1 Introduction

Dr. Charlotte Kestner
c.kestner@imperial.ac.uk
Huxley 620

1.1 Authorship Details

The majority of these notes were written by Dr. Charlotte Kestner (me), with a few edits here and there to update them by Robert Barham. The exception to this is this introduction, which Robert Barham completely rewrote, and I have decided to keep in the most part.

The problem sheets have underwent substantially more changes whilst Robert taught last year, and I have decided to keep the format he introduced.

Obviously the exams are not the same year on year!

Please refer to the Blackboard page for information about assessments, problem sheets and classes, and other things that are about the module, and not the material.

1.2 What is Linear Algebra?

Linear Algebra is (this is a lie, forgive me, but I'll fix it next page) the study of linear equations, with a particular emphasis on how to solve them. As you probably already know, a system of linear equations is a list of equations which look something like this:

$$\begin{array}{rcrcrcl} x & + & 2y & + & z & = & 1 \\ 2x & + & 3y & + & 4z & = & 0 \\ \pi x & + & \pi y & + & z & = & \frac{1}{2} \end{array}$$

Here, x, y, and z are our variables. We're allowed to multiply the variables by fixed coefficients, and we're allowed to add variables together, but we can't multiply variables together. A surprising number of problems can be formalised into linear equations.

When systems of linear equations do and do not have solutions, as well as how to find those solutions, is something that we can completely answer in this module. Not just for three variables, but for any number of them. We won't be stopping there, though, and we'll be taking this material further than this question.

Of course, like all mathematics, the ideas and techniques developed for one purpose turns out to be really good at a whole bunch of other things. The least surprising of these 'other things' is vector geometry, and much of this course can be visualised in terms of transformations of space.

1.3 History of Linear Algebra

This section is a summary of the information in Chapter 5 of the book 'History of Linear Algebra' by Isreal Kleiner (Kleiner I. (2007) History of Linear Algebra. In: Kleiner I. (eds) A History of Abstract Algebra. Birkhäuser Boston. https://doi.org/10.1007/978-0-8176-4685-1_5). Anything **not** in italics is taken directly from that book. It's worth a read; Imperial library has online access to it. I think you'll enjoy it more if you read it at the end of your degree than at the beginning, though.

Linear Algebra has been studied by the very earliest mathematicians. The Babylonians knew how to solve two simultaneous equations with two variables (2×2). The 'Nine Chapters of the Mathematical Arts (Jiǔzhāng Suànshù)' (One of the earliest surviving mathematics textbooks, the final version dates from the 2nd Century BCE, but first versions appeared in the 10th Century BCE.) contains a method to solve 3×3 systems. More modern contributions come from Leibniz (definition of the determinant), Cramer (Cramer's Rule), Euler (a study of the uniqueness of solutions), and Gauss (Gaussian elimination/Row Reduction to Echelon form).

The impression I get from Kleiner's book is that the ancient methods bear much more similarity to modern techniques than they do in other subjects. The Renaissance methods of solving polynomial equations look pretty strange to me, and I can't imagine doing dynamics or geometry without vectors, which are a 19th century invention. However, the methods that Kleiner describe feel very familiar.

The matrix notation we use today is sort-of present in Chinese texts, where they're used as an abbreviated way of writing out a system of linear equations. The idea of treating these arrays of numbers as objects in their own right, e.g. adding them, multiplying them, classifying them by shape, appears implicitly in the work of Gauss, but it's Cayley in 1850 who first explicitly champions that approach.

In this mid-19th Century period, a huge amount of work was done on matrices, but these results were not viewed as being all being parts of the same general theory. Peano gave the general framework that we use today, by defining the Vector Space (*I can now correct my earlier lie. Linear Algebra is the study of Vector Spaces*) in 1888, and the first modern usage of the term 'Linear Algebra' appears in the book 'Modern Algebra' by van der Waerden in 1930.

1.4 What are Groups?

That's next term. I can tell you now that they're fun!

2 Systems of Linear Equations

2.1 Introduction

This section is all about methods for solving systems of linear equations. A system of linear equations is a set of equations in the same variables. For example:

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

This system has three equations and three unknowns, but in general this could be different. For example:

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

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

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n &= b_2 \\ \vdots &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

Definition 2.1.1 *Given a system of m linear equations in n unknowns we can write this in matrix form as follows:*

$$\mathsf{AX} = \mathsf{B}$$

where $\mathsf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ *and* $\mathsf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ x_m \end{pmatrix}$ *are column matrices, and*
$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$ *is an $m \times n$ matrix.*

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

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

Example 2.1.2.

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

Could be written as

$$\begin{pmatrix} 1 & -1 & 1 & 2 \\ 1 & 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$$

The Augmented matrix would be:

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

Remark 2.1.3 *You should have seen some matrix multiplication already (e.g. see Problem sheet 0 on blackboard). Notice that matrix multiplication is defined precisely so that the above equation works out.*

2.2 Matrix Algebra

We will very briefly go over Matrix algebra. You should make sure you go over the exercises on Problem sheet 0. For the moment we will mostly assume that the matrices take their values in $\mathbb{R}$ (at the end of this section we will see that we could have chosen to take values from any *Field* F).

If we want to add two matrices, they must have the **same size and shape (the same order)**. Then we can simply **add corresponding elements**. Formally:

Definition 2.2.1. Given $m \times n$ matrices, $\mathsf{A} = [a_{ij}]_{m\times n}$ and if $\mathsf{B} = [b_{ij}]_{m\times n}$, then the **(matrix) sum of** A **and** B is the $m \times n$ matrix $\mathsf{C} = [c_{ij}]_{m\times n}$ where $\boxed{c_{ij} = a_{ij} + b_{ij}.}$

We can also multiply by a scalar product (any element of the field - here $\mathbb{R}$):

Definition 2.2.2. Let $\mathsf{A} = [a_{ij}]$ be any matrix, and let $\lambda \in \mathbb{R}$. Then the **scalar multiple of** A by λ, denoted by $\lambda\mathsf{A}$, is obtained by multiplying every element of A by λ. Thus if $\mathsf{A} = [a_{ij}]_{m\times n}$ then $\boxed{\lambda\mathsf{A} = [\lambda a_{ij}]_{m\times n}.}$

See the handout sheet for properties of matrix addition and scalar multiplication.

We can also multiply two matrices together.

Definition 2.2.3. Let $\mathsf{A} = (a_{ij})_{p\times q}$ and $\mathsf{B} = (b_{ij})_{q\times r}$. Then the **matrix product of** A **and** B, denoted by AB, is the matrix C, where

$$\mathsf{C} = (c_{ij})_{p\times r}, \qquad \text{where } \boxed{c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj}}$$

Hopefully you will have done lots of examples of this already. Let's have look at some properties of matrix multiplication.

Theorem 2.2.4. Matrix multiplication is associative. That is
Let A, B, C be matrices, and $\alpha \in \mathbb{R}$, then $(\mathsf{AB})\mathsf{C} = \mathsf{A}(\mathsf{BC})$.

Proof For $\mathsf{A}(\mathsf{BC})$ to be defined, we require the respective sizes of the matrices to be $m\times n, n\times p, p\times q$ in which case the product $\mathsf{A}(\mathsf{BC})$ is also defined. Calculating the (i,j)th element of this product, we obtain,

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

If we now calculate the (i,j)th element of $(\mathsf{AB})\mathsf{C}$ we obtain the same result:

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

Consequently, we see that $\mathsf{A}(\mathsf{BC}) = (\mathsf{AB})\mathsf{C}$.

Example 2.2.5. Matrix multiplication is not commutative (i.e. $AB \neq BA$)

Proof: To show this we just need one counterexample. Lets try to make it as simple as possible.

- 1×1 matrices - multiplying these is just like multiplying elements of $\mathbb{R}$ and that is commutative!
- So we have to look at the 2×2 matrices.

  $$\begin{aligned}
  AB &= \begin{pmatrix} a_{11} & a_{12} \\ * & * \end{pmatrix}\begin{pmatrix} * & b_{12} \\ * & b_{22} \end{pmatrix} = \begin{pmatrix} * & a_{11}b_{12} + a_{12}b_{22} \\ * & * \end{pmatrix} \\
  BA &= \begin{pmatrix} b_{11} & b_{12} \\ * & * \end{pmatrix}\begin{pmatrix} * & a_{12} \\ * & a_{22} \end{pmatrix} = \begin{pmatrix} * & b_{11}a_{12} + b_{12}a_{22} \\ * & * \end{pmatrix}
  \end{aligned}$$

  Set $a_{11} = b_{12} = a_{12} = b_{22} = b_{11} = 1$...get $AB = BA$ only if $a_{22} = 1$
- *****MENTIMETRE**********Is there another way of seeing this in full generality?*

Exercise 2.2.6. Let A, B be matrices with entries in $\mathbb{R}$. Show $\lambda AB = A(\lambda B)$.

Proof This splits into two cases:

- AB is not defined therefore λAB is not either. Also B is the same order as λB thus $A(\lambda B)$ is also not defined.
- In the case where AB is defined let $c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj}$ be such that c_{ij} is the ij^{th} entry of AB. Then the ij^{th} entry of λAB is λc_{ij}.

  The ij^{th} entry of $A(\lambda B)$ is

  $$\begin{aligned} d_{ij} &= \sum_{k=1}^{q} a_{ik}\lambda b_{kj} \\ &= \lambda \sum_{k=1}^{q} a_{ik}b_{kj} \\ &= \lambda c_{ij} \end{aligned}$$

  As the two matrices have the same entries they are equal.

2.3 Row Operations

Recall the definition of an Augmented Matrix from the first lecture. Here's an example to help.

Exercise 2.3.1. Find the Augmented matrix for the following system of linear equations

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

$$\begin{pmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 4 \end{pmatrix}$$

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

From School you know how to solve systems of linear equations. There are 3 operations you can do:

- multiply an equation by a non-zero factor.
- Add a multiple of one equation to another
- Swap equations around.

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

Definition 2.3.2. Elementary row operations (e.r.o's) are performed on an augmented matrix. There are three allowable operations:

- **Multiply** a row by any (non-zero) number
- **Add to any row a multiple of another row**
- **Interchange** two rows

Note that the elementary row operations amount to the actions we could take on the original equations.

Remark 2.3.3 *1. Performing row operations preserves the solutions of a linear system.*

2. Each row operation has an inverse row operation.

Example 2.3.4.

$$\begin{aligned} 3x - 2y + z &= -6 \quad &(1)\\ 4x + 6y - 3z &= 5 \quad &(2)\\ -4x + 4y &= 12 \quad &(3) \end{aligned}$$

First multiply (3) by $\frac{1}{4}$:

$$-x + y = 3 \qquad (4)$$

Then add $3 \times (4)$ to (1) and $4 \times (4)$ to (2)

$$\begin{aligned} y + z &= 3 \quad &(5)\\ 10y - 3z &= 17 \quad &(6) \end{aligned}$$

Then take $10 \times (5)$ from (6)

$$-13z = -13 \qquad (7)$$

So $z = 1$. Plug this into (5):

$$y + 1 = 3$$

So $y = 2$. Plug this into (4):

$$-x + 2 = 3$$

So $x = -1$

$$\left(\begin{array}{ccc|c} 3 & -2 & 1 & -6\\ 4 & 6 & -3 & 5\\ -4 & 4 & 0 & 12 \end{array}\right) \xrightarrow{R_3 \mapsto \frac{1}{4}R_3}$$

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

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3\\ 0 & 10 & -3 & 17\\ -1 & 1 & 0 & 3 \end{array}\right) \xrightarrow{R_2 \mapsto R_2 - 10R_1}$$

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

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

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

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2\\ 0 & 0 & 1 & 1\\ 1 & 0 & 0 & -1 \end{array}\right) \xrightarrow[R_3 \mapsto R_1]{R_1 \mapsto R_2, \quad R_2 \mapsto R_3}$$

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

We can read this off:

$$\begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x\\ y\\ z \end{pmatrix} = \begin{pmatrix} -1\\ 2\\ 1 \end{pmatrix}$$

So we get $x = -1$, $y = 2$, $z = 1$.

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

- They are both inconsistent.
- The augmented matrix of the first system can be obtained using row operations from the augmented matrix of the second system and vice versa.

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 set of solutions.*

If a row consists of mainly 0s and 1s it becomes easier to read off the solutions to the equations. For example:

Example 2.3.7. If we are working in unknowns x, y, z:

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

Corresponds to

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

Whereas

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

Corresponds to

$$\begin{aligned} y &= 2 \\ z &= 5 \end{aligned}$$

Definition 2.3.8. We say a matrix is in **echelon form (ef)** if must satisify the following:

- All of the zero rows are at the bottom.
- The first non-zero entry in each row is 1.
- The first non-zero entry in 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 (rref)** if it is in echelon form and:

- The first non-zero entry in row i appears in column j, then every other element in column j is zero.

Example 2.3.9.

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

2.4 Elementary matrices

Elementary row operations can be carried out using matrix multiplication.

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

There are three types of elementary matrix:

- The general form of the elementary matrix which multiplies a row by any (non-zero) number, α is of the form

$$\mathsf{E_r}(\alpha) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \alpha & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 1 \end{pmatrix}$$

  where all elements on row r is multiplied by α.

- The general form of the elementary matrix which adds a multiple of a row by any non-zero number α to another is of the form

$$\mathsf{E_{rs}}(\alpha) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & \dots & \alpha & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix}$$

  where all elements of s are multiplied by α and added to row r.

- The general form of the elementary matrix which interchanges two rows is of the form

$$\mathsf{E_{rs}} = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix}$$

  where r and s are the rows to interchange.

Example 2.4.2. Find the string of elementary matrices that correspond to the following row operations:

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 0 & -13 & -13 \\ -1 & 1 & 0 & 3 \end{array}\right) \xrightarrow{R_2 \mapsto -\frac{1}{13}R_2} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{13} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 3 \end{array}\right) \xrightarrow{R_1 \mapsto R_1 - R_2} \quad \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 3 \end{array}\right) \xrightarrow{R_3 \mapsto -R_3 + R_1} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & -1 \end{array}\right) \xrightarrow{R_1 \mapsto R_2, R_2 \mapsto R_3, R_3 \mapsto R_1} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

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

Exercise 2.4.3. Find the string of elementary matrices corresponding to the following row operations.

$$\left(\begin{array}{ccc|c} 3 & -2 & 1 & -6 \\ 4 & 6 & -3 & 5 \\ -4 & 4 & 0 & 12 \end{array}\right) \xrightarrow{R_3 \mapsto \frac{1}{4}R_3}$$

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

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 3 \\ 0 & 10 & -3 & 17 \\ -1 & 1 & 0 & 3 \end{array}\right) \xrightarrow{R_2 \mapsto R_2 - 10R_1}$$

*****Mentimeter***************

Theorem 2.4.4. Let A be a $m \times n$ matrix and let E be an elementary $m \times m$ matrix. The matrix multiplication EA applies the same elementary row operation on A that was performed on the identity matrix to obtain E.

Proof: exercise.

2.5 More matrices

Definition 2.5.1. We say a matrix is square if it has the same number of rows as it does columns (i.e. its a member of $M_{n\times n}(F)$ for some field F).

Definition 2.5.2.

A square matrix $\mathsf{A} = a_{ij} \in M_{n\times n}(F)$ is said to be:

1. **upper triangular** if $a_{ij} = 0$ wherever $\boxed{i > j}$. A has zeros for all its elements below the diagonal.
2. **lower triangular** if $a_{ij} = 0$ wherever $\boxed{i < j}$. A has zeros for all its elements above the diagonal.
3. **diagonal** if $a_{ij} = 0$ wherever $\boxed{i \neq j}$. That is to say A has zeros for all its elements except those on the main diagonal.

Example 2.5.3.

$$\underset{\text{Upper triangular}}{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix}} \quad \underset{\text{Lower triangular}}{\begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}} \quad \underset{\text{diagonal}}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}}$$

Definition 2.5.4. The $n \times n$ **identity matrix** is denoted by $\mathsf{I_n}$. An identity matrix has all of its diagonal entreis equal to 1 and all other entries equal to 0. It is called the identity matrix because it is the multiplicative identity matrix for $n \times n$ matrices, i.e.

For $A \in M_{n\times n}(\mathbb{R})$, $\mathsf{I_n}A = A\mathsf{I_n} = A$

Definition 2.5.5. If, for a square matrix B, if there exists another square matrix B^{-1} such that $\mathsf{BB}^{-1} = \mathsf{I} = \mathsf{B}^{-1}\mathsf{B}$, then we say that B is invertible, and B^{-1} **is an inverse of** B.

It is important to realise that **the matrix** B **might not have an inverse:** B^{-1} **might not exist**.

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

Example 2.5.7. Let $\mathsf{A} = \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}$, verify that it has an inverse: $\mathsf{B} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix}$.

$$\begin{aligned} \mathsf{AB} &= \begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \mathsf{BA} &= \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & -1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 1 & -1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Theorem 2.5.8. The inverse of a given matrix is unique. If there exist square matrices A, B, C such that $\mathsf{AB} = \mathsf{I} = \mathsf{CA}$, then $\mathsf{B} = \mathsf{C}$.

Proof Suppose that $\mathsf{AB} = \mathsf{BA} = \mathsf{I}$ and $\mathsf{AC} = \mathsf{CA} = \mathsf{I}$, then

$$\begin{aligned} \mathsf{B} &= \mathsf{BI} \\ &= \mathsf{B(AC)} \\ &= \mathsf{(BA)C} \\ &= \mathsf{IC} \\ &= \mathsf{C} \end{aligned}$$

This theorem shows that if a matrix A is invertible, we can talk about the inverse of A, denoted by A^{-1}. In some circumstances, we can say that a matrix is invertible, and we can find an expression for its inverse, without knowing exactly what the matrix is.

Exercise 2.5.9. Suppose $A, B \in M_{n\times n}(\mathbb{R})$ are both invertible. Show that AB is invertible by finding its inverse.

********MENTIMETER***************

(a) $A^{-1}B^{-1}$

(b) $B^{-1}A^{-1}$

(c) $BAA^{-1}B^{-1}B^{-1}A^{-1}$

Definition 2.5.10. If $\mathsf{A} = [a_{ij}]_{m\times n}$, then the **Transpose of** A is $\mathsf{A}^\mathsf{T} = [a_{ji}]_{n\times m}$.

Example 2.5.11. If

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

and we can see that the transpose of a 2×3 matrix must be a 3×2 matrix.

Exercise 2.5.12. Let $A \in M_{n\times m}(\mathbb{R})$, $B \in M_{m\times p}(\mathbb{R})$, $(AB)^T = B^T A^T$.

Proof:

First remark that $B^T A^T$ is defined and has order $p \times n$, not also AB has order $n \times p$ so $(AB)^T$ has order $p \times n$.

Let $\mathsf{A} = (a_{ij})$ and $\mathsf{B} = (b_{ij})$

- The ij^{th} entry of AB is $\sum_{k=1}^m a_{ik}b_{kj}$. This is the ji^{th} entry of $(AB)^T$
- The ji^{th} entry of $B^T A^T$ is $\sum_{k=1}^m (b^T)_{jk}(a^T)_{ki} = \sum_{k=1}^m (b)_{kj}(a)_{ik} = \sum_{k=1}^m a_{ik}b_{kj}$

Theorem 2.5.13. Given an invertible square matrix A, then A^T is also invertible, and $(\mathsf{A}^\mathsf{T})^{-1} = (\mathsf{A}^{-1})^\mathsf{T}$.

Proof From the definition of the inverse

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

Also

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

Equations 8 and 8 prove that $(\mathsf{A}^{-1})^\mathsf{T}$ is the (unique) inverse of A^T, as required.

2.6 Inverses using row operations

We can use Elementary matrices to find inverses of matrices (it they exist).

Theorem 2.6.1. Every elementary matrix is invertible and the inverse is also an elementary matrix.

Proof

Matrix multiplication can be used to check that

$$\begin{aligned} \mathsf{E_r}(\alpha)\mathsf{E_r}(\alpha^{-1}) = \mathsf{E_r}(\alpha^{-1})\mathsf{E_r}(\alpha) &= \mathsf{I} \\ \mathsf{E_{rs}}(\alpha)\mathsf{E_{rs}}(\alpha^{-1}) = \mathsf{E_{rs}}(\alpha^{-1})\mathsf{E_{rs}}(\alpha) &= \mathsf{I} \\ \mathsf{E_{rs}}(\alpha)\mathsf{E_{rs}}(\alpha) &= \mathsf{I} \end{aligned}$$

Alternatively, the results can be checked by considering the corresponding ero's. Hence

$$\mathsf{E_r}(\alpha)^{-1} = \mathsf{E_r}(\alpha^{-1}), \quad \mathsf{E_{rs}}(\alpha)^{-1} = \mathsf{E_{rs}}(-\alpha), \quad \mathsf{E_{rs}}^{-1} = \mathsf{E_{rs}}$$

Theorem 2.6.2. If the square matrix A can be row-reduced to an identity matrix by a sequence of elementary row operations, then A is invertible and the inverse of A is found by applying the same sequence of elementary row operations to I.

Proof

Let A be a square matrix, then A can be row-reduced to I by a sequence of elementary row operations. Let $\mathsf{E}_1, \mathsf{E}_2, \mathsf{E}_3 \dots, \mathsf{E}_r$ be the elementary matrices corresponding to the elementary row operations, so that

$$\mathsf{E}_r \dots \mathsf{E}_2 \mathsf{E}_1 \mathsf{A} = \mathsf{I} \tag{8}$$

But Theorem 2.6.1 states that $\mathsf{E}_r, \dots, \mathsf{E}_2, \mathsf{E}_1$ are invertible. Multiplying Equation 8 by $\mathsf{E}_1^{-1}\mathsf{E}_2^{-1} \dots \mathsf{E}_r^{-1}$ gives $\mathsf{A} = \mathsf{E}_1^{-1}\mathsf{E}_2^{-1} \dots \mathsf{E}_r^{-1}$. Since A is a product of elementary matrices, it is invertible (using Theorem 2.6.1) and

$$\mathsf{A}^{-1} = (\mathsf{E}_1^{-1}\mathsf{E}_2^{-1} \dots \mathsf{E}_r^{-1})^{-1} = (\mathsf{E}_r \dots \mathsf{E}_2 \mathsf{E}_1)\mathsf{I}$$

Example 2.6.3. Let $\mathsf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & 0 \\ 3 & 0 & 4 \end{pmatrix}$ find A^{-1}.

The method consists of writing the identity matrix I to the right of our given matrix, and then using the same elementary row operations on both matrices to turn the left-hand matrix into I. When this has been achieved, the right-hand matrix will have been transformed into the inverse matrix, A^{-1}.

First, we construct the augmented matrix $\mathsf{A}|\mathsf{I}$, by writing the identity matrix to the right of the matrix A,

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

After our row operations, this matrix will be transformed into $\mathsf{I}|\mathsf{A}^{-1}$.

The steps might be as follows:

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

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

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

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

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

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

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

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

$$\xrightarrow{R_2 \mapsto \frac{1}{2} R_2}$$

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

We have found the inverse of our matrix. We could check by doing the matrix multiplication:

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

as desired.

2.7 Geometric Interpretation

As you have seen in the introductory module vectors in $\mathbb{R}^2/\mathbb{R}^3$ can be represented as points in 2 or 3 dimensional space. In this section we will look 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:

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

Using row reduction we get $x_1 = -0.5$, $x_2 = -2.5$ $x_3 = 2$, which specifies a point. Whereas:

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

Using row reduction we get $x_1 = -2.5 - 0.5x_3$ and $x_2 = 1.5 - 0.5z$ giving the line

$$\begin{pmatrix} -2.5 \\ 1.5 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -0.5 \\ -0.5 \\ 1 \end{pmatrix} \text{ for } \lambda \in \mathbb{R}$$

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:

$$\begin{array}{cccc} A: & \mathbb{R}^n & \mapsto & \mathbb{R}^m \\ & A(v) & = & Av \end{array}$$

We can represent many different operations using matrices.

Example 2.7.2.

Consider $A = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$

Then $A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5x \\ 5y \end{pmatrix}$.

This is a stretch by a factor of 5.

Definition 2.7.3. Let T be a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ then we say T is a *linear transformation* if for every $v_1, v_2 \in \mathbb{R}^n$ and every $\alpha, \beta \in \mathbb{R}$ we have:

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$$

Proposition 2.7.4. Let $A \in M_{n\times m}(\mathbb{R})$ then seen as a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ A is a linear transformation.

Proof:

$$\begin{array}{rcll} A(\alpha v_1 + \beta v_2) & = & A(\alpha v_1) + A(\beta v_2) & \text{by distributivity of matrix multiplication} \\ & = & \alpha A(v_1) + \beta A(v_2) & \text{by exercise} \end{array}$$

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

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

(ii) $A^T A = I_n = AA^T$.

(iii) A preserves inner products (i.e. for all $x, y \in \mathbb{R}^n$ $(Px)\cdot(Py) = x \cdot y$.

Proof:

$(i) \Leftrightarrow (ii)$ is just by definition.

$(ii) \Leftrightarrow (iii)$First note that for $x, y \in \mathbb{R}^n$ $x \cdot y$ as defined in the intro to maths course is just $x^T y$ as matrix multiplication. So A preserves inner products if and only if:

$$\begin{array}{rcl} & (Px)\cdot(Py) = x\cdot y & \forall x, y \in \mathbb{R}^n \\ \iff & (Px)^T(Py) = x^T y & \forall x, y \in \mathbb{R}^n \\ \iff & x^T P^T P y = x^T I_n y & \forall x, y \in \mathbb{R}^n \\ \iff & x^T(P^T P - I_n)y = 0 & \forall x, y \in \mathbb{R}^n \end{array}$$

$(ii) \Rightarrow (iii)$ now trivial.

$(iii) \Rightarrow (ii)$ let $x_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ i.e. column vector with 0's everywhere except the i^{th} row where there is a 1. Then we know for each x_i $(x_i)^T(P^T P - I_n)y = 0$ so we can conclude that

$$(P^T P - I_n)y = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Similarly taking y_i to be the column vector with 0's everywhere except the i^{th} row where there is a 1 we get $(P^T P - I_n) = 0$ so $P^T P = I_n$.

Definition 2.7.6. A matrix $A \in M_{n\times n}$ is called *Orthogonal* if it is such that $A^{-1} = A^T$

Example 2.7.7.

1. Consider the matrix
$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
This matrix is orthogonal as $A^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
If we apply it to $\begin{pmatrix} x \\ y \end{pmatrix}$ we get $\begin{pmatrix} -y \\ x \end{pmatrix}$. This is a rotation through $\frac{\pi}{2}$ radians anti clockwise.
2. Consider the matrix
$$B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$
This matrix is orthogonal as $A^{-1} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$
If we apply it to $\begin{pmatrix} x \\ y \end{pmatrix}$ we get $\begin{pmatrix} -y \\ -x \end{pmatrix}$. This is a reflection through the line $y = -x$.

Exercise 2.7.8. Watch the linear Algebra video to help you.

1. Let R_θ be the anticlockwise rotation of $\mathbb{R}^2$ about the origin through θ radians. Using any school geometry or trigonometry you like, find the matrix representing R_θ.
Assuming R_θ is linear (see lectures!) the vector $e_1 = \binom{1}{0}$ is rotated to $\binom{\cos\theta}{\sin\theta}$ while $e_2 = \binom{0}{1}$ is rotated to $\binom{-\sin\theta}{\cos\theta}$ so the matrix is
$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

2.8 Fields

So far, for both matrices and linear equations, we have only been using entries in $\mathbb{R}$. However, we could have taken entries from any field.

Every field has distinguished elements 0 (additive identity) and 1 (multiplicative identity).

Fact 2.8.1. Over any field F we can define:

1. The null matrix (i.e. the additive identity matrix) for $M_{n\times m}(F)$ as

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

2. The (multiplicative) identity matrix for $M_{n\times n}(F)$ as

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

Remark 2.8.2 *It is important to know what field we are working in, and that we don't say take scalars from a different field to the one matrix entries are from. (e.g. the set of matrices $M_{n\times m}(\mathbb{Q})$ is not closed under scalar multiplication by elements from $\mathbb{R}$).*

Being able to work over a general field allows us to use finite fields.

Theorem 2.8.3. Let $\mathbb{F}_p = \{0, 1, ..., p-1\}$, consider $\mathbb{F}_p$ with addition defined by addition modulo p and multiplication as multiplication modulo p. Then the structure $(\mathbb{F}_p, +_{\pmod p}, \times_{\pmod p})$ is a field.

Proof:

A1-4 (Additive (commutative) group) obvious from properties of addition in $\mathbb{Z}$.

M1-3 (mulitplicative semigp with 1)obvious from properties of addition in $\mathbb{Z}$.

M4: inverses: obviously for $0 \leq x < p$ we have $\gcd(x, p) = 1$ by Intro to Uni Maths we have: $\exists s, t \in \mathbb{Z}$ such that $1 = sx + tp$ then take $x^{-1} = s \pmod p$.

D1 (distributive law) obvious from properties of addition in $\mathbb{Z}$.

Example 2.8.4. $\mathbb{F}_6$ defined as above is not a field. For example $3 \neq 0$ does not have an inverse.

3 Vector Spaces

3.1 Intro 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**
$$\begin{array}{rrcl} \oplus: & V \times V & \mapsto & V \\ & (v_1, v_2) & \mapsto & v_1 \oplus v_2 \end{array}$$

2. **Scalar Multiplication**
$$\begin{array}{rrcl} \odot: & F \times V & \mapsto & V \\ & (f, v_2) & \mapsto & f \odot v_2 \end{array}$$

$\oplus$ and $\odot$ must satisfy the following *Vector Space axioms:*

For Vector Addition:

A1 Associative law: $(\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} = \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w})$.

A2 Commutative law: $\mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v}$.

A3 Additive identity: $0_V \oplus \mathbf{v} = \mathbf{v}$, where 0_V is called the IDENTITY vector (or sometimes the *zero vector*).

A4 Additive inverse: $\mathbf{v} \oplus (\ominus \mathbf{v}) = 0_V$.

For scalar multiplication:

A5 Distributive law: $r \odot (\mathbf{v} \oplus \mathbf{w}) = (r \odot \mathbf{v}) \oplus (r \odot \mathbf{w})$.

A6 Distributive law: $(r + s) \odot \mathbf{v} = (r \odot \mathbf{v}) \oplus (s \odot \mathbf{v})$.

A7 Associative law: $r \odot (s \odot \mathbf{v}) = (rs) \odot \mathbf{v}$.

A8 Identity for scalar mult: $1 \odot \mathbf{v} = \mathbf{v}$.

From now on we will drop the $\oplus$ and $\odot$, and use $+$ and $\cdot$ the point was to emphasise that these are not the same as field addition and multiplication.

Definition 3.1.2. Let V be a vector space over F we call:

- Elements of V are called *vectors*.
- Elements of F are called *scalars*.
- We sometimes refer to V as an F-vector space.

Example 3.1.3. The following are examples of vector spaces over $\mathbb{R}$:

- The canonical example is the set of vectors $\mathbb{R}^n$ over $\mathbb{R}$, where $\oplus$ is normal vector addition and $\odot$ is multiplication by a scalar. The additive inverse of $\mathbf{v}$ is simply $-\mathbf{v}$
- The set $\mathsf{M_{mn}}$ of all $m \times n$ matrices. This is because addition of two $m \times n$ matrices produces an $m \times n$ matrix and multiplication of an $m \times n$ matrix by a scalar also produces a $m \times n$ matrix. The zero vector is the zero matrix, and for any matrix A, the matrix $-$A is the additive inverse. Properties of matrix arithmetic covered in Chapter 1, show that all properties in Definition 3.1.2 required of a vector space are satisfied. We will see this later in the course in detail.
- Define $\mathbb{R}^X$ to be the set of real valued functions on X (i.e. $\mathbb{R}^X := \{f : f \text{ a function}, f : X \to \mathbb{R}\}$). Then for $f, g \in \mathbb{R}^X$ and $\alpha \in \mathbb{R}$ define:

$$f \oplus g : X \to \mathbb{R} \qquad (\alpha \odot f) : X \to \mathbb{R}$$
$$(f \oplus g)(x) = f(x) + g(x) \qquad (\alpha \odot f)(x) = \alpha(f(x))$$

Exercise 3.1.4. Which of the following examples of vector spaces over $\mathbb{R}$:

1. The set of vectors

$$V = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} : a, b \in \mathbb{Z} \right\} \text{ with the usual vector addition and multiplication}$$

$$\text{No, because } \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in V, \quad \text{, but } \sqrt{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin V$$

2. The set of vectors:

$$V = \left\{ \begin{pmatrix} a+1 \\ 2 \end{pmatrix} : a \in \mathbb{R} \right\} \text{ with the usual vector addition and multiplication}$$

$$\text{No, because } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \notin V$$

3. $V = \mathbb{R}^2$ with the following addition and scalar multiplication operations:

$$\begin{pmatrix} x \\ y \end{pmatrix} \oplus \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix} \quad \text{and} \quad r \odot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ ry \end{pmatrix}$$

yes

3.2 Subspaces

Definition 3.2.1. A subset W of a vector space V is a **subspace** of V if

S1 W is not empty (i.e. $e \in W$)

S2 for $\mathbf{v}, \mathbf{w} \in W$, then $\mathbf{v} \oplus \mathbf{w} \in W$ *closed under vector addition*

S3 $\mathbf{v} \in W$ and $r \in F$, then $r \odot \mathbf{v} \in W$ *closed under scalar multiplication.*

N.B. Sometimes we use the notation $U \leq V$ to mean U is a subspace of V.

Remark 3.2.2 *Note that V and the zero subspace,* **0** *are always subspaces of V. Any other subspace of V is called a* ***proper subspace*** *of V.*

Proposition 3.2.3. Every subspace of an F-vector space V must contain the zero vector.

Proof:

Claim: For an F-vector space V with $0 \in F$ the field additive identity we have $0v = 0_V$ for all $v \in V$. *Proof of claim:* Exercise (note enought to show that $0v$ is the vector space additive identity.

Let $v \in V$ (V is non-empty) then $0_V = 0 \oplus v \in V$ as V is closed under scalar multiplication.

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 $\mathbb{R}^2$.

Worked Answer

S1 The vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \in X$, therefore X is non-empty.

S2 If $\mathbf{v} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} x_2 \\ 0 \end{pmatrix}$, then

$$\begin{aligned} \mathbf{v} \oplus \mathbf{w} &= \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} x_2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + x_2 \\ 0 \end{pmatrix} \in X \end{aligned}$$

S3 If $\mathbf{v} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ and $r \in \mathbb{R}$, then

$$\begin{aligned} r \odot \mathbf{v} &= r \odot \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} rx_1 \\ 0 \end{pmatrix} \in X \end{aligned}$$

Therefore, $X = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} ; x \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^2$.

Exercise 3.2.5. All subspaces of a vector space over F are vector spaces over F in their own right.

Theorem 3.2.6. Let U, W be subspaces of V. Then $U \cap W$ is a subspace of V. In general, the intersection of any set of subspaces of a vector space V is a subspace of V.

ProofLet C be a set of subspaces of V and T is their intersection. Then $T \neq \emptyset$ since every subspace of V (and therefore every subspace in C) contains the zero vector, and so does T.

Suppose that $x, y \in T$. Since x and y belong to every subspace W in C, so does $x \oplus y$, and therefore $x \oplus y \in T$.

If $x \in T$, then x belongs to every subspace W in the set C, and so does $r \odot x$ and so $r \odot x \in T$.

Therefore T is a subspace of V.

Example 3.2.7. Note that in general if U and W are subspaces of V, then $U \cup W$ is not a subspace of V. For example, let

$$U = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \in \mathbb{R} \right\}, W = \left\{ \begin{pmatrix} 0 \\ y \end{pmatrix} : y \in \mathbb{R} \right\} \quad \text{and} \quad V = \mathbb{R}^2.$$

Then

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

but

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

3.3 Spanning

Definition 3.3.1. Let V be an F-vector space. Let $u_1, ..., u_m \in V$ then:

- A *Linear Combination* of $u_1, ..., u_m$ is a vector of the form $\alpha_1 u_1 + ... + \alpha_m u_m$ for scalars $\alpha_1, ..., \alpha_m \in F$. Note we can also write $\alpha_1 u_1 + ... + \alpha_m u_m$ as $\sum_{i=1}^{m} \alpha_i u_i$.
- The *span* of $u_1, ..., u_m$ is the set of linear combinations of $u_1, ..., u_m$. i.e. $Span(u_1, ..., u_m) = \{\alpha_1 u_1 + ... + \alpha_m u_m \in V : \alpha_1, ..., \alpha_m \in F\}$.

NB: there are several different notations used for Span, e.g., $Sp(X)$, $< X >$.

Lemma 3.3.2.

Let V be an F vector space, and $u_1, ..., u_m \in V$ then $Span(u_1, ..., u_m)$ is a subspace of V.

Proof: Clearly $Span(u_1, ..., u_m) \subset V$ so we do the subspace test:

SS1 $u_1 \in Span(u_1, ..., u_m)$

SS2 Suppose $v, w \in Span(u_1, ..., u_m)$ then $v = \sum_{i=1}^{m} \alpha_i u_i$ and $w = \sum_{i=1}^{m} \beta_i u_i$ so

$$v + w = \sum_{i=1}^{m} (\alpha_i + \beta_i) u_i \in Span(u_1, ..., u_m) \text{ as } F \text{ closed under addition, i.e. } \alpha_i + \beta_i \in F$$

SS3 Suppose $v \in Span(u_1, ..., u_m)$ and $\lambda \in F$ then $v = \sum_{i=1}^{m} \alpha_i u_i$ do $v = \sum_{i=1}^{m} \lambda \alpha_i u_i \in Span(u_1, ..., u_m)$ as $\lambda \alpha_i \in F$ for each $i \in \{1, ..., m\}$

Remark 3.3.3.

- By convention we take the empty sum to be 0_V, so $Span\emptyset = \{0_V\}$
- For an infinite set S we still only take finite sums i.e.

$$Span(S) = \left\{ \sum_{s_i \in S'} \alpha_i s_i : S' \subset^{finite} S, \alpha_i \in F \right\}$$

Exercise 3.3.4. Show that for an infinite subset S of an F-vector space V, $Span(S)$ is a subspace of V.

Definition 3.3.5. Let V be an F vector space, and suppose $S \subset V$ is such that $Span(S) = V$ then we say S *spans* V, or equivalently S is a *spanning set* for V.

Example 3.3.6.

- $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ spans $\mathbb{R}^3$.
- $\mathbb{R}^{deg \leq n}[x]$ spanned by $\{1, x, x^2 ...., x^n\}$

Exercise 3.3.7. Which of the following sets span $\mathbb{R}^3$:

1. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$
2. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$
3. $\begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$
4. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

In the above exercise we see that we sometimes have “redundant” vectors in a spanning set. If as well as spanning the set is linearly independent, then this won’t happen.

3.4 Linear Independence

Definition 3.4.1. Let V be an F-vector space. We say $u_1, ..., u_m \in V$ are *linearly independent* if whenever

$$\alpha_1 u_1 + ... + \alpha_m u_m = 0_V,$$

then it must be that

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

We say $\{u_1, ..., u_m\}$ is a *linearly independent set.*

Alternatively, a set $\{u_1, ..., u_m\}$ is *linearly dependent* if $\alpha_1 u_1 + ... + \alpha_m u_m = 0_V$ where at least one $\alpha_i \neq 0$, and a set is linearly independent if it is not linearly dependent.

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 $\mathbb{R}^3$.
- Let $f, g : \mathbb{R} \mapsto \mathbb{R}$ be functions and suppose $f(x) = x$ and $g(x) = x^2$. The set $\{f, g\}$ is a linearly independent subset of $V = \mathbb{R}^{\mathbb{R}}$.
  Proof: Suppose $\alpha g + \beta f = 0_V$ now two functions are equal if they are equal on all of the domain. So consider $1, 2 \in \mathbb{R}$. Then we get

$$\begin{array}{lll} 0_V(1) & = & (\alpha g + \beta f)(1) \\ 0 & = & \alpha + \beta \\ & & \\ 0_V(2) & = & (\alpha g + \beta f)(2) \\ 0 & = & 2\alpha + 4\beta \end{array}$$

  So we have $\alpha = -\beta$ and $\alpha = -2\beta$ thus $\alpha = \beta = 0$.
- The set $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a linearly *dependent* subset of $\mathbb{R}^3$.

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

- For V and F-vector space then $\{0_V\}$ is linearly *dependent*
- For V and F-vector space $v \in V$ then $\{v\}$ is linearly independent iff $v \neq 0_V$.

Exercise 3.4.3. Which of the following sets are in linearly independent subsets of $\mathbb{R}^3$:

1. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

2. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

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

4. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

Lemma 3.4.4. Let $v_1, ..., v_n$ be linearly independent in an F-vector space V. Let v_{n+1} be such that $v_{n+1} \notin Span(v_1, ..., v_n)$. Then $v_1, ..., v_{n+1}$ is linearly independent.

Proof: Suppose $\alpha_1, ..., \alpha_{n+1} \in F$ are such that $\alpha_1 v_1 + ... + \alpha_{n+1} v_{n+1} = 0_V$.

If $\alpha_{n+1} \neq 0$ then $v_{n+1} = \frac{1}{\alpha_{n+1}}(\alpha_1 v_1 + ... + \alpha_n v_n) \in Span(v_1, ..., v_n)$. Contradiction.

So $\alpha_{n+1} = 0$ so $\alpha_1 v_1 + ... + \alpha_n v_n = 0_V$, but $v_1, ..., v_n$ are linearly independent, thus $\alpha_1 = ... = \alpha_n = 0$.

3.5 Bases

Definition 3.5.1.

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

Example 3.5.2.

- 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 $\mathbb{R}^3$. We have done linear independence, you must show $Span(B) = \mathbb{R}^3$.
- Let F be a field, then in F^n let e_i be the columnn vector with zeros everywhere except the i^{th} row. Then $\{e_1, ..., e_n\}$ forms a basis for F^n and is known as the *standard basis*.
- $\mathbb{R}[x]$ has basis $\{1, x, x^2, ...\}$.

Note: Not every vector space is finite dimensional. For example $\mathbb{R}[x]$ the set of real polynomials doesn't have a finite basis, but it does have infinite bases, e.g., $\{1, x, x^2, \dots\}$.

Exercise 3.5.3. Which of the following sets span $\mathbb{R}^3$:

1. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$
2. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$
3. $\begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$
4. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

Proposition 3.5.4. Let V be an F-vector space, let $S = \{u_1, ..., u_m\} \subseteq V$. Then S is a basis of V if and only if every vector in V has a unique expression as a linear combination of elements of S.

Proof:

($\Rightarrow$) Suppose S is a basis of V. Take $v \in V$.

[AIM: there are unique $\alpha_1, ..., \alpha_n \in F$ such that $v = \sum_{i=1}^m \alpha_i u_i$]

Since V is spanned by S we have some $\alpha_1, ..., \alpha_n \in F$ such that $v = \sum_{i=1}^m \alpha_i u_i$.

Suppose for contradiction the α_i's are not unique, i.e. there exist $\beta_1, ..., \beta_n \in F$ such that $v =$

$\sum_{i=1}^{m} \beta_i u_i$.

Then we have:

$$\sum_{i=1}^{m} \alpha_i u_i = \sum_{i=1}^{m} \beta_i u_i$$
$$\sum_{i=1}^{m} (\alpha_i - \beta_i) u_i = 0$$

As S is LI we get $\alpha_i - \beta_i = 0$ thus $\alpha_i = \beta_i$

($\Leftarrow$) Suppose conversely that for every $v \in V$ has there are unique $\alpha_1, ..., \alpha_m$ such that $v = \sum_{i=1}^{m} \alpha_i u_i$

[AIM: we need to show that S is spanning and LI.]

- *Spanning:* Let $v \in V$ then $v = \sum_{i=1}^{m} \alpha_i u_i \in Span(S)$
- *LI:* First remark that $0u_1 + ...0u_m = 0_V$ so if $\sum_{i=1}^{m} \lambda_i u_i = 0_V$ then by uniqueness we get $\alpha_i = 0$

So S is a basis for V.

Remark 3.5.5 *Let $B = \{u_1, ...u_m\}$ be a basis for an F-vector space V. By Proposition 3.5.4 we see that we have a bijective map f from V to F^m, for $v = \alpha_1 u_1 + ... + \alpha_m u_m$ we define $f(v) = (\alpha_1, ..., \alpha_m)$ we call $(\alpha_1, ..., \alpha_m)$ the co-ordinates of v*

Proposition 3.5.6. Let V be a non-trivial (i.e. not $\{0\}$) F-vector space and suppose V has a finite spanning set S then S contains a linearly independent spanning set.

I.e., if V has a finite spanning set it has a basis - for cases where there is no finite spanning set we would need something called the axiom of choice to show this (see LOGIC course in year 3)

Proof:

Consider T such that T is linearly independent subset of S, and for any LI subset of S, T' we have that $|T'| \leq |T|$. We can get such a T as we have at least some $v \in V$ so $\{v\}$ is linearly indpendent (i.e. $|T| \geq 1$).

Claim T is spanning.

Proof of Claim: Suppose not then there is a $v \in S \setminus Span(T)$ but by Lemma 3.4.4 $v \cup T$ is LI. Contradiction.

3.6 Dimension

Lemma 3.6.1. *Steinitz Exchange Lemma*
Let V be a vector space over F. Take $X \subseteq V$ and suppose $u \in Span(X)$ but $u \notin Span(X \setminus \{v\})$ for some $v \in X$. Let $Y = (X \setminus \{v\}) \cup \{u\}$ (i.e., we "exchange v for u"). Then $Span(X) = Span(Y)$.

Proof

Since $u \in Span(X)$ we have $\alpha_1, ..., \alpha_n \in F$ such that $v_1, ..., v_n \in X$ such $u = \alpha_1 v_1 + ... + \alpha_n v_n$. Now there is a $v \in X$ such that $u \notin Span(X \setminus \{v\})$ we may assume, WLOG, that $v = v_n$, thus $\alpha_n \neq 0$ so:

$$v = v_n = \frac{1}{\alpha_n}(u - \alpha_1 v_1 ... - \alpha_{n-1} v_{n-1})$$

Now if $w \in Span(Y)$ then for some $\beta_0, \beta_1, ..., \beta_m$ we have $v_1, ..., v_m \in X \setminus \{v\}$

$$\begin{array}{rcl} w & = & \beta_0 u + \sum_{i=0}^m \beta_i v_i \\ & = & \beta_0(\alpha_1 v_1 + ... + \alpha_n v_n) + \sum_{i=0}^m \beta_i v_i \quad \in Span(X \setminus \{v\} \cup \{v\}) = Span(X) \end{array}$$

So $Span(Y) \subseteq Span(X)$.

Similarly we have that if $w \in Span(X)$ the w is a linear combination of elements of X, now we can replace v_n with $\frac{1}{\alpha_n}(u - \alpha_1 v_1 ... - \alpha_{n-1} v_{n-1}$ so we can express w as a linear combination of elements of Y. So $Span(X) \subseteq Span(Y)$, thus $Span(Y) = Span(X)$.

This lemma is essential to being able to define the dimension of a vector space - and relies on being able to invert elements in the field.

Exercise 3.6.2. Verify the Steinitz exchange lemma where:

- $V = \mathbb{R}^3$
- $X = \{e_1, e_2\}$
- $u = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$

Theorem 3.6.3. Let V be a finite dimensional vector space over F. Let S, T be finte subsets of V. If S is LI and T spans V then $|S| \leq |T|$. That is, LI sets are at most as big as spanning sets.

Proof: Assume S is LI and T spans V and suppose:

$$S = \{s_1, ..., s_m\}$$
$$T = \{t_1, ..., t_n\}$$

Let $T = T_0$, since $Span(T_0) = V$ there is some I such that $s_1 \in Span(\{t_1, ..., t_i\})$, but $s_1 \notin Span(\{t_1, ..., t_{i-1}\})$.

Thus by SEL $Span(\{s_1, t_1, ..., t_{i-1}\}) = Span(\{t_1, ..., t_i\})$.

Let $T_1 = \{s_1, t_2, ..., t_{i-1}, t_{i+1}, ..., t_n\}$, then we have $Span(T_1) = Span(T_0) = V$. We continue this process inductively.

Suppose that for some j with $1 \leq j \leq m$ we have $T_j = \{s_1, ..., s_j, t_{i_1}, ..., t_{i_{n-j}}\}$, with $Span(T_j) = Span(T)$, and $t_{i_k} \in T$.

Now $s_{j+1} \in Span(T_j)$ so there is an i_k such that $s_{j+1} \in Span(\{s_1, ..., s_j, t_{i_1}, ..., t_{i_k}\})$, but $s_{j+1} \notin Span(\{s_1, ..., s_j, t_{i_1}, ..., t_{i_{k-1}}\})$.

Note S is LI so $s_{j+1} \notin Span(\{s_1, ..., s_j\})$ i.e. $t_{i_k} \in T$.

We let $T_{j+1} = \{s_1, ..., s_{j+1}, t_{i_1}, ..., t_{i_{k-1}}, t_{i_k}, ..., t_{i_{n-j}}\}$ and by SEL we have $Span(T_{j+1}) = Span(T_j) = Span(T) = V$, by relabeling the elements of T_{j+1} we can see we have a set of the form:

$$T_{j+1} = \{s_1, ..., s_{j+1}, t_{i_1}, ..., t_{i_{n-(j+1)}}$$

After j steps we have replaced j members of T with j members of S. We cannot run out of members of T before we run out of members of S; as otherwise a remaining element of S would be a linear combination of the elements of S already swapped into T, thus $m \leq n$.

Corollary 3.6.4. Let V be a finite dimensional vector space. Let S, T be bases of V, then S and T are both finite and $|S| = |T|$.

Proof: Since V is finite dimensional it has a finite basis B say. Suppose $|B| = n$. Now B is a spanning set and $|B| = n$ so by Theorem 3.6.3 any LI subset has size at most n.

Since S is LI we get $|S| \leq n$, similarly $|T| \leq n$ - so both sets are finite.

Also we have S is spanning and T is LI, so $|T| \leq |S|$, also T is spanning and S is LI, so $|S| \leq |T|$. Thus $|S| = |T|$.

Definition 3.6.5. Let V be a finite dimensional vector space. The *dimension of* V, written $\dim V$, is the size of any basis of V.

Remark 3.6.6 *Note that we needed Corollary 3.6.4 and thus the SEL to know that the size of a basis is unique (a basis certainly isn't).*

Example 3.6.7. In PS2 you were asked to describe all the subspaces of $\mathbb{R}^3$ this becomes much easier once we know about dimensions.$\mathbb{R}^3$ is an $\mathbb{R}$ vector space of dimension 3.

As subspaces are vector spaces in their own right so they also have dimensions, and these must be less than or equal to 3:

- $\dim 3$: the only subspace of dimension 3 is $\mathbb{R}^2$
- $\dim 2$: planes going through the origin
- $\dim 1$: lines going through the

- $\dim 0$: $\{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\}$

Lemma 3.6.8. Suppose that $\dim V = n$:

1. Any spanning set of size n is a basis.
2. Any linearly independent set of size n is a basis.
3. S is a spanning set if and only if it contains a basis (as a subset).
4. S is linearly independent if and only if it is contained in a basis (i.e. it's a subset of a basis).
5. Any subset of V of size $> n$ is linearly dependent.

Proof: Exercise.

3.7 More subspaces

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

- The *intersection of U and W* is:

$$U \cap W = \{v \in V : v \in W \text{ and } v \in U\}$$

- The *sum of U and W* is:

$$U + W = \{u + w : u \in U, w \in W\}$$

Remark 3.7.2. $U \subseteq U + W$ and $W \subseteq U + W$. This is because $0 \in U$ and $0 \in W$, so for every $u \in U$, $u = u + 0 \in U + W$. Similarly, for every $w \in W$, $w = 0 + w \in U + W$

Example 3.7.3. Let $V = \mathbb{R}^2$ over $\mathbb{R}$, $U = Span\{(1,0)\}$, $W = Span\{(0,1)\}$. *Claim* $U+W = \mathbb{R}^2$.

Proof: Let $(\lambda, \mu) \in \mathbb{R}^2$ then $(\lambda, 0) \in U$, $(0, \mu) \in W$ so

$$(\lambda, \mu) = (\lambda, 0) + (0, \mu) \in U + W$$

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

Proof:

1. $U + W$ is a subspace: Clearly $U + W \subset V$, so we can apply the subspace test:
   - $0 \in U$ and $0 \in W$ so $0 + 0 = 0 \in U + W$.
   - Suppose $v_1, v_2 \in U + W$ then $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$ for some $u_i \in U$ and $w_i \in W$. Consider

   $$\begin{aligned} v_1 + v_2 &= (u_1 + w_1) + (u_2 + w_2) \\ &= \underbrace{(u_1 + u_2)}_{\in U} + \underbrace{(w_1 + w_2)}_{\in W} && \begin{array}{l} + \text{ in } V \text{ is commutative and associative} \\ U, W \text{ closed under } + \end{array} \end{aligned}$$

   So $v_1 + v_2 \in U + W$
   - Let $\lambda \in \mathbb{R}$ and $v \in U + W$ then $v = u + w$ for some $u \in U$ and $w \in W$. Consider

   $$\begin{aligned} \lambda v &= \lambda(u + w) \\ &= \underbrace{\lambda u}_{\in U} + \underbrace{\lambda w}_{\in W} && \begin{array}{l} \text{by distributivity in } V \\ U, W \text{ closed under scalar} \times \end{array} \end{aligned}$$

   So $\lambda v \in U + W$
2. $U \cap W$ is a subspace. Exercise.

Proposition 3.7.5. Let V be a vector space over F. Let U and W be subspaces of V, suppose additionally:

- $U = Span\{u_1, ..., u_s\}$
- $W = Span\{w_1, ..., w_r\}$

Then $U + W = Span\{u_1, ..., u_s, w_1, ..., w_r\}$.

Proof:

1. Show $U + W \subseteq Span\{u_1, ..., u_s, w_1, ..., w_r\}$. Let $v \in U + W$ then $u = u + w$ for some $u \in U$ and $w \in W$. Therefore:
   - $u = \lambda_1 u_1 + ... + \lambda_s u_s$
   - $w = \mu_1 w_1 + ... + \mu_r w_r$

   So $v = \lambda_1 u_1 + ... + \lambda_s u_s + \mu_1 w_1 + ... + \mu_r w_r \in Span\{u_1, ..., u_s, w_1, ..., w_r\}$

2. Show $Span\{u_1, ..., u_s, w_1, ..., w_r\} \subseteq U + W$. Suppose $v \in Span\{u_1, ..., u_s, w_1, ..., w_r\}$ then:
   $$v = \underbrace{\lambda_1 u_1 + ... + \lambda_s u_s}_{\substack{\in Span\{u_1, ..., u_s\} \\ = U}} + \underbrace{\mu_1 w_1 + ... + \mu_r w_r}_{\substack{\in Span\{w_1, ..., w_r\} \\ = W}}$$

   So $v \in U + W$.

   Alternatively:

   - $u_i \in U \subseteq U + W$ for each $i \in \{1, ..., s\}$
   - $w_i \in W \subseteq U + W$ for each $i \in \{1, ..., r\}$

   So $\{u_1, ..., u_s, w_1, ..., w_r\} \in U + W$ so $Span\{u_1, ..., u_s, w_1, ..., w_r\} \in U + W$. As $u + W$ is closed under linear combinations.

Example 3.7.6. Let $V = \mathbb{R}^2$, let $U = Span\{(0, 1)\}$, $W = Span\{(1, 0)\}$. Then by proposition 3.7.5 we have $U + W = Span\{(0, 1), (1, 0)\} = \mathbb{R}^2$. Agrees with example 3.7.3.

Example 3.7.7. Let $V = \mathbb{R}^3$ and:
Let $U = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$
Let $W = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : -x_1 + 2x_2 + x_3 = 0\}$
Question: Find bases for $U, W, U \cap W, U + W$.

Answer:

- A general vector in $u \in U$ is of the form $u = (a, b, -a-b)$ for $a, b \in \mathbb{R}$. So $u = a(1, 0, -1) + b(0, 1, -1)$, therefore $\{(1, 0, -1), (0, 1, -1)\}$ is a spanning set for U, and as the vectors are linearly independent this is a basis for U.

- A general vector in $w \in W$ is of the form $w = (2a + b, a, b)$ for $a, b \in \mathbb{R}$. So $u = a(2,1,0) + b(1,0,1)$, therefore $\{(2,1,0),(1,0,1)\}$ is a basis for W, as they are clearly linearly independent.

- By proposition **??** we know that $\{(1,0,-1),(0,1,-1),(2,1,0),(1,0,1)\}$ is a spanning set for $U + W$, this is clearly not linearly independent, so we do row reduction to get an LI set:

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

  So a linearly independent spanning set is $\{(1,0,0),(0,1,0),(0,0,1)\}$. So $dim(U+W) = 3$ so as $U + W \subseteq \mathbb{R}^3$ we have $U + W = \mathbb{R}^3$.

- We want a basis for $U \cap W$. Let $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. We have:
  $x \in U$ iff $x_1 + x_2 + x_3 = 0$
  $x \in W$ iff $-x_1 + 2x_2 + x_3 = 0$
  So $x \in U \cap W$ iff $x_1 + x_2 + x_3 = -x_1 + 2x_2 + x_3 = 0$ (i.e. $U \cap W = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0 \text{ and } -x_1 + 2x_2 + x_3 = 0\}$)

  That is to say $2x_1 - x_2 = 0$, so $x_2 = 2x_1$, and therefore $x_3 = -x_1 - x_2 = -3x_1$. So x is of the form $(x_1, 2x_1, -3x_1)$. So a spanning set for $U \cap W$ is $\{(1,2,-3)\}$ which is clearly a basis.

Remark 3.7.8. A neater way of finding a basis for $U + W$ would have been to use the basis for $U \cap W$. Since $U \cap W \subset U$ we can find a basis for U containing out basis for $U \cap W$ and similarly for W. The union of these bases will be a basis for U_W.

For instance, a basis for U is $\{(1,0,-1),(1,2,-3)\}$, and a basis for W is $\{(1,0,1),(1,2,-3)\}$, so a basis for $U + W$ is $\{(1,0,1),(1,0,-1),(1,2,-3)\}$. Note that this has three elements, and $dim(U + W) = 3$ so as this is a spanning set it must be a basis.

Theorem 3.7.9. Let V be a vector space over F, U and W subspaces of V. Then

$$dim(U + W) = dimU + dimW - dim(U \cap W).$$

Proof: Suppose $dim(U \cap W) = m$, $dimU = r$ and $dimW = s$ (so we need to prove that $dim(U + W) = r + s - m$).

Now as $dim(U \cap W) = m$ we have a basis $B_{U \cap W}\{v_1, ..., v_m\}$ of $U \cap W$. Now as $U \cap W \subseteq U$ and $B_{U \cap W}$ is linearly independent it is contained in a basis $B_U = \{v_1, ..., v_m, u_{m+1}, ..., u_r\} \supseteq B_{U \cap w}$. Simiarly we have a basis $B_W = \{v_1, ..., v_m, w_{m+1}, ..., w_s\}$ containing $B_{U \cap W}$.

Claim $B_U \cup B_W = \{v_1, ..., v_m, u_{m+1}, ..., u_r, w_{m+1}, ..., w_s\}$ is a basis for $V + W$.
Proof of Claim:

Span: By proposition **??** $B_U \cup B_W$ is a spanning set.

LI: Suppose we have:

$$\lambda_1 v_1 + ... + \lambda_m v_m + \mu_{m+1} u_{m+1} + ... + \mu_r u_r + \nu_{m+1} w_{m+1} + ... + \nu_s w_s = 0$$

For $\lambda_i, \mu_i, \nu_i \in F$. [We need to show $\lambda_i = \mu_j = \nu_k = 0$ for all $i.j, k$.]

Now we have

$$\underbrace{\lambda_1 v_1 + ... + \lambda_m v_m + \mu_{m+1} u_{m+1} + ... + \mu_r u_r}_{\in U} = \underbrace{-\nu_{m+1} w_{m+1} - ... - \nu_s w_s}_{\in W}$$

Thus $\lambda_1 v_1 + ... + \lambda_m v_m + \mu_{m+1} u_{m+1} + ... + \mu_r u_r \in U \cap W$. So $\lambda_1 v_1 + ... + \lambda_m v_m + \mu_{m+1} u_{m+1} + ... + \mu_r u_r = \beta_1 v_1 + ...\beta_m v_m$ for some $\beta_i \in F$. Thus

$$\beta_1 v_1 + ...\beta_m v_m + \nu_{m+1} w_{m+1} + ... + \nu_s w_s = 0$$

As $\{v_1, ..., v_m, w_{m+1}, ..., w_s\}$ is a basis for W (thus linearly independent) we have $\beta_1 = ... = \beta_m = \nu_{m+1} = ...\nu_s = 0$.

Thus $\lambda_1 v_1 + ... + \lambda_m v_m + \mu_{m+1} u_{m+1} + ... + \mu_r u_r = 0$. As $\{v_1, ..., v_m, u_{m+1}, ...u_r\}$ is a basis for U we have $\lambda_1 = ... = \lambda = \mu_{m+1} = ...\mu_r = 0$.

So $\lambda_i = \mu_j = \nu_k = 0$ for all $i.j, k$, so $B_U \cup B_W$ is linearly independent.

$B_U \cup B_W$ is a spanning set for $U + W$ and is linearly independent thus it is a basis.

Now $|B_U \cap B_W| = r + s - m$, thus $dim(U + W) = r + s - m$.

□

3.8 Rank of a Matrix

Definition 3.8.1. Let A be an $m \times n$ matrix with entries from a field F. Define:

- The *Row Space of A (RSp(A))* as the span of the rows of A. This is a subspace of F^n.
- The *Row Rank of A* is $dim(RSp(A))$.
- The *Column Space of A (CSp(A))* as the span of the columns of A. This is a subspace of F^m.
- The *Column Rank of A* is $dim(CSp(A))$.

Example 3.8.2. Let $F = \mathbb{R}$ and $A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & -1 & 1 \end{pmatrix}$. Then,

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

Now the row vectors $(3 \quad 1 \quad 2)$ and $(0 \quad -1 \quad 1)$ are linearly independent so $dim(RSp(A)) = 2$, so the column rank is 2. The set

$$\left\{ \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

is linearly dependent as

$$\begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

So

$$CSp(A) = Span\{\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\},$$

which is linearly independent, so $dimCSp(A) = 2$.

Procedure 3.8.3.

Calculating the row rank of a matrix A.

- Step 1: Reduce A to row echelon form using row operations:

$$A_{ech} = \begin{pmatrix} 1 & * & * & * & * & \dots \\ 0 & 0 & 1 & * & * & \dots \\ 0 & 0 & 0 & 1 & * & *\dots \\ \vdots & & & & & \\ 0 & \dots & & & & \end{pmatrix}$$

  (Actually it doesn't matter whether the leading entries in each row are 1s or not.)

- Step 2: The row rank of A is the number of non-zero rows in A_{ech}. In fact it the non-zero rows of A_{ech} form a basis for $RSp(A)$.

Justification
It will be enough to show:

1. $RSp(A) = RSP(A_{ech})$
2. The rows of A_{ech} are linearly independent.

To show 1., not that to obtain A_{ech} from A we use row operations:

$$\begin{cases} r_i \mapsto r_i + \lambda r_j & \lambda \in F, \quad i \neq j \\ r_i \mapsto \lambda r_i & \lambda \in F \setminus \{0\} \\ r_i \mapsto r_j & i \neq j \end{cases}$$

Let A' be obtained from A by one row operation. then clearly every row of A' lies in $RSp(A)$ and so $RSp(A') \subseteq RSp(A')$. Also every row operation is invertibl;e by another row operation:

$$\begin{cases} r_i \mapsto r_i + \lambda r_j & \text{has inverse} & r_i \mapsto r_i - \lambda r_j \\ r_i \mapsto \lambda r_i & \text{has inverse} & r_i \mapsto \frac{1}{\lambda} r_i \\ r_i \mapsto r_j & \text{has inverse} & r_i \mapsto r_j \end{cases}$$

It follwos that A is obtained from A' by row operations, so $RSp(A) \subseteq RSp(A')$. Hence $RSp(A) = RSp(A')$.

In other words row operations have no effect on the row space. In particular $RSp(A) = RSp(A_{ech})$.

For 2. let $i_1, ..., i_k$ be the numbers of the columns of A_{ech} containing the leading entries:

$$\begin{array}{ccc} A_{ech} & = & \begin{pmatrix} 1 & * & * & * & * & ... \\ 0 & 0 & 1 & * & * & ... \\ 0 & 0 & 0 & 1 & * & *... \\ \vdots & & & & & \\ 0 & \dots & & & & \end{pmatrix} \\ & & \begin{array}{cccccc} i_1 & \quad & i_2 & i_3 & ... & \end{array} \end{array}$$

Let $r_1, ...r_k$ are the rows of A_{ech}. Suppose $\lambda_1 r_1 + ... + \lambda_k r_k = 0$ for scalars λ_i. We see that the i_1^{th} entry of $\lambda_1 r_1 + ... + \lambda_k r_k$ is $\lambda_1 \cdot 1 = \lambda_1$ hence $\lambda_1 = 0$. Therefore $\lambda_1 r_1 + ... + \lambda_k r_k = \lambda_2 r_2 + ... + \lambda_k r_k$, similarly the I_2^{th} entry of $\lambda_2 r_2 + ... + \lambda_k r_k$ is λ_2, so $\lambda_2 = 0$. By induction we can show that $a_i = 0$ for all i. So $\{r_1, ..., r_k\}$ is linearly independent.

Example 3.8.4. Find the row rank of $A = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 0 \\ -1 & 4 & 15 \end{pmatrix}$

Answer:

$$A \mapsto \begin{pmatrix} 1 & 2 & 5 \\ 0 & -3 & -10 \\ 0 & 6 & 20 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & \frac{10}{3} \\ 0 & 0 & 0 \end{pmatrix} = A_{ech}$$

A_{ech} has 2 non-zero rows, so the row rank of A is 2.

Example 3.8.5. Find the dimension of

$$W = Span\{(\ -1 \ \ 1 \ \ 0 \ \ 1 \ ), (\ 2 \ \ 3 \ \ 1 \ \ 0 \ ), (\ 0 \ \ 1 \ \ 2 \ \ 3 \ )\} \subseteq \mathbb{R}^4.$$

Answer

We can work this out by seeing our vectors as the rows of a matrix:

Let $A = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}$. The span we want is the row span of this matrix, which we work out:

$$\begin{aligned} A &\mapsto \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix} \mapsto \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 5 & 10 & 15 \end{pmatrix} \\ &\mapsto \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 5 & 1 & 2 \\ 0 & 0 & 9 & 13 \end{pmatrix} = A_{ech} \end{aligned}$$

A_{ech} has 3 non-zero rows so $RSp(A)$ has dimension 3. So $sim(W) = 3$.

We can find the column rank of a matrix in a very similar way to finding the row rank of a matrix.

Procedure 3.8.6. The columns of A are the rows of A^T so we can apply Procedure 3.8.3 to A^T.

Alternatively: use column operations to resuce A to "column echelon form and then count the non-zero columns.

Example 3.8.7. Let $A = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 0 \\ -1 & 4 & 15 \end{pmatrix}$. Find the column rank of A. This equals the row rank of A^T.

$$A^T = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 4 \\ 5 & 0 & 15 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & -1 \\ 0 & -3 & 6 \\ 0 & -10 & 20 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & -1 \\ 0 & -3 & 6 \\ 0 & 0 & 0 \end{pmatrix} = A^T_{ech}$$

So the column rank of A is 2. A basis for $RSp(A^T)$ is $\{\begin{pmatrix} 1 & 2 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -3 & 6 \end{pmatrix}\}$. So a basis for $CSp(A)$ is $\left\{ \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 6 \end{pmatrix} \right\}$

Theorem 3.8.8. For any matrix A the row rank of A is equal to the column rank of A.

Proof:

Let $A = (a_{ij}) \in M_{m\times n}(F)$. Let the rows of A be $r_1, ..., r_m$, so $r_i = (a_{i1}, ..., a_{in})$.Let the columns of A be $c_1, ..., c_n$, so $c_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$.

Let k be the row rank of A. Then $RSp(A)$ has a basis $\{v_1, ..., v_k\}$. Every row r_i is a linear combination of $v_1, ..., v_k$. Say:

$$r_i = \lambda_{i1}v_1 + ... + \lambda_{ik}v_k(\dagger)$$

Suppose that $v_i = (b_{i1}, b_{i2}, ..., b_{in})$ then looking at the j^{th} coordinate in ($\dagger$) we get:

$$a_{ij} = \lambda_{i1}b_{1j} + \lambda_{i2}b_{2j} + ... + \lambda_{ik}b_{kj}$$

Now

$$\begin{aligned} c_j &= \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} \lambda_{11}b_{1j} + \lambda_{12}b_{2j} + ... + \lambda_{1k}b_{kj} \\ \lambda_{21}b_{1j} + \lambda_{22}b_{2j} + ... + \lambda_{2k}b_{kj} \vdots \\ \lambda_{m1}b_{1j} + \lambda_{m2}b_{2j} + ... + \lambda_{mk}b_{kj} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_{11} \\ \vdots \\ \lambda_{m1} \end{pmatrix} b_{1j} + \begin{pmatrix} \lambda_{12} \\ \vdots \\ \lambda_{m2} \end{pmatrix} b_{2j} + ... + \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{mk} \end{pmatrix} b_{kj} \end{aligned}$$

So c_j is a linear combination of the vectors:

$$\begin{pmatrix} \lambda_{11} \\ \vdots \\ \lambda_{m1} \end{pmatrix}, \begin{pmatrix} \lambda_{12} \\ \vdots \\ \lambda_{m2} \end{pmatrix}, ..., \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{mk} \end{pmatrix}$$

Hence $CSp(A)$ is spanned by these vectors, thus $sim(CSp(A) \leq k = dim(RSP(A))$. Equally the column rank of A^T is at most the row rank of A^T (by the same argument). The column rank of A^T is the row rank of A, and the row rank of A^T is the Column rank of A. Thus we have $dim(RSp(A)) \leq dim(CSp(A))$, and hence $dim(RSp(A)) = dim(CSp(A))$.

Example 3.8.9. Let $A = \begin{pmatrix} 1 & 2 & -1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 3 & -1 & 1 \end{pmatrix}$

Note that $r_3 = r_1 + r_2$, so a basis for $RSp(A)$ is

$$\{\underbrace{(1, 2, -1, 0)}_{v_1} \quad , \quad \underbrace{(-1, 1, 0, 1)}_{v_2}\}$$

Write the rows as linear combinations of v_1 and v_2:

$$\begin{aligned} r_1 &= 1v_1 + 0v_2 \\ r_2 &= 0v_1 + 1v_2 \\ r_3 &= 1v_1 + 1v_2 \end{aligned}$$

These co-efficients are the λ_{ij}'s from the proof:

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

According to the proof, a spanning set for $CSp(A)$ is:

$$\begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Check this is really a spanning set for $CSP(A)$: Let $w_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $w_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

Now we have:

$$\begin{aligned} c_1 &= \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} &= w_1 - w_2 \\ c_2 &= \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} &= 2w_1 + w_2 \\ c_3 &= \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} &= -w_1 \\ c_4 &= \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} &= w_2 \end{aligned}$$

So it is indeed the case that $\{w_1, w_2\}$ spans $CSp(A)$.

Definition 3.8.10. Let A be a matrix. The *rank of* A written $rank(A)$ or $rk(A)$, is the row rank of A (or the column rank since they are the same).

Proposition 3.8.11. Let A be an $n \times n$ matrix with entried in F, then the following statements are equivalent:

1. $rank(A) = n$ ("A has full rank").
2. The rows of A form a basis for F^n.
3. The columns of A form a basis for F^n.
4. A is invertible (so $det(A) \neq 0$, etc.).

Proof:

- (1) $\Leftrightarrow$ (2):

$$\begin{aligned} rank(A) = n &\Leftrightarrow dim(RSp(A)) = n \\ &\Leftrightarrow RSp(A) = F^n \\ &\Leftrightarrow \text{the rows of } A \text{ form a basis for } F^n \end{aligned}$$

- (1) $\Leftrightarrow$ (3): The same, but with columns.

- (1) $\Leftrightarrow$ (4): $rank(A) = n$ if and only if $A_{ech} = \begin{pmatrix} 1 & & & & \\ & 1 & & * & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix}$

Now all of the $*$ entries can be eliminated using row operations and so A is reducible to Id using row operations. By 2.6.2 this is equivalent to A being invertible.

4 Linear Transformations

4.1 Introduction

Definition 4.1.1. Suppose V, W are vector spaces over a field F. Let $T : V \longrightarrow W$ be a function from V to W. We say:

- T *preserves addition* if for all $v_1, v_2 \in V$ we have $T(v_1 + v_2) = T(v_1) + T(v_2)$. (i.e. if $T(v_1) = w_1, T(v_2) = w_2$ for $w_1, w_2 \in W$ we have $T(v_1 + v_2) = w_1 + w_2$.
- T *preserves scalar multiplication* if for all $v \in V$, $\lambda \in F$, $T(\lambda v) = \lambda T(v)$.
- T is a *linear transformation* (or *linear map*) if it:
  1. preserves addition.
  2. preserves scalar multiplication

Example 4.1.2.

(a) The identity map $T : V \longrightarrow V$ is obviously a linear transformation.

(b) $T : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $T(x, y) = x + y$ is a linear transformation.
Check:

- $T((x_1, y_1) + (x_2, y_2)) = T((x_1 + x_2, y_1 + y_2)) = x_1 + x_2 + y_1 + y_2 = (x_1 + y_1) + (x_2 + y_2) = T((x_1, y_1)) + T((x_2, y_2))$ So T preserves addition.
- Let $\lambda \in \mathbb{R}$ then $T(\lambda(x, y)) = T((\lambda x, \lambda y)) = \lambda x + \lambda y = \lambda T((x, y))$. So T preserves scalar multiplication.

(c) Let V be the space of all polynomials in x over $\mathbb{R}$ (i.e. $V = \mathbb{R}[x]$). Defin $T : V \longrightarrow V$ by $T(f(x)) = \frac{d}{dx} f(x)$. Then T is a linear map.
Check:

- $T(f(x) + g(x)) = \frac{d}{dx}(f(x) + g(x)) = \frac{d}{dx}(f(x)) + \frac{d}{dx}(g(x)) = T(f(x)) + T(g(x))$ So T preserves addition.
- Let $\lambda \in \mathbb{R}$ then $T(\lambda f(x)) = \frac{d}{dx} \lambda f(x) = \lambda \frac{d}{dx} f(x) = \lambda T(f(x))$. So T preserves scalar multiplication.

(d) Let $V = \mathbb{C}$ (as a 1-dimensional vector space over $\mathbb{C}$). The map $T(z) = \bar{z}$ is ***not*** a linear map:

- $T(z_1 + z_2) = z_1 \bar{+} z_2 = \bar{z_1} + \bar{z_2} = T(z_1) + T(z_2)$ So T *does* preserve addition.
- $T(\lambda z) = \bar{\lambda z} = \bar{\lambda}\bar{z} \neq \lambda \bar{z} = \lambda T(z)$ for $\lambda \notin \mathbb{R}$. So T *does **not*** preserve scalar multiplication.

(e) Let $T : \mathbb{R}^3 \longrightarrow \mathbb{R}$ be given by $T(x, y, z) = (xyz)^{\frac{1}{3}}$ then:

- $T(\lambda(x, y)) = T((\lambda x, \lambda y)) = (\lambda^3 xyz)^{\frac{1}{3}} = \lambda T((x, y, z))$. So T preserves scalar multiplication.
- $T((x_1, y_1, z_1) + (x_2, y_2, z_2)) = T((x_1 + x_2, y_1 + y_2, z_1 + z_2)) = ((x_1 + x_2)(y_1 + y_2)(z_1 + z_2))^{\frac{1}{3}} \neq ((x_1 + y_1 + z_1)^{\frac{1}{3}} + (x_2 + y_2 + z_2))^{\frac{1}{3}} = T((x_1, y_1, z_1)) + T((x_2, y_2, z_2))$. So T *does **not*** preserve addition.

(f) Lots of functions preserve neither addition nor scalar multiplication, e.g., for $\mathbb{R} \to \mathbb{R}$ the functions taking $x \mapsto x+1$, $x \mapsto x^2$, and $x \mapsto e^x$.

Proposition 4.1.3. Let A be an $m \times n$ matrix over F. Define $T : F^n \longrightarrow F^m$ (spaces of column vectors), by $T(v) = Av$ (for $v \in F^n$. Then T is a linear transformation.
Proof: We need to check:

- Preserves addition: Let $v_1, v_2 \in F^n$

$$T(v_1 + v_2) = A(v_1 + v_2) = Av_1 + Av_2 = T(v_1) + T(v_2) \quad \text{by M1GLA}$$

- Preserves scalar multiplicaiton: Let $v \in V$, $\lambda \in F$ then:

$$T(\lambda v) = A(\lambda v) = \lambda A v = \lambda T(v)$$

Proposition 4.1.4. *Basic Properties of linear transformations*

Let $T : V \longrightarrow W$ be a linear map. Write 0_V, 0_W for the zero vectors in V and W respectively. We have:

1. $T(0_v) = 0_W$
2. Suppose $v = \lambda_1 v_1 + ... + \lambda_k v_k$ for $\lambda_i \in F$, $v_i \in V$. Then $T(v) = \lambda_1 T(v_1) + ... + \lambda_k T(v_k)$.

Proof:

1. Since T preserves scalar multiplication we have $T(\lambda 0_v) = \lambda T(0_v)$ for $\lambda \in F$. Taking $\lambda = 0$, we have $T(00_v) = 0T(0_v)$, but $0 \cdot 0_v = 0_v$ and $0 \cdot T(0_v) = 0_W$. Hence $T(0_v) = 0_W$.
2. Induction on k.
   Base case. The case where $k = 1$ just says T preserves scalar multiplication, so it true.

   Inductive step: Suppose we know $T(\lambda_1 v_1 + ... + \lambda_{k-1} v_{k-1}) = \lambda_1 T(v_1) + ... + \lambda_{k-1} T(v_{k-1})$. Now
$$\begin{aligned} T(\lambda_1 v_1 + ... + \lambda_k v_k) &= T(\lambda_1 v_1 + ... + \lambda_{k-1} v_{k-1}) + T(\lambda_k v_k) \\ &= T(\lambda_1 v_1 + ... + \lambda_{k-1} v_{k-1}) + \lambda_k T(v_k) \\ &= T(\lambda_1 v_1 + ... + \lambda_{k-1} v_{k-1}) + \lambda_k T(v_k) \end{aligned}$$

Example 4.1.5. *Question:* Find the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that $T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ and $T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$.

Answer: Note that $\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$ form a basis for $\mathbb{R}^2$, a general vector of $\mathbb{R}^2$ is $\begin{pmatrix} a \\ b \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. So we must have:

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

This map is linear as $T\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$, so a matrix transformation.

Proposition 4.1.6. Let V and W be vector spaces over F. Let $\{v_1, ..., v_n\}$ be a basis for V. Let $w_1, ..., w_n$ be any n vectors from W (these don't need to be distinct). Then there is a unique linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for all i.

Proof: Suppose that $v \in V$, then there exist $\lambda_1, ..., \lambda_n$ such that $v = \lambda_1 v_1 + ... + \lambda_n v_n$. Define the following map:

$$\begin{aligned} &T : V \to W \\ &T(v) = \lambda_1 w_1 + ... + \lambda_n w_n \end{aligned}$$

Claim: T is a linear transformation.

- *T preserves addition:.* Suppose $v, u \in V$, so we have $v = \lambda_1 v_1 + ... + \lambda_n v_n$ and $u = \mu_1 v_1 + ... + \mu_n v_n$. So:

$$\begin{aligned} T(v+u) &= T(\lambda_1 v_1 + ... + \lambda_n v_n + \mu_1 v_1 + ... + \mu_n v_n) \\ &= T((\lambda_1 + \mu_1)v_1 + ... + (\lambda_n + \mu_n)v_n) \\ &= (\lambda_1 + \mu_1)w_1 + ... + (\lambda_n + \mu_n)w_n \\ &= \lambda_1 w_1 + ... + \lambda_n w_n + \mu_1 w_1 + ... + \mu_n w_n \\ &= T(v) + T(u) \end{aligned}$$

- *T preserves scalar multiplication:* Suppose $v \in V$ and $\alpha \in F$, we have $v = \lambda_1 v_1 + ... + \lambda_n v_n$.

So

$$\begin{aligned} T(\alpha v) &= T(\alpha(\lambda_1 v_1 + ... + \lambda_n v_n)) \\ &= T(\alpha\lambda_1 v_1 + ... + \alpha\lambda_n v_n) \\ &= \alpha\lambda_1 w_1 + ... + \alpha\lambda_n w_n \\ &= \alpha(\lambda_1 w_1 + ... + \lambda_n w_n) \\ &= \alpha T(v) \end{aligned}$$

So it remains to check uniqueness. Suppose that we have a linear transformation S such that $S(v_i) = w_i$ for all i. Then we have:

$$\begin{aligned} S(\lambda_1 v_1 + ... + \lambda_n v_n) &= \lambda_1 S(v_1) + ... + \lambda_n S(v_n) \\ &= \lambda_1 w_1 + ... + \lambda_n w_n \end{aligned}$$

So $T = S$ proving uniqueness.

Remark 4.1.7. This shows that once we know what a linear transformation does to a basis we know what the transformation is.

Example 4.1.8. Let V be the space of all polynomials in x over $\mathbb{R}$ with degree less than or equal to 2. A basis for this is $\{1, x, x^2\}$. We can pick any three arbitary vectors in V for example:

$$\begin{aligned} &w_1 = 1 + x \\ &w_2 = x - x^2 \\ &w_3 = 1 + x^2 \end{aligned}$$

By Proposition 4.1.6 there is a linear transformation $T : V \to V$ such that $T(1) = w_1, T(x) = w_2$, $T(x^2) = w_3$.

We can work out what T does to a general element of V. A general element is of the form $v = a1 + bx + cx^2$, so

$$\begin{aligned} T(v) &= T(a1 + bx + cx^2) \\ &= a(1 + x) + b(x - x^2) + c(1 + x^2) \\ &= (a + c) + (a + b)x + (-b + c)x^2 \end{aligned}$$

4.2 Image and Kernel

Definition 4.2.1. Let $T : V \to W$ be a linear transformation:

- The *Image of* T is the set $Im\,T = \{T(v) \in W : v \in V\} \subseteq W$.
- The *Kernel of* T is the set $Ker\,T = \{v \in V : T(v) = 0_W\} \subseteq V$.

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

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

- The image of T is the set of all vectors in $\mathbb{R}^2$ of the form $\begin{pmatrix} 3x_1 + x_2 + 2x_3 \\ -x_1 + x_3 \end{pmatrix}$ for $x_1, x_2, x_3 \in \mathbb{R}$. This is the space:

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

- The kernel of T is the set of vectors in $\mathbb{R}^3$ such that $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0_W$ that is so say such that:

$$\begin{pmatrix} 3x_1 + x_2 + 2x_3 \\ -x_1 + x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Alternatively this is the solution space of $Ax = 0$. In this case the kernel is $Sp\begin{pmatrix} 1 \\ -5 \\ 1 \end{pmatrix}$.

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

1. $Im\,T$ is a subspace of W.
2. $Ker\,T$ is a subspace of V.

Note: In general we write $U \leq V$ to mean U is a subspace of V, so with this notation we are saying $Im\,T \leq W$ and $Ker\,T \leq V$.

Proof: For both we need to check the vector space criterion.

1.
   - Certainly $Im\,T \neq \emptyset$, since $T(0) \in Im\,T$.
   - Suppose $w_1, w_2 \in Im\,T$ then there exist $v_1, v_2 \in V$ such that $w_1 = T(v_1)$ and $w_2 = T(v_2)$. Now,
   $$T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2$$
   So $w_1 + w_2 \in Im\,T$.

- Suppose $w \in ImT$ and let $\lambda \in F$. We have $w = T(v)$ for some $v \in V$, now $T(\lambda v) = \lambda T(v) = \lambda w$. So $\lambda w \in ImT$

So $ImT \leq W$.

Example 4.2.4. Let V_n be the vector space of polynomials in x over $\mathbb{R}$ of degree $\leq n$. We have $V_0 \leq V_1 \leq V_2$.... Define:

$$T : V_n \to V_{n-1},$$
$$T(f(x)) = f'(x).$$

Note: T is linear.

$$\begin{aligned} KerT &= \{f(x) : f'(x) = 0\} \\ &= \{\text{constant polys}\} \\ &= V_0 \end{aligned}$$

Suppose $g(x)$ has degree $\leq n-1$. Then by integrating $g(x)$ we can find $f(x)$ such that $f'(x) = g(x)$ and $deg(f(x)) = 1 + deg(g(x))$, so $deg(f(x)) \leq n$. Hence $ImT = V_{n-1}$.

Of course the $f(x)$ such that $f'(x) = g(x)$ is not unique - if c is a constant then $f(x) + c$ also has this property. In fact we get the set $\{h(x) : h'(x) = g(x)\}$ consists of polynomials $f(x) + k(x)$ where $k(x) \in KerT$.

Proposition 4.2.5. Let $T : V \to W$ be a linear tranformation and let $v_1, v_2 \in V$. Then

$$T(v_1) = T(v_2) \text{ iff } v_1 - v_2 \in Ker\, T.$$

Proof:

$$\begin{aligned} T(v_1) = T(v_2) \quad &\text{iff} \quad T(v_1) - T(v_2) = 0 \\ &\text{iff} \quad T(v_1 - v_2) = 0 \\ &\text{iff} \quad v_1 - v_2 \in Ker\, T \end{aligned}$$

Proposition 4.2.6. Let $T : V \to W$ be a linear transformation. Suppose that $\{v_1, ..., v_n\}$ is a basis for V. Then $Im\, T = Span\{T(v_1), .., T(v_n)\}$.

Proof: Clearly $Span\{T(v_1), .., T(v_n)\} \subseteq Im\, T$. Conversely, let $w \in ImT$. Then $w = Tv$ for some $v \in V$. Since $\{v_1, ..., v_n\}$ is a basis for V we can find scalars λ_i such that

$$\begin{aligned} v &= \lambda_1 v_1 + ...\lambda_n v_n \\ w &= T(v) \\ &= T(\lambda_1 v_1 + ...\lambda_n v_n) \\ &= \lambda_1 T(v_1) + ...\lambda_n T(v_n) \in Span\{T(v_1), ..., T(v_n)\} \end{aligned}$$

Proposition 4.2.7. Let A be an $m \times n$ matrix. Let $T : F^n \to F^m$ be given by $T(v) = Av$. Then:

1. $Ker\, T$ is the solution space to $Av = 0$.
2. $Im\, T$ is the column space of A.
3. $dim(Im\, T) = rankA$.

Proof:

1. Immediate from definitions
2. Take the "standard" basis for F^n that is:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, ..., e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

   By proposition 4.2.6 we have $Im\, T = Span\{T(e_1), ..., T(e_n)\}$. Now $T(e_i) = Ae_i = c_i$ where c_i is the i^{th} column of A. SO $Im\, T = Span\{c_1, ..., c_n\} = CSp(A)$.
3. By (ii) $dim(Im\, T) = dim(CSp(A)) =$ column rank of $A = rk(A)$

Theorem 4.2.8. ***The rank nulity theorem**: We've seen that when $Tv = Av$, $rank(A) = dim(Im\, T)$. An old fashioned name for $dim(Ker\, T)$ is the nulity of A*

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

$$dim(Im\, T) + dim(Ker\, T) = dim(V)$$

Proof: Let $\{u_1, ...u_s\}$ be a basis for $\ker T$, and let $\{w_1, ..., w_r\}$ be a basis for $Im\, T$. For each $w_i \in ImT$, and so $\exists v_i \in V$ with $Tv_i = w_i$. We claim that $B = \{u_1, ..., u_s\} \cup \{v_1, ...v_r\}$ is a basis for V.

- *Spanning set:* Let $v \in V$ since $Tv \in Im\, T$ we can write $Tv = \lambda_1 w_1 + ...\lambda_r w_r$ for scalars λ_i. So

$$\begin{aligned} Tv &= \lambda_1 w_1 + ...\lambda_r w_r \\ &= T(\lambda_1 v_1 + ...\lambda_r v_r) \end{aligned}$$

  Now by proposition 4.2.5 $v - \lambda_1 v_1 + ...\lambda_r v_r \in \ker T$ so $v - \lambda_1 v_1 + ...\lambda_r v_r = \mu_1 u_1 + ... + \mu_s u_s$. Thus

$$v = \mu_1 u_1 + ... + \mu_s u_s + \lambda_1 v_1 + ...\lambda_r v_r \in span(B)$$

- *Linear independence* Suppose:

$$\lambda_1 v_1 + ...\lambda_r v_r + \mu_1 u_1 + ... + \mu_s u_s = 0$$

  By applying T we get:

$$\begin{aligned} 0 &= T(\lambda_1 v_1 + ...\lambda_r v_r + \mu_1 u_1 + ... + \mu_s u_s) \\ &= \lambda_1 T(v_1) + ...\lambda_r T(v_r) + \mu_1 T(u_1) + ... + \mu_s T(u_s) \\ &= \lambda_1 w_1 + ...\lambda_r w_r \end{aligned}$$

  Thus $\lambda_1 = ... = \lambda_r = 0$, so we get that $\mu_1 u_1 + ... + \mu_s u_s = 0$, so $\mu_1 = ... = \mu_s = 0$.

Example 4.2.9.

Let $a, b, c \in \mathbb{R}$, define $U = \{(x, y, z) \in \mathbb{R}^3 : ax + by + cz = 0\}$. U is a subspace of $\mathbb{R}^3$.

We can find dimension of U by defining:

$$\begin{aligned} &T : \mathbb{R}^3 \to \mathbb{R} \\ &T(x, y, z) = (a, b, c) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \end{aligned}$$

Now $U = \ker T$, and clearly $ImT = \mathbb{R}$ (as not all $a, b, c = 0$), thus $dim(ImT) = 1$. So

$$\begin{aligned} dimU &= dim(\ker T) \\ &= dim(\mathbb{R}^3) - dim(ImT) \\ &= 3 - 1 = 2 \end{aligned}$$

Corollary 4.2.10. A system of linear equations in n unknowns with co-efficients in F:

$$\begin{array}{llll} a_{11}x_1 & +a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 & +a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & +a_{m2}x_2 + a_{m3}x_3 + \ldots + a_{mn}x_n & = & b_m \end{array}$$

is called *homogeneous* if $b_1 = b_2 = ... = b_m = 0$.

We know in this case that we will always get at least a trivial solution to the system - and we saw in the test that the set of solutions forms a subspace of F^n, but what dimension will this subspace have?

We can use the rank-nulity theorem to work this out:

We know that if we let $A = (a_{ij})$, then this system of linear equations can be represented as $Ax = 0$. We also know that A can be seen as a linear transformation $A : F^n \mapsto F^m$.

By Proposition 4.2.7 the set of solutions in this case is $ker(A)$, and by the rank nulity we get

$dim(ker(A)) = dim(F^n) - dim(Im(A))$

Now the $dim(Im(A)) = rank(A)$ thus the we can work out how many solutions we have to a set of homogeneous equations with n unknowns:

- If $rank(A) \geq n$ we get one solution (the trivial one i.e. 0_V)
- If $rank(A) < n$ we get infinitely many solutions (assuming F is infinite)

Exercise 4.2.11. In this case the rank of the augmented matrix $(A|0)$ is the same as that of A.

How does this work for a non homogeneous system of linear equations?

Essentially almost the same except - but we are taking a coset of the system of equations and we have to account for the case were $rank(A) < rank(A|b)$

4.3 Representing vectors and transformations with respect to a basis

Let V be an n-dimensional v.s. over F, let $B = \{v_1, ..., v_n\}$ be a basis for V.

Definition 4.3.1. For $v \in V$ with $v = \lambda_1 v_1 + ... + \lambda_n v_n$ the *vector of* V *wrt* B is

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

This is well defined since v has a unique expression as a linear combination of $v_1, ..., v_n$.

Example 4.3.2.

(a) $V = \mathbb{R}^3$, $B = \{e_1, e_2, e_3\}$. Then

$$[\begin{pmatrix} a \\ b \\ c \end{pmatrix}]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ as } \begin{pmatrix} a \\ b \\ c \end{pmatrix} = ae_1 + be_3 + ce_3$$

(b) Let V be the v.s. of polys in x of degree ≤ 2

- $B = \{1, x, x^2\}$ then $[a + bx + cx^2]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.
- If instead we take $B = \{x^2, x, 1\}$ then $[a + bx + cx^2]_B = \begin{pmatrix} c \\ b \\ a \end{pmatrix}$
- Or $B = \{1, x + 1, x^2 + x + 1\}$ then:

$$a + bx + cx^2 = (a - b) + (b - c)(x + 1) + c(x^2 + x + 1)$$

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

Proposition 4.3.3. Let V be an n-dimensional vector space over F with a basis B. Then the map:

$$T : V \to F^n$$
$$T(v) = [v]_B$$

is a bijective linear transformation (i.e. a *linear isomorphism*).

Proof: Suppose $B = \{v_1, ..., v_n\}$

1. Linear Transformation:

   (a) Preserves Addition:
   LEt $u, v \in V$ then $u = \lambda_1 v_1 + ... + \lambda_n v_n$ and $v = \mu_1 v_1 + ... + \mu_n v_n$ so $u + v =$

$(\lambda_1 + \mu_1)v_1 + ... + (\lambda_n + \mu_n)v_n$.

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

Therefore

$$\begin{aligned} [u+v]_B &= [u]_B + [v]_B \\ T(u+v) &= T(u) + T(v) \end{aligned}$$

(b) Preserves scalar multiplication:
LEt $u \in V$ and $\alpha \in F$ so $u = \lambda_1 v_1 + ... + \lambda_n v_n$, now $\alpha u = (\alpha\lambda_1)v_1 + ... + (\alpha\lambda_n)v_n$

So $[u]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}, \quad [\alpha u]_B = \begin{pmatrix} \alpha\lambda_1 \\ \vdots \\ \alpha\lambda_n \end{pmatrix}$ So

$$\begin{aligned} [\alpha u]_B &= \alpha[u]_B \\ T(\alpha u) &= \alpha T(u) \end{aligned}$$

2. T is bijective:

(a) Injective:
Suppose $u, w \in V$ such that $T(u) = T(w)$ then $T(u - w) = 0$ as T is linear.

So $[u - w]_B = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ so $u - w = 0v_1 + ... + 0v_n = 0$ hence $u = w$

(b) Surjective:

Let $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in F^n$ now $[a_1 v_1 + ... + a_n v_n]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ So $T(a_1 v_1 + ... + a_n v_n) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, thus T is surjective.

Construction 4.3.4.

Now let V, W be finite dimensional vector spaces over F

- $B = \{v_1, ..., v_n\}$ a basis for V.
- $C = \{w_1, ...w_m\}$ a basis for W.

Let $T : V \mapsto W$ be a linear transformation, we have:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, leftrightarrow, "{[-]_B}"'] & W \arrow[d, leftrightarrow, "{[-]_C}"] \\ F^n \arrow[r, dashed] & F^m \end{tikzcd}$$

We can define a map $F^n \to F^m$ by following the diagram around. This map is linear as it is a composition of linear maps (exercise).

Now a linear map $F^n \mapsto F^m$ is a matrix transformation (by hand-in). Let A be the matrix for this transformation, then $A[v]_B = [Tv]_C$.

We calculate A by figuring out it's columns $c_1, ..., c_n$. To calculate c_i, we work out Tv_i and find

$$Tv_i = a_{1i}w_1 + ... + a_{mi}w_m,$$

so we get $c_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}$. We get:

$$c_i = Ae_i = A[v_i]_B = [Tv_i]_C.$$

Definition 4.3.5. The matrix A constructed above is *the matrix of T with respect to B and C* , we write this ${}_C[T]_B$, so ${}_C[T]_B[v]_B = [Tv]_C$. If $V = W$ and $B = C$ we sometimes write this simply as $[T]_B$.

Remark 4.3.6. If $T : V \mapsto V$ and B a basis for V then for all $v \in V$ $[Tv]_B = [T]_B[v]_B$

Example 4.3.7.

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \text{ defined by } T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 \\ x_1 + 2x_2 \end{pmatrix}$$

- Take $E = \{e_1, e_2\}$. Find $[T]_E = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$
- Let $B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$. Find $[T]_B = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$
- Find ${}_B[T]_E = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$.

Proposition 4.3.8. Let V be a vector space. Let $B = \{v_1, ..., v_n\}$ and $C = \{w_1, ..., w_n\}$ be bases for V. Then for $j \in \{1, ..., n\}$ we can write $v_j = \lambda_{ij} w_1 + ... + \lambda_{nj} w_n$.

Let P be the matrix $(\lambda_{ij}) = \begin{pmatrix} \lambda_{11} & ... & \lambda_{1n} \\ \vdots & & \vdots \\ \lambda_{n1} & ... & \lambda_{nn} \end{pmatrix}$. So the j^{th} column is $[v_j]_C$.

1. $P = [X]_C$ where $X : V \to V$ is the unique linear transformation such that $X(w_j) = v_j$ for all j.
2. For all $v \in V$, $P[v]_B = [v]_C$.
3. $P = {}_C[Id]_B$ where Id is the identity transformation of V.

Proof:

1. The j^{th} column of $[X]_C$ is the image $X(w_j)$ written as a vector in C. Now $X(w_j) = v_j$ so the j^{th} column is $[v_j]_C$ and this is the j^{th} column of P, so $[X]_C = P$.
2. For a basis vector $v_j \in B$ we have:

$$\begin{aligned} P[v_j]_B &= Pe_j \\ &= j^{th} \text{ Column of } P \\ &= [v_j]_C \end{aligned}$$

   So the claim is true for elements of the basis B, hence it is true for all $v \in V$.
3. Exercise (essentially part (ii) expressed differently).

Definition 4.3.9. P is the *change of basis matrix* from B to C.
******Warning*********** Confusing because of 1 in Prop 4.3.8 maps basis elements of C to those of B - sometimes described the other way around.

Proposition 4.3.10. Let V, B, C P as above. Then:

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

Proof:

1. Let Q be the change of basis matrix from C to B. Then:
   $Q[v]_C = [v]_B$ for all $v \in V$
   $P[v]_B = [v]_C$ for all $v \in V$

   Hence $QP[v]_B = Q[v]_C = [v]_B$. As v ranges over V, $[v]_B$ ranges over all of F^n. So $QPx = x$ for all $x \in F^n$. Therefore $QP = I_n$, hence P is invertible with inverse Q.

2.

$$\begin{array}{rcll} [T]_C[v]_C & = & [T(v)]_C & \text{for all } v \in V \\ \\ (P[T]_B P^{-1})[v]_C & = & (P[T]_B P^{-1})P[v]_B & \\ & = & (P[T]_B)(P^{-1})P)[v]_B & \\ & = & (P[T]_B)[v]_B & \\ & = & (P[T(v)]_B) & \\ & = & [T(v)]_C & \end{array}$$

AS this is for all $v \in V$ we have $(P[T]_B P^{-1}) = [T]_C$.

Example 4.3.11. $V = \mathbb{R}^2$, $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ -2x_1 + 3x_2 \end{pmatrix}$. Take bases $B = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ and $E = \{e_1, e_2\}$

Caluculate:

1. $[T]_E = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}$, $[T]_B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
2. $[P]$ the change of basis matrix from E to B (hint: find P^{-1})

Remark 4.3.12. It is a fact that if P is the change of basis matrix ${}_C[Id]_B$ from B to C and Q is the change of basis matrix ${}_D[Id]_C$ (where B, C, D are all basis for F^n, then $QP = {}_D[Id]_C\,{}_C[Id]_B = {}_D[Id]_B$, the change of basis matrix from B to D.

In Example 4.3.11, we saw that for any given basis B of F^n the matrix ${}_E[Id]_B$ was easy to calculate, since its columns are the elements of B. Now as

$$\begin{array}{rcl} {}_C[Id]_B & = & {}_C[Id]_E\,{}_E[Id]_B \\ & = & ({}_E[Id]_C)^{-1}\,{}_E[Id]_B \end{array}$$

This gives us a quick method of calculating chance of basis matrices for F^n.