) = -1$ and $\det(\mathbf{EA}) = -\det(\mathbf{A})$ by 5.1.8

3. If $\mathbf{E} = \mathbf{E}_r(\alpha)$ then $\det(\mathbf{E}) = \alpha$ and $\det(\mathbf{EA}) = \alpha \det(\mathbf{A})$ by (D1).

Lemma 5.2.2. *Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. Then*

1. **AB** *is singular iff either* **A** *or* **B** *is singular.*

2. $\det(\mathbf{AB}) = 0$ *iff either* $\det(\mathbf{A})$ *or* $\det(\mathbf{B})$ *are 0*

Proof: Exercise (use contrapositive - part 2 follows from part 1 by previous theorems)

Theorem 5.2.3. *If* $\mathbf{A}, \mathbf{B} \in M_n(F)$*, then* $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$*.*

Proof: If either **A** or **B** is singular, this follows from the previous lemma. So suppose both **A** and **B** are non-singular. Then by 5.2.0, $\exists$ elementary matrices $\mathbf{E}_1, .., \mathbf{E}_r$ and $\mathbf{E}'_1, ..., \mathbf{E}'_s$ such that

$$\begin{aligned} \mathbf{A} &= \mathbf{E}_1 \cdots \mathbf{E}_r \\ \mathbf{B} &= \mathbf{E}'_1 \cdots \mathbf{E}'_s \end{aligned}$$

then

$$\mathbf{AB} = \mathbf{E}_1 \cdots \mathbf{E}_r \cdot \mathbf{E}'_1 \cdots \mathbf{E}'_s$$

Applying 5.2.1 repeatedly,

$$\begin{aligned} \det(\mathbf{A}) &= (\det(\mathbf{E}_1))(\det(\mathbf{E}_2)) \cdots (\det(\mathbf{E}_r)) \\ \det(\mathbf{B}) &= (\det(\mathbf{E}'_1)) \cdots (\det(\mathbf{E}'_s)) \\ \det(\mathbf{AB}) &= (\det(\mathbf{E}_1))(\det(\mathbf{E}_2)) \cdots (\det(\mathbf{E}_r)) \cdot (\det(\mathbf{E}'_1)) \cdots (\det(\mathbf{E}'_s)) \quad = \det(\mathbf{A}) \cdot \det(\mathbf{B}) \end{aligned}$$

□

Theorem 5.2.4. *For* $\mathbf{A} \in M_n(F)$*,*

$$\det(\mathbf{A}^T) = \det(\mathbf{A})$$

Corollary. *Suppose* $1 \leq j \leq n$*, then*

$$\det(\mathbf{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij})$$

(this is expansion down column j*)*

Proof:

$$\begin{aligned} &\det(\mathbf{A}) = \det(\mathbf{A}^T) \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det((\mathbf{A}^T)_{ji}) \qquad \text{expanding along row } i \end{aligned}$$

since $(\mathbf{A}^T)_{ji} = (\mathbf{A}_{ij})^T$, this gives the result. □

Example: (Vandermonde Determinant): let $n >> 2, x_1, ..., x_n \in F$.

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} = \prod_{1 \leq i < j \leq n} (x_j - x_i)$$

Note: This is 0 if and only if $x_i = x_j$ for some $i \neq j$.

Proof: by induction on n. Base case $n = 2$ is true. Then assume true for $(n-1)\times(n-1)$. Use column operations to clear row 1: do $C_n - x_1C_{n-1}, C_{n-1} - x_1C_{n-2}, \cdots, C_2 - x_1C_1$:

$$\begin{aligned}
\det &= \det\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & \cdots & \cdots & (x_2 - x_1)x_2^{n-3} & (x_2 - x_1)x_2^{n-2} \\ \vdots & & & & & \vdots \\ 1 & (x_n - x_1) & \cdots & \cdots & (x_n - x_1)x_n^{n-3} & (x_n - x_1)x_n^{n-2} \end{pmatrix} \\
&= \det\begin{pmatrix} (x_2 - x_1) & \cdots & (x_2 - x_1)x_2^{n-3} & (x_2 - x_1)x_2^{n-2} \\ \vdots & & & \vdots \\ x_n - x_1 & \cdots & (x_n - x_1)x_n^{n-3} & (x_n - x_1)x_n^{n-2} \end{pmatrix} \\
&= (x_2 - x_1)\cdots(x_n - x_1)\det(\text{Some matrix left over}) \\
&= (x_2 - x_1)\cdots(x_n - x_1)\prod_{z\leq i<j\leq n}(x_j - x_i)
\end{aligned}$$

□

5.3 Inverting

Definition 5.3.1. Let $\mathbf{A} = (a_{ij}) \in M_n(F), 1 \leq i, j, \leq n$. The ij *cofactor* of $\mathbf{A}$ is

$$c_{ij} = (-1)^{i+j}\det(\mathbf{A}_{ij})$$

Let $\mathbf{C} = (c_{ij}) \in M_n(F)$. The *Adjugate* of $\mathbf{A}$ is $\text{adj}(\mathbf{A}) = \mathbf{C}^T$.

Theorem 5.3.2.

$$\text{adj}(\mathbf{A})\mathbf{A} = \det(\mathbf{A})\mathbf{I}_n$$

So if $\det(\mathbf{A}) \neq 0, a^{-1} = \frac{1}{\det(\mathbf{A})}\text{adj}(\mathbf{A})$

Proof: The ji entry of

$$\begin{aligned}
\mathbf{C}^T A &= \sum_{i=1}^{n} c_{ij}a_{ij} \\
&= \sum_{i=1}^{n}(-1)^{i+j}\det(\mathbf{A}_{ij})a_{ij} \\
&= \det(\mathbf{A})
\end{aligned}$$

If $j \neq k$ the jk entry of

$$\mathbf{C}^T\mathbf{A} = \sum_{j=1}^{n} c_{ij}a_{ik}$$

To compare this we never use the entries in column j of $\mathbf{A}$... so, for the purpose of the calculation, we can assume that column j is the same as column k in the original matrix. Then the formula is:

$$\sum_{i=1}^{n} c_{ij} a_{ij} = 0$$

as it's the determinant of a matrix with 2 columns... so we have the result. $\square$

Corollary. *If* $\mathbf{A}$ *is an* $n \times n$ *matrix of integers and* $\det(\mathbf{A}) = \pm 1$*, then* $\mathbf{A}^{-1}$ *is also a matrix of integers (because* $\operatorname{adj}(\mathbf{A})$ *is also a matrix of integers)*

5.4 The Determinant of a Linear Transformation

Suppose V is a finite dimensional vector space over a field F and $T : V \to V$ is a linear transformation. Let B be a basis of V and consider $\mathbf{M} = [T]_B$.

Define $\det(T) = \det(\mathbf{M})$. Why does this not depend on the choice of basis B?

Theorem 5.4.1. *The determinant of a linear transformation* T *does not depend on the choice of basis* B *from which you construct the matrix* $[T]_B$

Proof: Consider a basis B and a basis C of the vector space C. Then

$$[T]_B = \mathbf{P}^{-1}[T]_C\mathbf{P}$$

with $\mathbf{P}$ the change of basis matrix from B to C

$$\det([T]_B) = \det(\mathbf{P}^{-1}[T]_c\mathbf{P})$$

Since $\det(\mathbf{AB}) = \det(\mathbf{BA})$,

$$\det([T]_B) = \det(\mathbf{P}^{-1}\mathbf{P}[T]_C) = \det([T]_C)$$

$\square$ *Example:* Let V be a vector space of polynomials of degree ≤ 2 over $\mathbb{R}$. Let

$$T : V \to V, T(p(x)) = p(3x+1)$$

What is $\det(T)$?

Let a basis of V be $B = \{1, x, x^2\}$. Then

$$[T]_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 9 \end{pmatrix}$$

so $\det(T) = 27$

6 Eigenvalues and Eigenvectors

6.1 Definitions and Basics

Definition 6.1.1.
1. Suppose $\mathbf{A} \in M_n(F)$ and $\lambda \in F$. Say that λ is an *Eigenvalue* of $\mathbf{A}$ is there is a non zero $\mathbf{v} \in F^n$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

Such a $\mathbf{v}$ is called an *Eigenvector* of $\mathbf{A}$.

2. Suppose V is a vector space over a field F, and $T : V \to V$ is a linear map. Say $\lambda \in F$ is an eigenvalue of T if there is a non zero $\mathbf{v} \in V$ with $T(\mathbf{v}) = \lambda\mathbf{v}$. Such a $\mathbf{v}$ is called an *eigenvector* of T.

Example 6.1.2.

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 10 & -1 & -12 \\ 8 & 1 & -12 \\ 5 & -1 & -5 \end{pmatrix} \\ T_{\mathbf{A}} &: \mathbb{R}^3 \to \mathbb{R}^3 \\ T_{\mathbf{A}}(\mathbf{v}) &= \mathbf{A}\mathbf{v} \end{aligned}$$

Let

$$\mathbf{v_1} = \begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix}, \mathbf{v_2} = \begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix}, \mathbf{v_3} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} 1$$

Then

$$\begin{aligned} T_{\mathbf{A}}(\mathbf{v_1}) &= \mathbf{A}\mathbf{v_1} = 1 \cdot \mathbf{v_1} \\ T_{\mathbf{A}}(\mathbf{v_2}) &= \mathbf{A}\mathbf{v_2} = 2\mathbf{v_2} \\ T_{\mathbf{A}}(\mathbf{v_3}) &= A\mathbf{v_3} = 3\mathbf{v_3} \end{aligned}$$

So $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}$ are eigenvectors of $\mathbf{A}$ with corresponding eigenvalues $1, 2, 3$.

Proposition 6.1.3. Suppose V is a finite dimensional vector space over F and B is a basis. Let $T : V \to V$.

i The eigenvalues of T and the eigenvalues of the matrix $[T]_B$ are equal.

ii A vector $\mathbf{v} \in V$ is an eigenvector of $T \iff [\mathbf{v}]_B$ is an eigenvector of $[T]_B$.

Proof: Two observations:

1. $[\mathbf{v}]_B = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \iff \mathbf{v} = \mathbf{0}$

2. $T(\mathbf{v}) = \lambda\mathbf{v} \iff [T(\mathbf{v})]_B = [\lambda\mathbf{v}]_B \iff \lambda[\mathbf{v}]_B = [T]_B[\mathbf{v}]_B$.

□

6.2 The Characteristic Polynomial

Definition 6.2.1.

1. Suppose $\mathbf{A} \in M_n(F)$ and let x denote a variable. The *Characteristic Polynomial* of $\mathbf{A}$ is
$$\chi_{\mathbf{A}}(x) = \det(x\mathbf{I}_n - \mathbf{A})$$

2. Suppose V is a finite dimensional vector space over F and $T : V \to V$ be linear, and B a basis of V. Define the *Characteristic Polynomial* of T to be
$$\chi_T(x) = \chi_C(x)$$
where $\mathbf{C} = [T]_B$.

Remark 6.2.2.

1. Some people use the characteristic polynomial as $\det(\mathbf{A} - x\mathbf{I}_n)$ instead of the other way around.

2. BUT: $\det(x\mathbf{I}_n - \mathbf{A})$ is a polynomial of degree n, and the coefficient of x^n is 1.

In part 2. of Definition 6.2.1, $\chi_T(x)$ does not depend on the choice of basis B. The proof of this is similar to the proof of Theorem 5.4.1

Theorem 6.2.3.

1. *If $\mathbf{A} \in M_n(F)$ and $\lambda \in F$ then λ is an eigenvalue of $\mathbf{A}$ iff $\chi_{\mathbf{A}}(\lambda) = 0$*

2. *If V is a finite dimensional vector space over F, and $T : V \to V$, then for $\lambda \in F, \lambda$ is an eigenvalue of T iff $\chi_T(\lambda) = 0$.*

Corollary. *If $\mathbf{A} \in M_n(F)$ then $\mathbf{A}$ has $\leq n$ eigenvalues.*

Proof:

1.
$$\begin{aligned} &\lambda \text{ is an eigenvalue of} A \\ \iff &\exists \mathbf{v} \in V, \mathbf{v} \neq \mathbf{0} : \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \\ \iff &(\lambda\mathbf{I}_n - \mathbf{A})\mathbf{v} = 0 \\ \iff &\det(\lambda\mathbf{I}_n - \mathbf{A}) = 0 \end{aligned}$$

2. By (1) and Prop 6.1.3.

Notation: If $\mathbf{A} \in M_n(F), \lambda \in F$, let
$$\begin{aligned} E_\lambda &= \{\mathbf{v} \in F^n : \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\} \\ &= \{\mathbf{v} \in F^n : (\lambda\mathbf{I}_n - \mathbf{A})\mathbf{v} = 0\} \end{aligned}$$

This is a subspace of F^n. Sometimes known as the *Eigenspace*. □

Example 6.2.4.

(1)

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$$
$$\chi_{\mathbf{A}}(\lambda) = \det(\lambda \mathbf{I}_2 - \mathbf{A}) = (\lambda - 1)^2$$

The only eigenvalue is therefore 1. Now find the eigenvector(s)

$$\mathbf{A} - 1 \cdot \mathbf{I}_2 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}$$

So the eigenvectors are all in the span of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(2)

$$\mathbf{A} = \begin{pmatrix} 10 & -1 & -12 \\ 3 & 1 & -12 \\ 5 & -1 & -5 \end{pmatrix} \in M_3(\mathbb{R})$$
$$\chi_{\mathbf{A}}(\lambda) = \det \begin{pmatrix} \lambda - 10 & 1 & 12 \\ -8 & \lambda - 1 & 12 \\ -5 & 1 & \lambda + 5 \end{pmatrix} = \cdots = (\lambda - 1)(\lambda - 2)(\lambda - 3)$$

So the eigenvalues are $\lambda = 1, 2, 3$. To find the eigenvectors, consider each λ in turn:

$$\begin{aligned} \mathbf{A} - \mathbf{I}_3 &= \begin{pmatrix} 9 & -1 & -12 \\ 8 & 0 & -12 \\ 5 & -1 & -6 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & -1 & - \\ 0 & 8 & -12 \\ 0 & 4 & -6 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -3 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Which gives the solution

$$\mathbf{v} = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} \cdot \alpha \in \mathbb{R}$$

The other eigenvectors can be found using a similar method.

(3) Let V be the vector space of polynomials of degree ≤ 2 over $\mathbb{R}$. Let $T : V \to V$ be a linear transformation with

$$T(p(t)) = p(3t + 1)$$

Let B be a basis with $B = \{1, t, t^2\}$ and

$$[T]_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 9 \end{pmatrix} = \mathbf{A}$$

Then

$$\begin{aligned} \chi_T(\lambda) &= \det \begin{pmatrix} x-1 & -1 & -1 \\ 0 & x-3 & -6 \\ 0 & 0 & x-9 \end{pmatrix} \\ &= (x-1)(x-3)(x-9) \\ \lambda &= 1, 3, 9 \end{aligned}$$

Now find the eigenvectors for these values.

$$\begin{aligned} A - 3I_3 &= \begin{pmatrix} -2 & 1 & 1 \\ 0 & 0 & 6 \\ 0 & 0 & 6 \end{pmatrix} \rightarrow \begin{pmatrix} -2 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\ \therefore E_3 &= \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \alpha & \alpha \in \mathbb{R} \end{aligned}$$

So the eigenvectors of T are $\alpha(1 + 2t)$

6.3 Diagonalisation

Definition 6.3.1.

(1) A linear transformation $T : V \to V$ is *Diagonalisable* if there is a basis of V consisting of eigenvectors of T.

(2) A matrix $\mathbf{A} \in M_n(F)$ is *diagonalisable* if there is a basis of F^n consisting of eigenvectors of $\mathbf{A}$.

So if $\mathbf{A} \in M_n(F)$ let $T_\mathbf{A} : F^n \to F^n, T_\mathbf{A}(\mathbf{v}) = \mathbf{A}\mathbf{v}$. Then $\mathbf{A}$ is diagonalisable $\iff$ $T_\mathbf{A}$ is diagonalisable.

Example:

1) The matrix $\mathbf{A}$ in 6.2.4 is diagonalisable.

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

in 6.2.4 isn't diagonalisable (only eigenvectors are multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$).

3) The linear transformation T in 6.2.4 (3) isn't diagonalisable since $B : \{1, 1 + 2t, 1 + 4t + 4t^2\}$ is a basis of V.

Theorem 6.3.2.

(1) *Suppose V is a f.d vector space over F and $T : V \to V$ is a linear transformation. Then T diagonalisable iff there is a basis $B : \mathbf{v}_1, ..., \mathbf{v}_n$ of V such that $[T]_B$ is a diagonal matrix*

(2) *$\mathbf{A} \in M_n(F)$ is diagonalisable iff there is an invertible $\mathbf{P} \in M_n(F)$ such that $\mathbf{P}^{-1}\mathbf{AP}$ is a diagonal matrix. In this case, the columns of $\mathbf{P}$ consist of eigenvectors.*

Proof:

(1) Let $B : \mathbf{v}_1, ..., \mathbf{v}_n$ be a basis. Note: $\mathbf{v}_i \neq 0$. Let $\mathbf{D} = [T]_B$ then $\mathbf{D}$ is a diagonal matrix:

$$\begin{aligned} &\iff \forall j \leq n, T(\mathbf{v}_j) = d_{jj}\mathbf{v}_j \\ &\iff \text{each } \mathbf{v}_j \text{ is an eigenvector of } T \end{aligned}$$

(2) Suppose $\mathbf{P}$ is invertible. The columns $\mathbf{v}_1, ...\mathbf{v}_n$ of $\mathbf{P}$ are a basis B of F^n and $\mathbf{P} =_E [Id]_B$ where E is the standard basis.

$$\begin{aligned} \mathbf{P}^{-1}\mathbf{AP} &=_B [Id]_{EE}[T_\mathbf{A}]_{EE}[Id]_B \\ &=_B [T_\mathbf{A}]_B \end{aligned}$$

This is the diagonal matrix

$$\begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}$$

$$\begin{aligned} &\iff T_(\mathbf{v}_j) = d_j\mathbf{v}_j, \forall j \leq n \\ &\iff \mathbf{v}_1, ..., \mathbf{v}_n \text{ eigenvectors of } \mathbf{A} \end{aligned}$$

□

Example 6.3.3.

(1) Let $\mathbf{A} \in M_2(\mathbb{R})$,

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

$$\chi_\mathbf{A}(x) = x^2 + 1$$

(2)

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{C})$$

$$\chi_\mathbf{A}(x) = x^2 + 1 = (x + i)(x - i)$$

So the eigenvalues are $i, -i$

$$\begin{array}{c|c} i & \begin{pmatrix} 1 \\ -i \end{pmatrix} \\ -i & \begin{pmatrix} 1 \\ i \end{pmatrix} \end{array}$$

These are linearly independent so they're diagonalisable over $\mathbb{C}$.

Example 6.3.4.

① *Powers and roots of matrices.* Let $\mathbf{A} \in M_n(F)$, suppose $\mathbf{P} \in M_n(F)$ and $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \text{Diag}\{d_1, ..., d_n\}$. Then for $k \in \mathbb{N}$:

$$\begin{aligned} (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^k &= \mathbf{P}^{-1}\mathbf{A}^k\mathbf{P} \\ \mathbf{D}^k &= \text{Diag}(d_1^k, ..., d_n^k) \\ \text{so } \mathbf{A}^k &= \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1} \end{aligned}$$

Then if $c_1, ..., c_n \in F$ and $c_i^k = d_i, \forall i = 1, ..., n$, then let $\mathbf{E} = \text{Diag}(c_1, ..., c_n), \mathbf{E}^k = \mathbf{D}$.

$$\begin{aligned} (\mathbf{P}\mathbf{E}\mathbf{P}^{-1})^k &= \mathbf{P}\mathbf{E}^k\mathbf{P}^{-1} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \\ &= \mathbf{A} \end{aligned}$$

② *Recurrence relations* The sequences $(L_n)_{n\geq 0}, (T_n)_{n\geq 0}$ of real numbers satisfy $L_0 = 1000, T_0 = 8.$ and

$$3L_n = 2L_{n-1} + T_{n-1}$$
$$3T_n = 4L_{n-1} + 2T_{n-1}$$

Find a general expression for L_n and T_n.

$$\begin{aligned} \begin{pmatrix} L_n \\ T_n \end{pmatrix} &= \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} L_{n-1} \\ T_{n-1} \end{pmatrix} \\ \begin{pmatrix} L_n \\ T_n \end{pmatrix} &= \frac{1}{3^n}\mathbf{A}^n\begin{pmatrix} L_0 \\ T_0 \end{pmatrix} \end{aligned}$$

So we can find the characteristic polynomial of $\mathbf{A}$:

$$\chi_A(x) = x^2 - 4x = x(x-4)$$

So the eigenvalues are $\lambda = 0, 4$

Theorem 6.3.5. *Suppose V is a vector space over F and $T : V \to V$ is linear. Suppose $\mathbf{v}_1, ..., \mathbf{v}_n$ are eigenvectors of T with $T(\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for $i \leq n, \lambda_i \neq \lambda_j, \forall i \neq j$. Then the $\mathbf{v}_i$ are linearly independent.*

Corollary.

(1) Suppose V is finite dimensional, and $\dim(V) = n$, and T has n distinct eigenvalues in F. Then T is diagonalisable

(2) If $\mathbf{A} \in M_n(F), \chi_{\mathbf{A}}(x)$ has n distinct roots in F, then $\mathbf{A}$ is diagonalisable over F.

Proof of theorem: By induction on n.
$n = 1$: $\mathbf{v}_1 \neq 0$, as $\mathbf{v}_1$ is an eigenvector.
Inductive step: Suppose $n > 1$ and the result is true for all $< n$. Suppose for a contradiction that $\mathbf{v}_1, ...\mathbf{v}_n$ are linearly dependent, so there exist $\alpha_1, ..., \alpha_n \in F$ not all 0, with

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$$

By the induction hypothesis, we have $\alpha_i \neq 0$, otherwise there is a smaller subset of $\mathbf{v}_1, ...\mathbf{v}_n$ which is linearly dependent. So, if we divide by α_1, we can assume $\alpha_1 = 1$, so

$$\mathbf{v_1} + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0} \tag{1.1}$$

Then applying T to (1.1):

$$\begin{aligned}\mathbf{0} = T(\mathbf{0}) &= T(\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n)\\ &= \lambda_1\mathbf{v}_1 + \lambda_2\alpha_2\mathbf{v}_2 + \cdots + \lambda_n\alpha_n\mathbf{v}_n\end{aligned}$$

Example 6.3.6. Given V a finite dimensional vector space over F and $T : V \to V$ a linear map, we check if T is diagonalisable.

① Compute $\chi_T(x)$ and find the eigenvalues $\lambda_1, ...\lambda_r \in F$

② for each $i \leq r$ find a basis B_i for

$$E_{\lambda_i} = \{\mathbf{v} \in V : T(\mathbf{V}) = \lambda_i\mathbf{v}\}$$

③ if

$$\sum_{i=1}^{r} \dim(E_{\lambda_i}) < \dim(V)$$

Then T is not diagonalisable.

④ If

$$\sum_{i=1}^{r} \dim(E_{\lambda_i}) > \dim(V)$$

Then we have equality here and the union of the B_i gives a basis of V. Therefore T is diagonalisable.

Proof of ④: Write

$$B_i : \mathbf{v}_{i_1}, ..., \mathbf{v}_{i_{n(i)}}$$

We need to show that the $\mathbf{v}_{ij}$ are linearly independent, then ④ follows. Suppose $\alpha_{ij} \in F$ and

$$\sum_{i=1}^{r} \left(\sum_{j=1}^{n(i)} \alpha_{ij} \mathbf{v}_{ij} \right) = 0$$

Then let $\mathbf{w}_i = \sum_{j=1}^{n(i)} \alpha_{ij} \mathbf{v}_{ij}$. So $\mathbf{w}_i \in E_{\lambda_i}$ and

$$\mathbf{w}_1 + \cdots + \mathbf{w}_r = 0$$

As $\lambda_i \neq \lambda_{i'}$ if $i \neq i'$. Then 6.3.5 gives $\mathbf{w}_i = 0, \forall i \leq r$. Thus, as $\mathbf{v}_{i_1}, ..., \mathbf{v}_{i_{n(i)}}$ are linearly independent, we obtain from the def. of the $\mathbf{w}_i$ that

$$\alpha_{ij} = 0, \forall i \leq r, \forall 1 \leq j \leq n(i)$$

$\square$

6.4 Orthogonal vectors in $\mathbb{R}^n$

Definition 6.4.1. If

$$\mathbf{u} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

The *inner product* of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} \alpha_i \beta_i$$

Say that $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if their inner product is 0. The *norm* of $\mathbf{u}$ is

$$\begin{aligned} \|\mathbf{u}\| &= \sqrt{\mathbf{u} \cdots \mathbf{u}} \\ &= \left(\sum_{i=1}^{n} \alpha_i^2 \right)^{\frac{1}{2}} \\ \|\mathbf{u} - \mathbf{v}\| &= \left(\sum_{i=1}^{n} (\alpha_i - \beta_i)^2 \right)^{\frac{1}{2}} \\ &= \text{ distance of } \mathbf{u} from \mathbf{v} \end{aligned}$$

Note also that

(1) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$

(2) $\|\mathbf{u}\| = 0 \iff \mathbf{u} = 0$

(3) $\|\alpha \mathbf{u}\| = |\alpha| \|\mathbf{u}\|$. So, if $\mathbf{u} \neq 0$ then

$$\left\| \frac{\mathbf{u}}{\|\mathbf{u}\|} \right\| = 1$$

Theorem 6.4.2.

1) (Cauchy-Schwarz) If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

$$\|\mathbf{u}\| \cdot \|\mathbf{v}\| \geq |\mathbf{u} \cdot \mathbf{v}|$$

There is equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

2) (Triangle Inequality)

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

3) (Metric Triangle Inequality)

$$\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u} - \mathbf{w}\| + \|\mathbf{w} - \mathbf{v}\|$$

Proof:

1) Wlog assume $\mathbf{u} \neq \mathbf{0}$. Then consider $\|\lambda \mathbf{u} - \mathbf{v}\|^2$. We know that $0 \leq \|\lambda \mathbf{u} - \mathbf{v}\|^2$ and so by expanding this, we get $0 \leq \lambda^2 \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\lambda(\mathbf{u} \cdot \mathbf{v})$. We now want the value of λ which minimises the right hand side of this expression, and after some calculus (which we can do because this is a quadratic in λ), we get $\lambda = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2}$, and using this value of λ, we get the right hand side to be $\frac{(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{u}\|^2} + \|\mathbf{v}\|^2 - 2\frac{(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{u}\|^2}$, and after some rearrangement with the previous expression, we come to the statement $\|\mathbf{u}\|^2 \, \|\mathbf{v}\|^2 \geq (\mathbf{u} \cdot \mathbf{v})^2$ and from there, the required inequality follows immediately

2) We can use (1) to get the expression $\|\mathbf{u} + \mathbf{v}\|^2 \leq (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$

3) The metric triangle inequality follows directly from (2) by considering $\mathbf{u} - \mathbf{v}$ as $(\mathbf{u} - \mathbf{w} + \mathbf{w} - \mathbf{v})$ and then applying (2)

Say vectors $\mathbf{u}_1, ..., \mathbf{u}_n \in \mathbb{R}^n$ form an *orthonormal set* if $\|\mathbf{u}_i\| = 1$ and $\mathbf{u}_i \cdot \mathbf{u}_j = 0, i \neq j$.

Definition 6.4.3. $\mathbf{P} \in M_n(\mathbb{R})$ is an orthogonal matrix if $\mathbf{P}^T \mathbf{P} = \mathbf{I}_n$

Lemma 6.4.4. *$\mathbf{P} \in M_n(\mathbb{R})$ is an orthogonal matrix iff the columns of $\mathbf{P}$ form an orthonormal set in $\mathbb{R}^n$.*

Proof: The ij entry of $\mathbf{P}^T \mathbf{P}$ is the inner product of the columns i and j of $\mathbf{P}$.

Theorem 6.4.5. *(Gram-Schmidt) Let $\mathbf{v}_1, ...\mathbf{v}_r$ be a set of linearly independent vectors in $\mathbb{R}^n$. Then there exists an orthonormal set $\mathbf{u}_1, ...\mathbf{u}_r \in \mathbb{R}^n$ such that for $i \leq r$*

$$\text{Span}(\mathbf{v}_1, ...\mathbf{v}_i) = \text{Span}(\mathbf{u}_1, ..., \mathbf{u}_i)$$

Corollary.

1) *If U is a subspace of $\mathbb{R}^n$ there is an orthonormal basis $\mathbf{u}_1, ...\mathbf{u}_r$ of U.*

2) *If $\mathbf{v} \in \mathbb{R}^n$ and $\|\mathbf{v}\| = 1$ there is an orthogonal matrix $\mathbf{P}$ with first column $\mathbf{v}$.*

Proof:

1) *Take $\mathbf{v}_1, ...\mathbf{v}_r$ a basis of U and apply Gram-Schmidt*

2) *Extend $\mathbf{v}$ to a basis*
$$\mathbf{v}_1 = \mathbf{v}, ..., \mathbf{v}_n$$
of $\mathbb{R}^n$. Apply Gram-Schmidt to obtain $\mathbf{u}_1, ..., \mathbf{u}_n$ with $\mathbf{u}_1 = \mathbf{v}$ and $\mathbf{u}_1, ...\mathbf{u}_n$ an orthonormal set. Then take $\mathbf{u}_1, ...\mathbf{u}_n$ as the columns of $\mathbf{P}$.

Gram-Schmidt

Given $\mathbf{v}_1, ..., \mathbf{v}_r \in \mathbb{R}^n$ linearly independent, we *find* orthogonal vectors $\mathbf{w}_1, ...\mathbf{w}_r \in \mathbb{R}^n$ with $\mathrm{Span}(\mathbf{v}_1, ..., \mathbf{v}_i) = \mathrm{Span}(\mathbf{w}_1, ...\mathbf{w}_i), \forall i \leq r$.
Then, *normalise* the vectors,
$$\mathbf{u}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$$
Then $\mathbf{u}_1, ..., \mathbf{u}_r \in \mathbb{R}^n$ are orthonormal and the span is equal to the span of the $\mathbf{v}_i$.

We define the $\mathbf{w}_i$ inductively.
$$\begin{aligned} \mathbf{w}_1 &= \mathbf{v}_1 \\ &\vdots \\ \mathbf{w}_i &= \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\mathbf{w}_j \cdot \mathbf{v}_i}{\mathbf{w}_j \cdot \mathbf{w}_j} \mathbf{w}_j \end{aligned}$$

We prove by induction that:

(a) $\mathbf{w}_i \neq 0$

(b) $\mathrm{Span}(\mathbf{v}_1, ...\mathbf{v}_i) = \mathrm{Span}(\mathbf{w}_1, ..., \mathbf{w}_i)$

(c) if $k < i$ then $\mathbf{w}_k \cdot \mathbf{w}_i = 0$

Proof: Inductive step. Assume that a,b and c above are true for smaller i.

(a) If $\mathbf{w}_i = 0$ then by the inductive hypothesis
$$\mathbf{v}_i \in \mathrm{Span}(\mathbf{w}_1, ...\mathbf{w}_{i-1}) \overset{\text{ind. hyp.}}{=} \mathrm{Span}(\mathbf{v}_1, ..., \mathbf{v}_{i-1})$$
which is a contradiction since they are LI.

(b) We have that $\mathbf{w}_{i+1} = \mathbf{v}_{i+1} - (\alpha_1\mathbf{w}_1 + \cdots + \alpha_i\mathbf{w}_i)$ where the final linear combination is from previous applications of Gram-Schmidt. If we let $\mathbf{u} = \alpha_1\mathbf{w}_1 + \cdots + \alpha_i\mathbf{w}_i$ then we have $\mathbf{w}_{i+1} = \mathbf{v}_{i+1} - \mathbf{u}$ then $\mathbf{w}_{i+1} \in \text{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_{i+1})$ and the same relationship for $\mathbf{v}_{i+1}$ can be reached through the same logic.

(c)
$$\mathbf{w}_k \cdot \mathbf{w}_i = \mathbf{w}_k \cdot \mathbf{v}_i - \sum_{j=1}^{i-1} \frac{\mathbf{w}_j \cdot \mathbf{v}_i}{\mathbf{w}_j \cdot \mathbf{w}_j}\mathbf{w}_j$$

Then $k, j < i$ so by the inductive hypothesis, $\mathbf{w}_k \cdot \mathbf{w}_j = 0$ unless $\mathbf{w}_k = 0$, so
$$\mathbf{w}_k \cdot \mathbf{w}_i = 0$$

□ *Example*: Find an orthogonal matrix $\mathbf{P} \in M_3(\mathbb{R})$ with the first column
$$\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

Apply Gram-Schmidt to
$$\mathbf{v}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

Solution:
$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

Then we obtain
$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}$$
$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}-1\\2\\-1\end{pmatrix}$$

It's easier to take $\mathbf{w}_2 = \begin{pmatrix}-1\\2\\-1\end{pmatrix}$. Then
$$\mathbf{w}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix} - \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{-1}{6}\begin{pmatrix}-1\\2\\-1\end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix}$$

Now we normalise the vectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \mathbf{u}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}-1\\2\\-1\end{pmatrix}$$

$$\mathbf{u}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\0\\1\end{pmatrix}$$

$$\mathbf{P} = \frac{1}{\sqrt{6}}\begin{pmatrix}\sqrt{2} & -1 & -\sqrt{3}\\ \sqrt{2} & 2 & 0\\ \sqrt{2} & -1 & \sqrt{3}\end{pmatrix}$$

6.5 Real symmetric matrices

If $\mathbf{A} \in M_n(\mathbb{R})$ then we have $\mathbf{A}^T = \mathbf{A}$. A key property of this is, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then

$$\begin{aligned}&(\mathbf{Au}) \cdot \mathbf{v}\\ &= (\mathbf{u}^T\mathbf{A}^T)\mathbf{v} = \mathbf{u}^T(\mathbf{A}^T\mathbf{v})\\ &= \mathbf{u}^T(\mathbf{Av})\\ &= \mathbf{u} \cdot (\mathbf{Av})\end{aligned}$$

So the linear map given by $\mathbf{A}$ is *self-adjoint*.

Fact: (Fundamental theorem of Algebra, C.F.Gauss) Suppose $p(x)$ is a non-constant polynomial with coefficients in $\mathbb{C}$, then there is a root $\alpha \in \mathbb{C}$. (i.e., $p(\alpha) = 0$ for some $\alpha \in \mathbb{C}$.

Proof: See Complex Analysis in 2nd year

Lemma 6.5.1. *Suppose $A \in M_n(\mathbb{R})$ is symmetric. Suppose $\lambda \in \mathbb{C}$ is a root of $\chi_A(x)$. Then $\lambda \in \mathbb{R}$.*

By 6.5.1 and the FTA, we have

Corollary. *If $A \in M_n(\mathbb{R})$ is symmetric, then there is an eigenvalue $\lambda \in \mathbb{R}$ of A.*

Proof of 6.5.1: Think of $\mathbf{A} \in M_n(\mathbb{C})$. So λ is an eigenvalue of $\mathbf{A}$. So there is $0 \neq \mathbf{v} \in \mathbb{C}^n$ with

$$\mathbf{Av} = \lambda\mathbf{v}$$

So let

$$\bar{\mathbf{v}} = \begin{pmatrix}\bar{\alpha}_1\\ \vdots\\ \bar{\alpha}_n\end{pmatrix}$$

when

$$\mathbf{v} = \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}$$

So

$$\begin{aligned}\bar{\mathbf{v}}^T(\mathbf{A}\mathbf{v}) &= \bar{\mathbf{v}}^T(\lambda\mathbf{v})\\ &= \lambda\bar{\mathbf{v}}^T\mathbf{v}\end{aligned}$$

Note that $\mathbf{A} = \bar{\mathbf{A}} = \bar{\mathbf{A}}^T$. So

$$\begin{aligned}\bar{\mathbf{v}}^T(\mathbf{A}\mathbf{v}) &= (\bar{\mathbf{V}}^T\bar{\mathbf{A}}^T)\mathbf{v}\\ &= (\overline{\mathbf{v}^T\mathbf{A}^T})\mathbf{v}\\ &= (\overline{\mathbf{A}\mathbf{v}}^T)\mathbf{v}\\ &= (\overline{\lambda\mathbf{v}})^T\mathbf{v} = \bar{\lambda}\bar{\mathbf{v}}^T\mathbf{v}\\ \bar{\mathbf{v}}^T\mathbf{v} &= \sum_{j=1}^{n}|\alpha_j|^2 \neq 0\end{aligned}$$

so

$$\lambda = \bar{\lambda}$$

So $\lambda \in \mathbb{R}$.

Lemma 6.5.2. *Suppose* $\mathbf{A} \in M_n(\mathbb{R})$ *is symmetric and* $\lambda, \mu \in \mathbb{R}$ *are distinct eigenvalues of* $\mathbf{A}$. *Suppose* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *are eigenvectors with corresponding eigenvalues* λ, μ. *Then* $\mathbf{u}\cdot\mathbf{v} = 0$.

Proof: As $\mathbf{A}$ is symmetric,

$$(\mathbf{A}\mathbf{u})\cdot\mathbf{v} = \mathbf{u}\cdot(\mathbf{A}\mathbf{v})$$

Thus $\lambda\mathbf{u}\cdot\mathbf{v} = \mu\mathbf{u}\cdot\mathbf{v}$. As $\lambda \neq \mu$ we get $\mathbf{u}\cdot\mathbf{v} = 0$

Theorem 6.5.3. *Suppose* $\mathbf{A} \in M_n(\mathbb{R})$ *is symmetric. Then there exists an orthogonal matrix* $\mathbf{P} \in M_n(\mathbb{R})$ *with* $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$*, a diagonal matrix.*

Proof: By induction on n. $n = 1$ is trivial. Then suppose we have the result for $(n-1)\times(n-1)$ real matrices.

By 6.5.2 there is an eigenvalue $\lambda_1 \in \mathbb{R}$ of $\mathbf{A}$, and let $\mathbf{v}_1$ be the corresponding eigenvector with $\|\mathbf{v}_1\| = 1$.

Let $\mathbf{P}_1$ be an orthogonal $n \times n$ matrix with first column $\mathbf{v}_1$. So

$$\mathbf{P}_1 = \begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix}$$

Then $\mathbf{P}_1^{-1} = \mathbf{P}_1^T$ and

$$\mathbf{P}_1^T\mathbf{A}\mathbf{P}_1 = \begin{pmatrix}\mathbf{v}_1^T\\ \vdots\\ \mathbf{v}_n^T\end{pmatrix}\begin{pmatrix}\mathbf{A}\mathbf{v}_1 & \cdots & \mathbf{A}\mathbf{v}_n\end{pmatrix}$$

Noting that $\mathbf{A}\mathbf{v}_n = \lambda_n \mathbf{v}_n$

$$\mathbf{P}_1^T \mathbf{A} \mathbf{P}_1 = \begin{pmatrix} \lambda_1 & * \\ 0 & \\ \vdots & \mathbf{A'} \\ 0 & \end{pmatrix}$$

This matrix is symmetric as

$$(\mathbf{P}_1^T \mathbf{A} \mathbf{P}_1)^T = \mathbf{P}_1^T \mathbf{A}^T \mathbf{P}_1 = \mathbf{P}_1^T \mathbf{A} \mathbf{P}_1$$

So

$$\mathbf{P}_1^T \mathbf{A} \mathbf{P}_1 = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{A'} & \\ 0 & & & \end{pmatrix}$$

And $\mathbf{A'}$ is symmetric. By the inductive hypothesis there is an orthogonal matrix $\mathbf{P'} \in M_{n-1}(\mathbb{R})$ with $(\mathbf{P'})^T \mathbf{A'} \mathbf{P'}$ diagonal. Now, let

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{P'} & \\ 0 & & & \end{pmatrix} \qquad \in M_n(\mathbb{R})$$

Easily, $\mathbf{P}_2$ is orthogonal, and

$$\begin{aligned} \mathbf{P}_2^T (\mathbf{P}_1^T \mathbf{A} \mathbf{P}_1) \mathbf{P}_2 &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{P'}^T \mathbf{A'} \mathbf{P'} & \\ 0 & & & \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \end{aligned}$$

Let $\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2$. Then $\mathbf{P}$ is orthogonal and

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{P}_2^T \mathbf{P}_1^T \mathbf{A} \mathbf{P}_1 \mathbf{P}_2 \tag{1.1}$$
$$= \operatorname{diag}(\lambda_1, ..., \lambda_n) \tag{1.2}$$

Method for finding P

① Compute eigenvalues $\lambda_1, ...\lambda_r \in \mathbb{R}$ of $\mathbf{A}$

② For each $i \leq r$ find a basis of

$$I_{\lambda_i} = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{A}\mathbf{v} = \lambda_i \mathbf{v}\}$$

Use Gram-Schmidt to obtain an orthonormal basis of E_{λ_i}.

③ Take all of those bases together: we have a basis for $\mathbb{R}^n$. By 6.5.3 it is an orthonormal basis. Take this as the columns of $\mathbf{P}$.

Example 6.5.4. Find an orthogonal matrix $\mathbf{P} \in M_3(\mathbb{R})$ such that $\mathbf{P}^T\mathbf{A}\mathbf{P}$ is diagonal where

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

Solution: The characteristic polynomial is:

$$\begin{aligned} \det \begin{pmatrix} x-1 & 1 & 1 \\ 1 & x-1 & 1 \\ 1 & 1 & x-1 \end{pmatrix} &= (x-1)^3 - (x-1) - (x-2) - (x-2) \\ &= x^3 - 3x^2 + 4 = (x+1)(x-2) \end{aligned}$$

So our eigenvalues are 2 and -1. The eigenspace E_{-1} :

spanned by $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Normalise: $\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Eigenspace E_2:

$$\begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

We obtain 2 linearly independent solutions

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Now we use G-S to get an orthonormal basis.

$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

$$\begin{aligned} \mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{1}_1 \cdot \mathbf{v}_2}{\mathbf{w}_1 \cdot \mathbf{w}_1} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} &= \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \end{aligned}$$

Now normalise

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

Let

$$\mathbf{P} = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{pmatrix}$$

Then

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Remark 6.5.5. If $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ then $\det(\mathbf{P}) = \pm 1$

Chapter 2

Groups

1 Groups + Subgroups

1.1 Binary operations; groups; basic facts

Definition 1.1.1. Suppose S a set. A *binary operation* on S assigns to each ordered pair (a, b) of S an element $(a * b)$ of S

Formally, $*$ is a function

$$S \times S \to S$$

Examples Let $S = M_2(\mathbb{R})$:

1) $a * b$, $*$ is matrix multiplication

2) $*$ matrix addition

3) $a * b = a$

4) $a * b = ab - ba$

5) Let $S_1 \subset S$

$$S_1 = \{a \in M_2(\mathbb{R}) : a \text{ invertible}\}$$

Let $a * b =$ matrix multiplication - binary operation on S_1 as $a * b \in S_1$

Definition 1.1.2. A binary operation $*$ on S is associative if

$$\forall a, b, c \in S, (a * b) * c = a * (b * c)$$

Associativity means that we can unambiguously write an expression such as

$$((a_1 * a_2) * (a_3 * a_4)) * a_5$$

as

$$a_1 * a_2 * a_3 * a_4 * a_5$$

Definition 1.1.3. A *group* $(G, *)$ consists of a set G with a binary operation $*$ on G satisfying:

G1. (Associativity)

$$\forall g, h, k \in G, g * (h * k) = (g * h) * k$$

G2. (Identity axiom)

$$\exists e \in G \text{ such that } \forall g \in G, e * g = g * e = g$$

There is a unique such e, which we will prove and call the identity element of the group.

G3. (Existence of inverses) With e as in G2:

$$\forall g \in G, \exists h \in G \text{ such that } g * h = h * g = e$$

We will show that h here is uniquely determined by g: call h the *inverse* of g, denoted g^{-1}.

Notation and Terminology

1. More common to use $\cdot$ instead of $*$ for the group operation, i.e. write $g \cdot h$. Often we omit it and just write gh. Call the operation the product.

2. A group $(G, *)$ is *abelian* or *commutative* if

$$\forall g, h \in G, g * h = h * g$$

In such cases we sometimes write the operation as $+$; the identity 0 and inverse if a as $-a$.

Justification of 1.3

Suppose $(G, *)$ is a group.

1. If $e, e' \in G$ and

$$\begin{array}{cc} \forall g \in G & e \cdot g = g \cdot e = g \\ \text{and} & e' \cdot g = g \cdot e' = g \end{array}$$

then $e = e'$

Proof: $e = e \cdot e' = e'$ by the above equations.

② If $g, g;, g;; \in G$ and

$$\begin{aligned} gg' &\overset{①}{=} e \overset{②}{=} g'g \\ \text{and } gg'' &\overset{③}{=} e \overset{④}{=} g''g \end{aligned}$$

then $g' = g''$.

Proof:

$$\begin{aligned} (g'g)g'' &\overset{②}{=} eg'' = g'' \\ g'(gg'') &\overset{③}{=} g'e = g' \end{aligned}$$

So by associativity, $g' = g''$

Lemma 1.1.4. (Equations in Groups) *Suppose* $(G, *)$ *is a group and* $g, h \in G$

① *for* $x \in G, gx = h \iff x = g^{-1}h$

② *for* $y \in G, yg = h \iff y = hg^{-1}$

Proof:

① $\Longrightarrow$:

$$\begin{aligned} gx = h &\implies g^{-1}(gx) = g^{-1}h \\ &\implies (g^{-1}g)x = g^{-1}h \\ &\implies ex = g^{-1}h \implies x = g^{-1}h \end{aligned}$$

$\Longleftarrow$: Similar logic in reverse.

② Similar, but multiply on the right by g^{-1}

Lemma 1.1.5. (Inverse of a product) *Suppose* $(G, \cdot)$ *is a group. Then*

① *If* $g, h \in G$ *then:*

$$(gh)^{-1} = h^{-1}g^{-1}$$

② *if* $g_1, ..., g_n \in G$ *then*

$$(g_1 \cdots g_n)^{-1} = g_n^{-1} \cdots g_1^{-1}$$

Proof: Trivial (Good starter exercise)

Example 1.1.6. (From Fields)

① $\mathbb{R}$ with the operation $+$ on a group. The identity element is 0 and inverse of $a \in \mathbb{R}$ is $-a$.

② $\mathbb{R}^x = \mathbb{R} \backslash \{0\}$ with the operation $\cdot$ is a group

③ ① and ② work in any field

④ if F is a field ad $n \in \mathbb{N}$ then $(F^n, +)$ is a group

⑤ if V is a vector space over F then $(V, +)$ is a group.

⑥ Let $n \in \mathbb{N}$ and F a field. The general linear group

$$\mathrm{GL}(n, F)$$

is the set G of $n \times n$ invertible matrices and operation matrix multiplication. This is a group:

- Binary operation: if $A, B \in G$, check $AB \in G$. (Because the inverse of AB is $B^{-1}A^{-1}$)
- Associativity (G1): property of matrix multiplication as defined.
- Existence of Identity (G2): The identity matrix $\mathbf{I}_n$ which is also invertible.
- Existence of Inverses (G3): By definition of G.

1.2 The Symmetric Groups

Definition 1.2.1. Suppose X is any non-zero set. A *Permutation* of X is a bijection $\alpha : X \to X$.

E.g. if $X = \{1, 2, 3, 4\}$,

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

More general notation: if $X = \{1, 2, ..., n\}$ denote a bijection $\alpha : X \to X$ by

$$\alpha = \begin{pmatrix} 1 & 2 & \cdots & n \\ \alpha(1) & \alpha(2) & \cdots & \alpha(n) \end{pmatrix}$$

Example 1.2.2. If $\alpha, \beta : X \to X$ are permutations, so are their compositions:

$$(\alpha \circ \beta)(x) = \alpha(\beta(x))$$

Some more examples can be found in the other notes or if anyone wants to copy them hahaha Let $\mathrm{Sym}(X)$ denote the set of all permutations of X.

Theorem 1.2.3. $\mathrm{Sym}(X)$ *is a group, called the symmetric group on* X *if* $X = \{1, ..., n\}$*, otherwise denoted as* $\mathrm{Sym}(n)$ *or* S_n*.*

Proof: We have a binary operation by the examples. Check the axioms:

G1. (Associativity) Composition of functions is associative. If $\alpha, \beta, \gamma \in \mathrm{Sym}(X)$ then $\alpha \circ (\beta \circ \gamma)$ and $(\alpha \circ \beta) \circ \gamma$ are the same.

G2. (Identity element) the identity function

$$1 : X \to X$$
$$1(x) = x$$

G3. (Existence of Inverses) If $\alpha \in \mathrm{Sym}(X)$ it is a bijection, so has an inverse by the introduction module.

We often write $\alpha\beta$ instead of $\alpha \circ \beta$.

Example 1.2.4. Consider the following elements of S_6

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix}$$
$$\beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 5 & 4 & 3 & 2 \end{pmatrix}$$

Compute

$$\alpha\beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 6 & 5 & 4 & 3 \end{pmatrix}$$
$$\beta\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}$$
$$\alpha^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$
$$\beta^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 5 & 4 & 3 & 2 \end{pmatrix}$$
$$\alpha\beta\alpha^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 6 & 5 & 4 \end{pmatrix}$$

Definition 1.2.5. Say that a group $(G, \cdot)$ is a finite group if the set G is a finite set. In this case the *order* of this group is $|G|$.

Lemma 1.2.6. *If $n \in \mathbb{N}$, then $|S_n| = n!$*

Proof: We have to count the permutations.

$$\alpha = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

$a_1, ...a_n$ are $1, ..., n$ in some order. There are n choices for a_1, $n-1$ choices for a_2, and so on.. so there are $n!$ total possibilities for α.

1.3 Powers and subgroups

Definition 1.3.1. Suppose $(G, \cdot)$ is a group. For $g \in G$, we let

$$g^0 = e, g^1 = g, g^2 = g \cdot g, ...$$

More precisely, for $n \in \mathbb{N}$, we define it inductively:

$$g^0 = e, g^1 = g, g^{n+1} = g^n \cdot g$$

We also define $g^{-n} = (g^{-1})^n$.

Lemma 1.3.2. *With the notation if $m, n \in \mathbb{Z}$, then*

(i) $g^{m+n} = g^m \cdot g^n$

(ii) $(g^m)^{-1} = g^{-m}$

(iii) $g^{mn} = (g^m)^n$

E.g.

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} \in S_4$$
$$g^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$
$$g^3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

So, if $g^4 = 1, g^5 = g, ..., g^{19} = g^3$ since $19 \equiv 3 \mod 4$, and equivalently if $n \equiv k \mod 4$ then $g^n = g^k$.

Proof: (of lemma 1.3.2)

(i) Proof is by induction on n. Our base case $n = 0$:

$$\begin{aligned} g^{m+0} &= g^m \\ g^m g^0 &= g^m e = g^m && \text{as required} \end{aligned}$$

Inductive step: suppose we know $g^{m+n} = g^m g^n$, then

$$\begin{aligned} g^{m+(n+1)} &= g^{(m+n)+1} && \stackrel{\text{def}}{=} g^{m+n} g \\ &= (g^m g^n) g = g^m (g^n g) \\ &\stackrel{\text{def}}{=} g^m g^{n+1} && \text{as required} \end{aligned}$$

We still need to prove the negative case, however.

(ii) By (i)

(iii) Similar, using (i)

Remark 1.3.3. (on additive addition): If our group is $(G, +)$ write $g + g + \cdots + g$ as ng, not g^n.

Definition 1.3.4. Suppose $(G, \cdot)$ a group and $H \subset G$. Say that H is a *subgroup* of $(G, \cdot)$ if H with the binary operation it inherits from G is a group, i.e.

$$\forall h_1, h_2 \in H, h_1 \cdot h_2 \in H$$

and $(H, \cdot)$ satisfies axioms G1, G2, G3.

Example 1.3.5. ① G is a subgroup of G

② $\{e\}$ is a subgroup of G.

Theorem 1.3.6. *(Test for a subgroup.) Suppose $(G, \cdot)$ is a group and $H \subset G$. Then H is a subgroup if and only if*

① $H \neq \emptyset$

② $\forall h_1, h_2 \in H, h_1 \cdot h_2 \in H$ *(Closed under $\cdot$)*

③ $\forall h \in H, h^{-1} \in H$

Example If $g \in G$, let $H = \{g^m : m \in \mathbb{Z}\}$. This is a subgroup of G by the previous theorem and lemma 1.3.2. *Proof of 1.3.3* $\Longleftarrow$: Suppose ① ② ③ hold. By ②, we have a binary operation on H given by $\cdot$. So we check if $(H, \cdot)$ satisfies G1,G2 and G3.

G1. Follows from associativity in G.

G2. Enough to show that $e_G \in H$. By ① there is some $h \in H$. By ③, $h^{-1} \in H$. So then we have $e_G = h^{-1}h \in H$.

G3. Follows from ③

$\Longrightarrow$: if H is a subgroup of G, then ② holds by definition. By G2, $H \neq \emptyset$ so ① holds. For ③, first show $e_G \in H$. Let $h \in H$. As H is a subgroup, there is some $x \in H$ with $hx = h$. But the only solution to this equation in G is $x = e$, so $x = e \in H$. Similarly, the only solution to $hh' = e$ in G is $h' = h^{-1}$. So as H is a group, $h^{-1} \in H$.

Remark 1.3.7. $\Longrightarrow$ shows that if H is a subgroup of G, then e_G is in H and inverses are the same in H as they are in G.

Definition 1.3.8.

i) Suppose $(G, \cdot)$ a group[and $g \in G$. The *cyclic subgroup* generated by G is $< g >= \{g^m : m \in \mathbb{Z}\}$.

ii) G is *cyclic* if there is some $g \in G$ with $< g >= G$. g is a *generator* of G.

Example 1.3.9. ① Let $G = GL(n, \mathbb{R})$. ($n \times n$ invertible matrices). Here are some subgroups:

1) Let $H = \{g \in G : \det(g) = 1\}$. Check:

- $H \neq \emptyset$ as $I_n \in H$.
- H closed under $\cdot$:

  $$\det(g_1, g_2) = \det(g_1)\det(g_2)$$

  so if $g_1, g_2 \in H$, then $g_1 g_2 \in H$
- H closed under inverses:

  $$\det(g^{-1}) = \frac{1}{\det(g)}$$

  So if $h \in H$ then $h^{-1} \in H$.

② $K = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\} \leq GL(2, \mathbb{R})$, or the group of rotations about O. This group is not cyclic, but it is abelian. It's also uncountable.

③ We also have

$$\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq (\mathbb{C}, +)$$

We have $\mathbb{Z} =< 1 >$ and therefore is cyclic.

④ $U = \{e^{i\theta} : \theta \in \mathbb{R}\} \leq (\mathbb{C}^x, \cdot)\}$ U can also be written as $\{z \in \mathbb{C} : |z| = 1\}$. Since $z\bar{z} = 1$, we have $z^{-1} = \bar{z}$. Group isn't cyclic, but it is abelian.

⑤ Let $n \in \mathbb{N}$ and $\zeta = e^{2\pi i/n}$:

$$< \zeta >= \{1, \zeta, \zeta^2, ..., \zeta^{n-1}\} \leq U(\leq \mathbb{C}$$

The entries of $< \zeta >$ form the vertices of a regular n-gon around the unit circle, like roots of unity. For every $n \in \mathbb{N}$, there is a cyclic group of order n!.

⑥ Let F be a field and consider $(F^n, +)$. Any subspace of this is also a subgroup. But the converse is not necessarily true, e.g.:

$$(\mathbb{Q}^2, +) \leq (\mathbb{R}^2, +)$$

But The former isn't a subspace because it's not closed under scalar multiplication.

⑦ Let

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$
$$\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$
$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$
$$V = \{1, \alpha, \beta, \gamma\} \leq S_4.$$

This group isn't cyclic, but it is abelian.

Notation: We often write 'G is a group' rather than '$(G, \cdot)$ is a group' and assume the multiplication operation. If $H \subset G$ and H is a subgroup, we indicate this by $H \leq G$.

1.4 Orders of Elements

Definition 1.4.1. Suppose G a group and $g \in G$. Say that g has *finite order* if there is no $n \in \mathbb{N}$ such that

$$g^n = e \, (\mathbb{N} = \{1, 2, ...\})$$

In this case the smallest $n \in \mathbb{N}$ with $g^n = e$ is called the *order* of g. (Denoted by $\mathrm{ord}(g)$). If there is no such n, we say g has infinite order.

Examples:

① $e \in G$ has order 1

②

$$g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \in S_3$$

has order 3

$$g = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \in S_3$$

has order 2.

③ $2 \in (\mathbb{R}^x, \cdot)$ has infinite order. -1 has order 2

④ $\zeta = e^{\frac{2\pi i}{n}} \in (\mathbb{C}^x, \cdot)$ has order n

⑤

$$g = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \in GL_3(\mathbb{R})$$

has order 3.

Theorem 1.4.2. *Suppose G is a finite group.*

1) *Every $g \in G$ has finite order*

2) *If $H \subset G$ and*

   i) *$H \neq \emptyset$*

   ii) *if $h_1, h_2 \in H$ then $h_1 h_2 \in H$*

   then H is a subgroup of G.

Proof:

1) Consider
$$g, g^2, ... \in G$$
As $|G|$ is finite there are $0 < m < n$ with
$$g^m = g^n \implies g^{n-m} = e$$
So g has finite order.

2) We have to show H is closed under inverses. Let $h \in H$. By part 1) there is $n \in \mathbb{N}$ such that $h^n = e$. We want to show that $h^{-1} \in H$. Can assume $h \neq e$, so $n > 1$. Then
$$h^{-1} = h^{n-1}$$
and by ii) $h^{n-1} \in H$ (as $h \in H$).

Some things to come later:

Ⓐ If G is a finite group and $g \in G$, then $\mathrm{ord}(g)$ divides $|G|$

Ⓑ This gives a nice way of computing orders of elements in S_n

1.5 More on cyclic groups

Theorem 1.5.1. *Suppose $(G, \cdot)$ a cyclic group, and $G = < g >$.*

① *If $H \leq G$ then H is cyclic.*

② *Suppose $|G| = n$ (i.e. g has order n), and $m \in \mathbb{Z}$. Let $d = \gcd(m, n)$ then*
$$< g^m > = < g^d >$$
and
$$| < g^d > | = \frac{n}{d}$$
(So $< g^m > = G \iff d = 1 \iff \gcd(m, n) = 1 \iff m, n$ co-prime)

③ *if $|G| = n$ and $k \leq n$, then G has a subgroup of order k iff $k|n$. In this case, the subgroup is $< g^{\frac{n}{k}}$*

Example:
$$g = e^{\frac{2\pi i}{6}} \in (\mathbb{C}^x, \cdot)$$
has order 6. The subgroups of $G = < g >$ are of orders 1,2,3,6:
$$\{1\}$$
$$< g^3 > = \{1, -1\}$$
$$< g^2 > = \{1, e^{\frac{2\pi i}{3}}, e^{\frac{4\pi i}{3}}\}$$
$$< g > \text{- order 6}$$

Proof:

① May assume $H \neq \{e\}$. Let d be the least element of

$$\{n \in \mathbb{N} : g^n \in H\}$$

Claim: $H =< g^d >$. As $g^d \in H$ and $H \leq G$, we have $< g^d > \subset H$. Let $h \in H$. So $h = g^m$ for some $m \in \mathbb{Z}$.

Write $m = qd + r$ where $q, r \in \mathbb{Z}$ and $0 \leq r < d$, then

$$\begin{aligned} h &= g^m = g^{qd+r} \\ &= (g^D)^q g^r \\ \therefore g^r &= h(g^d)^{-q} \in H \end{aligned}$$

As $h \in H$ and $g^d \in H$, we have by minimality of d that $r = 0$. So

$$\begin{aligned} h &= g^{qd} \\ &= (g^d)^q \\ \text{i.e. } h &\in < g^d > \end{aligned}$$

② As $d = \gcd(m, n)$ there are $k, l \in \mathbb{Z}$ with

$$d = km + ln$$

To show $< g^m >=< g^D >$ it is enough to prove $g^m \in < g^d >$ and $g^d \in < g^m >$. Now, as $d|m, g^m$ is a power of g^d, so the first is true. For the second,

$$\begin{aligned} g^d &= g^{km+ln} \\ &= (g^m)^k (g^n)^l \\ &= g^m)^k && \text{as n =ord}(g) \\ &\in < g^m > \end{aligned}$$

As $d|n$ we can write $n = df, f \in \mathbb{N}$. Then $< g^d >= \{g^0, g^d, .., g^{(f-1)d}\}$ $g^0, g^d, ..., g^{(f-1)d}$ are distinct as $d, ..., (f-1)d$ are $< n$.

$$\therefore | < g^d > | = f = \frac{n}{d}$$

③ By ① and ②, the unique subgroup with $\frac{n}{d}$ elements is $< g^d >$.

Application: For $n \in \mathbb{N}$, the *Euler totient function* is

$$\begin{aligned} \phi &: \mathbb{N} \to \mathbb{N} \\ \phi(n) &\mapsto |\{k \in \mathbb{N} : 1 \leq k \leq n \wedge \gcd(k, n) = 1\}| \end{aligned}$$

Theorem 1.5.2.

$$\sum_{1 \leq d \leq n, d|n} \phi(d) = n$$

Proof: let G be a cyclic group of order n. By 1.5.1, if $d|n$ then G has a unique subgroup G_d which contains every element of G of order d. This is cyclic by 1, and by 2 G_d has $\phi(d)$ elements of order d. Thus G has $\phi(d)$ elements of order G. Then any element of G has order dividing $n = |G|$, so

$$\sum_{d|n} \phi(d) = n$$

By counting elements of G according to their possible orders.

Definition 1.5.3. Suppose $(G, \cdot)$ a group and $S \subset G (S \neq \emptyset)$. Let $S^{-1} := \{g^{-1} : g \in S\}$ and

$$< S >= \{g_1, g_2, ..., g_k : k \in \mathbb{N} \wedge g_1, ..., g_k \in S \cup S^{-1}\}$$

Or, the set of all possible products of elements of S and their inverses. This allows repetitions.

Lemma 1.5.4. *With this notation:*

1) $< S >$ is a subgroup of G

2) if $H \leq G$ and $S \subset H$ then $H \geq < S >$

So $< S >$ is the smallest subgroup of G containing S. It is called the subgroup *generated* by S.

If $S = \{x_1, ..., x_r\}$ write $< S >$ as $< x_1, ..., x_r >$. If G is abelian, then

$$< x_1, ..., x_r >= \{x_1^{k_1}, x_2^{k_2}, ..., x_r^{k_r} : k_1, ..., k_r \in \mathbb{Z}\}$$

2 Lagrange's Theorem + Cosets

Theorem 2.0.1. (Lagrange)*: Suppose $(G, \cdot)$ a finite group and $H \leq G$. The $|H|$ divides $|G|$.*

Example: S_5 has order $5! = 120$ has no subgroup of order 50.

Theorem 2.0.2. *Suppose G is a finite group and $|G| = n$. Let $g \in G$, then*

1 *The order of g divides n,*

2 $g^n = e$

Proof:

1 The order of g is $| < g > |$. $< g >$ is a subgroup of G and so this follows from Lagrange's theorem.

2 Suppose $\text{ord}(g) = k$. By 1, $k|n$. Then

$$g^n = (g^k)^{\frac{n}{k}} = e^{\frac{n}{k}} = e$$

Corollary. *(Fermat's Little Theorem): Suppose p is any prime number. If $x \in \mathbb{Z}$ and $p \nmid x$ then $x^{p-1} \cong 1 \mod p$.*

Proof: Let $\mathbb{F}_p$ be the field with p elements. Consider $(\mathbb{F}_p^x, \cdot)$ the multiplicative group of non-zero elements

$$|\mathbb{F}_p^x| = p - 1$$

So for every $g \in \mathbb{F}_p^x$,

$$g^{p-1} = [1]_p$$

The residue class of $1 \mod p$. Then if $n \in \mathbb{Z}$ and $p \nmid x$ then $[x]_p \neq [0]_p$. So take $g = [x]_p$, we obtain

$$[x]_p^{p-1} = [x^{p-1}]_p = [1]_p$$

i.e. $x^{p-1} \equiv 1 \mod p$

Example 2.0.3. $G = S_3$. Let

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

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

Then $< \alpha, \beta >= G$. We have $\mathrm{ord}(\alpha) = 3, \mathrm{ord}(\beta) = 2$. Then $< \alpha, \beta >$ has subgroups:

$$\begin{array}{ll} < \alpha > & \text{of order } 3 \\ < \beta > & \text{of order } 2 \end{array}$$

Then by Lagrange's theorem we have $2 || < \alpha, \beta > |$ and $3 || < \alpha, \beta > |$ so $6 || < \alpha, \beta > |$. And as $|G| = 6$, we have $< \alpha, \beta >= G$

Theorem 2.0.4. *Suppose p is a prime and $(G, \cdot)$ is a group of order p. Then G is cyclic. In fact, if $g \in G, g \neq e$, then $< g >= G$.*

Proof: Let $g \in G, g \neq e$. Then $| < g > |$ divides $p = |G|$ (Lagrange). And, $| < g > | \geq 2$ (as $e, g \in < g >$) So $| < g > | = p$.

2.1 Cosets

Definition 2.1.1. Suppose $(G, \cdot)$ a group and $H \leq G$. Let $g \in G$. The subset

$$g^H := \{gh : h \in H\} \subset G$$

Is called a *left coset* of H in G. (Sometimes called *H-coset*

If $H = \{h_1, ..., h_r\}$ then $g^H := \{gh_1, ..., gh_r\}$.
In past papers, questions worked with right cosets:

$$Hg := \{hg : h \in H\}$$

But we'll stick with left cosets for now.

Example 2.1.2. ①

$$\begin{aligned} G &= (\mathbb{C}^x, \cdot) \\ H &= \{z \in \mathbb{C}^x : |z| = 1\} \end{aligned}$$

let $g = 2$

$$2H = \{2e^{i\theta} : \theta \in \mathbb{R}\} \qquad\qquad = \{z \in \mathbb{C}^x : |z| = 2\}$$

Generally, if $w \in \mathbb{C}^x$, then $wH = \{z \in \mathbb{C}^x : |z| = |w|\}$

② Let $G = (\mathbb{Z}, +), H = \{5m : m \in \mathbb{Z}\}$. Write the cosets additively:

$$\begin{aligned} 0 + H &= H \\ 1 + h &= \{1 + 5m : m \in \mathbb{Z}\} \\ &= \{k \in \mathbb{Z} : k \equiv 1 \mod 5\} \qquad\qquad = [1]_5 \\ &\vdots \\ 4 + H &= [4]_5 \\ &\vdots \\ 6 + H &= \{6 + 5m : m \in \mathbb{Z}\} \\ &= \{1 + 5m : m \in \mathbb{Z}\} \\ &= [1]_5 \end{aligned}$$

So there are exactly 5 left cosets: $0 + H, 1 + H, 2 + H, ..., 4 + H$

③ Let

$$\begin{aligned} \mathbf{A} &\in M_{m \times n}(\mathbb{R}) \\ W &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = 0_m\} \leq \mathbb{R}^n \end{aligned}$$

Suppose $\mathbf{b} \in \mathbb{R}^m$ and there is $\mathbf{c} \in \mathbb{R}^n$ with $\mathbf{Ac} = \mathbf{b}$

$$\begin{aligned} \mathbf{Ax} = \mathbf{b} &\iff \mathbf{A}(\mathbf{x} - \mathbf{c}) = 0 \\ &\iff \mathbf{x} - \mathbf{c} \in W \\ &\iff \mathbf{x} \in \mathbf{c} + W \end{aligned}$$

So the solutions to $\mathbf{Ax} + \mathbf{b}$ are a coset of W in $\mathbb{R}^n$

Lemma 2.1.3. *Suppose $(G, \cdot)$ a group and $H \leq G$.*

① *If $_1, g_2 \in G$ and $g_2 \in g_1 H$ then $g_2 H = g_1 H$.*

② *If $g, h \in G$ and $gH \cap kH \neq \emptyset$ then $gH = kH$*

Proof:

① First, prove that if $g_2 \in g_1H$ then $g_2H \subseteq g_1H$. As $g_2 \in g_1H$ there is $h \in H$ with $g_2 = g_1H$. Any element of g_2H is of the form g_2h' for some $h' \in H$. Then

$$g_2h' = (g_1h)h' = g_1(hh')$$

as $H \leq G, hh' \in H$ So $g_2h' \in g1'H \therefore g_2H \subseteq g_1H$

Also, $g_1 = g_2h^{-1}$ as $h^{-1} \in H$. The same argument gives

$$g_1H \subseteq g_2H$$

② let $x \in gH \cap kH$. By ① twice:

$$gH = xH = kH$$

Lemma 2.1.4. *Suppose* $(G, \cdot)$ *a group and* $H \leq G$. *If* $g \in G$, *the map*

$$H \to gH$$

given by

$$h \mapsto gh$$

is a bijection. So if H *is finite, then* $|H| = |gH|$.

Proof: By definition, the map is surjective. If $gh_1 = gh_2$ then multiplying by g^{-1} gives us $h_1 = h_2$. So map is also injective.

Proof of Lagrange's theorem: We have $(g, \cdot)$ a finite group and $H \leq G$. Then we want to prove $|H| \mid |G|$.

Consider the left cosets of H in G. Any one of these has $|H|$ elements. Also, any two of them are disjoint. Any $g \in G$ lies in some H-coset, namely g^H, so $|G| = |H| \times$ number of distinct H-cosets in G So $|H|$ divides $|G|$.

Definition 2.1.5. The number of left cosets of H in G is called the *index* of H in G.

Another proof of Lagrange:

Theorem 2.1.6. *Suppose* $(G\cdot)$ *is a group and* $H \leq G$. *define the relation on* G *by*

$$g \sim k \iff g^{-1}k \in H$$

① *is an equivalence relation*

② $g\ k \iff g^H = k^H$

Proof:

① Question sheet 6

②

$$\begin{aligned} & g^{-1}k \in H \\ \iff & g^{-1}kH = H \\ \iff & kH = gH \end{aligned}$$

The equivalence classes are the left H-cosets.

Example 2.1.7. Let $G = S_3$, and $H =< \alpha >, \alpha = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$. What are the left H-cosets?

e	$\beta = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$	$\gamma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$
α	$\beta\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$	$\gamma\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$

Solution: We have

$$\begin{aligned} & \beta H = \{\beta, \beta\alpha\} \\ & H\beta = \{\beta, \alpha\beta\} \\ & \beta H \neq H\beta \end{aligned}$$

3 Homomorphisms

Definition 3.0.1. ① Suppose $(G, \cdot)$ and $(H, \cdot)$ are groups. A function $\phi : G \to H$ is called a *homomorphism* if

$$\forall g_1, g_2 \in G, \phi(g_1 g_2) = \phi(g_1)\phi(g_2)$$

② The image of ϕ is

$$\mathrm{Im}\phi = \{\phi(g) : g \in G\}$$

and then kernel of ϕ is

$$\mathrm{Ker}\phi = \{g \in G : \phi(g) = e_h\}$$

③ If the homomorphism ϕ is a bijection, say ϕ is an *isomorphism*. For groups G, H if there exists an isomorphism $\phi : G \to H$ then we say G, H are *isomorphic*, or write $G \cong H$.

Lemma 3.0.2. *Suppose G, H are groups, and $\phi : G \to H$ is a homomorphism. Then*

i) $\phi(e_G) = e_H$

ii) $\phi(g^{-1}) = (\phi(g))^{-1} \forall g \in G$

iii) $\text{Im}\phi \leq H, \text{Ker}\phi \leq G$

Proof:

i)

$$\begin{aligned}\phi(e_G) &= \phi(e_G e_G) \\ &= \phi(e_G)\phi(e_G) \\ (h = hh &\implies e_H = h)\end{aligned}$$

so

$$\phi(e_G) = e_H$$

ii)

$$\begin{aligned}e_H \overset{\text{(i)}}{=} \phi(e_H) &= \phi(gg^{-1}) \\ &= \phi(g)\phi(g^{-1}) \\ \therefore \phi(g^{-1}) &= (\phi(g))^{-1}\end{aligned}$$

iii) use the test for a subgroup (1.3.6)

Example 3.0.3.

① Trivial Examples:

$$\begin{aligned}i : G &\to G \\ i(g) &= g\end{aligned}$$

Is the identity homomorphism.

$$\begin{aligned}\psi : G &\to H \\ \psi(g) &= e_H \; \forall g \in G\end{aligned}$$

② F is a field, $G = \text{GL}_n(F)$. then

$$\det : \text{GL}_n(F) \to (F^x, \cdot)$$

is a homomorphism:

$$\det(g_1 g_2) = \det(g_1)\det(g_2)$$

③ Suppose $(H, \cdot)$ is any group and $h \in H$. Define

$$\begin{aligned}\phi : (\mathbb{Z}, +) &\to H \\ \phi(n) &= h^n \\ \phi(n+m) &= h^{n+m} = h^n h^m \\ &= \phi(n)\phi(m)\end{aligned}$$

Therefore ϕ is a homomorphism. If h has infinite order, then

$$\text{Ker}\phi = \{0\}$$

If h has finite order, then

$$\text{Ker}\phi = n\mathbb{Z}$$

$$\text{Im}\phi = < h >$$

④

$$\exp : (\mathbb{R}, +) \to (\mathbb{R}^{>0}, \cdot)$$

Is a homomorphism:

$$\exp(x + y) = \exp(x)\exp(y)$$

This is an isomorphism

⑤

$$| \, |(\mathbb{C}^x, \cdot) \to (\mathbb{R}^x, \cdot)$$

(Modulus)

$$|z_1 z_2| = |z_1||z_2|$$

Which means it's a homomorphism. The kernel is

$$\text{Ker}| \, | = \{z \in \mathbb{C}^x : |z| = 1\}$$

Lemma 3.0.4.

i) A homomorphism $\phi : G \to H$ is injective if and only if

$$\text{Ker}\phi = \{e_G\}$$

ii) If $\phi : G \to H$ and $\psi : H \to K$ are homomorphisms, then

$$\psi \circ \phi : G \to K$$

is a homomorphism.

iii) If $\phi : G \to H$ is an isomorphism, then $\phi^{-1} : H \to G$ is an isomorphism.

Proof: " $\implies$ " We know

$$\phi(e_G) = e_H$$

So $e_G \in \text{Ker}\phi$. Injectivity says that $|\text{Ker}\phi| \leq 1$.

" $\impliedby$ ": Suppose $\text{Ker}\phi = \{e_G\}$.

$$\begin{aligned} \phi(g_1) = \phi(g_2) & \underset{3.0.2}{\implies} \phi(g_1 g_2^{-1}) \in \text{Ker}\phi \\ & \implies g_1 g_2^{-1} \in \text{Ker}\phi \\ & \implies g_1 g_2^{-1} = e^G \\ & \implies g_1 = g_2 \end{aligned}$$

□

Theorem 3.0.5.

① *Suppose G, H are cyclic groups of the same order. The there is an isomorphism $\alpha : G \to H$.*

② *If V_1, V_2 are non-cyclic groups of order 4, then $V_1 \cong V_2$*

Proof:

① Case 1: Suppose G, H are finite cyclic groups of order n.

$$G =< g >$$
$$H =< h >$$

Define $\alpha : G \to H$ by $\alpha(g^k) = h^k$ for $k \in \mathbb{Z}$. This is well-defined, i.e.

$$\begin{aligned} g^k = g^l &\implies h^k = h^l \\ g^k = g^l &\implies g^{k-l} = e^g \\ &\implies n \mid k - l \end{aligned}$$

as g, h have the same order n

$$\begin{aligned} &\implies h^{k-l} = e_H \\ &\implies h^k = h^l \end{aligned}$$

This is injective too as all arrows reverse :) trust me :)))) This gives α is a bijection, and

$$\begin{aligned} \alpha(g^k g^l) &= \alpha(g^{k+l}) \\ &= h^{k+l} = h^k h^l \\ &= \alpha(g^k)\alpha(g^l) \end{aligned}$$

so α is a homomorphism $\implies$ isomorphism.

② Case 2: G, H are of infinite order. The only thing we need to change is the proof that α is well defined:

$$g^k = g^l \implies g^{k-l} = e_G$$

But as g has infinite order, this implies $k - l = 0$, i.e. $k = l$.

□

So, this ends the notes that I typed up at college cause this is where social distancing came into effect and lectures were cancelled.

Acknowledgements

Many, many thanks to Amir, Rahul, and others for helping me type out this document and for making some of the Tikz diagrams earlier on. Everyone's effort is really appreciated!