Linear Algebra I

Vicky Neale

Michaelmas Term 2019

Contents

1 Introduction

These notes are to accompany the 2019–20 Oxford Prelims course Linear Algebra I.

This course is an introduction to Linear Algebra. We'll meet matrices, and how we can use them to solve systems of simultaneous linear equations. We'll also meet vector spaces, and explore some of their properties.

There are several resources that will help you as you study the course:

- the lectures
- these notes
- the problems sheets, with starter, main course and pudding problems
- the solutions to the starter and pudding problems

- each other

- tutorials in college.

In places, you will find blanks in these notes. They are there deliberately! We'll fill in these blanks during lectures.

I'll regularly be uploading new versions of these notes during the term, both as I review material for later lectures, and as I fix any typos we spot along the way.

I have included the notes from last year's course, so that you can look ahead if you want to, but please be aware that things might change significantly. I have clearly marked the point where we switch from updated 2019 notes to the old 2018 notes. To find this point in an electronic copy of these notes, search for "2018 course" and you'll find my note.

If I find (or am told about) further typos then I will continue to fix them. The version on the course materials website will always be the most up-to-date version, so if you think that you have found a typo then please check there before emailing me.

Acknowledgements

These notes, and the lectures they accompany, are extremely closely based on those produced by Dr Peter Neumann, which in turn built on notes by previous lecturers. The same applies to the problems sheets.

I would like these notes to be as useful as possible. If you (whether student or tutor) think that you've noticed a typo, or mistake, or part that is unclear, please check the current, up-to-date, notes on the website, to see whether I've already fixed it. If not, please email me (vicky.neale@maths) and I'll do something about it, and (with your permission) thank you here.

Thanks to Jacob Armstrong, Harry Best, Dan Claydon, Raymond Douglas, Liam Hopson, Dominik Koller, Amrit Lohia, Gianmarco Luppi, Shaun Marshall and Jingjie Yang for helping to fix glitches in these notes, problems sheets and solutions.

2 Linear equations and matrices

2.1 Systems of linear equations

You already have experience from school of solving systems of simultaneous equations such as

$$\begin{cases} 3x & + & 5y & = & -1 \\ 4x & - & y & = & 10 \end{cases} \tag{1}$$

In this course, we'll explore this in more generality. What if there are more equations? What if there are more unknowns? For example

$$\begin{cases} 8x & - & 7y & + & 6z & = & 59 \\ x & & & + & 2z & = & 9 \\ 3x & + & 2y & - & z & = & 11 \end{cases} \tag{2}$$

or

$$\begin{cases} x & + & y & + & z & + & w & = & 2 \\ -x & + & y & - & z & + & w & = & 0 \\ 3x & + & 2y & & & - & 10w & = & 7 \end{cases} \tag{3}$$

How can we tell whether there is a solution? How can we tell whether there are many solutions? How can we (efficiently) find all the solutions?

Solving systems of simultaneous linear equations is important across the whole of mathematics, as well as in many other disciplines that use tools from mathematics. You'll find the ideas that we develop in this course being useful in many other courses.

You already have strategies for solving systems of equations such as those above. Such strategies may not be practical if the system has hundreds of equations in hundreds of variables. One of our goals is to understand strategies that can be effectively employed by a computer.

Matrices and the beginnings of matrix algebra

Writing a system of equations in the form (1) or (2) or (3) is visually convenient for seeing what is going on, but is not always convenient for manipulation. We can use *matrices* to record the key information—the coefficients.

For example, the key information in (1) is the matrix $\begin{pmatrix} 3 & 5 \\ 4 & -1 \end{pmatrix}$ recording the coefficients from the left-hand side, and the matrix $\begin{pmatrix} -1 \\ 10 \end{pmatrix}$ recording the data from the right-hand side.

Similarly, the data from (2) can be recorded using the two matrices

$$\begin{pmatrix} 8 & -7 & 6 \\ 1 & 0 & 2 \\ 3 & 2 & -1 \end{pmatrix} \text{ and } \begin{pmatrix} 59 \\ 9 \\ 11 \end{pmatrix}$$

and from (3) using

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 3 & 2 & 0 & -10 \end{pmatrix} \text{ and } \begin{pmatrix} 2 \\ 0 \\ 7 \end{pmatrix}.$$

Definition. For m, $n \geq 1$, an $m \times n$ *matrix* is a rectangular array with m rows and n columns. For us, the entries will be real or complex numbers (it is possible to be more general).

We number the rows as 1, 2, ..., m from top to bottom, and the columns 1, 2, ..., n from left to right.

We refer to the entry in row i and column j as the (i, j) entry. If the matrix is A, then we record the (i, j) entry as a_{ij}, and often write the matrix as $A = (a_{ij})$.

Remark. There are variations on this notation, depending on context and on personal preference. The same is true for other pieces of mathematical notation too!

Definition. A $1 \times n$ matrix is called a *row vector*. An $m \times 1$ matrix is called a *column vector*. An $n \times n$ matrix is called a *square matrix*. If $A = (a_{ij})$ is a square matrix and $a_{ij} = 0$ whenever $i \neq j$, then we say that A is a *diagonal matrix*.

Example. A row vector

A column vector

A square matrix that is not diagonal

A square matrix that is diagonal

Definition. The *zero matrix* has all entries 0: if $A = (a_{ij})$ is the $m \times n$ zero matrix then $a_{ij} = 0$ for $1 \leq i \leq m$, $1 \leq j \leq n$. We write $A = 0$ (or maybe $A = 0_{m\times n}$).

Definition. The entries of a matrix are called *scalars*. We often write $\mathbb{F}$ for the set from which the entries come. (Secretly, $\mathbb{F}$ will usually be a *field*.) For us, usually $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, but occasionally $\mathbb{F} = \mathbb{Q}$ or another set.

We write $\mathcal{M}_{m\times n}(\mathbb{F}) = \{A : A \text{ is an } m \times n \text{ matrix with entries from } \mathbb{F}\}$.

We sometimes write $\mathbb{F}^n$ for $\mathcal{M}_{1\times n}(\mathbb{F})$ or $\mathbb{F}^m$ for $\mathcal{M}_{m\times 1}(\mathbb{F})$. Whether we mean row vectors or column vectors will depend on context.

We introduced matrices as a way to record data from systems of equations. But what is really important about matrices is not just that they record information tidily, but rather that (if the conditions are right) we can manipulate matrices: we can, sometimes, add and multiply matrices, for example.

Definition (Addition of matrices)**.** Let $A = (a_{ij})$, $B = (b_{ij})$ be $m \times n$ matrices. We define the sum $A + B = (c_{ij})$ to have (i, j) entry $c_{ij} = a_{ij} + b_{ij}$. We describe the addition as *coordinatewise*.

Remark. We have defined addition of matrices only when the matrices have the same size. We can't add any old pair of matrices.

Definition (Scalar multiplication of matrices)**.** Let $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$ and take $\lambda \in \mathbb{F}$. Define λA to be the $m \times n$ matrix with (i, j) entry λa_{ij}. We call this *scalar multiplication*.

Addition and scalar multiplication of matrices have some familiar, and useful, properties.

Proposition 1. *Take A, B, $C \in \mathcal{M}_{m\times n}(\mathbb{F})$ and λ, $\mu \in \mathbb{F}$. Then*

(i) $A + 0_{m\times n} = 0_{m\times n} + A = A$;

(ii) $A + B = B + A$ *(addition is* commutative*)*;

(iii) $A + (B + C) = (A + B) + C$ *(addition is* associative*)*;

(iv) $\lambda(\mu A) = (\lambda\mu)A$;

(v) $(\lambda + \mu)A = \lambda A + \mu A$;

(vi) $\lambda(A + B) = \lambda A + \lambda B$.

Proof. Exercise. □

We would like to multiply matrices. The definition of matrix multiplication can look a little mysterious at first. As we'll see, matrices can be interpreted as describing certain geometric transformations. We then want matrix multiplication to correspond to composition of these transformations (doing one and then another). This informs why we define multiplication in the way we do.

In order to multiply matrices, we need to take care that the sizes of the matrices are compatible.

Definition. Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, $B \in \mathcal{M}_{n\times p}(\mathbb{F})$. Then AB is the $m \times p$ matrix with (i, j) entry

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

Remark. I visualise matrix multiplication using a hand gesture. I cannot capture this in written lecture notes, but I shall demonstrate and describe it in the lecture!

Remark. Note that the number of *columns* of A must equal the number of *rows* of B.

Example.

$$\begin{pmatrix} 8 & -7 & 6 \\ 1 & 0 & 2 \\ 3 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 3 & 2 & 0 & -10 \end{pmatrix} = \begin{pmatrix} 33 & 13 & 15 & -59 \\ & & & \\ & & & 15 \end{pmatrix}$$

Remark. We have defined matrix multiplication in such a way that we can see how to implement it on a computer. But how long will it take for a computer to run such a calculation?

To multiply two $n \times n$ matrices in this way, for each of the n^2 entries we must multiply n pairs and carry out $n - 1$ additions. So the process takes

around n^3 multiplications and $n^2(n-1)$ additions. When n is large, these are very large numbers!

In 1969, Strassen gave a faster algorithm, which has since been improved on. It is not known whether these algorithms give the fastest possible calculations. Such research falls into the field of *computational complexity*, drawing on ideas from both mathematics and computer science.

Like addition, multiplication of matrices also has some useful properties.

Proposition 2. *Let A, $A' \in \mathcal{M}_{m\times n}(\mathbb{F})$, B, $B' \in \mathcal{M}_{n\times p}(\mathbb{F})$, $C \in \mathcal{M}_{p\times q}(\mathbb{F})$, and $\lambda \in \mathbb{F}$. Then*

(i) $A(BC) = (AB)C$ *(multiplication is associative);*

(ii) $(A + A')B = AB + A'B$;

(iii) $A(B + B') = AB + AB'$;

(iv) $(\lambda A)B = A(\lambda B) = \lambda(AB)$.

Remark. (ii) and (iii) together are known as *distributivity* of multiplication over addition.

Proof. Exercise. □

Definition. Let A, $B \in \mathcal{M}_{n\times n}(\mathbb{F})$. We say that A and B *commute* if $AB = BA$.

Remark. Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, $B \in \mathcal{M}_{n\times m}(\mathbb{F})$. Then we can define both AB and BA. Note that AB is an $m \times m$ matrix, and BA is an $n \times n$ matrix, so if $m \neq n$ then certainly $AB \neq BA$. If $m = n$, then we may or may not have $AB = BA$—it depends.

Definition. Let $A = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. We say that A is

diagonal if $a_{ij} = 0$ whenever $i \neq j$;

upper triangular if $a_{ij} = 0$ whenever $i > j$;

lower triangular if $a_{ij} = 0$ whenever $i < j$.

Example. diagonal (this is also both upper and lower triangular)

upper triangular

lower triangular

Definition. The $n \times n$ *identity matrix* I_n has (i, j) entry

$$\begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Example.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Lemma 3. *Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, $B \in \mathcal{M}_{n\times p}(\mathbb{F})$. Then $AI_n = A$ and $I_nB = B$.*

Proof. Exercise. □

Definition. We say that $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ is *invertible* if there exists $B \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $AB = I_n = BA$.

Lemma 4. *If $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ is invertible, then it has a unique inverse $B \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $AB = BA = I_n$.*

Proof. Suppose that B, $C \in \mathcal{M}_{n\times n}(\mathbb{F})$ are both inverses for A. [Secret aim: $B = C$]

Then $AB = BA = I_n$ and $AC = CA = I_n$
so $B = BI_n = B(AC) = (BA)C = I_nC = C$. □

Definition. If $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ is invertible, then we write A^{-1} for its inverse.

Proposition 5. *Let A, B be invertible $n\times n$ matrices. Then AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. Exercise. □

Definition. The *transpose* of $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$ is the $n \times m$ matrix A^T with (i, j) entry a_{ji}.

Example.

$$\begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix}^T = \begin{pmatrix} & \\ & \end{pmatrix}.$$

Remark. We sometimes write a column vector as the transpose of a row vector:

$$\begin{pmatrix} 1 \\ -3 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & -3 & 0 & 2 \end{pmatrix}^T.$$

Definition. We say that $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is *orthogonal* if $AA^T = I_n = A^TA$. Equivalently, A is invertible and $A^{-1} = A^T$.

We say that $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is *unitary* if $A\overline{A}^T = I_n = \overline{A}^TA$. (By $\overline{A}$ we mean the matrix obtained from A by replacing each entry by its complex conjugate.)

Remark. We'll explore orthogonal (and perhaps unitary) matrices later in the course. Look out for them in other courses, such as Geometry!

2.2 Solving systems of simultaneous linear equations using matrices

Let's revisit the examples from earlier:

$$\begin{cases} 3x + 5y = -1 \\ 4x - y = 10 \end{cases} \qquad (1)$$

$$\begin{cases} 8x - 7y + 6z = 59 \\ x + 2z = 9 \\ 3x + 2y - z = 11 \end{cases} \qquad (2)$$

$$\begin{cases} x + y + z + w = 2 \\ -x + y - z + w = 0 \\ 3x + 2y - 10w = 7 \end{cases} \qquad (3)$$

We can represent each system of equations as an equation involving matrices.

$$\begin{pmatrix} 3 & 5 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 10 \end{pmatrix} \qquad (1)$$

$$\begin{pmatrix} 8 & -7 & 6 \\ 1 & 0 & 2 \\ 3 & 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 59 \\ 9 \\ 11 \end{pmatrix} \qquad (2)$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 3 & 2 & 0 & -10 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 7 \end{pmatrix} \qquad (3)$$

How might we go about solving such an equation?

Here's how I might try to solve the system (2), without using matrices.

$$\begin{cases} 8x - 7y + 6z = 59 \\ x + 2z = 9 \\ 3x + 2y - z = 11 \end{cases} \qquad (2)$$

I can divide the top equation by 8: the given system of equations is equivalent to

$$\begin{cases} x - \frac{7}{8}y + \frac{3}{4}z = \frac{59}{8} \\ x + 2z = 9 \\ 3x + 2y - z = 11 \end{cases} \qquad (2')$$

Now the first equation gives x in terms of y and z. By subtracting suitable multiples of the first equation from the second and third, I can eliminate x from each of them: our system of equations is equivalent to

$$\begin{cases} x - \frac{7}{8}y + \frac{3}{4}z = \frac{59}{8} \\ \frac{7}{8}y + \frac{5}{4}z = \frac{13}{8} \\ \frac{37}{8}y - \frac{13}{4}z = -\frac{89}{8} \end{cases} \qquad (2'')$$

(In this case, we might have subtracted multiples of the second equation from the first and third at the outset, but I'm trying to avoid strategies that rely on particular features of the specific equations.)

Now we have two equations in just two variables—we can solve these to find y and z, and then use the first equation to determine x. How do we solve them? Well, "do the same again"! Concretely, I can divide the second equation by the coefficient of y, $\frac{7}{8}$. The system of equations is equivalent to

$$\begin{cases} x - \frac{7}{8}y + \frac{3}{4}z = \frac{59}{8} \\ y + \frac{10}{7}z = \frac{13}{7} \\ \frac{37}{8}y - \frac{13}{4}z = -\frac{89}{8} \end{cases} \qquad (2''')$$

Now I can subtract an appropriate multiple of the second equation from the third, to eliminate y: the system of equations is equivalent to

$$\begin{cases} x & - & \frac{7}{8}y & + & \frac{3}{4}z & = & \frac{59}{8} \\ & & y & + & \frac{10}{7}z & = & \frac{13}{7} \\ & & & - & \frac{69}{7}z & = & -\frac{138}{7} \end{cases} \qquad (2'''')$$

Now the third equation tells us that $z = 2$, so the second tells us that $y = \frac{13}{7} - \frac{10}{7}z = -1$, and finally the first tells us that $x = \frac{59}{8} + \frac{7}{8}y - \frac{3}{4}z = 5$.

Excitingly,

(a) this method generalises nicely; and

(b) this method looks nicer when recorded with matrices!

The strategy is called *Gaussian elimination*, although Gauss was not the first person to use it (see the Historical interlude).

Here are a couple of important subtleties:

1. At two points I divided by a coefficient. Happily, this coefficient was not 0. If it had been 0, I definitely wouldn't have been able to divide by it. But... if the coefficient had been 0 then I'd have had an equation in which I'd effectively already eliminated that variable, which would only have made life easier. Executive summary: this is not a problem, we just have to think about our general strategy in a way that takes account of the possibility.

2. At each stage I carefully wrote "the system of equations is equivalent to". So x, y, z satisfy the original equations *if and only if* $x = 5$, $y = -1$, $z = 2$. This is true because each of my steps was reversible. (Did you check? If not, go back and check now!)

We might summarise the general strategy for a system of m equations in variables $x_1, \ldots, x_n$ as follows.

- Swap equations if necessary to make the coefficient of x_1 in the first equation nonzero.
- Divide through the first equation by the coefficient of x_1.
- Subtract appropriate multiples of the first equation from all other equations to eliminate x_1 from all but the first equation.

- Now the first equation will tell us the value of x_1 once we have determined the values of $x_2, \ldots, x_n$, and we have $m-1$ other equations in $n-1$ variables.
- Use the same strategy to solve these $m-1$ equations in $n-1$ variables.

How can we carry out this process conveniently using matrices?

2.3 Elementary Row Operations (EROs)

Definition. Take a system of linear equations $Ax = b$. To get the *augmented matrix* $A \mid b$, take the $m \times n$ matrix A and adjoin b as an $(n+1)^{\text{th}}$ column.

Remark. The augmented matrix $A \mid b$ records all the data from the system $Ax = b$.

Definition. There are three *elementary row operations* (*EROs*) on an augmented matrix $A \mid b$:

- for some $1 \leq r < s \leq m$, interchange rows r and s;
- for some $1 \leq r \leq m$ and $\lambda \neq 0$, multiply (every entry of) row r by λ;
- for some $1 \leq r, s \leq m$ with $r \neq s$ and $\lambda \in \mathbb{F}$, add λ times row r to row s.

Remark. Notice that each type of elementary row operation does not change the set of solutions of the corresponding system of equations. Each ERO is invertible.

These three (categories of) elementary row operations can be used to carry out Gaussian elimination. By applying them in an appropriate order, we can put any augmented matrix $A \mid b$ into a form $E \mid d$ from which it is easier to determine whether the system has any solutions.

Definition. We say that an $m \times n$ matrix E is in *echelon form* if

(i) if row r of E has any nonzero entries, then the first of these is 1;

(ii) if $1 \leq r < s \leq m$ and rows r, s of E contain nonzero entries, the first of which are e_{rj} and e_{sk} respectively, then $j < k$ (the leading entries of lower rows occur to the right of those in higher rows);

(iii) if row r of E contains nonzero entries and row s does not (that is, $e_{sj} = 0$ for $1 \leq j \leq n$), then $r < s$ (zero rows, if any exist, appear below all nonzero rows).

Remark. We can use EROs not only to reduce the matrix of coefficients A to echelon form, but also to reduce the augmented matrix $A \mid b$ to echelon form, so we do not particularly distinguish between these.

Example.

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 4 & 2 \\ 2 & 3 & 4 & 5 & 0 \end{pmatrix} \xrightarrow[R_1 \leftrightarrow R_2]{} \begin{pmatrix} 1 & 2 & 3 & 4 & 2 \\ 0 & 1 & 2 & 3 & 0 \\ 2 & 3 & 4 & 5 & 0 \end{pmatrix}$$

$$\xrightarrow[R_3 \to R_3 - 2R_1]{} \begin{pmatrix} 1 & 2 & 3 & 4 & 2 \\ & & & & \\ & & & & \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 4 & 2 \\ & & & & \\ & & & & \end{pmatrix}$$

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

This calculation was on the augmented matrix of the equations

$$\begin{aligned} y &+ 2z &+ 3w &= 0 \\ x &+ 2y &+ 3z &+ 4w &= 2 \\ 2x &+ 3y &+ 4z &+ 5w &= 0 \end{aligned} . \tag{4}$$

Each ERO that we applied to the augmented matrix did not change the set of solutions. So the solutions to (4) are precisely the solutions to

$$\begin{aligned} x &+ 2y &+ 3z &+ 4w &= 2 \\ y &+ 2z &+ 3w &= 0 \\ 0 &= 1 \end{aligned} . \tag{5}$$

Clearly there are no solutions to (5), because of the third equation, so there are no solutions to (4)—we say that the system of equations is *inconsistent*.

Definition. Let $E \mid d$ be the $m \times (n+1)$ augmented matrix of a system of equations, where E is in echelon form. We say that variable x_j is *determined* if there is i such that e_{ij} is the leading entry of row i of E (so $e_{ij} = 1$). Otherwise we say that x_j is *free*.

Remark. If the last rows of $E \mid d$ are zero, then they can be deleted (they record that $0 = 0$). So we focus on the remaining (nonzero) rows.

If the final row records that $0 = 1$, then the equations are inconsistent. If not, then the equations have at least one solution. We can choose arbitrary values for the free variables. Now work up from the last equation to the first. If row i has its leading 1 in position (i, j), then the corresponding equation records

$$x_j + \sum_{k=j+1}^{n} e_{ik} x_k = d_i$$

and we already know the values of x_k for $j + 1 \leq k \leq n$, so we can read off the value of x_j.

Example. If a system of linear equations has been reduced to the echelon form

$$\begin{array}{rcrcrcrcl} x_1 & - & x_2 & + & 2x_3 & - & 2x_4 & = & 0 \\ & & x_2 & - & 7x_3 & + & 5x_4 & = & 2 \\ & & & & & & x_4 & = & 0 \end{array}$$

with corresponding matrix

$$\left(\begin{array}{cccc|c} 1 & -1 & 2 & -2 & 0 \\ 0 & 1 & -7 & 5 & 2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

then variables x_1, x_2, x_4 are determined while x_3 is free.

The solutions are

$$\begin{aligned} x_4 &= 0 \\ x_2 &= 2 + 7x_3 - 5x_4 = 2 + 7x_3 \\ x_1 &= x_2 - 2x_3 + 2x_4 = (2 + 7x_3) - 2x_3 = 2 + 5x_3 \end{aligned}$$

.

2.4 Historical interlude 1

It's good to know something about the history of our subject.

You can find an overview of the history of matrices and determinants on the very useful MacTutor website:
`http://www-history.mcs.st-and.ac.uk/HistTopics/Matrices_and_determinants.html`

This describes how the ideas of "Gaussian" elimination go back to China in the first century BCE. This is described in a work called *Jiuzhang suanshu*, or *Nine Chapters on the Mathematical Art*—see for example

http://www-history.mcs.st-and.ac.uk/HistTopics/Nine_chapters.html
and
http://www-history.mcs.st-and.ac.uk/HistTopics/Chinese_overview.html.
Carl Friedrich Gauss
http://www-history.mcs.st-and.ac.uk/Biographies/Gauss.html
wrote about what we now call Gaussian elimination in the context of solving equations as part of his study of the orbit of the asteroid Pallas.

Apparently the first person to use the word 'matrix' in this mathematical context was James Joseph Sylvester
http://www-history.mcs.st-and.ac.uk/Biographies/Sylvester.html
who was for a while the Savilian Professor of Geometry here in Oxford. (The current Savilian Professor of Geometry is Frances Kirwan.) MacTutor has a picture of the page on which Sylvester introduced the word 'matrix'
http://www-history.mcs.st-and.ac.uk/Bookpages/Sylvester9.gif

2.5 Reduced row echelon (RRE) form

Definition. We say that an $m \times n$ matrix is in *reduced row echelon form* (*RRE form*) if it is in echelon form and if each column containing the leading entry of a row has all other entries 0.

Example. We started with matrix

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

earlier, and put it into echelon form

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

That leaves some nonzero entries in columns where we don't want them. We can clear them by using EROs—specifically by subtracting suitable multiples of rows from other rows.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 2 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[R_1\to]{} \begin{pmatrix} 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[R_1\to]{} \begin{pmatrix} 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Remark. We can always take a matrix in echelon form and put it in RRE form using EROs.

Let E be an $m \times n$ matrix in echelon form. Take a row with leading entry $e_{ij} = 1$. For $1 \leq k \leq i-1$, subtract e_{kj} times row i from row k. These are EROs, and make all the entries above e_{ij} into 0. Doing this for each leading entry produces a matrix in RRE form.

Remark. It is handy to have an augmented matrix in echelon form if we want to see whether the system of equations has any solutions. Putting the matrix into reduced row echelon form makes it particularly convenient to read off the solutions (if there are any).

Question What is the RRE form of an invertible square matrix?

Theorem 6. *An invertible $n \times n$ matrix can be reduced to I_n using EROs.*

Proof. Take $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ with A invertible.

Let E be an RRE form of A. [Secret aim: $E = I_n$. To show this, it's enough to show that every variable is determined, there are no free variables.]

We can obtain E from A by EROs, and EROs do not change the solution set of the system of equations $Ax = 0$. If $Ax = 0$, then $x = I_n x = (A^{-1}A)x = A^{-1}(Ax) = A^{-1}0 = 0$, so the only $n \times 1$ column vector x with $Ax = 0$ is $x = 0$. (Here 0 is the $n \times 1$ column vector of zeros.) So the only solution of $Ex = 0$ is $x = 0$.

We can read off solutions to $Ex = 0$. We could choose arbitrary values for the free variables—but the only solution is $x = 0$, so there are no free variables. So all the variables are determined, so each column must contain the leading entry of a row (which must be 1). Since the leading entry of a row comes to the right of leading entries of rows above, it must be the case that $E = I_n$. □

2.6 Elementary matrices and the use of EROs to compute inverses

Definition. For an ERO on an $m \times n$ matrix, we define the corresponding *elementary matrix* to be the result of applying that ERO to I_m.

Remark.
- The ERO of interchanging rows r and s has elementary matrix

—that is, I_m but with 0 in positions (r, r) and (s, s) and 1 in positions (r, s) and (s, r).

- The ERO of multiplying row r by $\lambda \neq 0$ has elementary matrix

—that is, I_m but with λ in position (r, r).

- The ERO of adding λ times row r to row s has elementary matrix

—that is, I_m but with λ in position (s, r).

- Since each ERO is invertible, so are the corresponding elementary matrices. The inverse of an ERO is an ERO, and the inverse of an elementary matrix is another elementary matrix.

Lemma 7. *Let A be an $m \times n$ matrix, let B be obtained from A by applying an ERO. Then $B = EA$, where E is the elementary matrix for that ERO.*

Proof. Exercise. □

Theorem 8. *Let A be an invertible $n \times n$ matrix. Let X_1, X_2, ..., X_k be a sequence of EROs that take A to I_n. Let B be the matrix obtained from I_n by this same sequence of EROs. Then $B = A^{-1}$.*

Remark. The sequence of EROs X_1, X_2, ..., X_k that take A to I_n exists by Theorem 6.

Proof. Let E_i be the elementary matrix corresponding to ERO X_i.
Then applying X_1, X_2, ..., X_k to A gives matrix $E_k \cdots E_2 E_1 A = I_n$,
and applying X_1, X_2, ..., X_k to I_n gives matrix $E_k \cdots E_2 E_1 = B$.
So $BA = I_n$, so $B = A^{-1}$. □

Remark. In the proof of Theorem 8, we show that $BA = I_n$, and claim that this means that $B = A^{-1}$. But our definition of the inverse would require us also to show that $AB = I_n$. In the context of this proof, we don't need to do any more. Why is this? Well, we know that $E_1, \ldots, E_k$ are invertible, and so $B = E_k \cdots E_2 E_1$ is a product of invertible matrices and hence invertible, say with inverse B^{-1}. Now we know that $BA = I_n$ (we proved this), and we can premultiply both sides by B^{-1} and postmultiply both sides by B to get $AB = I_n$.

Example. Let's find the inverse of

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$$

(note the special form of this matrix—it is an example of a *Vandermonde matrix*).

We use EROs to reduce the matrix to RRE form, and for convenience apply these same EROs to I_4 at the same time.

$$\left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 0 & 1 & 0 & 0 \\ 1 & 4 & 9 & 16 & 0 & 0 & 1 & 0 \\ 1 & 8 & 27 & 64 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow[R_4 \to R_4 - R_1]{R_2 \to R_2 - R_1,\ R_3 \to R_3 - R_1} \left(\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 4 & -\frac{13}{3} & \frac{3}{2} & -\frac{1}{6} \\ 0 & 1 & 0 & 0 & -6 & \frac{19}{2} & -4 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 4 & -7 & \frac{7}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & -1 & \frac{11}{6} & -1 & \frac{1}{6} \end{array}\right)$$

so

$$A^{-1} = \begin{pmatrix} 4 & -\frac{13}{3} & \frac{3}{2} & -\frac{1}{6} \\ -6 & \frac{19}{2} & -4 & \frac{1}{2} \\ 4 & -7 & \frac{7}{2} & -\frac{1}{2} \\ -1 & \frac{11}{6} & -1 & \frac{1}{6} \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 24 & -26 & 9 & -1 \\ -36 & 57 & -24 & 3 \\ 24 & -42 & 21 & -3 \\ -6 & 11 & -6 & 1 \end{pmatrix}$$

(and a quick check would confirm that multiplying this by A gives I_4—useful when you're doing computations by hand and want to know whether you've made an arithmetic error!).

Remark. This is not hugely exciting for a human to carry out, but it is very efficient for a computer. Next term, in Linear Algebra II, you'll meet determinants of matrices (if you haven't already done so), and it is possible to use these to compute inverses. But it is staggeringly inefficient to find inverses this way. Using EROs, we can invert an $n \times n$ matrix using no more than $2n^3$ arithmetical operations, which even for say a 30×30 matrix takes a fraction of a second on a computer. See the problems sheet for a question relating to this. The Numerical Analysis course in Part A goes into more detail about efficient computations with matrices—this is extremely important in modern life, as well as mathematically interesting.

3 Vector spaces

3.1 Familiar vectors

You might have met vectors previously, perhaps in geometry or mechanics. In these contexts, we often work with vectors in two or three dimensions, and a vector has both length and direction. We can add (and subtract) vectors, and we can multiply vectors by scalars (real numbers).

A vector space is a generalisation of this idea. We work more abstractly, concentrating on the properties that we want vectors to have. The theory then applies to familiar vectors in two and three dimensions, but also to other contexts: real vector spaces in higher dimensions, complex vector spaces, spaces of functions, vector spaces over finite fields, ... We need to identify the key properties that a vector space should have—called axioms—and then we can explore theorems that apply to any object satisfying the axioms.

As with other aspects of mathematics, it is important that we use the formal definitions when proving formal statements—we need a precise definition of a vector space, and an object is a vector space if and only if it satisfies that definition, so the only properties we can use when proving results are either in the definition or are properties that we've already proved are consequences of the definition. But, again as with other aspects of mathematics, we also need an informal intuition about what a vector space is, to help us "sniff out" possible results that we might seek to prove, to help us to understand and internalise theorems, and to help us have a broad understanding of what the concept means. So we'll explore plenty of examples, not only the familiar cases of 2D and 3D real vectors.

3.2 What is a vector space?

To define a vector space, we need an underlying *field* $\mathbb{F}$. In a field, we can add, subtract, multiply, and divide by nonzero elements, and arithmetic works as we expect. A field always contains an additive identity 0 and a multiplicative identity 1. For us, generally $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ (other fields are available).

Definition. Let $\mathbb{F}$ be a field. A *vector space* over $\mathbb{F}$ is a non-empty set V together with a map $V \times V \to V$ given by $(v, v') \mapsto v + v'$ (called *addition*) and a map $\mathbb{F} \times V \to V$ given by $(\lambda, v) \mapsto \lambda v$ (called *scalar multiplication*) that satisfy the *vector space axioms*

- $u + v = v + u$ for all u, $v \in V$ (addition is *commutative*);
- $u + (v + w) = (u + v) + w$ for all u, v, $w \in V$ (addition is *associative*);

- there is $0_V \in V$ such that $v + 0_V = v = 0_V + v$ for all $v \in V$ (existence of *additive identity*);
- for all $v \in V$ there exists $w \in V$ such that $v+w = 0_V = w+v$ (existence of *additive inverses*);
- $\lambda(u + v) = \lambda u + \lambda v$ for all u, $v \in V$, $\lambda \in \mathbb{F}$ (*distributivity* of scalar multiplication over vector addition);
- $(\lambda + \mu)v = \lambda v + \mu v$ for all $v \in V$, λ, $\mu \in \mathbb{F}$ (distributivity of scalar multiplication over field addition);
- $(\lambda\mu)v = \lambda(\mu v)$ for all $v \in V$, λ, $\mu \in \mathbb{F}$ (scalar multiplication interacts well with field multiplication);
- $1v = v$ for all $v \in V$ (identity for scalar multiplication).

Example. We write $\mathbb{R}^n$ for the set of n-tuples $(v_1, \ldots, v_n)$ with v_1, $\ldots$, $v_n \in \mathbb{R}$. Then $\mathbb{R}^n$ is a real vector space under componentwise addition and scalar multiplication:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_y)$$
$$\text{and } \lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

These satisfy the vector space axioms.

We think of $\mathbb{R}^2$ as the Cartesian plane, and $\mathbb{R}^3$ as three-dimensional space. We can also consider $n = 1$: $\mathbb{R}^1$ is a real vector space, which we think of as the real line. We tend to write it simply as $\mathbb{R}$.

Example. The field $\mathbb{C}$ is a real vector space, it is essentially the same as $\mathbb{R}^2$.

Example. For m, $n \geq 1$, the set $\mathcal{M}_{m\times n}(\mathbb{R})$ is a real vector space (see Proposition 1).

Definition. Elements of V are called *vectors*.

Elements of $\mathbb{F}$ are called *scalars*.

If V is a vector space over $\mathbb{R}$, then we say that V is a *real vector space*.

If V is a vector space over $\mathbb{C}$, then we say that V is a *complex vector space*.

If V is a vector space over $\mathbb{F}$, then we say that V is an $\mathbb{F}$ vector space.

Remark. Our main focus in this course will be real vector spaces. Most of the theory works over any field, but sometimes it makes a difference (as we'll see, for example for inner products).

Lemma 9. *Let V be a vector space over $\mathbb{F}$. Then there is a unique additive identity element 0_V.*

Proof. Exercise. □

Remark. Where it will not be ambiguous, we often write 0 for 0_V.

Lemma 10. *Let V be a vector space over $\mathbb{F}$. Take $v \in V$. Then there is a unique additive inverse for v. That is, if there are w_1, $w_2 \in V$ with $v + w_1 = 0_V = w_1 + v$ and $v + w_2 = 0_V = w_2 + v$, then $w_1 = w_2$.*

Proof. Exercise. □

Remark. Using the notation of Lemma 10, we write $-v$ for the unique additive inverse of v.

Proposition 11. *Let V be a vector space over a field $\mathbb{F}$. Take $v \in V$, $\lambda \in \mathbb{F}$. Then*

(i) $\lambda 0_V = 0_V$;

(ii) $0v = 0_V$;

(iii) $(-\lambda)v = -(\lambda v) = \lambda(-v)$;

(iv) if $\lambda v = 0_V$ *then* $\lambda = 0$ *or* $v = 0_V$.

Proof. (i) We have

$$\begin{aligned}\lambda 0_V &= \lambda(0_V + 0_V) \text{ (definition of additive identity)}\\ &= \lambda 0_V + \lambda 0_V \text{ (distributivity of scalar } \cdot \text{ over vector } +).\end{aligned}$$

Adding $-(\lambda 0_V)$ to both sides, we have

$$\lambda 0_V + (-(\lambda 0_V)) = (\lambda 0_V + \lambda 0_V) + (-(\lambda 0_V))$$

so $0_V = \lambda 0_V$ (using definition of additive inverse, associativity of addition, definition of additive identity).

(ii) Exercise (hint: in $\mathbb{F}$ we have $0 + 0 = 0$).

(iii) We have

$$\begin{aligned}\lambda v + \lambda(-v) &= \lambda(v + (-v)) \text{ (distributivity of scalar } \cdot \text{ over vector } +)\\ &= \lambda 0_V \text{ (definition of additive inverse)}\\ &= 0_V \text{ (by (ii).}\end{aligned}$$

So $\lambda(-v)$ is the additive inverse of λv (by uniqueness), so $\lambda(-v) = -(\lambda v)$.

Similarly, we see that $\lambda v + (-\lambda)v = 0_V$ and so $(-\lambda)v = -(\lambda v)$.

(iv) Suppose that $\lambda v = 0_V$, and that $\lambda \neq 0$. [Secret aim: $v = 0_V$]

Then λ^{-1} exists in $\mathbb{F}$, and

$$\lambda^{-1}(\lambda v) = \lambda^{-1} 0_V$$

so

$$(\lambda^{-1}\lambda)v = 0_V \text{ (scalar } \cdot \text{ interacts well with field } \cdot\text{, and by (i))}$$

so

$$1v = 0_V$$

so $v = 0_V$ (identity for scalar multiplication).

□

3.3 Historical interlude 2

We have just met a lovely, clean, polished definition of a vector space. This reflects how we think about vector spaces today, but says nothing about how vector spaces were first studied. The development of mathematics is complicated: people have ideas, build on others' ideas, overlook or misunderstand ideas, have ideas that others have already had, and so on. When you stand on the top of a hill, it is easy to look around and survey the landscape, to pick out the most interesting features, and to look down to see what your route up should be next time, but when you stand at the foot of a hill and are surveying the landscape, there is lots to explore and it might not even be apparent that there *is* a hill, let alone that it would be an enlightening hill to climb. Successive generations of mathematicians are like pioneers of the hills, passing on tips and maps to those who follow (or occasionally losing the map altogether), but fashions change, technology improves, and some expeditions that seemed important at the time turn out to be less significant generations later.

This course may be about Linear Algebra, but the development of the subject is anything but linear (and many of the ideas were studied in detail long before the phrase "linear algebra" started to be used). You can read an overview of the development of the ideas behind vector spaces on MacTutor.
`http://www-groups.dcs.st-and.ac.uk/history/HistTopics/Abstract_linear_spaces.htm`

The MacTutor biography of Alexandre-Théophile Vandermonde is interesting, not least because it suggests that there is no evidence that Vandermonde studied the determinants named after him.
`http://www-groups.dcs.st-and.ac.uk/history/Biographies/Vandermonde.html`

3.4 Subspaces

Whenever we have a mathematical object with some structure, we want to consider subsets that also have that same structure.

Definition. Let V be a vector space over $\mathbb{F}$. A *subspace* of V is a non-empty subset of V that is *closed* under addition and scalar multiplication, that is, a subset $U \subseteq V$ such that

(i) $U \neq \emptyset$ (U is non-empty);

(ii) $u_1 + u_2 \in U$ for all $u_1, u_2 \in U$ (U is closed under addition);

(iii) $\lambda u \in U$ for all $u \in U$, $\lambda \in \mathbb{F}$ (U is closed under scalar multiplication).

Definition. Note that the sets $\{0_V\}$ and V are always subspaces of V. The subspace $\{0_V\}$ is sometimes called the *zero subspace* or the *trivial subspace*. Subspaces other than V are called *proper subspaces*.

Theorem 12 (Subspace Test)**.** *Let V be a vector space over $\mathbb{F}$, let U be a subset of V. Then U is a subspace if and only if*

(i) $0_V \in U$*; and*

(ii) $\lambda u_1 + u_2 \in U$ *for all* $u_1, u_2 \in U$ *and* $\lambda \in \mathbb{F}$.

Proof. ($\Rightarrow$) Assume that U is a subspace of V.

- $0_V \in U$: Since U is a subspace, it is non-empty, so there exists $u_0 \in U$.

  Since U is closed under scalar multiplication, $0u = 0_V \in U$ (using Proposition 11).

- $\lambda u_1 + u_2 \in U$ for all $u_1, u_2 \in U$ and all $\lambda \in \mathbb{F}$: Take $u_1, u_2 \in U$, and $\lambda \in \mathbb{F}$. Then $\lambda u_1 \in U$ because U is closed under scalar multiplication, so $\lambda u_1 + u_2 \in U$ because U is closed under addition.

($\Leftarrow$) Assume that $0_V \in U$ and that $\lambda u_1 + u_2 \in U$ for all $u_1, u_2 \in U$ and $\lambda \in \mathbb{F}$.

- U is non-empty: have $0_V \in U$.
- U is closed under addition: for $u_1, u_2 \in U$ have $u_1 + u_2 = 1 \cdot u_1 + u_2 \in U$.
- U is closed under scalar multiplication: for $u \in U$ and $\lambda \in \mathbb{F}$, have $\lambda u = \lambda u + 0_V \in U$.

So U is a subspace of V. □

Notation If U is a subspace of the vector space V, then we write $U \leq V$. (Compare with $U \subseteq V$, which means that U is a subset of V but we do not know whether it is a subspace.)

Proposition 13. *Let V be a vector space over $\mathbb{F}$, and let $U \leq V$. Then*

(i) U is a vector space over $\mathbb{F}$; and

(ii) if $W \leq U$ then $W \leq V$ ("a subspace of a subspace is a subspace").

Proof. (i) We need to check the vector space axioms, but first we need to check that we have legitimate operations.

Since U is closed under addition, the operation $+$ restricted to U gives a map $U \times U \to U$.

Since U is closed under scalar multiplication, that operation restricted to U gives a map $\mathbb{F} \times U \to U$.

Now for the axioms.

Commutativity and associativity of addition are inherited from V.

There is an additive identity (by the Subspace Test).

There are additive inverses: if $u \in U$ then multiplying by $-1 \in \mathbb{F}$ and applying Proposition 11(iii) shows that $-u \in U$.

The other four properties are all inherited from V.

(ii) This is immediate from the definition of a subspace.

□

Definition. Let V be a vector space over $\mathbb{F}$. Take A, $B \subseteq V$ and take $\lambda \in \mathbb{F}$. We define $A + B := \{a + b : a \in A, b \in B\}$ and $\lambda A := \{\lambda a : a \in A\}$.

Proposition 14. *Let V be a vector space. Take U, $W \leq V$. Then $U + W \leq V$ and $U \cap W \leq V$.*

Proof. Exercise. □

3.5 More examples of vector spaces

Example. Consider a system of *homogeneous* linear equations with real coefficients:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ \vdots \qquad\qquad & \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0. \end{aligned} \tag{6}$$

(We say this is homogeneous because all the real numbers on the right are 0.)

Let V be the set of real solutions of (6). Then V is a real vector space. This becomes more apparent if we write the equations in matrix form. We see that (6) corresponds to $Ax = 0$, where $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{R})$, x is an $n \times 1$ column vector of variables, and 0 is shorthand for $0_{n\times 1}$. Then each element of V can be thought of as an $n \times 1$ column vector of real numbers.

To show that V is a vector space, we show that it is a subspace of $\mathbb{R}^n$.

Clearly V is non-empty, because $0 \in V$.

For v_1, $v_2 \in V$, we have $Av_1 = 0$ and $Av_2 = 0$, so $A(v_1+v_2) = Av_1+Av_2 = 0 + 0 = 0$, so $v_1 + v_2 \in V$. So V is closed under addition.

For $v \in V$ and $\lambda \in \mathbb{F}$, we have $A(\lambda v) = \lambda(Av) = \lambda 0 = 0$, so $\lambda v \in V$. So V is closed under scalar multiplication.

So $V \leq \mathbb{R}^n$, so V is a vector space (by Proposition 13).

Example. Let n be a non-negative integer. The set of polynomials $c_n x^n + \cdots + c_1 x + c_0$ with c_0, c_1, …, $c_n \in \mathbb{R}$ (that is, real polynomials with degree $\leq n$) is a real vector space.

Example. Let X be a set. Define $\mathbb{R}^X := \{\text{functions } f \text{ with } f : X \to \mathbb{R}\}$, the set of real-valued functions on X. This is a real vector space with operations of pointwise addition and pointwise multiplication by a real number: for $x \in X$, we define $(f + g)(x) = f(x) + g(x)$ and $(\lambda f)(x) = \lambda f(x)$.

One important example is $\mathbb{R}^{\mathbb{R}}$, the vector space of all real-valued functions of a real variable.

Example. The set of real sequences (a_n) forms a real vector space. We add sequences term by term, and we multiply a sequence by a scalar λ by multiplying each term by λ. This is essentially the vector space $\mathbb{R}^{\mathbb{N}}$.

One interesting subspace of $\mathbb{R}^{\mathbb{N}}$ is the space of convergent sequences. You will define and study convergence of real sequences in the Analysis I course.

Example. In Analysis II, you will learn about continuity and differentiability of functions $\mathbb{R} \to \mathbb{R}$. It turns out that $\{f \in \mathbb{R}^{\mathbb{R}} : f \text{ is continuous}\}$ and $\{f \in \mathbb{R}^{\mathbb{R}} : f \text{ is differentiable}\}$ are important subspaces of $\mathbb{R}^{\mathbb{R}}$.

Example. We can study the solutions of a homogeneous linear second-order differential equation. These are twice-differentiable functions y that satisfy an equation $y'' + a(x)y' + b(x)y = 0$. This equation is *linear* because y and its derivatives occur only to the first power and are not multiplied together. And it is *homogeneous* because of the 0 on the right-hand side. Such equations are important in many applications of mathematics.

The set S of solutions of this homogeneous linear second-order differential equation is a vector space, a subspace of $\mathbb{R}^{\mathbb{R}}$. Indeed, S is clearly non-empty (it contains the 0 function), and if $w = u + \lambda v$ where u, $v \in S$ and $\lambda \in \mathbb{R}$, then

$$\begin{aligned} w'' + a(x)w' + b(x)w &= (u'' + \lambda v'') + a(x)(u' + \lambda v') + b(x)(u + \lambda v) \\ &= (u'' + a(x)u' + b(x)u) + \lambda(v'' + a(x)v' + b(x)v) \\ &= 0 \end{aligned}$$

so $w \in S$. So, by the Subspace Test, $S \leq \mathbb{R}^{\mathbb{R}}$.

This generalises to homogeneous linear differential equations of any order.

3.6 Subspaces of $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$

Example. What are the subspaces of $\mathbb{R}$?

Let $V = \mathbb{R}$, let U be a non-trivial subspace of V. [Thinking geometrically, we might have a theory about what U looks like.]

Then there exists $u_0 \in U$ with $u_0 \neq 0$. Take $x \in \mathbb{R}$. Let $\lambda = \frac{x}{u_0}$. Then $x = \lambda u_0 \in U$, because U is closed under scalar multiplication. So $U = V$.

So $\mathbb{R}$ has no non-zero proper subspaces.

Example. What are the subspaces of $\mathbb{R}^2$?

Let $V = \mathbb{R}^2$, let U be a non-trivial subspace of V. [Thinking geometrically, we might have a theory about what U looks like.]

Then there exists $u_0 \in U$ with $u_0 \neq 0$, say $u_0 = (a, b)$. We have $\text{Span}(u_0) = \{\lambda u_0 : \lambda \in \mathbb{R}\} \subseteq U$ (see the next section for more on the span of a set).

Case 1 $\text{Span}(u_0) = U$.

If $a \neq 0$, then let $m = \frac{b}{a}$. Then $\text{Span}(u_0) = \{(x, y) \in \mathbb{R}^2 : y = mx\}$.

If $a = 0$, then $\text{Span}(u_0) = \{(0, y) : y \in \mathbb{R}\}$.

So if U is the span of a single element, then geometrically it is a line in $\mathbb{R}^2$ through the origin, and every such line in $\mathbb{R}^2$ through the origin corresponds to a subspace.

Case 2 $\text{Span}(u_0) \neq U$.

Then there is some $u_1 = (c, d) \in U \setminus \text{Span}(u_0)$.

Consider the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Applying any sequence of EROs to this matrix gives a matrix whose rows are in U. The matrix must have RRE form $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. So U contains the vectors $(1, 0)$ and $(0, 1)$, and hence $U = \mathbb{R}^2$.

So the only non-zero proper subspaces of $\mathbb{R}^2$ correspond geometrically to lines in $\mathbb{R}^2$ through the origin.

Example. The only non-trivial proper subspaces of $\mathbb{R}^3$ correspond geometrically to lines and planes through the origin. Exercise: prove this!

4 Bases

One key goal of this section is to develop a sensible notion of the 'dimension' of a vector space. In order to do this, we need to develop some theory that is in itself both important and interesting.

4.1 Spanning sets

Lemma 15. *Let V be a vector space over $\mathbb{F}$, take $u_1, u_2, \ldots, u_m \in V$. Define $U := \{\alpha_1 u_1 + \cdots + \alpha_m u_m : \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}$. Then $U \leq V$.*

Proof. Idea: use Subspace Test.

- $0_V \in U$: have $0_V = 0u_1 + \cdots + 0u_m \in U$.
- $\lambda v_1 + v_2 \in U$: take $v_1, v_2 \in U$, say $v_1 = \alpha_1 u_1 + \cdots + \alpha_m u_m$ and $v_2 = \beta_1 u_1 + \cdots + \beta_m u_m$, where $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in \mathbb{F}$. Take $\lambda \in \mathbb{F}$.

  Then $\lambda v_1 + v_2 = (\lambda\alpha_1 + \beta_1)u_1 + \cdots + (\lambda\alpha_m + \beta_m)u_m \in U$.

So, by the Subspace Test, $U \leq V$. □

Definition. Let V be a vector space over $\mathbb{F}$, take $u_1, u_2, \ldots, u_m \in V$. A *linear combination* of $u_1, \ldots, u_m$ is a vector $\alpha_1 u_1 + \cdots + \alpha_m u_m$ for some $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$.

We define the *span* of $u_1, \ldots, u_m$ to be $\mathrm{Span}(u_1, \ldots, u_m) := \{\alpha_1 u_1 + \cdots + \alpha_m u_m : \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}$. "the smallest subspace of V that contains $u_1, \ldots, u_m$"

There are other notations for the span of a set of vectors, for example you might see $\mathrm{Sp}(u_1, \ldots, u_m)$ or $\langle u_1, \ldots, u_m \rangle$.

More generally, we can define the *span* of any set $S \subseteq V$ (even a potentially infinite set S). We define $\mathrm{Span}(S) := \{\alpha_1 s_1 + \cdots + \alpha_m s_m : m \geq 0, s_1, \ldots, s_m \in S, \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}$.

Remark. Note that a linear combination only ever involves finitely many elements of S, even if S is infinite.

Remark. The convention is that $\sum_{i \in \emptyset} \alpha_i u_i$ is 0_V (the 'empty sum'), so $\mathrm{Span}\, \emptyset = \{0_V\}$.

Remark. For any $S \subseteq V$, we have $\mathrm{Span}(S) \leq V$. Lemma 15 shows this in the case that S is finite. Exercise: prove it in the general case.

Definition. Let V be a vector space over $\mathbb{F}$. If $S \subseteq V$ is such that $V = \mathrm{Span}(S)$, then we say that S *spans* V, and that S is a *spanning set* for V.

Example. { , } spans $\mathbb{R}^2$. So does { , }. But { , } does not.

4.2 Linear independence

Definition. Let V be a vector space over $\mathbb{F}$. We say that $v_1, \ldots, v_m \in V$ are *linearly dependent* if there are $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$, not all 0, such that $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$. If $v_1, \ldots, v_m$ are not linearly dependent, then we say that they are *linearly independent*.

We say that $S \subseteq V$ is *linearly independent* if every finite subset of S is linearly independent.

Remark. So $v_1, \ldots, v_m \in V$ are linearly independent if and only if the only linear combination of them that gives 0_V is the trivial combination, that is, if and only if $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$ implies $\alpha_1 = \cdots = \alpha_m = 0$.

Example.

- $\{$, $\} \subseteq \mathbb{R}^2$ is linearly independent.

- $\{$, $\} \subseteq \mathbb{R}^2$ is linearly dependent.

- For V a real vector space, if $S \subseteq V$ and $0_V \in S$, then S is linearly dependent.
- For V a real vector space, v_1, v_1, v_2, …, v_m are linearly dependent.
- For V a real vector space, $\{v\}$ is linearly independent if and only if $v \neq 0_V$.

Lemma 16. *Let v_1, …, v_m be linearly independent in an $\mathbb{F}$-vector space V. Let $v_{m+1} \in V$ be such that $v_{m+1} \notin \mathrm{Span}(v_1, \ldots, v_m)$. Then v_1, v_2, …, v_m, v_{m+1} are linearly independent.*

Proof. Take $\alpha_1, \ldots, \alpha_{m+1} \in \mathbb{F}$ such that $\alpha_1 v_1 + \cdots + \alpha_{m+1} v_{m+1} = 0$.

[Secret aim: $\alpha_1 = \cdots = \alpha_{m+1} = 0$.]

If $\alpha_{m+1} \neq 0$, then we have

$$v_{m+1} = -\frac{1}{\alpha_{m+1}}(\alpha_1 v_1 + \cdots + \alpha_m v_m) \in \mathrm{Span}(v_1, \ldots, v_m),$$

which is a contradiction.

So $\alpha_{m+1} = 0$, so $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0$.

But v_1, …, v_m are linearly independent, so this means that $\alpha_1 = \cdots = \alpha_m = 0$. □

4.3 Bases

Definition. Let V be a vector space. A *basis* of V is a linearly independent spanning set.

If V has a finite basis, then we say that V is a *finite-dimensional* vector space.

Remark. Not every vector space is finite-dimensional. For example, the space of real sequences does not have a finite basis. But in this course we'll generally study finite-dimensional vector spaces. The courses on Functional Analysis in Part B will explore the theory of infinite-dimensional vector spaces. Where possible, we work with general vector spaces, but sometimes we'll need to specialise to the finite-dimensional case.

Example. Let's work in $\mathbb{R}^n$. For $1 \leq i \leq n$, let e_i be the row vector with coordinate 1 in the i^{th} entry and 0 elsewhere.

Then $e_1, \ldots, e_n$ are linearly independent: if $\alpha_1 e_1 + \cdots + \alpha_n e_n = 0$ then by looking at the i^{th} entry we see that $\alpha_i = 0$ for all i.

Also, $e_1, \ldots, e_n$ span $\mathbb{R}^n$, because $(a_1, \ldots, a_n) = a_1 e_1 + \cdots + a_n e_n$.

So $e_1, \ldots, e_n$ is a basis of $\mathbb{R}^n$. We call it the *standard basis* of $\mathbb{R}^n$.

Example. Consider $V = \mathcal{M}_{m\times n}(\mathbb{R})$. For $1 \leq i \leq m$ and $1 \leq j \leq n$, let E_{ij} be the matrix with a 1 in entry (i, j) and 0 elsewhere. Then $\{E_{ij} : 1 \leq i \leq m,\ 1 \leq j \leq n\}$ is a basis for V, called the *standard basis* of $\mathcal{M}_{m\times n}(\mathbb{R})$.

Proposition 17. *Let V be a vector space over $\mathbb{F}$, let $S = \{v_1, \ldots, v_n\} \subseteq V$. 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 that S is a basis of V.

Take $v \in V$.

[Secret aim: there are unique $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$.]

Since S spans V, there are $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$.

Suppose that also we have $\beta_1, \ldots, \beta_n \in \mathbb{F}$ such that $v = \beta_1 v_1 + \cdots + \beta_n v_n$.

[Secret aim: $\alpha_i = \beta_i$ for all i.]

Then $\alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n$,

so $(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n = 0_V$.

But S is linearly independent, so $\alpha_1 - \beta_1 = \cdots = \alpha_n - \beta_n = 0$,

that is, $\alpha_i = \beta_i$ for all i.

So the linear combination is unique.

($\Leftarrow$) Suppose that every vector in V has a unique expression as a linear combination of elements of S.

- S spanning set: for any $v \in V$ we can write v as a linear combination of elements of S. So $\text{Span}(S) = V$.
- S linearly independent: for $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, if $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 = 0v_1 + \cdots + 0v_n$, then by uniqueness we have $\alpha_i = 0$ for all i.

So S is a basis for V. □

Remark. Proposition 17 gives a very helpful way to understand the idea of a basis!

Remark. Proposition 17 tells us that S is a basis of V if and only if the map $\mathbb{F}^n \to V$ given by $(\alpha_1, \ldots, \alpha_n) \mapsto \alpha_1 v_1 + \cdots + \alpha_n v_n$ is bijective. We can think of $(\alpha_1, \ldots, \alpha_n)$ as the *coordinates* of $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ with respect to the

basis $v_1, \ldots, v_n$. Here the order of the basis vectors $v_1, \ldots, v_n$ is important, so we omit the curly braces.

Where possible, it is generally nicer to avoid picking a basis—often it is more convenient to avoid using coordinates. But sometimes coordinates are good.

Example. Consider $V = \mathcal{M}_{2\times 2}(\mathbb{R})$. The standard basis for this space is $\mathcal{E} = \{E_{11}, E_{12}, E_{21}, E_{22}\}$, where

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Here we also consider a different basis $\mathcal{F} = \{B_1, B_2, B_3, B_4\}$ where

$$B_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ B_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ B_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ B_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let's consider $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$. With respect to the standard basis, A has coordinate vector $\begin{pmatrix} 2 \\ 1 \\ -1 \\ 0 \end{pmatrix}$. With respect to the basis $\mathcal{F}$, A has coordinate vector $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$.

Question Does a vector space always have a basis?

Proposition 18. *Let V be a vector space over $\mathbb{F}$. Suppose that V has a finite spanning set S. Then S contains a linearly independent spanning set.*

Remark. That is, if V has a finite spanning set, then V has a basis. We say nothing here about what happens if V does not have a finite spanning set. This question is addressed in the Part B course on Set Theory (using the Axiom of Choice).

Proof. Let S be a finite spanning set for V.

Take $T \subseteq S$ such that T is linearly independent, and T is a largest such set (any linearly independent subset of S has size $\leq |T|$).

[Secret aim: T is linearly independent and spanning.]

Suppose, for a contradiction, that $\mathrm{Span}(T) \neq V$.

Then, since $\mathrm{Span}(S) = V$, there must exist $v \in S \setminus \mathrm{Span}(T)$.

Now by Lemma 16 we see that $T \cup \{v\}$ is linearly independent, and $T \cup \{v\} \subseteq S$, and $|T \cup \{v\}| > |T|$, which contradicts our choice of T.

So T spans V, and by our choice is linearly independent. □

4.4 Dimension

Theorem 19 (Steinitz Exchange Lemma)**.** *Let V be a vector space over $\mathbb{F}$. Take $X \subseteq V$. Suppose that $u \in \mathrm{Span}(X)$ but that $u \notin \mathrm{Span}(X \setminus \{v\})$ for some $v \in X$. Let $Y = (X \setminus \{v\}) \cup \{u\}$ ("exchange u for v"). Then $\mathrm{Span}(Y) = \mathrm{Span}(X)$.*

Proof. Since $u \in \mathrm{Span}(X)$, there are $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ and $v_1, \ldots, v_n \in X$ such that $u = \alpha_1 v_1 + \cdots + \alpha_n v_n$.

There is $v \in X$ such that $u \notin \mathrm{Span}(X \setminus \{v\})$. Without loss of generality, we may assume that $v = v_n$. Since $u \notin \mathrm{Span}(X \setminus \{v_n\})$, we see that $\alpha_n \neq 0$.

So we can divide by α_n and rearrange, to obtain

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

Now if $w \in \mathrm{Span}(Y)$ then we have an expression of w as a linear combination of elements of Y. We can replace u by $\alpha_1 v_1 + \cdots + \alpha_n v_n$ to express w as a linear combination of elements of X. So $\mathrm{Span}(Y) \subseteq \mathrm{Span}(X)$.

And if $w \in \mathrm{Span}(X)$ then we have an expression of w as a linear combination of elements of X. We can replace v_n by $\frac{1}{\alpha_n}(u - \alpha_1 v_1 - \cdots - \alpha_{n-1} v_{n-1})$ to express w as a linear combination of elements of Y. So $\mathrm{Span}(Y) \supseteq \mathrm{Span}(X)$. □

Remark. The Steinitz Exchange Lemma is called a lemma, which sounds unimportant, and it looks a bit like a niche technical result. But in fact it is completely fundamental to defining the dimension of a vector space. For the purposes of Prelims, you can safely ignore the rest of this remark, but perhaps some of you will be interested to peek ahead. If you choose the Rings and Modules course in Part A, you will (unsurprisingly) learn about modules. A module is a bit like a vector space, but the scalars come from a ring. In a ring, we can add, subtract and multiply, but there is no requirement about being able to divide. For example, the integers form a ring (and every field is also a ring). It turns out that the Steinitz Exchange Lemma doesn't in general work for modules (can you spot the crucial moment in the proof where it mattered that the scalars came from a field?), and that has all sorts of interesting consequences for modules. Anyway, back to Prelims work.

Theorem 20. *Let V be a vector space. Let S, T be finite subsets of V. If S is linearly independent and T spans V, then $|S| \leq |T|$. "linearly independent sets are at most as big as spanning sets"*

Proof. Assume that S is linearly independent and that T spans V.

List the elements of S as $u_1, \ldots, u_m$ and the elements of T as $v_1, \ldots, v_n$.

[Idea: use the Steinitz Exchange Lemma to swap out elements of T and replace them by elements of S, while still having a spanning set of V. We cannot run out of elements of T before we run out of elements of S, or a remaining element of S would have to be a linear combination of the ones swapped into T.]

Let T_0 be the list $v_1, \ldots, v_n$.

Since $\mathrm{Span}(T_0) = V$, there is some i such that $u_1 \in \mathrm{Span}(v_1, \ldots, v_i)$. Choose the least such i.

Then $u_1 \in \mathrm{Span}(v_1, \ldots, v_i)$ but $u_1 \notin \mathrm{Span}(v_1, \ldots, v_{i-1})$.

Let T_1 be the list $u_1, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$.

Then the Steinitz Exchange Lemma shows that $\mathrm{Span}(T_1) = \mathrm{Span}(T_0) = \mathrm{Span}(T)$.

We continue inductively.

Suppose that for some j with $1 \leq j < m$ we have a list T_j where the first j elements are $u_j, \ldots, u_1$, the remaining elements are in T, and $\mathrm{Span}(T_j) = \mathrm{Span}(T)$.

Now $u_{j+1} \in \mathrm{Span}(T)$, so $u_{j+1} \in \mathrm{Span}(T_j)$, so there is some v in the list T_j such that u_{j+1} is in the span of v and its predecessors, but u_{j+1} is not in the span of the predecessors of v.

Note that v cannot be any of $u_1, \ldots, u_j$, because $u_{j+1} \notin \mathrm{Span}(u_1, \ldots, u_j)$ (since S is linearly independent), so $v \in T$.

Let T_{j+1} be the list obtained from T_j by removing v, and adding u_{j+1} at the start.

Then the Steinitz Exchange Lemma shows that $\mathrm{Span}(T_{j+1}) = \mathrm{Span}(T_j) = \mathrm{Span}(T)$.

After j steps, we have replaced j members of T by j members of S.

We cannot run out of members of T before we run out of members of S: otherwise a remaining element of S would be a linear combination of the elements of S already swapped into the list T_n. So we must have $m \leq n$. □

Corollary 21. *Let V be a finite-dimensional vector space. Let S, T be bases of V. Then S and T are finite, and $|S| = |T|$.*

Proof. Since V is finite-dimensional, it has a finite basis B. Say $|B| = n$.

Now B is a spanning set and $|B| = n$, so by Theorem 20 any finite linearly independent subset of V has size at most n.

Since S is a basis of V, it is linearly independent, so every finite subset of S is linearly independent.

So in fact S must be finite, and $|S| \leq n$. Similarly, T is finite and $|T| \leq n$.

Now S is linearly independent and T is spanning, so by Theorem 20 $|S| \leq |T|$.

Applying Theorem 20 with the roles of S and T reversed shows that $|S| \geq |T|$.

So $|S| = |T|$. □

Remark. Corollary 21 is crucial for allowing us to define the notion of dimension. It all relies on the Steinitz Exchange Lemma!

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

Example. The vector space $\mathbb{R}^n$ has standard basis $e_1, \ldots, e_n$, and hence has dimension n.

Example. The vector space $\mathcal{M}_{m\times n}(\mathbb{R})$ has dimension mn (see the standard basis from earlier).

4.5 Row rank

Here is an important example of the dimension of a vector space.

Definition. Let A be an $m \times n$ matrix over $\mathbb{F}$. We define the *row space* of A to be the span of the subset of $\mathbb{F}^n$ consisting of the rows of A, and we denote it by $\mathrm{rowsp}(A)$. We define the *row rank* of A to be $\mathrm{rowrank}(A) := \dim \mathrm{rowsp}(A)$.

Remark. We'll revisit this in detail later in the course.

Remark. We define the column space and column rank of a matrix analogously.

Lemma 22. *Let A be an $m \times n$ matrix, and let B be a matrix obtained from A by a finite sequence of EROs. Then* $\mathrm{rowsp}(A) = \mathrm{rowsp}(B)$. *In particular,* $\mathrm{rowrank}(A) = \mathrm{rowrank}(B)$.

Proof. Exercise. Hint: check that each of the three types of ERO does not change the row space. □

4.6 Historical interlude 3

MacTutor has a biography of Ernst Steinitz, after whom the Steinitz Exchange Lemma is named.

`http://www-groups.dcs.st-and.ac.uk/history/Biographies/Steinitz.html`

I mentioned the Axiom of Choice in the context of proving that every vector space has a basis. You can read more about this, and about the history of set theory, on MacTutor.

http://www-groups.dcs.st-and.ac.uk/history/HistTopics/Beginnings_of_set_theory.h

I said that in this course we'll mostly concentrate on vector spaces over $\mathbb{R}$ and $\mathbb{C}$, but that other fields are available. For example, there are finite fields. You'll start to meet some of these in the Prelims Groups and Group Actions course, and can study them further in Rings and Modules in Part A. Vector spaces over finite fields are fertile sources for interesting mathematics. For example, they are important in information theory and coding theory (which you can choose to study in Part B). The subject was pioneered by Claude Shannon.
http://www-groups.dcs.st-and.ac.uk/history/Biographies/Shannon.html

5 Bases and subspaces

5.1 Bases of subspaces

Proposition 23. *Let U be a subspace of a finite-dimensional vector space V. Then*

(a) U is finite-dimensional, and $\dim U \leq \dim V$; and

(b) if $\dim U = \dim V$, then $U = V$.

Proof. Let $n = \dim V$.

(a) By Theorem 20, every linearly independent subset of V has size at most n.

Let S be a largest linearly independent set contained in U, so $|S| \leq n$.

[Secret aim: S spans U.]

Suppose, for a contradiction, that $\mathrm{Span}(S) \neq U$.

Then there exists $u \in U \setminus \mathrm{Span}(S)$.

Now by Lemma 16 $S \cup \{u\}$ is linearly independent, and $|S \cup \{u\}| > |S|$, which contradicts our choice of S.

So $U = \mathrm{Span}(S)$ and S is linearly independent, so S is a basis of U, and as we noted earlier $|S| \leq n$.

(b) If $\dim U = \dim V$, then there is a basis S of U with $\dim U$ elements. Then S is a linearly independent subset of V with size $\dim V$. Now adding any vector to S must give a linearly dependent set as every linearly independent subset of V has size at most n, so S must span V. So $V = \mathrm{Span}(S) = U$.

□

Remark. In (b) here we used the very useful fact that in an n-dimensional vector space, any linearly independent set of size n is a basis. Similarly, any spanning set of size n is a basis.

Theorem 24. *Let U be a subspace of a finite-dimensional vector space V. Then every basis of U can be extended to a basis of V. That is, if $u_1, \ldots, u_m$ is a basis of U, then there are $v_{m+1}, \ldots, v_n \in V$ such that $u_1, \ldots, u_m, v_{m+1}, \ldots, v_n$ is a basis of V.*

Health warning. Theorem 24 does *not* say that if $U \leq V$ and if we have a basis of V then there is a subset that is a basis of U. The reason it does not say this is that in general this is false.

Let $V = \mathbb{R}^2$. Pick a basis of V. Is there a subset of your basis that is a basis for $U_1 = \mathrm{Span}((1,0))$? Is there a subset of your basis that is a basis for $U_2 = \mathrm{Span}((-3,142))$? Now pick a basis of U_1 or U_2. Can you extend it to a basis of V?

It is *very* easy to assume that if you have a basis of V then there is a subset that is a basis of U. But it is not true. Please try not to do this!

Proof. Let $u_1, \ldots, u_m$ be a basis of U, let $v_1, \ldots, v_n$ be a basis of V.

Idea: start with $u_1, \ldots, u_m$, and add vectors v_i till we reach a basis of V.

Let $S_0 = \{u_1, \ldots, u_m\}$.

Then S_0 is certainly linearly independent.

For $1 \leq i \leq n$, we define a new set S_i in such a way that it is still linearly independent. We define

$$S_i = \begin{cases} S_{i-1} & \text{if } v_i \in \mathrm{Span}(S_{i-1}) \\ S_{i-1} \cup \{v_i\} & \text{if } v_i \notin \mathrm{Span}(S_{i-1}). \end{cases}$$

If S_{i-1} is linearly independent, then so is S_i (immediately in one case, and using Lemma 16 in the other case). So, by induction, S_n is linearly independent.

Also, we always have $v_i \in \mathrm{Span}(S_i)$, and $\mathrm{Span}(S_i) \subseteq \mathrm{Span}(S_n)$, so $v_i \in \mathrm{Span}(S_n)$ for $1 \leq i \leq n$. So $\mathrm{Span}(S_n) = V$.

So S_n is a basis of V, and by construction $u_1, \ldots, u_m \in S_n$. □

Question Let S be a finite set of vectors in $\mathbb{R}^n$. How can we (efficiently) find a basis of $\mathrm{Span}(S)$?

Example. Let $S = \{(0,1,2,3),(1,2,3,4),(2,3,4,5)\} \subseteq \mathbb{R}^4$. Let's find a basis for $\mathrm{Span}(S)$.

Let A be the matrix with rows from S: define $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \end{pmatrix}$. So $\mathrm{Span}(S) = \mathrm{rowsp}(A)$.

Applying EROs to A does not change the row space (Lemma 22).

Reduce A to echelon form. Conveniently, we did this in a previous example, and found echelon form

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

Now $\mathrm{Span}(S) = \mathrm{rowsp}(E) = \mathrm{Span}\{(1,2,3,4),(0,1,2,3)\}$.

And $\{(1,2,3,4),(0,1,2,3)\}$ is clearly linearly independent: if we take λ, $\mu \in \mathbb{R}$ such that $\lambda(1,2,3,4) + \mu(0,1,2,3) = (0,0,0,0)$ then $\lambda = 0$ (look at first coordinate) and so $\mu = 0$ (look at second coordinate).

So $\{(1,2,3,4),(0,1,2,3)\}$ is a basis for $\mathrm{Span}(S)$.

General strategy Let $m = |S|$. Write the m elements of S as the rows of an $m \times n$ matrix A.

Use EROs to reduce A to matrix E in echelon form. Then $\mathrm{rowsp}(E) = \mathrm{rowsp}(A) = \mathrm{Span}(S)$, by Lemma 22.

The nonzero rows of E are certainly linearly independent. So the nonzero rows of E give a basis for $\mathrm{Span}(S)$.

5.2 Sums and intersections of subspaces, and the dimension formula

We previously claimed that the sum and intersection of two subspaces is a subspace, but the proof was an exercise. This result is now central to what we are doing, so this is a good time to prove it carefully. This is a recap of Proposition 14.

Proposition. *Let V be a vector space over $\mathbb{F}$. Take U, $W \leq V$. Then $U + W \leq V$ and $U \cap W \leq V$.*

Proof. Idea: use Subspace Test.
$\underline{U + W}$

- 0_V: We have $0_V \in U$ and $0_V \in W$ (since U, $W \leq V$),
  so $0_V = 0_V + 0_V \in U + W$.

- $\lambda v_1 + v_2$: Take $v_1, v_2 \in U + W$ and $\lambda \in \mathbb{F}$.

  Then $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$ for some $u_1, u_2 \in U$ and $w_1, w_2 \in W$.

  Now $\lambda u_1 + u_2 \in U$ and $\lambda w_1 + w_2 \in W$, because $U, W \leq V$, so $\lambda v_1 + v_2 = (\lambda u_1 + u_2) + (\lambda w_1 + w_2) \in U + W$.

So, by the Subspace Test, $U + W \leq V$.
$\underline{U \cap W}$

- 0_V: We have $0_V \in U$, $0_V \in W$ (since $U, W \leq V$)

  so $0_V \in U \cap W$.

- $\lambda v_1 + v_2$: Take $v_1, v_2 \in U \cap W$ and $\lambda \in \mathbb{F}$.

  Then $v_1, v_2 \in U$ so $\lambda v_1 + v_2 \in U$,

  and $v_1, v_2 \in W$ so $\lambda v_1 + v_2 \in W$,

  so $\lambda v_1 + v_2 \in U \cap W$.

So, by the Subspace Test, $U \cap W \leq V$. □

The next result is particularly useful.

Theorem 25 (Dimension Formula)**.** *Let U, W be subspaces of a finite-dimensional vector space V over $\mathbb{F}$. Then* $\dim(U + W) + \dim(U \cap W) = \dim U + \dim W$.

Proof. Take a basis $v_1, \ldots, v_m$ of $U \cap W$.

Now $U \cap W \leq U$ and $U \cap W \leq W$, so by Theorem 24 we can extend this basis to a basis $v_1, \ldots, v_m, u_1, \ldots, u_p$ of U, and a basis $v_1, \ldots, v_m, w_1, \ldots, w_q$ of W.

With this notation, we see that $\dim(U \cap W) = m$, $\dim U = m + p$ and $\dim W = m + q$.

Claim. $v_1, \ldots, v_m, u_1, \ldots, u_p, w_1, \ldots, w_q$ is a basis of $U + W$.

Proof of claim. Call this collection of vectors S.

Note that all these vectors really are in $U+W$ (for example, $u_1 = u_1 + 0_V \in U + W$).

spanning: Take $x \in U + W$. Then $x = u + w$ for some $u \in U$, $w \in W$.

Since $v_1, \ldots, v_m, u_1, \ldots, u_p$ span U, there are $\alpha_1, \ldots, \alpha_m, \alpha'_1, \ldots, \alpha'_p \in \mathbb{F}$ such that $u = \alpha_1 v_1 + \cdots + \alpha_m v_m + \alpha'_1 u_1 + \cdots + \alpha'_p u_p$.

Similarly, there are $\beta_1, \ldots, \beta_m, \beta'_1, \ldots, \beta'_q \in \mathbb{F}$ such that $w = \beta_1 v_1 + \cdots + \beta_m v_m + \beta'_1 w_1 + \cdots + \beta'_q w_q$.

Then $x = u + w = (\alpha_1 + \beta_1)v_1 + \cdots + (\alpha_m + \beta_m)v_m + \alpha'_1 u_1 + \cdots + \alpha'_p u_p + \beta'_1 w_1 + \cdots + \beta'_q w_q \in \mathrm{Span}(S)$.

And certainly $\mathrm{Span}(S) \subseteq U + W$.

So $\mathrm{Span}(S) = U + W$.

lin indep: Take $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q \in \mathbb{F}$ such that

$$\alpha_1 v_1 + \cdots + \alpha_m v_m + \beta_1 u_1 + \cdots + \beta_p u_p + \gamma_1 w_1 + \cdots + \gamma_q w_q = 0. \quad (7)$$

Then $\alpha_1 v_1 + \cdots + \alpha_m v_m + \beta_1 u_1 + \cdots + \beta_p u_p = -(\gamma_1 w_1 + \cdots + \gamma_q w_q)$.

The vector on the left-hand side is in U, and the vector on the right-hand side is in W. So they are both in $U \cap W$.

So there are $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ such that $-(\gamma_1 w_1 + \cdots + \gamma_q w_q) = \lambda_1 v_1 + \cdots + \lambda_m v_m$,

that is, $\gamma_1 w_1 + \cdots + \gamma_q w_q + \lambda_1 v_1 + \cdots + \lambda_m v_m = 0$.

But $\{v_1, \ldots, v_m, w_1, \ldots, w_q\}$ is linearly independent, so $\gamma_1 = \cdots = \gamma_q = 0$.

Returning to equation 7, this means that

$$\alpha_1 v_1 + \cdots + \alpha_m v_m + \beta_1 u_1 + \cdots + \beta_p u_p = 0.$$

But $\{v_1, \ldots, v_m, u_1, \ldots, u_p\}$ is linearly independent, so $\alpha_1 = \cdots = \alpha_m = \beta_1 = \cdots = \beta_p = 0$.

So S is linearly independent.

This proves the claim.

So S is a basis of $U + W$, so

$$\begin{aligned} \dim(U + W) &= m + p + q = (m + p) + (m + q) - m \\ &= \dim U + \dim W - \dim(U \cap W). \end{aligned}$$

□

Example. Let V be a vector space of dimension 10. Let X, Y be subspaces of dimension 6. Then $X + Y \leq V$ so $\dim(X + Y) \leq \dim V = 10$ (using Proposition 23). So, by the dimension formula,

$$\dim(X \cap Y) = \dim(X) + \dim(Y) - \dim(X + Y) \geq 6 + 6 - 10 = 2.$$

5.3 Direct sums of subspaces

Definition. Let U, W be subspaces of a vector space V. If $U \cap W = \{0_V\}$ and $U + W = V$, then we say that V is the *direct sum* of U and W, and we write $V = U \oplus W$.

In this case, we say that W is a *direct complement* of U in V (and vice versa).

Proposition 26. *Let U, W be subspaces of a finite-dimensional vector space V. The following are equivalent:*

(i) $V = U \oplus W$;

(ii) *every $v \in V$ has a unique expression as $u+w$ where $u \in U$ and $w \in W$;*

(iii) $\dim V = \dim U + \dim W$ *and* $V = U + W$;

(iv) $\dim V = \dim U + \dim W$ *and* $U \cap W = \{0_V\}$;

(v) *if $u_1, \ldots, u_m$ is a basis for U and $w_1, \ldots, w_n$ is a basis for W, then $u_1, \ldots, u_m, w_1, \ldots, w_n$ is a basis for V.*

Proof. Exercise.

Hint: (i) $\Leftrightarrow$ (ii) follows from the definition of direct sum.

Try using the dimension formula to prove that (i)/(ii) are equivalent to (iii), (iv), (v). □

Health warning. It is *not* the case that if $V = U \oplus W$ then every basis of V is the union of a basis of U and a basis of W. Can you find a counterexample to show this?

5.4 Historical interlude 4

This course is about linear algebra. The word 'algebra' comes from the title of the work *Hisab al-jabr w'al-muqabala* by Abu Ja'far Muhammad ibn Musa Al-Khwarizmi. You can read about his life and work on MacTutor. Our word 'algorithm' comes from the title of a Latin translation of a work by Al-Khwarizmi on Hindu-Arabic numerals (the word derives from Al-Khwarizmi's name).

`http://www-groups.dcs.st-and.ac.uk/history/Biographies/Al-Khwarizmi.html`

We have spent time in this course discussing equations. This would all look very different had Robert Recorde not invented the equals sign =, which he did in a book called *The Whetstone of Witte*, in 1557. As MacTutor describes, he wrote that he had chosen the symbol, using two parallel line

segments, "bicause noe 2 thynges can be moare equalle". You can read more about Recorde (who studied at Oxford) on MacTutor, where you can also find an image of the famous page from *The Whetstone of Witte*.
`http://www-groups.dcs.st-and.ac.uk/history/Biographies/Recorde.html`
`http://www-groups.dcs.st-and.ac.uk/history/Bookpages/Recorde4.jpeg`

6 Linear transformations

We have objects with some structure (vector spaces). This section is about structure-preserving maps between these objects. You will see a similar phenomenon in lots of other contexts too—whenever we have objects with some kind of structure, we can ask about structure-preserving maps between objects. (This can lead to further abstraction, which is explored in Category Theory, an interesting part of mathematics and currently a Part C course.)

6.1 What is a linear transformation?

Definition. Let V, W be vector spaces over $\mathbb{F}$. We say that a map $T : V \to W$ is *linear* if

(i) $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$ (preserves additive stucture); and

(ii) $T(\lambda v) = \lambda T(v)$ for all $v \in V$ and $\lambda \in \mathbb{F}$ (respects scalar multiplication).

We call T a *linear transformation* or a *linear map*.

Proposition 27. *Let V, W be vector spaces over $\mathbb{F}$, let $T : V \to W$ be linear. Then $T(0_V) = 0_W$.*

Remark. The additive identity is an important part of the vector space structure, and Proposition 27 confirms that a linear map respects the additive identity too. In particular, if $T : V \to W$ has $T(0_V) \neq 0_W$, then T is certainly not linear. This can be very useful in practice!

Remark. A closer examination of Proposition 27 shows that either one of the two conditions for linearity is enough to guarantee that $T(0_V) = 0_W$. That is, if T is any map that preserves additive structure then $T(0_V) = 0_W$, and if T is any map that respects scalar multiplication then $T(0_V) = 0_W$.

Proof. Let $z = T(0_V) \in W$.

Then $z + z = T(0_V) + T(0_V) = T(0_V + 0_V) = T(0_V) = z$ (using the assumption to see that $T(0_V) + T(0_V) = T(0_V + 0_V)$),

so $z = 0_W$. □

Proposition 28. *Let V, W be vector spaces over $\mathbb{F}$, let $T : V \to W$. The following are equivalent:*

(i) *T is linear;*

(ii) *$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$ for all $v_1, v_2 \in V$ and $\alpha, \beta \in \mathbb{F}$;*

(iii) *for any $n \geq 1$, if $v_1, \ldots, v_n \in V$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ then $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$.*

Proof. Exercise. □

Remark. In practice, I most often think about linear maps using (ii) from Proposition 28.

Example.

- Let V be a vector space. Then the *identity map* $\mathrm{id}_V : V \to V$ given by $\mathrm{id}_V(v) = v$ for all $v \in V$ is a linear map.
- Let V, W be vector spaces. The *zero map* $0 : V \to W$ that sends every $v \in V$ to 0_W is a linear map. (In particular, there is at least one linear map between any pair of vector spaces.)
- For m, $n \geq 1$, let $V = \mathbb{R}^n_{\mathrm{col}} = \mathcal{M}_{n\times 1}(\mathbb{R})$ and $W = \mathbb{R}^m_{\mathrm{col}} = \mathcal{M}_{m\times 1}(\mathbb{R})$. Take $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Define the left multiplication map $L_A : V \to W$ by $L_A(v) = Av$ for $v \in V$. This is a linear map. (We proved this in Proposition 2.)

  Similarly, we have a right multiplication map $R_A : \mathbb{R}^m \to \mathbb{R}^n$ sending v to vA (row vectors).
- Take m, n, $p \geq 1$. Let $V = \mathcal{M}_{n\times p}(\mathbb{R})$, let $W = \mathcal{M}_{m\times p}(\mathbb{R})$. Take $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The left multiplication map $V \to W$ sending X to AX is a linear map.
- Let V be a vector space over $\mathbb{F}$ with subspaces U, W such that $V = U \oplus W$. For $v \in V$ there are unique $u \in U$, $w \in W$ such that $v = u + w$ (see Proposition 26). Define $P : V \to V$ by $P(v) = w$.

  Claim. P is a linear map.

  Proof. Take v_1, $v_2 \in V$ and α_1, $\alpha_2 \in \mathbb{F}$.

  [Secret aim: $P(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 P(v_1) + \alpha_2 P(v_2)$.]

  Then there are u_1, $u_2 \in U$, w_1, $w_2 \in W$ such that $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$.

Now

$$\begin{aligned}\alpha_1 v_1 + \alpha_2 v_2 &= \alpha_1(u_1 + w_1) + \alpha_2(u_2 + w_2) \\ &= (\alpha_1 u_1 + \alpha_2 u_2) + (\alpha_1 w_1 + \alpha_2 w_2)\end{aligned}$$

where the first bracket is in U and the second is in W, so $P(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 w_1 + \alpha_2 w_2 = \alpha_1 P(v_1) + \alpha_2 P(v_2)$. □

The linear map P is called the *projection of V onto W along U*.

- For $A = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{R})$, we define the *trace* of A to be $\mathrm{tr}(A) := a_{11} + a_{22} + \cdots + a_{nn}$ (the sum of the entries on the main diagonal of A). The map $\mathrm{tr} : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$ is a linear map.

- Let $\mathbb{R}_n[x]$ be the vector space of polynomials of degree at most n. Define $D : \mathbb{R}_n[x] \to \mathbb{R}_n[x]$ by $p(x) \mapsto p'(x)$, that is, $D(a_n x^n + \cdots + a_1 x + a_0) = n a_n x^{n-1} + \cdots + a_1$. This is a linear map from $\mathbb{R}_n[x]$ to $\mathbb{R}_n[x]$. We could also think of it as a linear map $\mathbb{R}_n[x]$ to $\mathbb{R}_{n-1}[x]$.

- Let $C^1(\mathbb{R})$ be the subspace of $\mathbb{R}^{\mathbb{R}}$ consisting of differentiable functions $f : \mathbb{R} \to \mathbb{R}$. The differential operator $D : C^1(\mathbb{R}) \to \mathbb{R}^{\mathbb{R}}$ sending f to f' is a linear map.

- Let $C^\infty(\mathbb{R})$ be the subspace of $\mathbb{R}^{\mathbb{R}}$ consisting of differentiable functions $f : \mathbb{R} \to \mathbb{R}$ that are infinitely differentiable. The differential operator $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ sending f to f' is a linear map.

- Let X be a set, let $V = \mathbb{R}^X$. For $a \in X$, the *evaluation map* $E_a : V \to \mathbb{R}$ sending f to $f(a)$ is a linear map.

6.2 Useful ways to combine linear transformations

We can add linear transformations (pointwise), and we can multiply a linear transformation by a scalar (pointwise).
Question Do linear transformations themselves form a vector space?

Proposition 29. *Let V, W be vector spaces over $\mathbb{F}$. For S, $T : V \to W$ and $\lambda \in \mathbb{F}$, define $S + T : V \to W$ by $(S + T)(v) = S(v) + T(v)$ for $v \in V$, and define $\lambda S : V \to W$ by $(\lambda S)(v) = \lambda S(v)$ for $v \in V$. With these operations (and the zero map $0 : V \to W$ we saw in an earlier example), the set of linear transformations $V \to W$ forms a vector space.*

Proof. Exercise. □

We can also compose linear transformations.

Proposition 30. *Let U, V, W be vector spaces over $\mathbb{F}$. Let $S : U \to V$ and $T : V \to W$ be linear. Then $T \circ S : U \to W$ is linear.*

Proof. Take u_1, $u_2 \in U$ and λ_1, $\lambda_2 \in \mathbb{F}$. Then

$$\begin{aligned}(T \circ S)(\lambda_1 u_1 + \lambda_2 u_2) &= T(S(\lambda_1 u_1 + \lambda_2 u_2)) \text{ (definition of composition)} \\ &= T(\lambda_1 S(u_1) + \lambda_2 S(u_2)) \text{ (S is linear)} \\ &= \lambda_1 T(S(u_1)) + \lambda_2 T(S(u_2)) \text{ (T is linear)} \\ &= \lambda_1 (T \circ S)(u_1) + \lambda_2 (T \circ S)(u_2) \text{ (defn of composition)}\end{aligned}$$

so $T \circ S$ is linear by Proposition 28. □

Remark. We often write $T \circ S$ as TS. The notation $T \circ S$ removes any possible ambiguity about the order of the functions.

Definition. Let V, W be vector spaces, let $T : V \to W$ be linear. We say that T is *invertible* if there is a linear transformation $S : W \to V$ such that $ST = \mathrm{id}_V$ and $TS = \mathrm{id}_W$ (where id_V and id_W are the identity maps on V and W respectively). In this case, we call S the *inverse* of T, and write it as T^{-1}.

Remark. T is a function, so if it is invertible then it has a unique inverse (you saw this in the Introduction to University Maths course), so there is no ambiguity in writing T^{-1}.

Proposition 31. *Let V, W be vector spaces. Let $T : V \to W$ be linear. Then T is invertible if and only if T is bijective.*

Proof. ($\Rightarrow$) If T is invertible, then it is certainly bijective (see the Introduction to University Maths course).
($\Leftarrow$) Assume that T is bijective.

Then T has an inverse function $S : W \to V$.

[Secret aim: S is linear.]

Take w_1, $w_2 \in W$ and λ_1, $\lambda_2 \in \mathbb{F}$.

[Secret aim: $S(\lambda_1 w_1 + \lambda_2 w_2) = \lambda_1 S(w_1) + \lambda_2 S(w_2)$.]

Let $v_1 = S(w_1)$, $v_2 = S(w_2)$. Then $T(v_1) = TS(w_1) = w_1$ and $T(v_2) = TS(w_2) = w_2$.

Now

$$\begin{aligned}S(\lambda_1 w_1 + \lambda_2 w_2) &= S(\lambda_1 T(v_1) + \lambda_2 T(v_2)) \\ &= S(T(\lambda_1 v_1 + \lambda_2 v_2)) \text{ since } T \text{ is linear} \\ &= \lambda_1 v_1 + \lambda_2 v_2 \text{ as } S \text{ is inverse to } T \\ &= \lambda_1 S(w_1) + \lambda_2 S(w_2).\end{aligned}$$

So S is linear. □

Proposition 32. *Let U, V, W be vector spaces. Let $S : U \to V$ and $T : V \to W$ be invertible linear transformations. Then $TS : U \to W$ is invertible, and $(TS)^{-1} = S^{-1}T^{-1}$.*

Proof. Exercise. □

6.3 Rank and nullity

Definition. Let V, W be vector spaces. Let $T : V \to W$ be linear. We define the *kernel* (or *null space*) of T to be

$$\ker T := \{v \in V : T(v) = 0_W\}.$$

We define the *image* of T to be

$$\operatorname{Im} T := \{T(v) : v \in V\}.$$

Remark. This definition of image is the same for any function between any two sets.

Lemma 33. *Let V, W be vector spaces. Let $T : V \to W$ be linear. For v_1, $v_2 \in V$, $T(v_1) = T(v_2)$ if and only if $v_1 - v_2 \in \ker T$.*

Proof. For v_1, $v_2 \in V$, we have

$$T(v_1) = T(v_2) \Leftrightarrow T(v_1) - T(v_2) = 0_W \Leftrightarrow T(v_1 - v_2) = 0_W \Leftrightarrow v_1 - v_2 \in \ker T.$$

□

Here is a very useful corollary, one that is very helpful in practice.

Corollary 34. *Let V, W be vector spaces. Let $T : V \to W$ be linear. Then T is injective if and only if $\ker T = \{0_V\}$.*

Proof. ($\Leftarrow$) Assume that $\ker T = \{0_V\}$.
Take v_1, $v_2 \in V$ with $T(v_1) = T(v_2)$.
Then, by Lemma 33, $v_1 - v_2 \in \ker T$, so $v_1 = v_2$.
So T is injective.
($\Rightarrow$) Assume that $\ker T \neq \{0_V\}$. Then there is $v \in \ker T$ with $v \neq 0_V$.
Then $T(v) = T(0_V)$, so T is not injective. □

Here are some useful properties of kernels and images.

Proposition 35. *Let V, W be vector spaces over $\mathbb{F}$. Let $T : V \to W$ be linear. Then*

(i) $\ker T$ *is a subspace of* V *and* $\operatorname{Im} T$ *is a subspace of* W*;*

(ii) if A *is a spanning set for* V*, then* $T(A)$ *is a spanning set for* $\operatorname{Im} T$*; and*

(iii) if V *is finite-dimensional, then* $\ker T$ *and* $\operatorname{Im} T$ *are finite-dimensional.*

Proof. (i) Note that $\ker T \subseteq V$ and $\operatorname{Im} T \subseteq W$.

$\underline{\ker T}$ Note that $T(0_V) = 0_W$ so $0_V \in \ker T$.

Take v_1, $v_2 \in \ker T$ and $\lambda \in \mathbb{F}$, so $T(v_1) = T(v_2) = 0_W$.

Then $T(\lambda v_1 + v_2) = \lambda T(v_1) + T(v_2) = \lambda 0_W + 0_W = 0_W$, so $\lambda v_1 + v_2 \in \ker T$.

So, by the Subspace Test, $\ker T \leq V$.

$\underline{\operatorname{Im} T}$ We have $T(0_V) = 0_W$ so $0_W \in \operatorname{Im} T$.

Take w_1, $w_2 \in \operatorname{Im} T$ and $\lambda \in \mathbb{F}$. Then there are v_1, $v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$.

Then $\lambda w_1 + w_2 = \lambda T(v_1) + T(v_2) = T(\lambda v_1 + v_2) \in \operatorname{Im} T$.

So, by the Subspace Test, $\operatorname{Im} T \leq W$.

(ii) Let A be a spanning set for V.

Take $w \in \operatorname{Im} T$. Then $w = T(v)$ for some $v \in V$.

Now there are $v_1, \ldots, v_n \in A$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$.

So (by Proposition 28)

$$w = T(v) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n),$$

so $w \in \operatorname{Span}(T(A))$.

So $T(A)$ spans $\operatorname{Im} T$.

(iii) Assume that V is finite-dimensional. Then $\ker T \leq V$ so $\ker T$ is finite-dimensional by Proposition 23. Also, $\operatorname{Im} T$ is finite-dimensional by (ii). □

Definition. Let V, W be vector spaces with V finite-dimensional. Let $T : V \to W$ be linear. We define the *nullity* of T to be $\operatorname{null}(T) := \dim(\ker T)$, and the *rank* of T to be $\operatorname{rank}(T) := \dim(\operatorname{Im} T)$.

The next theorem is very important!

Theorem 36 (Rank-Nullity Theorem)**.** *Let V, W be vector spaces with V finite-dimensional. Let $T : V \to W$ be linear. Then* $\dim V = \operatorname{rank}(T) + \operatorname{null}(T)$.

Proof. Take a basis $v_1, \ldots, v_n$ for $\ker T$, where $n = \operatorname{null}(T)$.

Since $\ker T \leq V$, by Theorem 24 this can be extended to a basis $v_1, \ldots, v_n, v'_1, \ldots, v'_r$ of V.

Then $\dim(V) = n + r$.

For $1 \leq i \leq r$, let $w_i = T(v'_i)$.

Claim. $\{w_1, \ldots, w_r\}$ is a basis for $\operatorname{Im} T$.

Proof of claim

spanning:

By Proposition 35, $T(v_1), \ldots, T(v_n), T(v'_1), \ldots, T(v'_r)$ span $\operatorname{Im} T$.

But $v_1, \ldots, v_n \in \ker T$, so $T(v_1) = \cdots = T(v_n) = 0_W$, so these vectors do not contribute.

So in fact $w_1, \ldots, w_r$ span $\operatorname{Im} T$.

linearly independent:

Take $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$ such that $\alpha_1 w_1 + \cdots + \alpha_r w_r = 0_W$,

that is, $\alpha_1 T(v'_1) + \cdots + \alpha_r T(v'_r) = 0_W$,

but T is linear so this becomes $T(\alpha_1 v'_1 + \cdots + \alpha_r v'_r) = 0_W$.

So $\alpha_1 v'_1 + \cdots + \alpha_r v'_r \in \ker T$,

so there are $\beta_1, \ldots, \beta_n \in \mathbb{F}$ such that $\alpha_1 v'_1 + \cdots + \alpha_r v'_r = \beta_1 v_1 + \cdots + \beta_n v_n$,

that is, $\beta_1 v_1 + \cdots + \beta_n v_n - \alpha_1 v'_1 - \cdots - \alpha_r v'_r = 0_V$.

But $v_1, \ldots, v_n, v'_1, \ldots, v'_r$ are linearly independent, so $\beta_1 = \cdots = \beta_n = \alpha_1 = \cdots = \alpha_r = 0$.

So $w_1, \ldots, w_r$ are linearly independent.

This proves the claim.

Now using the claim we have $\operatorname{rank} T = r$,

and so $\dim(V) = n + r = \operatorname{null}(T) + \operatorname{rank}(T)$. □

Here are a couple of useful results in their own right that also illustrate the usefulness of the Rank-Nullity Theorem.

Corollary 37. *Let V be a finite-dimensional vector space. Let $T : V \to V$ be linear. The following are equivalent:*

(i) T is invertible;

(ii) $\operatorname{rank} T = \dim V$*;*

(iii) $\operatorname{null} T = 0$.

Proof. (i)⇒(ii):

Assume that T is invertible.

Then T is bijective (by Proposition 31), so is surjective, so $\operatorname{Im} T = V$, so $\operatorname{rank} T = \dim V$.

(ii)⇒(iii):

Assume that $\operatorname{rank} T = \dim V$.

Then, by Rank-Nullity, $\operatorname{null} T = 0$.

(iii)⇒(i):

Assume that $\operatorname{null} T = 0$.

Then $\ker T = \{0_V\}$,

so T is injective (by Corollary 34).

Also, by Rank-Nullity, $\operatorname{rank} T = \dim V$ and $\operatorname{Im} T \leq V$, so $\operatorname{Im} T = V$, so T is surjective.

So T is bijective, so T is invertible (by Proposition 31). □

The next result is important, and we'll use it again later in the course.

Corollary 38. *Let V be a finite-dimensional vector space. Let $T : V \to V$ be linear. Then any one-sided inverse of T is a two-sided inverse, and so is unique.*

Proof. Suppose that T has a right inverse $S : V \to V$, so $T \circ S = \mathrm{id}_V$.

Since id_V is surjective, T is surjective, so $\operatorname{rank} T = \dim V$.

So, by Corollary 37, T is invertible, say with two-sided inverse S'.

Then $S' = S' \circ \mathrm{id}_V = S' \circ (T \circ S) = (S' \circ T) \circ S = \mathrm{id}_V \circ S = S$.

So S is the (unique) two-sided inverse.

If instead we suppose that T has a left inverse $S : V \to V$, so $S \circ T = \mathrm{id}_V$, then T is injective so $\operatorname{null} T = 0$, and the argument is similar to the previous one. □

Lemma 39. *Let V and W be vector spaces, with V finite-dimensional. Let $T : V \to W$ be linear. Let $U \leq V$. Then $\dim U - \operatorname{null} T \leq \dim T(U) \leq \dim U$. In particular, if T is injective then $\dim T(U) = \dim U$.*

Proof. Let $S : U \to W$ be the restriction of T to U (that is, $S(u) = T(u)$ for all $u \in U$).

Then S is linear, and $\ker S \leq \ker T$ so $\operatorname{null} S \leq \operatorname{null} T$. Also, $\operatorname{Im} S = T(U)$.

By Rank-Nullity, $\dim T(U) = \dim \operatorname{Im} S = \dim U - \operatorname{null} S \leq \dim U$ and $\dim T(U) = \dim U - \operatorname{null} S \geq \dim U - \operatorname{null} T$.

If T is injective, then $\operatorname{null} T = 0$, so $\dim T(U) = \dim U$. □

7 Linear transformations and matrices

7.1 The matrix of a linear map with respect to given bases

We saw examples of linear maps arising from multiplying by a matrix: for $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, we defined $L_A : \mathbb{R}^n_{\text{col}} \to \mathbb{R}^m_{\text{col}}$ by $L_A(v) = Av$, and we defined $R_A : \mathbb{R}^m \to \mathbb{R}^n$ by $R_A(v) = vA$.
Question Does every linear map between real vector spaces have this form for a suitable matrix A?

Definition. Let V be an n-dimensional vector space over $\mathbb{F}$, let $v_1, \ldots, v_n$ be a basis of V. Let W be an m-dimensional vector space over $\mathbb{F}$, let $w_1, \ldots, w_m$ be a basis of W. Let $T : V \to W$ be a linear transformation. We define an $m \times n$ matrix for T as follows. For $1 \leq j \leq n$, $T(v_j) \in W$ so $T(v_j)$ is uniquely expressible as a linear combination of $w_1, \ldots, w_m$: there are unique a_{ij} (for $1 \leq i \leq m$) such that $T(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$. That is,

$$\begin{aligned} T(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ T(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\ &\vdots \quad\quad \vdots \\ T(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m. \end{aligned}$$

We say that $M(T) = (a_{ij})$ is the *matrix for T with respect to these ordered bases for V and W.*

Remark.

- The j^{th} column of $M(T)$ lists the coefficients of $T(v_j)$ with respect to the chosen basis for W.
- The matrix $M(T)$ above depended on our choice of bases. Different bases will in general give different matrices. This turns out to be interesting, and we'll explore it further soon.
- The order of the basis vectors mattered too.
- Sometimes it can be helpful to record the bases more explicitly. If B_V is an ordered basis for V and B_W is an ordered basis for W, then mathematicians sometimes write ${}_{B_V}M_{B_W}(T)$ (or other variants on this) for the matrix for T with respect to these ordered bases.
- If $V = W$ and we use the same ordered basis for both domain and codomain of $T : V \to V$, then we talk about the *matrix for T with respect to this basis.*

- If we have a linear map $T : V \to W$, then by writing vectors in V and W using coordinates with respect to given ordered bases, we can interpret T as left multiplication by $M(T)$. See Proposition 40 for more on this.

Example. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(x, y, z) = (0, x, y)$. This is linear (exercise!).

What is the matrix for T with respect to the standard basis for $\mathbb{R}^3$? We have

$$\begin{aligned} T(1,0,0) &= (\quad,\quad,\quad) \\ T(0,1,0) &= (\quad,\quad,\quad) \\ T(0,0,1) &= (\quad,\quad,\quad) \end{aligned}$$

so the matrix is

$$M(T) = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}.$$

Note that $T^2 \neq 0$ but that $T^3 = 0$ (exercise!). And $M(T)^2 \neq 0$ but $M(T)^3 = 0$ (exercise!).

Example. Take $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$ and consider $L_A : \mathbb{F}^n_{\text{col}} \to \mathbb{F}^m_{\text{col}}$ defined by $L_A(v) = Av$.

Take the standard basis $e_1, \ldots, e_n$ for $\mathbb{F}^n_{\text{col}}$ (where e_i is the n-vector that has 1 in position i and 0 elsewhere), and the standard basis $f_1, \ldots, f_m$ of $\mathbb{F}^m_{\text{col}}$ (where f_j is the m-vector that has 1 in position j and 0 elsewhere).

Then for $1 \leq i \leq n$ we have

$$L_A(e_i) = Ae_i = i^{\text{th}} \text{ column of } A = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} = a_{1i}f_1 + \cdots + a_{mi}f_m.$$

So $M(L_A) = A$ — we obtain the matrix A from which we started.

Proposition 40. *Let V be an n-dimensional vector space over $\mathbb{F}$, let B_V be an ordered basis for V. Let W be an m-dimensional vector space over $\mathbb{F}$, let B_W be an ordered basis for W. Then*

(i) the matrix of $0 : V \to W$ *is* $0_{m\times n}$*;*

(ii) the matrix of $\mathrm{id}_V : V \to V$ *is* I_n*;*

(iii) if $S : V \to W$, $T : V \to W$ *are linear and* α, $\beta \in \mathbb{F}$, *then* $M(\alpha S+\beta T) = \alpha M(S) + \beta M(T)$.

Moreover, let $T : V \to W$ *be linear, with matrix* A *with respect to* B_V *and* B_W. *Take* $v \in V$ *with coordinates* $x^T = (x_1, \ldots, x_n)^T$ *with respect to* B_V. *Then* Ax *is the coordinate vector of* $T(v)$ *with respect to* B_W.

Proof. (i), (ii), (iii): Exercise.

For the last part, say that basis B_V is $v_1, \ldots, v_n$, and basis B_W is $w_1, \ldots, w_m$. Saying that v has coordinates $(x_1, \ldots, x_n)^T$ means that $v = \sum_{j=1}^n x_j v_j$.

Then

$$\begin{aligned} T(v) &= T(\sum_{j=1}^{n} x_j v_j) \\ &= \sum_{j=1}^{n} x_j T(v_j) \text{ as } T \text{ linear} \\ &= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} w_i \right) \text{ by definition of } A \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) w_i \end{aligned}$$

so with respect to basis B_W, we see that $T(v)$ has i^{th} coordinate $\sum_{j=1}^n a_{ij}x_j = (Ax)_i$. □

Proposition 41. *Let* U, V, W *be finite-dimensional vector spaces over* $\mathbb{F}$, *with ordered bases* B_U, B_V, B_W *respectively. Say* B_U *has size* m, B_V *has size* n, B_W *has size* p. *Let* $S : U \to V$ *and* $T : V \to W$ *be linear. Let* A *be the matrix of* S *with respect to* B_U *and* B_V. *Let* B *be the matrix of* T *with respect to* B_V *and* B_W. *Then the matrix of* $T \circ S$ *with respect to* B_U *and* B_W *is* BA.

Proof. Note that A is an $n \times m$ matrix, and B is a $p \times n$ matrix, so the product matrix BA is defined, and is a $p \times m$ matrix.

Let B_U be $u_1, \ldots, u_m$

B_V be $v_1, \ldots, v_n$

B_W be $w_1, \ldots, w_p$.

As usual, write $A = (a_{ij})$ and $B = (b_{ij})$.
By definition of A and B, we have

$$S(u_i) = \sum_{j=1}^{n} a_{ji} v_j \text{ for } 1 \leq i \leq m \text{ and } T(v_j) = \sum_{k=1}^{p} b_{kj} w_k \text{ for } 1 \leq j \leq n.$$

Now for $1 \leq i \leq m$ we have

$$\begin{aligned}
(T \circ S)(u_i) = T(S(u_i)) &= T\left(\sum_{j=1}^{n} a_{ji} v_j\right) \\
&= \sum_{j=1}^{n} a_{ji} T(v_j) \text{ since } T \text{ is linear} \\
&= \sum_{j=1}^{n} a_{ji} \sum_{k=1}^{p} b_{kj} w_k \\
&= \sum_{k=1}^{p} \left(\sum_{j=1}^{n} b_{kj} a_{ji}\right) w_k
\end{aligned}$$

but $\sum_{j=1}^{n} b_{kj} a_{ji}$ is both the (k, i) entry of the matrix for $T \circ S$ with respect to B_U and B_W, and also the (k, i) entry of the matrix BA. □

Remark. This is why we define multiplication of matrices in the way that we do!

Remark. As we are about to see, this gives a relatively clear and painless proof that matrix multiplication is associative (see Proposition 2).

Corollary 42. *Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, take $B \in \mathcal{M}_{n\times p}(\mathbb{F})$, take $C \in \mathcal{M}_{p\times q}(\mathbb{F})$. Then $A(BC) = (AB)C$.*

Proof. We consider the left multiplication maps $L_A : \mathbb{F}^n_{\text{col}} \to \mathbb{F}^m_{\text{col}}$ and $L_B : \mathbb{F}^p_{\text{col}} \to \mathbb{F}^n_{\text{col}}$ and $L_C : \mathbb{F}^q_{\text{col}} \to \mathbb{F}^p_{\text{col}}$.

With respect to the standard bases of these spaces, the matrix of L_A is A, the matrix of L_B is B, and the matrix of L_C is C.

Hence, by Proposition 41, $A(BC)$ is the matrix of $L_A \circ (L_B \circ L_C) : \mathbb{F}^q_{\text{col}} \to \mathbb{F}^m_{\text{col}}$, and $(AB)C$ is the matrix of $(L_A \circ L_B) \circ L_C : \mathbb{F}^q_{\text{col}} \to \mathbb{F}^m_{\text{col}}$ with respect to the standard bases of the relevant spaces.

But composition of functions is associative, so $L_A \circ (L_B \circ L_C) = (L_A \circ L_B) \circ L_C$,

so $A(BC) = (AB)C$. □

Corollary 43. *Let V be a finite-dimensional vector space. Let $T : V \to V$ be an invertible linear transformation. Let $v_1, \ldots, v_n$ be a basis of V. Let A be the matrix of T with respect to this basis (for both domain and codomain). Then A is invertible, and A^{-1} is the matrix of T^{-1} with respect to this basis.*

Proof. Exercise. □

Corollary 44. *Let A be an $n \times n$ matrix. Any one-sided inverse of A is a two-sided inverse.*

Proof. This follows from Corollary 38 (the corresponding result for linear transformations) and Proposition 41 (relating matrices and linear transformations). □

7.2 Change of basis

Question Take two matrices for the same linear transformation with respect to different bases. How are the matrices related?

Example. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x, y) = (2x + y, 3x - 2y)$.

What is the matrix of T with respect to the standard basis e_1, e_2?

We have

$$T(1, 0) = (2, 3)$$
$$T(0, 1) = (1, -2)$$

so the matrix for T with respect to this basis is

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

Let $f_1 = (1, -2)$ and $f_2 = (-2, 5)$. Then f_1, f_2 is a basis of $\mathbb{R}^2$.

What is the matrix of T with respect to the basis f_1, f_2?

We have

$$T(f_1) = (0, 7) = 14f_1 + 7f_2$$
$$T(f_2) = (1, -16) = -27f_1 - 14f_2$$

so the matrix for T with respect to this basis is

$$B = \begin{pmatrix} 14 & -27 \\ 7 & -14 \end{pmatrix}.$$

How are these two matrices related?

Here is a blank space for a (hopefully helpful) diagram that I plan to draw in the lecture.

We can take f_1, f_2 and write them with respect to e_1, e_2: we have

$$\begin{aligned} f_1 &= e_1 - 2e_2 \\ f_2 &= -2e_1 + 5e_2 \end{aligned}$$

so we get a 'change of basis matrix'

$$P = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}.$$

Then we apply T using matrix A, which takes as input a vector written with respect to e_1, e_2 and returns a vector written with respect to e_1, e_2.

Then we rewrite the answer in terms of f_1, f_2, using P^{-1}.

This process corresponds to the matrix product $P^{-1}AP$.

Note that

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

so

$$\begin{aligned} P^{-1}AP &= \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 7 & -16 \end{pmatrix} \\ &= \begin{pmatrix} 14 & -27 \\ 7 & -14 \end{pmatrix} = B. \end{aligned}$$

Can we do something like this in general?

Theorem 45 (Change of basis theorem)**.** *Let V, W be finite-dimensional vector spaces over $\mathbb{F}$. Let $T : V \to W$ be linear. Let $v_1, \ldots, v_n$ and $v'_1, \ldots, v'_n$ be bases for V. Let $w_1, \ldots, w_m$ and $w'_1, \ldots, w'_m$ be bases for W. Let $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$ be the matrix for T with respect to $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$. Let $B = (b_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{F})$ be the matrix for T with respect to $v'_1, \ldots, v'_n$ and $w'_1, \ldots, w'_m$. Take p_{ij}, $q_{ij} \in \mathbb{F}$ such that $v'_i = \sum_{j=1}^{n} p_{ji} v_j$ and $w'_i = \sum_{j=1}^{m} q_{ji} w_j$. Let $P = (p_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $Q = (q_{ij}) \in \mathcal{M}_{m\times m}(\mathbb{F})$.*

Then $B = Q^{-1}AP$.

Proof. By definition of A and B, for $1 \le i \le n$ we have

$$T(v_i) = \sum_{j=1}^{m} a_{ji} w_j \text{ and } T(v'_i) = \sum_{j=1}^{m} b_{ji} w'_j.$$

Note that Q must be invertible. Let $(r_{ij}) = Q^{-1}$, so $w_i = \sum_{j=1}^{m} r_{ji} w'_j$.

Now

$$\begin{aligned}
T(v'_i) &= T\left(\sum_{j=1}^{n} p_{ji} v_j\right) \\
&= \sum_{j=1}^{n} p_{ji} T(v_j) \text{ as } T \text{ is linear} \\
&= \sum_{j=1}^{n} p_{ji} \sum_{k=1}^{m} a_{kj} w_k \\
&= \sum_{k=1}^{m} \left(\sum_{j=1}^{n} a_{kj} p_{ji}\right) w_k \\
&= \sum_{k=1}^{m} \left(\sum_{j=1}^{n} a_{kj} p_{ji}\right) \left(\sum_{\ell=1}^{m} r_{\ell k} w'_\ell\right) \\
&= \sum_{\ell=1}^{m} \left(\sum_{k=1}^{m}\sum_{j=1}^{n} r_{\ell k} a_{kj} p_{ji}\right) w'_\ell.
\end{aligned}$$

The coefficient of w'_ℓ in this sum is the (ℓ, i) entry of B, the matrix for T with respect to $v'_1, \ldots, v'_n$ and $w'_1, \ldots, w'_m$, but we have also just seen that it is the (ℓ, i) entry of $Q^{-1}AP$.

So $B = Q^{-1}AP$.

□

The following result follows immediately from the Change of basis theorem, but is sufficiently useful that it is worth recording it separately.

Corollary 46 (Change of basis theorem, version 2)**.** *Let V be a finite-dimensional vector space. Let $T : V \to V$ be linear. Let $v_1, \ldots, v_n$ and $v'_1, \ldots, v'_n$ be bases for V. Let A be the matrix of T with respect to $v_1, \ldots, v_n$. Let B be the matrix of T with respect to $v'_1, \ldots, v'_n$. Let P be the change of basis matrix, that is, the $n \times n$ matrix (p_{ij}) such that $v'_i = \sum_{j=1}^{n} p_{ji} v_j$.*

Then $B = P^{-1}AP$.

Proof. Immediate from the Change of basis theorem. □

Remark. The change of basis matrix P is the matrix of the identity map $\mathrm{id}_V : V \to V$ with respect to the basis $v'_1, \ldots, v'_n$ for V as domain and the basis $v_1, \ldots, v_n$ for V as codomain.

Definition. Take $A, B \in \mathcal{M}_{n\times n}(\mathbb{F})$. If there is an invertible $n \times n$ matrix P such that $P^{-1}AP = B$, then we say that A and B are *similar*.

Remark. So two matrices representing the same linear transformation from a finite-dimensional vector space to itself, but with respect to different bases, are similar.

7.3 Historical interlude 5

In a course such as this, the lecturer's job is to try to present the material in a way that is comprehensible, arranged in a sensible order, and where the theorems are all true. But mathematics does not arrive like that—proving theorems that have not been proved before is difficult, and mathematicians try many ideas that do not work before they find ones that do. The mathematician Julia Robinson summed this up eloquently when she was asked to describe her typical week:

> "Monday — tried to prove theorem
>
> Tuesday — tried to prove theorem
>
> Wednesday — tried to prove theorem
>
> Thursday — tried to prove theorem
>
> Friday — theorem false"

You can read more about the life and work of Julia Robinson on MacTutor, or in the book *Julia: A life in mathematics* by Constance Reid (published by the Mathematical Association of America, 1996), which is where I found the above quote.
`http://www-history.mcs.st-andrews.ac.uk/Biographies/Robinson_Julia.html`

7.4 Matrices and rank

For a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, we have defined the row space and row rank, and analogously the column space and column rank.
Question Are rowrank(A) and colrank(A) related?

Remark. From the definitions, we see that colsp(A) = rowsp(A^T) and so colrank(A) = rowrank(A^T). Similarly, rowsp(A) = colsp(A^T) and so rowrank(A) = colrank(A^T).

Proposition 47. *Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$, let $r = \text{colrank}(A)$. Then there are invertible matrices $P \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $Q \in \mathcal{M}_{m\times m}(\mathbb{F})$ such that $Q^{-1}AP$ has the block form $\begin{pmatrix} I_r & 0_{r\times s} \\ 0_{t\times r} & 0_{t\times s} \end{pmatrix}$ where $s = n - r$ and $t = m - r$.*

Proof. Idea: consider L_A, and find suitable bases for domain and codomain with respect to which L_A has matrix in the given block form.

Consider $L_A : \mathbb{F}^n_{\text{col}} \to \mathbb{F}^m_{\text{col}}$ defined by $L_A(v) = Av$. This is linear.

We have seen that with respect to the standard bases of $\mathbb{F}^n_{\text{col}}$ and $\mathbb{F}^m_{\text{col}}$ the matrix for L_A is A.

We have $\text{Im}\, L_A = \text{colsp}\, A$ (because if e_i is a standard basis vector of $\mathbb{F}^n_{\text{col}}$ then $L_A(e_i)$ is the i^{th} column of A), so $\text{rank}(L_A) = \text{colrank}(A) = r$.

By Rank-Nullity, $\text{null}(L_A) = n - r = s$. Take a basis $v_1, \ldots, v_s$ of $\ker L_A$, and extend to a basis $v_1, \ldots, v_s, v'_1, \ldots, v'_r$ of $\mathbb{F}^n_{\text{col}}$.

For $1 \le j \le r$, let $w_j = L_A(v'_j)$. Then, as in the proof of Rank-Nullity, $w_1, \ldots, w_r$ is a basis of $\text{Im}(L_A)$, and we can extend this to a basis $w_1, \ldots, w_r, w_{r+1}, \ldots, w_m$ of $\mathbb{F}^m_{col}$.

Take the ordered bases $v'_1, \ldots, v'_r, v_1, \ldots, v_s$ for $\mathbb{F}^n_{\text{col}}$ and $w_1, \ldots, w_r, w_{r+1}, \ldots, w_m$ of $\mathbb{F}^m_{\text{col}}$.

What is the matrix for L_A with respect to these?

We have $L_A(v'_j) = w_j$ for $1 \le j \le r$

and $L_A(v_i) = 0$ for $1 \le i \le s$,

so the matrix is

$$\begin{pmatrix} I_r & 0_{r\times s} \\ 0_{t\times r} & 0_{t\times s} \end{pmatrix}.$$

So, by the Change of basis theorem (Theorem 45), there are invertible $P \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $Q \in \mathcal{M}_{m\times m}(\mathbb{F})$ such that $Q^{-1}AP$ has this form. □

Proof. (Sketch of an alternative argument.)

Just as we have elementary row operations (EROs), we can define elementary column operations (ECOs). And analogously to reduced row echelon form (RRE form), we can consider reduced column echelon form (RCE form).

Take A, and use EROs to reduce it to E in RRE form. Now use ECOs to reduce E to F in RCE form. We consider the structure of F.

Any zero rows of F are below any nonzero rows, and any zero columns are to the right of any nonzero columns.

If a column contains the leading entry of a row, then it has exactly one 1 and all other entries 0.

Also, if a row contains the leading entry of a column, then it has exactly one 1 and all other entries 0.

So F has the form

$$\begin{pmatrix} I_r & 0_{r\times s} \\ 0_{t\times r} & 0_{t\times s} \end{pmatrix}.$$

□

Lemma 48. *Take $A \in \mathcal{M}_{m\times n}(\mathbb{F})$. Let R be an invertible $m \times m$ matrix, let P be an invertible $n \times n$ matrix. Then*

(i) $\mathrm{rowsp}(RA) = \mathrm{rowsp}(A)$ *and so* $\mathrm{rowrank}(RA) = \mathrm{rowrank}(A)$*;*

(ii) $\mathrm{colrank}(RA) = \mathrm{colrank}(A)$*;*

(iii) $\mathrm{colsp}(AP) = \mathrm{colsp}(A)$ *and so* $\mathrm{colrank}(AP) = \mathrm{colrank}(A)$*;*

(iv) $\mathrm{rowrank}(AP) = \mathrm{rowrank}(A)$.

Proof. Write $R = (r_{ij})$.

(i) Let $x_1, \ldots, x_m \in \mathbb{F}^n_{\mathrm{row}}$ be the m rows of A.

Then the i^{th} row of RA is $r_{i1}x_1 + \cdots + r_{im}x_m$ — a linear combination of the rows of A, so is in $\mathrm{rowsp}(A)$.

So $\mathrm{rowsp}(RA) \leq \mathrm{rowsp}(A)$.

But R is invertible, so we can apply the same argument to see that $\mathrm{rowsp}(A) = \mathrm{rowsp}(R^{-1}(RA)) \leq \mathrm{rowsp}(RA)$.

So $\mathrm{rowsp}(RA) = \mathrm{rowsp}(A)$, and so $\mathrm{rowrank}(RA) = \mathrm{rowrank}(A)$.

(ii) Let $y_1, \ldots, y_n \in \mathbb{F}^m_{\mathrm{col}}$ be the n columns of A.

Then the i^{th} column of RA is Ry_i,

so $\text{colsp}(RA) = L_R(\text{colsp}(A))$.

But R is invertible, so $\text{null}(L_R) = 0$, so Corollary 37 tells us that $\dim(L_R(\text{colsp}(A)) = \dim \text{colsp}(A)$.

So $\text{colrank}(RA) = \text{colrank}(A)$.

(iii) and (iv) Since P is invertible, P^T is also invertible, with inverse $(P^{-1})^T$. Now applying the above argument to $(AP)^T = P^T A^T$ shows both that $\text{colsp}(AP) = \text{colsp}(A)$ so $\text{colrank}(AP) = \text{colrank}(A)$, and also that $\text{rowrank}(AP) = \text{rowrank}(A)$.

□

Theorem 49. *Let A be an $m \times n$ matrix. Then* $\text{colrank}(A) = \text{rowrank}(A)$.

Proof. By Proposition 47, there are an invertible $n \times n$ matrix P and an invertible $m \times m$ matrix Q such that $Q^{-1}AP = B$, where B has the block form

$$\begin{pmatrix} I_r & 0_{r\times s} \\ 0_{t\times r} & 0_{t\times s} \end{pmatrix}.$$

By Lemma 48, $\text{rowrank}(Q^{-1}AP) = \text{rowrank}(A)$ and $\text{colrank}(Q^{-1}AP) = \text{colrank}(A)$.

But $\text{rowrank}(B) = \text{colrank}(B) = r$,

so $\text{rowrank}(A) = \text{colrank}(A)$. □

Definition. Let A be an $m \times n$ matrix. The *rank* of A, written $\text{rank}(A)$, is the row rank of A (which we have just seen is also the column rank of A).

Remark. Let $T : V \to W$ be linear. Let B_V, B_W be ordered bases of V, W respectively. Let A be the matrix for T with respect to B_V and B_W. Then $\text{rank}(A) = \text{rank}(T)$.

Proposition 50. *Let A be an $m \times n$ matrix. Let x be the $n \times 1$ column vector of variables $x_1, \ldots, x_n$. Let S be the solution space of the system $Ax = 0$ of m homogeneous linear equations in $x_1, \ldots, x_n$, that is, $S = \{v \in \mathbb{F}^n_{col} : Av = 0\}$. Then* $\dim S = n - \text{colrank}\, A$.

Proof. Consider $L_A : \mathbb{F}^n_{\text{col}} \to \mathbb{F}^m_{\text{col}}$ defined by $L_A(v) = Av$.

Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n_{\text{col}}$.

Then, as in the proof of Lemma 48, $\text{Im}(L_A)$ is spanned by $Ae_1, \ldots, Ae_n$, and Ae_i is the i^{th} column of A, so $\text{Im}(L_A) = \text{colsp}(A)$, so $\text{rank}(L_A) = \text{colrank}\, A$.

By definition, $\ker(L_A) = S$.

So, by Rank-Nullity, $\dim S + \text{colrank}\, A = \dim \mathbb{F}^n_{\text{col}} = n$. □

8 Inner product spaces

As you might have seen in geometrical contexts, there are various types of product of two real vectors. The cross product (written $x \times y$ or $x \wedge y$) is powerful in $\mathbb{R}^3$, but does not naturally generalise. The scalar product (written $x \cdot y$) *does* generalise nicely. Some of the theory works over an arbitrary field $\mathbb{F}$. Sometimes we need to specialise to $\mathbb{R}$ or to $\mathbb{C}$. Scalar products and related ideas are important in geometry, dynamics, quantum physics, ...

8.1 Bilinear forms

Definition. Let V be a vector space over $\mathbb{F}$. A *bilinear form* on V is a function of two variables from V taking values in $\mathbb{F}$, often written $\langle -, - \rangle : V \times V \to \mathbb{F}$, such that

(i) $\langle \alpha_1 v_1 + \alpha_2 v_2, v_3 \rangle = \alpha_1 \langle v_1, v_3 \rangle + \alpha_2 \langle v_2, v_3 \rangle$ for all v_1, v_2, $v_3 \in V$ and α_1, $\alpha_2 \in \mathbb{F}$; and

(ii) $\langle v_1, \alpha_2 v_2 + \alpha_3 v_3 \rangle = \alpha_2 \langle v_1, v_2 \rangle + \alpha_3 \langle v_1, v_3 \rangle$ for all v_1, v_2, $v_3 \in V$ and α_2, $\alpha_3 \in \mathbb{F}$.

Remark. Condition (i) says that $\langle -, - \rangle$ is linear in the first variable when we fix the second variable, and condition (ii) similarly the other way round.

Example. For $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{F}^n$, we define $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$. This gives a bilinear form.

In $\mathbb{R}^2$ and $\mathbb{R}^3$, this is the familiar dot product, or scalar product, often written $x \cdot y$. We use this terminology in higher dimensions too.

Example. Take $A \in \mathcal{M}_{n \times n}(\mathbb{F})$. For x, $y \in \mathbb{F}^n$, define $\langle x, y \rangle = xAy^T$. This gives a bilinear form on $\mathbb{F}^n$.

Remark. Officially, xAy^T is a 1×1 matrix, not an element of $\mathbb{F}$. But it is completely natural to identify the 1×1 matrix with the corresponding scalar (to think of them as the same).

Remark. Note that the usual scalar product from the previous example is an example of this in the special case that $A = I_n$, because $x \cdot y = xy^T$.

Definition. Let V be a vector space over $\mathbb{F}$. Let $\langle -, - \rangle$ be a bilinear form on V. Take $v_1, \ldots, v_n \in V$. The *Gram matrix* of $v_1, \ldots, v_n$ with respect to $\langle -, - \rangle$ is the $n \times n$ matrix $(\langle v_i, v_j \rangle) \in \mathcal{M}_{n \times n}(\mathbb{F})$.

Proposition 51. *Let V be a finite-dimensional vector space over $\mathbb{F}$. Let $\langle -, - \rangle$ be a bilinear form on V. Let $v_1, \ldots, v_n$ be a basis for V. Let $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ be the associated Gram matrix. For $u, v \in V$, let $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{F}^n$ be the unique coordinate vectors such that $u = x_1v_1 + \cdots + x_nv_n$ and $v = y_1v_1 + \cdots + y_nv_n$. Then $\langle u, v \rangle = xAy^T$.*

Remark. So the bilinear form of the second example above essentially describes *all* linear forms on finite-dimensional vector spaces.

Proof. We have

$$\begin{aligned}
\langle u, v \rangle &= \langle \sum_{i=1}^{n} x_i v_i, \sum_{j=1}^{n} y_j v_j \rangle \\
&= \sum_{i=1}^{n} x_i \langle v_i, \sum_{j=1}^{n} y_j v_j \rangle \text{ using linearity in the first entry} \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j \langle v_i, v_j \rangle \text{ using linearity in the second entry} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j a_{ij} \\
&= xAy^T.
\end{aligned}$$

□

Definition. We say that a bilinear form $\langle -, - \rangle : V \times V \to \mathbb{F}$ is *symmetric* if $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$ for all $v_1, v_2 \in V$.

8.2 Inner product spaces

Definition. Let V be a real vector space. We say that a bilinear form $\langle -, - \rangle : V \times V \to \mathbb{R}$ is *positive definite* if $\langle v, v \rangle \geq 0$ for all $v \in V$, with $\langle v, v \rangle = 0$ if and only if $v = 0$.

Definition. An *inner product* on a real vector space V is a positive definite symmetric bilinear form on V.

We say that a real vector space is an *inner product space* if it is equipped with an inner product. Unless otherwise specified, we write the inner product as $\langle -, - \rangle$.

Definition. Let V be a real inner product space. For $v \in V$, we define the *norm* (or *magnitude* or *length*) of v to be $\|v\| := \sqrt{\langle v, v \rangle}$.

Remark. You might have seen that in $\mathbb{R}^2$ or $\mathbb{R}^3$ we have $x \cdot y = \|x\|\|y\| \cos\theta$, where θ is the angle between the vectors x and y. In general, we can use this idea to *define* a notion of angle in an abstract inner product space V: we define the *angle* between nonzero vectors x, $y \in V$ to be $\cos^{-1}\left(\frac{\langle x,y\rangle}{\|x\|\|y\|}\right)$, where this is taken to lie in the interval $[0, \pi]$. We'll see later that $\left|\frac{\langle x,y\rangle}{\|x\|\|y\|}\right| \leq 1$ for any nonzero vectors x, y in an inner product space, and so this definition does make sense.

Example. The dot product on $\mathbb{R}^n$ is an inner product. We noted earlier that it is a bilinear form, and it is clearly symmetric. If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $x \neq 0$, then $x \cdot x = x_1^2 + \cdots + x_n^2 > 0$, so the dot product is also positive definite.

The inner product space consisting of $\mathbb{R}^n$ equipped with the dot product is known as n-dimensional *Euclidean space*. The dot product also turns $\mathbb{R}^n_{\text{col}}$ into an inner product space.

Example. Let $V = \mathbb{R}_n[x]$, the vector space of polynomials of degree $\leq n$ with real coefficients. For f, $g \in V$, define

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx.$$

Then $\langle -, -\rangle$ is bilinear and symmetric.

If $f \in V$ and $f \neq 0$, then $f(a) = 0$ for only finitely many a in $[0, 1]$, and $[f(x)]^2 > 0$ at other x, and we find that $\int_0^1 f^2(x)dx > 0$. So $\langle -, -\rangle$ is positive definite.

So $\langle -, -\rangle$ is an inner product on V.

Proposition 52. *Let V be a finite-dimensional real inner product space. Take $u \in V \setminus \{0\}$. Define $u^\perp := \{v \in V : \langle v, u\rangle = 0\}$. Then $u^\perp$ is a subspace of V, and $\dim u^\perp = \dim V - 1$, and $V = \text{Span}(u) \oplus u^\perp$.*

Proof. Consider $f : V \to \mathbb{R}$ given by $f(v) = \langle v, u\rangle$.

Since $\langle -, -\rangle$ is bilinear, f is linear.

By definition, $u^\perp = \ker(f)$, so $u^\perp$ is a subspace of V.

Also, $\text{Im}(f) \leq \mathbb{R}$ and $\dim(\text{Im}(f)) > 0$ (because $\langle u, u\rangle > 0$ so $\text{Im}(f)$ contains a nonzero real number), so $\dim \text{Im}(f) = 1$.

So, by Rank-Nullity, $\dim V = \dim(u^\perp) + 1$.

Take $v \in \text{Span}(u) \cap u^\perp$. Then $v = \lambda u$ for some $\lambda \in \mathbb{R}$, and $\langle v, u\rangle = 0$.

So $0 = \langle v, u\rangle = \langle \lambda u, u\rangle = \lambda\langle u, u\rangle$, but $\langle u, u\rangle > 0$ so this gives $\lambda = 0$.

So $\text{Span}(u) \cap u^\perp = \{0\}$.

So, by Proposition 26, $V = \text{Span}(u) \oplus u^\perp$. □

Definition. Let V be an inner product space. We say that $\{v_1, \ldots, v_n\} \subseteq V$ is an *orthonormal set* if for all i, j we have

$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Lemma 53. *Let $\{v_1, \ldots, v_n\}$ be an orthonormal set in an inner product space V. Then $v_1, \ldots, v_n$ are linearly independent.*

Proof. Take $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$.

Then for $1 \leq i \leq n$ we have

$$\begin{aligned} 0 = \langle 0, v_i \rangle &= \langle \alpha_1 v_1 + \cdots + \alpha_n v_n, v_i \rangle \\ &= \alpha_1 \langle v_1, v_i \rangle + \cdots + \alpha_n \langle v_n, v_i \rangle \\ &= \alpha_i \end{aligned}$$

so $\alpha_1 = \cdots = \alpha_n = 0$. □

Remark. So a set of n orthonormal vectors in an n-dimensional vector space is a basis.

Theorem 54. *Let V be an n-dimensional real inner product space. Then there is an orthonormal basis $v_1, \ldots, v_n$ of V.*

Proof. By induction on $n = \dim V$.

$\underline{n = 0}$: nothing to prove

$\underline{n = 1}$: Take $v \in V$ with $v \neq 0$. Let $v_1 = \frac{v}{\|v\|}$.

Then

$$\langle v_1, v_1 \rangle = \frac{\langle v, v \rangle}{\|v\|^2} = 1$$

and $\{v_1\}$ is a basis of V,

so $\{v_1\}$ is an orthonormal basis of V.

$\underline{\text{induction step}}$: Fix $n \geq 1$, and suppose that the result holds for real inner product spaces with dimension $n - 1$.

Take $v \in V$ with $v \neq 0$. Let $v_1 = \frac{v}{\|v\|}$, then $\|v_1\| = 1$ as above.

Let $U = v_1^{\perp}$. Then, by Proposition 52, $V = \text{Span}(v_1) \oplus U$.

The restriction of $\langle -, - \rangle$ to $U \times U$ makes U an inner product space, and $\dim U = n - 1$ (by Proposition 52). So, by the induction hypothesis, there is an orthonormal basis $v_2, \ldots, v_n$ of U.

Since $V = \text{Span}(v_1) \oplus U$, we see that $v_1, v_2, \ldots, v_n$ is a basis for V.

Now for $2 \leq i \leq n$ we have $\langle v_1, v_i \rangle = 0$ (because $U = v_1^{\perp}$),

so $v_1, v_2, \ldots, v_n$ is an orthonormal basis of V. □

8.3 Orthogonal matrices

In a previous definition, we said that a matrix $X \in \mathcal{M}_{n\times n}(\mathbb{R})$ is *orthogonal* if $XX^T = I_n = X^TX$. Equivalently, X is orthogonal if X is invertible and $X^{-1} = X^T$.

Lemma 55. *Take $X \in \mathcal{M}_{n\times n}(\mathbb{R})$. Consider $\mathbb{R}^n$ equipped with the usual inner product $\langle x, y\rangle = x \cdot y$. The following are equivalent:*

(i) $XX^T = I_n$;

(ii) $X^TX = I_n$;

(iii) the rows of X form an orthonormal basis of $\mathbb{R}^n$;

(iv) the columns of X form an orthonormal basis of $\mathbb{R}^n_{col}$;

(v) for all x, $y \in \mathbb{R}^n$, we have $xX \cdot yX = x \cdot y$.

Proof. (i) $\Leftrightarrow$ (ii): For any A, $B \in \mathcal{M}_{n\times n}(\mathbb{R})$, we have $AB = I_n$ if and only if $BA = I_n$ (Corollary 44).
(i) $\Leftrightarrow$ (iii): Say the rows of X are $x_1, \ldots, x_n$.
Note that the (i, j) entry of XX^T is $x_i \cdot x_j$.
But $XX^T = I_n$ if and only if the (i, j) entry of XX^T is δ_{ij}.
(ii) $\Leftrightarrow$ (iv): Say the columns of X are $y_1, \ldots, y_n$.
We see that the (i, j) entry of X^TX is $y_i \cdot y_j$.
(i) $\Rightarrow$ (v): We can think of $x \cdot y$ as xy^T.
Assume that $XX^T = I_n$.
Take x, $y \in \mathbb{R}^n$.
Then

$$\begin{aligned}(xX) \cdot (yX) &= (xX)(yX)^T \\ &= (xX)(X^Ty^T) \\ &= x(XX^T)y^T \\ &= xI_ny^T \\ &= xy^T \\ &= x \cdot y.\end{aligned}$$

(v) $\Rightarrow$ (iii): Assume that $xX \cdot yX = x \cdot y$ for all x, $y \in \mathbb{R}^n$.
Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$.
Then e_iX is the i^{th} row of X.
We have $e_iX \cdot e_jX = e_i \cdot e_j = \delta_{ij}$

so $e_1X, \ldots, e_nX$ is an orthonormal set of n vectors in $\mathbb{R}^n$ and hence a basis.

So the rows of X form an orthonormal basis of $\mathbb{R}^n$. □

Remark. Condition (v) says that the map $R_X : \mathbb{R}^n \to \mathbb{R}^n$ sending x to xX preserves the inner product, and hence preserves length and angle. Such a map is called an *isometry* of the Euclidean space $\mathbb{R}^n$. So Lemma 55 says that X is orthogonal if and only if the map R_X is an isometry.

8.4 The Cauchy-Schwarz Inequality

Theorem 56 (Cauchy-Schwarz Inequality)**.** *Let V be a real inner product space. Take v_1, $v_2 \in V$. Then $|\langle v_1, v_2\rangle| \leq \|v_1\|\|v_2\|$, with equality if and only if v_1, v_2 are linearly dependent.*

Proof. If $v_1 = 0$ then the inequality is clear, so assume that $v_1 \neq 0$.

For $t \in \mathbb{R}$, consider $\langle tv_1 + v_2, tv_1 + v_2\rangle$. Since $\langle -, -\rangle$ is bilinear and symmetric, we have

$$\begin{aligned}\langle tv_1 + v_2, tv_1 + v_2\rangle &= t^2\langle v_1, v_1\rangle + 2t\langle v_1, v_2\rangle + \langle v_2, v_2\rangle \\ &= \|v_1\|^2t^2 + 2t\langle v_1, v_2\rangle + \|v_2\|^2.\end{aligned}$$

This is a quadratic in t.

But also $\langle tv_1+v_2, tv_1+v_2\rangle = \|tv_1+v_2\|^2 \geq 0$ for all $t \in \mathbb{R}$, so the quadratic has non-positive discriminant.

So $(2\langle v_1, v_2\rangle)^2 - 4\|v_1\|^2\|v_2\|^2 \leq 0$

so $|\langle v_1, v_2\rangle| \leq \|v_1\|\|v_2\|$.

When do we have equality?

If $|\langle v_1, v_2\rangle| = \|v_1\|\|v_2\|$, then the discriminant is 0, so the quadratic has a repeated root. So there is some $\alpha \in \mathbb{R}$ such that $\|\alpha v_1 + v_2\| = 0$ and so $\alpha v_1 + v_2 = 0$ (as $\langle -, -\rangle$ is positive definite), so v_1 and v_2 are linearly dependent.

Conversely, if v_1, v_2 are linearly dependent and $v_1 \neq 0$ then $v_2 = \lambda v_1$ for some $\lambda \in \mathbb{R}$,

so $|\langle v_1, v_2\rangle| = |\lambda|\langle v_1, v_1\rangle = |\lambda|\|v_1\|^2$

and $\|v_1\|\|v_2\| = \sqrt{\langle v_1, v_1\rangle}\sqrt{\langle v_2, v_2\rangle} = |\lambda|\|v_1\|^2$

so we have equality.

If $v_1 = 0$ then we clearly have equality. □

8.5 Complex inner product spaces

Definition. Let V be a complex vector space. A function $\langle -, -\rangle : V \times V \to \mathbb{C}$ is a *sesquilinear form* if

(i) $\langle \alpha_1 v_1 + \alpha_2 v_2, v_3 \rangle = \alpha_1 \langle v_1, v_3 \rangle + \alpha_2 \langle v_2, v_3 \rangle$ for all v_1, v_2, $v_3 \in V$ and α_1, $\alpha_2 \in \mathbb{C}$; and

(ii) $\langle v_1, v_2 \rangle = \overline{\langle v_2, v_1 \rangle}$ for all v_1, $v_2 \in V$.

In particular, we have $\langle v, v \rangle \in \mathbb{R}$ for all $v \in V$. We say that a sesquilinear form is *positive definite* if $\langle v, v \rangle \geq 0$ for all $v \in V$, with $\langle v, v \rangle = 0$ if and only if $v = 0$.

A *complex inner product space* is a complex vector space equipped with a positive definite sesquilinear form.

Remark. Positive definite sesquilinear forms are often called *Hermitian forms*, and complex inner product spaces are often called *Hermitian spaces*.

You will explore inner product spaces further in future Linear Algebra courses.

8.6 Historical interlude 6

We've seen a few named mathematicians in the last few lectures, so here are some MacTutor biographies.

Jorgen Gram
`http://www-history.mcs.st-andrews.ac.uk/Biographies/Gram.html`

Augustin-Louis Cauchy
`http://www-history.mcs.st-andrews.ac.uk/Biographies/Cauchy.html`

Hermann Schwarz
`http://www-history.mcs.st-andrews.ac.uk/Biographies/Schwarz.html`

(Note that there have been mathematicians called Schwartz—with a t—but this Schwarz has no t.)

Hermitian forms are named after Charles Hermite
`http://www-history.mcs.st-andrews.ac.uk/Biographies/Hermite.html`

Incidentally, I was interested in the word 'sesquilinear', so I looked up its etymology. It turns out that the prefex 'sesqui-' comes from Latin, and means "one and a half".

If you are interested in the history of mathematics, then you might be interested in the Part B course on History of Maths. You can find reading recommendations on the webpage for that course. One good place to start might be

Jacqueline Stedall, *The history of mathematics: a very short introduction* (Oxford University Press, 2012).

To be continued... (in Linear Algebra II)