Linear Algebra I

Richard Earl

Michaelmas Term 2023

0. PRELIMINARY MATERIAL

0.1 Syllabus

Systems of linear equations. Matrices and the beginnings of matrix algebra. Use of matrices to describe systems of linear equations. Elementary Row Operations (EROs) on matrices. Reduction of matrices to echelon form. Application to the solution of systems of linear equations. [2.5]

Inverse of a square matrix. The use of EROs to compute inverses; computational efficiency of the method. Transpose of a matrix; orthogonal matrices. [1]

Vector spaces: definition of a vector space over a field (such as $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$). Subspaces. Many explicit examples of vector spaces and subspaces. [1.5]

Span of a set of vectors. Examples such as row space and column space of a matrix. Linear dependence and independence. Bases of vector spaces; examples. The Steinitz Exchange Lemma; dimension. Application to matrices: row space and column space, row rank and column rank. Coordinates associated with a basis of a vector space. [2]

Use of EROs to find bases of subspaces. Sums and intersections of subspaces; the dimension formula. Direct sums of subspaces. [1.5]

Linear transformations: definition and examples (including projections associated with direct-sum decompositions). Some algebra of linear transformations; inverses. Kernel and image, Rank-Nullity Theorem. Applications including algebraic characterisation of projections (as idempotent linear transformations). [2]

Matrix of a linear transformation with respect to bases. Change of Bases Theorem. Applications including proof that row rank and column rank of a matrix are equal. [2]

Bilinear forms; real inner product spaces; examples. Mention of complex inner product spaces. Cauchy–Schwarz inequality. Distance and angle. The importance of orthogonal matrices. [1.5]

0.2 Reading list

(1) Gilbert Strang, Introduction to linear algebra (Fifth edition, Wellesley-Cambridge 2016). http://math.mit.edu/~gs/linearalgebra/
(2) T.S. Blyth and E.F. Robertson, Basic linear algebra (Springer, London, 1998).

Further Reading:
(3) Richard Kaye and Robert Wilson, Linear algebra (OUP, Oxford 1998), Chapters 1-5 and 8.
(4) Charles W. Curtis, Linear algebra - an introductory approach (Springer, London, Fourth edition, reprinted 1994).
(5) R. B. J. T. Allenby, Linear algebra (Arnold, London, 1995).
(6) D. A. Towers, A guide to linear algebra (Macmillan, Basingstoke, 1988).
(7) Seymour Lipschutz and Marc Lipson, Schaum's outline of linear algebra (McGraw Hill, New York & London, Fifth edition, 2013).

1. LINEAR SYSTEMS AND MATRICES

1.1 Systems of linear equations

Definition 1 *(a) By a* ***linear system****, or* ***linear system of equations****, we will mean a set of m simultaneous equations in n real variables $x_1, x_2, \ldots, x_n$ which are of the form*

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1; \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2; \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m, \end{array} \tag{1.1}$$

where a_{ij} and b_i are real constants.

(b) Any vector $(x_1, x_2, \ldots, x_n)$ which satisfies (1.1) is said to be a ***solution****; if the linear system has one or more solutions then it is said to be* ***consistent****. The* ***general solution*** *to the system is any description of all the solutions of the system. We will see, in due course, that such linear systems can have zero, one or infinitely many solutions.*

(c) We will often write the linear system (1.1) as the ***augmented matrix*** $(A \,|\, \mathbf{b})$ *where*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

For now, we won't consider a matrix (such as A) or vector (such as $\mathbf{b}$) to be anything more than an array of numbers.

Consider as a first example the following linear system of 3 equations in 3 variables.

Example 2 *Determine the solutions (if any) to the following equations.*

$$3x + y - 2z = -2; \qquad x + y + z = 2; \qquad 2x + 4y + z = 0.$$

Solution. We can substitute $z = 2 - x - y$ from the second equation into the first and third to find

$$\begin{aligned} 3x + y - 2(2 - x - y) = 5x + 3y - 4 = -2 \quad &\Longrightarrow \quad 5x + 3y = 2; \\ 2x + 4y + (2 - x - y) = x + 3y + 2 = 0 \quad &\Longrightarrow \quad x + 3y = -2. \end{aligned}$$

Subtracting the second of these equations from the first gives $4x = 4$ and so we see

$$x = 1, \qquad y = (-2 - x)/3 = -1, \qquad z = 2 - x - y = 2. \tag{1.2}$$

Thus there is a unique solution $(x, y, z) = (1, -1, 2)$. We can verify easily that this is indeed a solution (just to check that the system contains no contradictory information elsewhere that we haven't used). ■

Whilst we solved the above rigorously – we showed of necessity $(1, -1, 2)$ was the only possible solution and then verified it is a solution – our approach was a little *ad hoc*; at least, it's not hard to appreciate that if we were presented with 1969 equations in 2021 variables then we would need a much more systematic approach to treat them – or more likely we would need to be more methodical while programming our computers to determine any solutions for us. We introduce such a process called *row-reduction* here.

We first improve the notation, writing the system as an augmented matrix.

$$\left(\begin{array}{ccc|c} 3 & 1 & -2 & -2 \\ 1 & 1 & 1 & 2 \\ 2 & 4 & 1 & 0 \end{array}\right). \tag{1.3}$$

All that has been lost in this representation are the names of the variables, but these names are unchanging and unimportant in the actual handling of the equations. The advantages, we shall see, are that we will be able to progress systematically towards any solution and at each stage we shall retain all the information that the system contains – any redundancies (superfluous, unnecessary equations) or contradictions will naturally appear as part of the calculation.

This process is called *row-reduction.* It relies on three types of operation, called *elementary row operations* or *EROs*, which importantly do not affect the set of solutions of a linear system as we apply them.

Definition 3 *Given a linear system of equations, an* ***elementary row operation*** *or* ***ERO*** *is an operation of one of the following three kinds.*

(a) The ordering of two equations (or rows) may be swapped – for example, one might reorder the writing of the equations so that the first equation now appears third and vice versa.

(b) An equation may be multiplied by a non-zero scalar – for example, one might replace $2x - y + z = 3$ *by* $x - \frac{1}{2}y + \frac{1}{2}z = \frac{3}{2}$ *from multiplying both sides of the equation by* $\frac{1}{2}$.

(c) A multiple of one equation might be added to a different equation – for example, one might replace the second equation by the second equation plus twice the third equation.

Notation 4 *(a) Let* S_{ij} *denote the ERO which swaps rows* i *and* j *(or equivalently the* i*th and* j*th equations).*

(b) Let $M_i(\lambda)$ *denote the ERO which multiplies row* i *by* $\lambda \neq 0$ *(or equivalently both sides of the* i*th equation).*

(c) For $i \neq j$*, let* $A_{ij}(\lambda)$ *denote the ERO which adds* λ *times row* i *to row* j *(or does the same to the equations).*

Note this is not standard notation in any way, but I've introduced it here for convenience.

All these operations may well seem uncontroversial (their validity will be shown in Corollary 40) but it is probably not yet clear that these three simple operations are powerful enough to *reduce* any linear system to a point where any solutions can just be read off (Proposition 44, Theorem 47). Before treating the general case, we will see how the three equations in (1.3) can be solved using EROs to get an idea of the process.

Example 5 *Find all solutions of the linear system (1.3).*

Solution. If we use S_{12} to swap the first two rows the system becomes

$$\left(\begin{array}{ccc|c} 3 & 1 & -2 & -2 \\ 1 & 1 & 1 & 2 \\ 2 & 4 & 1 & 0 \end{array}\right) \xrightarrow{S_{12}} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 3 & 1 & -2 & -2 \\ 2 & 4 & 1 & 0 \end{array}\right).$$

Now subtract three times the first row from the second, i.e. $A_{12}(-3)$ and follow this by subtracting twice the first row from the third, i.e. $A_{13}(-2)$, so that

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 3 & 1 & -2 & -2 \\ 2 & 4 & 1 & 0 \end{array}\right) \xrightarrow{A_{12}(-3)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 2 & 4 & 1 & 0 \end{array}\right) \xrightarrow{A_{13}(-2)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 2 & -1 & -4 \end{array}\right). \quad (1.4)$$

We can now divide the second row by -2, i.e. $M_2(-1/2)$ to find

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 2 & -1 & -4 \end{array}\right) \xrightarrow{M_2(-1/2)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 2 & -1 & -4 \end{array}\right).$$

We then subtract the second row from the first, i.e. $A_{21}(-1)$, and follow this by subtracting twice the second row from the third, i.e. $A_{23}(-2)$, to obtain

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 2 & -1 & -4 \end{array}\right) \xrightarrow{A_{21}(-1)} \left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 2 & -1 & -4 \end{array}\right) \xrightarrow{A_{23}(-2)} \left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right). \quad (1.5)$$

If we divide the third row by -6, i.e. $M_3(-1/6)$, the system becomes

$$\left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right) \xrightarrow{M_3(-1/6)} \left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 0 & 1 & 2 \end{array}\right).$$

Finally, we subtract $2\frac{1}{2}$ times the third row from the second, i.e. $A_{32}(-2\frac{1}{2})$, and follow this by adding $1\frac{1}{2}$ times the third row to the first, i.e. $A_{31}(1\frac{1}{2})$.

$$\left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 2\frac{1}{2} & 4 \\ 0 & 0 & 1 & 2 \end{array}\right) \xrightarrow{A_{32}(-5/2)} \left(\begin{array}{ccc|c} 1 & 0 & -1\frac{1}{2} & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right) \xrightarrow{A_{31}(3/2)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right).$$

The rows of the final matrix represent the equations $x = 1$, $y = -1$, $z = 2$ as expected from (1.2). ■

Remark 6 *In case the systematic nature of the previous example isn't apparent, note that the first three operations S_{12}, $A_{12}(-3)$, $A_{13}(-2)$ were chosen so that the first column became $(1, 0, 0)^T$ in (1.4). There were many other ways to achieve this: for example, we could have begun with $M_1(1/3)$ to divide the first row by 3, then used $A_{12}(-1)$ and $A_{13}(-2)$ to clear the rest of the column. Once done, we then produced a similar leading entry of 1 in the second row with $M_2(-1/2)$ and used $A_{21}(-1)$ and $A_{23}(-2)$ to turn the second column into $(0, 1, 0)^T$ in (1.5). The final three EROs were chosen to transform the third column to $(0, 0, 1)^T$ at which point we could simply read off the solutions.*

Here are two slightly different examples, the first where we find that there are infinitely many solutions, whilst in the second example we see that there are no solutions.

Example 7 *Find the general solution of the following systems of equations in variables* x_1, x_2, x_3, x_4.

(a) $x_1 - x_2 + x_3 + 3x_4 = 2; \qquad 2x_1 - x_2 + x_3 + 2x_4 = 4; \qquad 4x_1 - 3x_2 + 3x_3 + 8x_4 = 8.$

(b) $x_1 + x_2 + x_3 + x_4 = 4; \qquad 2x_1 + 3x_2 - 2x_3 - 3x_4 = 1; \qquad x_1 + 5x_3 + 6x_4 = 1.$

Solution. (a) This time we will not spell out at quite so much length which EROs are being used. But we continue in a similar vein to the previous example and proceed by the method outlined in Remark 6.

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 3 & 2 \\ 2 & -1 & 1 & 2 & 4 \\ 4 & -3 & 3 & 8 & 8 \end{array}\right) \overset{\substack{A_{12}(-2)\\ A_{13}(-4)}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & -1 & 1 & 3 & 2 \\ 0 & 1 & -1 & -4 & 0 \\ 0 & 1 & -1 & -4 & 0 \end{array}\right) \overset{\substack{A_{21}(1)\\ A_{23}(-1)}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 2 \\ 0 & 1 & -1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

We have manipulated our system of three equations to two equations equivalent to the original system, namely

$$x_1 - x_4 = 2; \qquad x_2 - x_3 - 4x_4 = 0. \tag{1.6}$$

The presence of the zero row in the last matrix means that there was some redundancy in the system. Note, for example that the third equation can be deduced from the first two (it's the second equation added to twice the first) and so it provides no new information. As there are now only two equations in four variables, it's impossible for each column to contain a row's leading entry. In this example, the third and fourth columns lack such an entry. To describe all the solutions to a consistent system, we assign parameters to the columns/variables without leading entries. In this case that's x_3 and x_4 and we'll assign parameters by setting $x_3 = s$, $x_4 = t$, and then use the two equations in (1.6) to read off x_1 and x_2. So

$$x_1 = t + 2, \qquad x_2 = s + 4t, \qquad x_3 = s, \qquad x_4 = t, \tag{1.7}$$

or we could write

$$(x_1, x_2, x_3, x_4) = (t + 2, s + 4t, s, t) = (2, 0, 0, 0) + s(0, 1, 1, 0) + t(1, 4, 0, 1). \tag{1.8}$$

For each choice of s and t we have a solution as in (1.7) and this is one way of representing the general solution. (1.8) makes more apparent that these solutions form a plane in $\mathbb{R}^4$, a plane which passes through $(2, 0, 0, 0)$ is parallel to $(0, 1, 1, 0)$ and $(1, 4, 0, 1)$ with s, t parametrizing the plane.

(b) Applying EROs again in a like manner, we find

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 2 & 3 & -2 & -3 & 1 \\ 1 & 0 & 5 & 6 & 1 \end{array}\right) \overset{\substack{A_{12}(-2)\\ A_{13}(-1)}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 1 & -4 & -5 & -7 \\ 0 & -1 & 4 & 5 & -3 \end{array}\right)$$

$$\overset{A_{23}(1)}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 1 & -4 & -5 & -7 \\ 0 & 0 & 0 & 0 & -10 \end{array}\right) \overset{\substack{M_3(-1/10)\\ A_{21}(-1)}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 0 & -5 & -6 & -11 \\ 0 & 1 & -4 & -5 & -7 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right).$$

Note that any $\mathbf{x} = (x_1, x_2, x_3, x_4)$ which solves the final equation must satisfy

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 1.$$

There clearly are no such x_i and so there are no solutions to this equation. Any solution to the system has, in particular, to solve the third equation and so this system has no solutions. In fact, this was all apparent once the third row had become $\left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & -10 \end{array}\right)$ as the equation it represents is clearly insolvable also. The final two EROs were simply done to put the matrix into what is called *reduced row echelon form* (see Definition 41). ■

Examples 5, 7(a) and 7(b) are specific examples of the following general cases.

- *A linear system can have no, one or infinitely many solutions.*

We shall prove this in due course (Proposition 44). We finish our examples though with a linear system that involves a parameter – so really we have a family of linear systems, one for each value of that parameter. What EROs may be permissible at a given stage may well depend on the value of the parameter and so we may see (as below) that such a family can exhibit all three of the possible scenarios just described.

Example 8 *Consider the system of equations in x, y, z,*

$$x + z = -5; \qquad 2x + \alpha y + 3z = -9; \qquad -x - \alpha y + \alpha z = \alpha^2,$$

where α is a constant. For which values of α has the system one solution, none or infinitely many?

Solution. Writing this system in matrix form and applying EROs we can argue as follows.

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 2 & \alpha & 3 & -9 \\ -1 & -\alpha & \alpha & \alpha^2 \end{array}\right) \overset{\substack{A_{12}(-2)\\A_{13}(1)}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & -\alpha & \alpha+1 & \alpha^2-5 \end{array}\right) \overset{A_{23}(1)}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & 0 & \alpha+2 & \alpha^2-4 \end{array}\right). \tag{1.9}$$

At this point, which EROs are permissible depends on the value of α. We would like to divide the second equation by α and the third by $\alpha + 2$. Both these are permissible provided that $\alpha \neq 0$ and $\alpha \neq -2$. We will have to treat separately those particular cases but, assuming for now that $\alpha \neq 0, -2$, we obtain

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & 0 & \alpha+2 & \alpha^2-4 \end{array}\right) \overset{\substack{M_2(1/\alpha)\\M_3(1/(\alpha+2))}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 1 & 1/\alpha & 1/\alpha \\ 0 & 0 & 1 & \alpha-2 \end{array}\right) \overset{\substack{A_{31}(-1)\\A_{32}(-1/\alpha)}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -\alpha-3 \\ 0 & 1 & 0 & 3/\alpha-1 \\ 0 & 0 & 1 & \alpha-2 \end{array}\right)$$

and we see that the system has a unique solution when $\alpha \neq 0, -2$. Returning though to the last matrix of (1.9) for our two special cases, we would proceed as follows.

$$\alpha = 0\text{:} \qquad \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & -4 \end{array}\right) \overset{\substack{A_{23}(-2)\\M_3(-1/6)}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

$$\alpha = -2\text{:} \qquad \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \overset{M_2(-1/2)}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

We see then that the system is inconsistent when $\alpha = 0$ (because of the insolvability of the third equation) whilst there are infinitely many solutions $x = -5-t$, $y = (t-1)/2$, $z = t$, when $\alpha = -2$. We assign a parameter, here t, to the variable z as the third column has no leading entry. ■

Before we treat linear systems more generally, we will first need to discuss matrices and their algebra.

1.2 Matrices and matrix algebra

At its simplest, a *matrix* is just a two-dimensional array of numbers; for example

$$\begin{pmatrix} 1 & 2 & -3 \\ \sqrt{2} & \pi & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ -1.2 \\ -1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{1.10}$$

are all matrices. The examples above are respectively a 2×3 matrix, a 3×1 matrix and a 2×2 matrix (read '2 by 3' etc.); the first figure refers to the number of horizontal *rows* and the second to the number of vertical *columns* in the matrix. Row vectors in $\mathbb{R}^n$ are $1 \times n$ matrices and columns vectors in $\mathbb{R}^n_{\text{col}}$ are $n \times 1$ matrices.

Definition 9 *Let m, n be positive integers. An $m \times n$* ***matrix*** *is an array of real numbers arranged into m* ***rows*** *and n* ***columns****.*

Example 10 *Consider the first matrix above. Its second row is* $\begin{pmatrix} \sqrt{2} & \pi & 0 \end{pmatrix}$ *and its third column is* $\begin{pmatrix} -3 \\ 0 \end{pmatrix}$.

Definition 11 *The numbers in a matrix are its* ***entries****. Given an $m \times n$ matrix A, we will write a_{ij} for the entry in the ith row and jth column. Note that i can vary between 1 and m, and that j can vary between 1 and n. So*

$$i\text{th row} = (a_{i1}, \ldots, a_{in}) \qquad \text{and} \qquad j\text{th column} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Notation 12 *We shall denote the set of real $m \times n$ matrices as M_{mn}. Note that $M_{1n} = \mathbb{R}^n$ and that $M_{n1} = \mathbb{R}^n_{\text{col}}$.*

Example 13 *If we write A for the first matrix in (1.10) then we have $a_{23} = 0$ and $a_{12} = 2$.*

There are three important operations that can be performed with matrices: *matrix addition, scalar multiplication* and *matrix multiplication*. As with vectors, not all pairs of matrices can be meaningfully added or multiplied.

Definition 14 ***Addition*** *Let $A = (a_{ij})$ be an $m \times n$ matrix (recall: m rows and n columns) and $B = (b_{ij})$ be a $p \times q$ matrix. As with vectors, matrices are added by adding their corresponding entries. So, as with vectors, to add two matrices they have to be the same size – that is, to add A and B, we must have $m = p$ and $n = q$. If we write $C = A + B = (c_{ij})$ then*

$$c_{ij} = a_{ij} + b_{ij} \qquad \text{for } 1 \leqslant i \leqslant m \text{ and } 1 \leqslant j \leqslant n.$$

Example 15 *Let*

$$\underbrace{A = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}}_{2\times 2}, \qquad \underbrace{B = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}}_{2\times 3}, \qquad \underbrace{C = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}}. \tag{1.11}$$

Of the possible sums involving these matrices, only $A + C$ and $C + A$ make sense as B is a different size. Note that

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

Remark 16 *In general, matrix addition is* ***commutative*** *as for matrices M and N of the same size we have*

$$M + N = N + M.$$

Addition of matrices is also ***associative*** *as*

$$L + (M + N) = (L + M) + N$$

for any matrices of the same size.

Definition 17 *The $m \times n$* ***zero matrix*** *is the matrix with m rows and n columns whose every entry is 0. This matrix is simply denoted as 0 unless we need to specify its size, in which case it is written 0_{mn}. For example,*

$$0_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A simple check shows that $A + 0_{mn} = A = 0_{mn} + A$ for any $m \times n$ matrix A.

Definition 18 ***Scalar Multiplication*** *Let $A = (a_{ij})$ be an $m \times n$ matrix and k be a real number (a scalar). Then the matrix kA is defined to be the $m \times n$ matrix with (i,j)th entry equal to ka_{ij}.*

Example 19 *Show that $2(A + B) = 2A + 2B$ for the following matrices:*

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}; \qquad B = \begin{pmatrix} 0 & -2 \\ 5 & 1 \end{pmatrix}.$$

Solution. Here we are checking the **distributive law** in a specific example. We note that

$$A + B = \begin{pmatrix} 1 & 0 \\ 8 & 5 \end{pmatrix}, \qquad \text{and so} \qquad 2(A + B) = \begin{pmatrix} 2 & 0 \\ 16 & 10 \end{pmatrix};$$

$$2A = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}, \quad \text{and} \quad 2B = \begin{pmatrix} 0 & -4 \\ 10 & 2 \end{pmatrix}, \quad \text{so} \quad 2A + 2B = \begin{pmatrix} 2 & 0 \\ 16 & 10 \end{pmatrix}.$$

■

Remark 20 *More generally the following identities hold. Let A, B, C be $m \times n$ matrices and λ, μ be real numbers.*

$$\begin{array}{ccc} A + 0_{mn} = A; & A + B = B + A; & 0A = 0_{mn}; \\ A + (-A) = 0_{mn}; & (A + B) + C = A + (B + C); & 1A = A; \\ (\lambda + \mu)A = \lambda A + \mu A; & \lambda(A + B) = \lambda A + \lambda B; & \lambda(\mu A) = (\lambda\mu)A. \end{array}$$

These are readily verified and show that M_{mn} is a real vector space. ■

Based on how we added matrices then you might think that we multiply matrices in a similar fashion, namely multiplying corresponding entries, but we do not. At first glance the rule for multiplying matrices is going to seem rather odd but, in due course, we will see why matrix multiplication is done as follows and that this is natural in the context of matrices representing linear maps.

Definition 21 ***Matrix Multiplication*** *We can multiply an $m \times n$ matrix $A = (a_{ij})$ with an $p \times q$ matrix $B = (b_{ij})$ if $n = p$. That is, A must have as many columns as B has rows. If this is the case then the product $C = AB$ is the $m \times q$ matrix with entries*

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{for } 1 \leqslant i \leqslant m \text{ and } 1 \leqslant j \leqslant q. \tag{1.12}$$

It may help to write the rows of A as $\mathbf{r}_1, \ldots, \mathbf{r}_m$ and the columns of B as $\mathbf{c}_1, \ldots, \mathbf{c}_q$. Rule (1.12) then states that

$$\text{the } (i,j)\text{th entry of } AB = \mathbf{r}_i \cdot \mathbf{c}_j \qquad \text{for } 1 \leqslant i \leqslant m \text{ and } 1 \leqslant j \leqslant q. \tag{1.13}$$

We dot (i.e. take the scalar product of) the rows of A with the columns of B; specifically to find the (i,j)th entry of AB we dot the ith row of A with the jth column of B.

Remark 22 *We shall give full details later as to why it makes sense (and, in fact, is quite natural) to multiply matrices as in (1.12). For now, it is worth noting the following. Let A be an $m \times n$ matrix and B be $n \times p$ so that AB is $m \times p$. There is a map L_A from $\mathbb{R}^n_{\text{col}}$ to $\mathbb{R}^m_{\text{col}}$ associated with A, as given an $n \times 1$ column vector $\mathbf{v}$ in $\mathbb{R}^n_{\text{col}}$ then $A\mathbf{v}$ is a $m \times 1$ column vector in $\mathbb{R}^m_{\text{col}}$. (Here the L denotes that we are multiplying on the left or premultiplying.) So we have associated maps*

$$L_A \text{ from } \mathbb{R}^n_{\text{col}} \text{ to } \mathbb{R}^m_{\text{col}}, \qquad L_B \text{ from } \mathbb{R}^p_{\text{col}} \text{ to } \mathbb{R}^n_{\text{col}}, \qquad L_{AB} \text{ from } \mathbb{R}^p_{\text{col}} \text{ to } \mathbb{R}^m_{\text{col}}.$$

Multiplying matrices as we have, it turns out that

$$L_{AB} = L_A \circ L_B.$$

This is equivalent to $(AB)\mathbf{v} = A(B\mathbf{v})$ which follows from the associativity of matrix multiplication. So matrix multiplication is best thought of as composition: performing L_{AB} is equal to the performing L_B then L_A. ■

Example 23 *Calculate the possible products of the pairs of matrices in (1.11).*

Solution. Recall that a matrix product MN makes sense if M has the same number of columns as N has rows. A, B, C are respectively 2×2, 2×3, 2×2 matrices. so the products we can form are AA, AB, AC, CA, CB, CC. Let's slowly go through the product AC.

$$\begin{pmatrix} \boxed{\begin{array}{cc} 1 & 2 \end{array}} \\ \begin{array}{cc} -1 & 0 \end{array} \end{pmatrix} \begin{pmatrix} \boxed{\begin{array}{c} 1 \end{array}} & -1 \\ \boxed{\begin{array}{c} 1 \end{array}} & -1 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 2 \times 1 & ?? \\ ?? & ?? \end{pmatrix} = \begin{pmatrix} 3 & ?? \\ ?? & ?? \end{pmatrix}.$$

This is how we calculate the $(1,1)$th entry of AC. We take the first row of A and the first column of C and dot them together. We complete the remainder of the product as follows:

$$\begin{pmatrix} \boxed{\begin{array}{cc} 1 & 2 \end{array}} \\ \begin{array}{cc} -1 & 0 \end{array} \end{pmatrix} \begin{pmatrix} 1 & \boxed{\begin{array}{c} -1 \end{array}} \\ 1 & \boxed{\begin{array}{c} -1 \end{array}} \end{pmatrix} = \begin{pmatrix} 2 & \boxed{1 \times (-1) + 2 \times (-1)} \\ ?? & ?? \end{pmatrix} = \begin{pmatrix} 3 & \boxed{-3} \\ ?? & ?? \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 \\ \boxed{\begin{array}{c} -1 \end{array}} & \boxed{\begin{array}{c} 0 \end{array}} \end{pmatrix} \begin{pmatrix} \boxed{\begin{array}{c} 1 \end{array}} & -1 \\ \boxed{\begin{array}{c} 1 \end{array}} & -1 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ \boxed{(-1) \times 1 + 0 \times 1} & ?? \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ \boxed{-1} & ?? \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 \\ \boxed{\begin{array}{c} -1 \end{array}} & \boxed{\begin{array}{c} 0 \end{array}} \end{pmatrix} \begin{pmatrix} 1 & \boxed{\begin{array}{c} -1 \end{array}} \\ 1 & \boxed{\begin{array}{c} -1 \end{array}} \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & \boxed{(-1) \times (-1) + 0 \times (-1)} \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ -1 & \boxed{1} \end{pmatrix}.$$

So finally

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

We complete the remaining examples more quickly but still leaving a middle stage in the calculation to help see the process.

$$AA = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1-2 & 2+0 \\ -1+0 & -2+0 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -1 & -2 \end{pmatrix};$$

$$AB = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1+6 & 2+4 & 3+2 \\ -1+0 & -2+0 & -3+0 \end{pmatrix} = \begin{pmatrix} 7 & 6 & 5 \\ -1 & -2 & -3 \end{pmatrix};$$

$$CA = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1+1 & 2-0 \\ 1+1 & 2-0 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix};$$

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

$$CC = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1-1 & -1+1 \\ 1-1 & -1+1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

■

Definition 24 *The $n \times n$ **identity matrix** I_n is the $n \times n$ matrix with entries*

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

For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

*The identity matrix will be simply denoted as I unless we need to specify its size. The (i,j)th entry of I is denoted as δ_{ij} which is referred to as the **Kronecker delta.***

Remark 25 *(**Sifting Property of the Kronecker Delta**) Let $x_1, \ldots, x_n$ be n real numbers, and $1 \leqslant k \leqslant n$. Then*

$$\sum_{i=1}^{n} x_i \delta_{ik} = x_k.$$

This is because $\delta_{ik} = 0$ when $i \neq k$ and $\delta_{kk} = 1$. Thus the above sum sifts out (i.e. selects) the kth element x_k. ■

There are certain important points to highlight from Example 23, some of which make matrix algebra crucially different from the algebra of real numbers.

Proposition 26 *(**Properties of Matrix Multiplication**) (a) For an $m \times n$ matrix A and positive integers l, p,*

$$A0_{np} = 0_{mp}; \qquad 0_{lm}A = 0_{ln}; \qquad AI_n = A; \qquad I_m A = A.$$

*(b) Matrix multiplication is **not commutative**; $AB \neq BA$ in general, even if both products meaningfully exist and have the same size.*
*(c) Matrix multiplication is **associative**; for matrices A, B, C, which are respectively $m \times n$, $n \times p$ and $p \times q$ we have*

$$A(BC) = (AB)C.$$

*(d) The **distributive** laws hold for matrix multiplication; whenever the following products and sums make sense,*

$$A(B + C) = AB + AC, \qquad \text{and} \qquad (A + B)C = AC + BC.$$

*(e) In Example 23 we saw $CC = 0$ even though $C \neq 0$ – so one **cannot** conclude from $MN = 0$ that either matrix M or N is zero.*

Proof. (a) To find an entry of the product $A0_{np}$ we dot a row of A with a zero column of 0_{np} and likewise in the product $0_{lm}A$ we are dotting with zero rows. Also, by the sifting property,

$$\text{the } (i,j)\text{th entry of } AI_n = \sum\nolimits_{k=1}^{n} a_{ik}\delta_{kj} = a_{ij};$$
$$\text{the } (i,j)\text{th entry of } I_n A = \sum\nolimits_{k=1}^{n} \delta_{ik}a_{kj} = a_{ij}.$$

(b) In Example 23, we saw that $AC \neq CA$. More generally, if A is $m \times n$ and B is $n \times p$ then the product AB exists but BA doesn't even make sense as a matrix product unless $m = p$.
(c) Given i, j in the ranges $1 \leqslant i \leqslant m, 1 \leqslant j \leqslant q$, we see

$$\text{the } (i,j)\text{th entry of } (AB)C = \sum_{r=1}^{p} \left(\sum_{s=1}^{n} a_{is}b_{sr} \right) c_{rj};$$
$$\text{the } (i,j)\text{th entry of } A(BC) = \sum_{s=1}^{n} a_{is} \left(\sum_{r=1}^{p} b_{sr}c_{rj} \right).$$

These are equal as the order of finite sums may be swapped.
(d) This is left as an exercise. ■

Because matrix multiplication is not commutative, we need to be clearer than usual in what we might mean by a phrase like 'multiply by the matrix A'; typically we need to give some context as to whether we have multiplied on the left or on the right.

Definition 27 *Let A and M be matrices.*
(a) To ***premultiply*** *M by A is to form the product AM – i.e. premultiplication is multiplication on the left.*
(b) To ***postmultiply*** *M by A is to form the product MA – i.e. postmultiplication is multiplication on the right.*

Notation 28 *We write A^2 for the product AA and similarly, for n a positive integer, we write A^n for the product*

$$\underbrace{AA\cdots A}_{n \text{ times}}.$$

Note that A must be a square matrix for this to make sense. We also define $A^0 = I$. Note that $A^m A^n = A^{m+n}$ for natural numbers m, n. Given a polynomial $p(x) = a_k x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0$, then we define

$$p(A) = a_k A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I.$$

Example 29 *Let*

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{1.14}$$

Then $A^2 = I_2$ for any choice of α. Also there is no matrix C (with real or complex entries) such that $C^2 = B$. This shows that the idea of a square root is a much more complicated issue for matrices than for real or complex numbers. A square matrix may have none or many, even infinitely many, different square roots.

Solution. We note for any α that

$$A^2 = \begin{pmatrix} \cos^2\alpha + \sin^2\alpha & \cos\alpha\sin\alpha - \sin\alpha\cos\alpha \\ \sin\alpha\cos\alpha - \cos\alpha\sin\alpha & \sin^2\alpha + (-\cos\alpha)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

To show B has no square roots, say a, b, c, d are real (or complex) numbers such that

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & bc + d^2 \end{pmatrix}.$$

Looking at the $(2,1)$ entry, we see $c = 0$ or $a + d = 0$. But $a + d = 0$ contradicts $b(a + d) = 1$ from the $(1,2)$ entry and so $c = 0$. From the $(1,1)$ entry we see $a = 0$ and from the $(2,2)$ entry we see $d = 0$, but these lead to the same contradiction. ■

Let's look at a simple case of simultaneous equations: 2 linear equations in two variables, such as

$$ax + by = e; \qquad cx + dy = f. \tag{1.15}$$

Simple algebraic manipulations show that *typically* there is a unique solution (x, y) given by

$$x = \frac{de - bf}{ad - bc}; \qquad y = \frac{af - ce}{ad - bc}. \tag{1.16}$$

However if $ad - bc = 0$ then this solution is meaningless. It's probably easiest to appreciate geometrically why this is: the equations in (1.15) represent lines in the xy-plane with gradients $-a/b$ and $-c/d$ respectively, and hence the two lines are parallel if $ad - bc = 0$. (Notice that this is still the correct condition when $b = d = 0$ and the lines are parallel and vertical.) If the lines are parallel then there cannot be a unique solution.

We can represent the two scalar equations in (1.15) and (1.16) by a single vector equation in each case:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}; \tag{1.17}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} e \\ f \end{pmatrix}. \tag{1.18}$$

Equation (1.17) is just a rewriting of the linear system (1.15). Equation (1.18) is a similar rewriting of the unique solution found in (1.16) and something we *typically* can do. It also introduces us to the notion of the *inverse* of a matrix. Note that

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (ad - bc)I_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{1.19}$$

So if $ad - bc \neq 0$ and we set

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \text{and} \qquad B = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

then $BA = I_2$ and $AB = I_2$.

Definition 30 *Let A be a square matrix. We say that B is an **inverse** of A if $BA = AB = I$. We refer to a matrix with an inverse as **invertible** and otherwise the matrix is said to be **singular**.*

Proposition 31 ***(Properties of Inverses)***
*(a) (**Uniqueness**) If a square matrix A has an inverse, then it is unique. We write A^{-1} for this inverse.*
*(b) (**Product Rule**) If A, B are invertible $n \times n$ matrices then AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$.*
*(c) (**Involution Rule**) If A is invertible then so is A^{-1} with $(A^{-1})^{-1} = A$.*

Proof. (a) Suppose B and C were two inverses for an $n \times n$ matrix A then

$$C = I_n C = (BA)C = B(AC) = BI_n = B$$

as matrix multiplication is associative. Part (b) is left as Sheet 1, Exercise S3. To verify (c) note that

$$\left(A^{-1}\right) A = A(A^{-1}) = I$$

and so $(A^{-1})^{-1} = A$ by uniqueness. ■

Definition 32 *If A is $m \times n$ and $BA = I_n$ then B is said to be a **left inverse**; if C satisfies $AC = I_m$ then C is said to be a **right inverse**.*

- If A is $m \times n$ where $m \neq n$ then A cannot have both left and right inverses. (This is non-trivial. We will prove this later.)
- If A, B are $n \times n$ matrices with $BA = I_n$ then, in fact, $AB = I_n$ (Proposition 168).

Inverses, in the 2×2 case, are a rather simple matter to deal with.

Proposition 33 *The matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an inverse if and only if $ad - bc \neq 0$. If $ad - bc \neq 0$ then*

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Remark 34 *The scalar quantity $ad - bc$ is called the determinant of A, written $\det A$. It is a non-trivial fact to show that a square matrix is invertible if and only if its determinant is non-zero. This will be proved in Linear Algebra II next term.* ■

Proof. We have already seen in (1.19) that if $ad - bc \neq 0$ then $AA^{-1} = I_2 = A^{-1}A$. If however $ad - bc = 0$ then

$$B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

satisfies $BA = 0$. If an inverse C for A existed then, by associativity, $0 = 0C = (BA)C = B(AC) = BI_2 = B$. So each of a, b, c and d would be zero and consequently $A = 0$ which contradicts $AC = I_2$. ■

We conclude this section with the following theorem. The proof demonstrates the power of the sigma-notation for matrix multiplication introduced in (1.12) and that of the Kronecker delta. In this proof we will make use of the *standard basis for matrices.*

Notation 35 *For I, J in the range $1 \leqslant I \leqslant m$, $1 \leqslant J \leqslant n$, we denote by E_{IJ} the $m \times n$ matrix with entry 1 in the Ith row and Jth column and 0s elsewhere. Then*

$$\textit{the } (i,j) \textit{ th entry of } E_{IJ} = \delta_{Ii}\delta_{Jj}$$

*as $\delta_{Ii}\delta_{Jj} = 0$ unless $i = I$ and $j = J$ in which case it is 1. These matrices form the **standard basis** for M_{mn}.*

Theorem 36 *Let A be an $n \times n$ matrix such that $AM = MA$ for all $n \times n$ matrices M. i.e. A commutes with all $n \times n$ matrices. Then $A = \lambda I_n$ for some real number λ.*

Proof. As A commutes with every $n \times n$ matrix, then in particular it commutes with each of the n^2 basis matrices E_{IJ}. So the (i,j)th entry of AE_{IJ} equals that of $E_{IJ}A$ for every I, J, i, j. Using the sifting property

$$\text{the } (i,j)\text{ th entry of } AE_{IJ} = \sum_{k=1}^{n} a_{ik}\delta_{Ik}\delta_{Jj} = a_{iI}\delta_{Jj};$$
$$\text{the } (i,j)\text{ th entry of } E_{IJ}A = \sum_{k=1}^{n} \delta_{Ii}\delta_{Jk}a_{kj} = \delta_{Ii}a_{Jj}.$$

Hence for all I, J, i, j,

$$a_{iI}\delta_{Jj} = \delta_{Ii}a_{Jj}. \tag{1.20}$$

Let $i \neq j$. If we set $I = J = i$, then (1.20) becomes $0 = a_{ij}$ showing that the non-diagonal entries of A are zero. If we set $I = i$ and $J = j$, then (1.20) becomes $a_{ii} = a_{jj}$, which shows that all the diagonal entries of A are equal – call this shared value λ and we have shown $A = \lambda I_n$. This shows that any such M is necessarily of the form λI_n, and conversely such matrices do indeed commute with every other $n \times n$ matrix. ■

1.3 Reduced Row Echelon Form

Now looking to treat linear systems more generally, we will first show that the set of solutions of a linear system does not change under the application of EROs. We shall see that applying any ERO to a linear system $(A|\mathbf{b})$ is equivalent to premultiplying by an invertible *elementary* matrix E to obtain $(EA|E\mathbf{b})$, and it is the invertibility of elementary matrices that means the set of solutions remains unchanged when we apply EROs.

Proposition 37 ***(Elementary Matrices)*** *Let A be an $m \times n$ matrix. Applying any of the EROs S_{IJ}, $M_I(\lambda)$ and $A_{IJ}(\lambda)$ is equivalent to pre-multiplying A by certain matrices which we also denote as S_{IJ}, $M_I(\lambda)$ and $A_{IJ}(\lambda)$. Specifically these matrices have entries*

$$\textit{the } (i,j)\textit{th entry of } S_{IJ} = \begin{cases} 1 & i = j \neq I, J, \\ 1 & i = J,\ j = I, \\ 1 & i = I,\ j = J, \\ 0 & \textit{otherwise.} \end{cases}$$
$$\textit{the } (i,j)\textit{th entry of } M_I(\lambda) = \begin{cases} 1 & i = j \neq I, \\ \lambda & i = j = I, \\ 0 & \textit{otherwise.} \end{cases}$$
$$\textit{the } (i,j)\textit{th entry of } A_{IJ}(\lambda) = \begin{cases} 1 & i = j, \\ \lambda & i = J,\ j = I, \\ 0 & \textit{otherwise} \end{cases}$$

The above matrices are known as ***elementary matrices****.*

Proof. The proof is left as an exercise. ■

Example 38 *When $m = 3$ we see*

$$S_{21} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_3(7) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{pmatrix}, \qquad A_{31}(-2) = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that these elementary matrices are the results of performing the corresponding EROs $S_{21}, M_3(7), A_{31}(-2)$ on the identity matrix I_3. This is generally true of elementary matrices.

Proposition 39 *Elementary matrices are invertible.*

Proof. This follows from noting that

$$(S_{ij})^{-1} = S_{ji} = S_{ij}; \qquad (A_{ij}(\lambda))^{-1} = A_{ij}(-\lambda); \qquad (M_i(\lambda))^{-1} = M_i(\lambda^{-1}),$$

whether considered as EROs or their corresponding matrices. ■

Corollary 40 *(**Invariance of Solution Space under EROs**) Let $(A|\mathbf{b})$ be a linear system of m equations and E an elementary $m \times m$ matrix. Then $\mathbf{x}$ is a solution of $(A|\mathbf{b})$ if and only if $\mathbf{x}$ is a solution of $(EA|E\mathbf{b})$.*

Proof. The important point here is that E is invertible. So if $A\mathbf{x} = \mathbf{b}$ then $EA\mathbf{x} = E\mathbf{b}$ follows by premultiplying by E. But likewise if $EA\mathbf{x} = E\mathbf{b}$ is true then it follows that $A\mathbf{x} = \mathbf{b}$ by premultiplying by E^{-1}. ■

So applying an ERO, or any succession of EROs, won't alter the set of solutions of a linear system. The next key result is that, systematically using EROs, it is possible to reduce any system $(A|\mathbf{b})$ to *reduced row echelon form*. Once in this form it is simple to read off the system's solutions.

Definition 41 *A matrix A is said to be in **reduced row echelon form** (or simply **RRE form**) if*

(a) the first non-zero entry of any non-zero row is 1;

(b) in a column that contains such a leading 1, all other entries are zero;

(c) the leading 1 of a non-zero row appears to the right of the leading 1s of the rows above it;

(d) any zero rows appear below the non-zero rows.

Definition 42 *The process of applying EROs to transform a matrix into RRE form is called **row-reduction**, or just simply **reduction.** It is also commonly referred to as **Gauss-Jordan elimination**.*

Example 43 *Of the following matrices*

$$\begin{pmatrix} 0 & 1 & 2 & 0 & -4 \\ 0 & 0 & 0 & 1 & \pi \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & \sqrt{2} & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 & \sqrt{3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix},$$

the first three are in RRE form. The fourth is not as the second column contains a leading 1 *but not all other entries of that column are* 0*. The fifth matrix is not in RRE form as the leading entry of the third row is not* 1*.*

We have yet to show that any matrix can be uniquely put into RRE form using EROs (Theorem 122) but – as we have already seen examples covering the range of possibilities – it seems timely to prove the following result here.

Proposition 44 ***(Solving Systems in RRE Form)*** *Let* $(A|\mathbf{b})$ *be a matrix in RRE form which represents a linear system* $A\mathbf{x} = \mathbf{b}$ *of* m *equations in* n *variables. Then*

(a) the system has no solutions if and only if the last non-zero row of $(A|\mathbf{b})$ *is*

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

(b) the system has a unique solution if and only if the non-zero rows of A *form the identity matrix* I_n*. In particular, this case is only possible if* $m \geqslant n$*.*

(c) the system has infinitely many solutions if $(A|\mathbf{b})$ *has as many non-zero rows as* A*, and not every column of* A *contains a leading* 1*. The set of solutions can be described with* k *parameters where* k *is the number of columns not containing a leading* 1*.*

Proof. If $(A|\mathbf{b})$ contains the row $\left(\begin{array}{ccccc|c} 0 & 0 & \cdots & 0 & 1 \end{array}\right)$ then the system is certainly inconsistent as no $\mathbf{x}$ satisfies the equation

$$0x_1 + 0x_2 + \cdots + 0x_n = 1.$$

As $(A|\mathbf{b})$ is in RRE form, then this is the only way in which $(A|\mathbf{b})$ can have more non-zero rows than A. We will show that whenever $(A|\mathbf{b})$ has as many non-zero rows as A then the system $(A|\mathbf{b})$ is consistent.

Say, then, that both $(A|\mathbf{b})$ and A have r non-zero rows, so there are r leading 1s within these rows and we have $k = n - r$ columns without leading 1s. By reordering the numbering of the variables $x_1, \ldots, x_n$ if necessary, we can assume that the leading 1s appear in the first r columns. So, ignoring any zero rows, and remembering the system is in RRE form, the system now reads as the r equations:

$$x_1 + a_{1(r+1)}x_{r+1} + \cdots + a_{1n}x_n = b_1; \qquad \cdots \qquad x_r + a_{r(r+1)}x_{r+1} + \cdots + a_{rn}x_n = b_r.$$

We can see that if we assign $x_{r+1}, \ldots, x_n$ the k parameters $s_{r+1}, \ldots, s_n$, then we can read off from the r equations the values for $x_1, \ldots, x_r$. So for any values of the parameters we have a solution $\mathbf{x}$. Conversely though if $\mathbf{x} = (x_1, \ldots, x_n)$ is a solution, then it appears amongst the

solutions we've just found when we assign values $s_{r+1} = x_{r+1}, \ldots, s_n = x_n$ to the parameters. We see that we have an infinite set of solutions associated with $k = n-r$ independent parameters when $n > r$ and a unique solution when $r = n$, in which case the non-zero rows of A are the matrix I_n. ■

Remark 45 *Note we showed in this proof that*

- *a system $(A|\mathbf{b})$ in RRE form is consistent if and only if $(A|\mathbf{b})$ has as many non-zero rows as A;*
- *all the solutions of a consistent system can be found by assigning parameters to the variables corresponding to the columns without leading 1s.* ■

Example 46

$$\underset{\text{no solutions}}{\left(\begin{array}{cccc|c} 1 & -2 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)}, \quad \underset{\text{unique solution}}{\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)},$$

$$\underset{\substack{\text{one parameter family of solutions} \\ (3-2s,s,2,1)}}{\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)}, \quad \underset{\substack{\text{two parameter family of solutions} \\ (3+2s-2t,s,-2-t,t)}}{\left(\begin{array}{cccc|c} 1 & -2 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)}.$$

Theorem 47 ***(Existence of RRE Form)***
Every $m \times n$ matrix A can be reduced by EROs to a matrix in RRE form.

Proof. Note that a $1 \times n$ matrix is either zero or can be put into RRE form by dividing by its leading entry. Suppose, as our inductive hypothesis, that any matrix with fewer than m rows can be transformed with EROs into RRE form. Let A be an $m \times n$ matrix. If A is the zero matrix, then it is already in RRE form. Otherwise there is a first column $\mathbf{c}_j$ which contains a non-zero element α. With an ERO we can swap the row containing α with the first row and then divide the first row by $\alpha \neq 0$ so that the $(1, j)$th entry now equals 1. Our matrix now takes the form

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \tilde{a}_{1(j+1)} & \ldots & \tilde{a}_{1n} \\ 0 & \cdots & 0 & \tilde{a}_{2j} & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \tilde{a}_{mj} & \tilde{a}_{m(j+1)} & \cdots & \tilde{a}_{mn} \end{pmatrix},$$

for some new entries $\tilde{a}_{1(j+1)}, \ldots, \tilde{a}_{mn}$. Applying consecutively $A_{12}(-\tilde{a}_{2j})$, $A_{13}(-\tilde{a}_{3j}), \ldots, A_{1m}(-\tilde{a}_{mj})$ leaves column $\mathbf{c}_j = \mathbf{e}_1^T$ so that our matrix has become

$$\left(\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & \tilde{a}_{1(j+1)} & \ldots & \tilde{a}_{1n} \\ 0 & \cdots & 0 & 0 & \multicolumn{3}{c}{\multirow{3}{*}{B}} \\ \vdots & \cdots & \vdots & \vdots & & & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right).$$

By induction, the $(m-1)\times(n-j)$ matrix B can be put into some RRE form by means of EROs. Applying these same EROs to the bottom $m-1$ rows of the above matrix we would have reduced A to

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \tilde{a}_{1(j+1)} & \cdots & \tilde{a}_{1n} \\ 0 & \cdots & 0 & 0 & & & \\ \vdots & \cdots & \vdots & \vdots & & \mathrm{RRE}(B) & \\ 0 & \cdots & 0 & 0 & & & \end{pmatrix}.$$

To get the above matrix into RRE form we need to make zero any of $\tilde{a}_{1(j+1)},\ldots,\tilde{a}_{1n}$ which are above a leading 1 in $\mathrm{RRE}(B)$; if $\tilde{a}_{1k}$ is the first such entry to lie above a leading 1 in row l then $A_{l1}(-\tilde{a}_{1k})$ will make the required edit and in due course we will have transformed A into RRE form. The result follows by induction. ■

2. INVERSES AND TRANSPOSES

Definition 48 *A **square** matrix is a matrix with an equal number of rows and columns. The **diagonal** of an $n \times n$ matrix A comprises the entries $a_{11}, a_{22}, \ldots, a_{nn}$ – that is, the n entries running diagonally from the top left to the bottom right. A **diagonal** matrix is a square matrix whose non-diagonal entries are all zero. We shall write* $\mathrm{diag}(c_1, c_2, \ldots, c_n)$ *for the $n \times n$ diagonal matrix whose (i,i)th entry is c_i.*

Definition 49 *Given an $m \times n$ matrix A, then its **transpose** A^T is the $n \times m$ matrix such that the (i,j)th entry of A^T is the (j,i)th entry of A.*

Proposition 50 *(**Properties of Transpose**)*
*(a) (**Addition and Scalar Multiplication Rules**) Let A, B be $m \times n$ matrices and λ a real number. Then*

$$(A+B)^T = A^T + B^T; \qquad (\lambda A)^T = \lambda A^T.$$

*(b) (**Product Rule**) Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. Then $(AB)^T = B^T A^T$.*
*(c) (**Involution Rule**) Let A be an $m \times n$ matrix. Then $(A^T)^T = A$.*
*(c) (**Inverse Rule**) A square matrix A is invertible if and only if A^T is invertible. In this case $(A^T)^{-1} = (A^{-1})^T$.*

Proof. These are left to Sheet 2, Exercise 3. ■

Definition 51 *A square matrix $A = (a_{ij})$ is said to be*

- ***symmetric** if $A^T = A$.*
- ***skew-symmetric** (or **antisymmetric**) if $A^T = -A$.*
- ***upper triangular** if $a_{ij} = 0$ when $i > j$. Entries below the diagonal are zero.*
- ***strictly upper triangular** if $a_{ij} = 0$ when $i \geqslant j$. Entries on or below the diagonal are zero.*
- ***lower triangular** if $a_{ij} = 0$ when $i < j$. Entries above the diagonal are zero.*
- ***strictly lower triangular** if $a_{ij} = 0$ when $i \leqslant j$. Entries on or above the diagonal are zero.*
- ***triangular** if it is either upper or lower triangular.*

Example 52 *Let*

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & -1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Then

$$A^T = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}, \qquad B^T = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & -1 \end{pmatrix}, \qquad C^T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad D^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Note that A is upper triangular and so A^T is lower triangular. Also C and C^T are skew-symmetric. And D is diagonal and so also symmetric, upper triangular and lower triangular.

We return now to the issue of determining the invertibility of a square matrix. There is no neat expression for the inverse of an $n \times n$ matrix in general – we have seen that the $n = 2$ case is easy enough (Proposition 33) though the $n = 3$ case is already messy – but the following method shows how to determine efficiently, using EROs, whether an $n \times n$ matrix is invertible and, in such a case, how to find the inverse.

Algorithm 53 *(**Determining Invertibility**) Let A be an $n \times n$ matrix. Place A side-by-side with I_n as an augmented $n \times 2n$ matrix $(A \,|\, I_n)$. There are EROs that will reduce A to a matrix R in RRE form. We will simultaneously apply these EROs to both sides of $(A \,|\, I_n)$ until we arrive at $(R \,|\, P)$.*

- *If $R = I_n$ then A is invertible and $P = A^{-1}$.*
- *If $R \neq I_n$ then A is singular.*

Proof. Denote the elementary matrices representing the EROs that reduce A as $E_1, E_2, \ldots, E_k$, so that $(A \,|\, I_n)$ becomes

$$(E_k E_{k-1} \cdots E_1 A | E_k E_{k-1} \cdots E_1) = (R \,|\, P) \tag{2.1}$$

and we see that $R = PA$ and $E_k E_{k-1} \cdots E_1 = P$. If $R = I_n$ then

$$(E_k E_{k-1} \cdots E_1)A = I_n \quad \Longrightarrow \quad A^{-1} = E_k E_{k-1} \cdots E_1 = P$$

as elementary matrices are (left and right) invertible. If $R \neq I_n$ then, as R is in RRE form and square, R must have at least one zero row. It follows that $(1, 0, \ldots, 0)(PA) = \mathbf{0}$. As P is invertible, if A were also invertible, we could postmultiply by $A^{-1}P^{-1}$ to conclude $(1, 0, \ldots, 0) = \mathbf{0}$, a contradiction. Hence A is singular; indeed we can see from this proof that as soon as a zero row appears when reducing A then we know that A is singular. ■

Example 54 *Determine whether the following matrices are invertible, finding any inverses that exist.*

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 3 & -1 & 0 \\ 0 & 2 & 1 & 1 \\ 3 & 1 & 2 & 1 \\ 0 & 1 & 5 & 3 \end{pmatrix}.$$

Solution. Quickly applying a sequence of EROs leads to

$$(A|I_3) \overset{\substack{A_{12}(-2)\\A_{13}(-1)}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -3 & -2 & -2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

$$\overset{\substack{A_{31}(-2)\\A_{32}(3)\\S_{23}}}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & 0 & -2 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & -5 & 1 & 3 \end{array}\right) \overset{\substack{M_3(-1/2)\\A_{31}(-1)}}{\longrightarrow} \left(\begin{array}{ccc|ccc} & & & 1/2 & 1/2 & -1/2 \\ & I_3 & & -1 & 0 & 1 \\ & & & 5/2 & -1/2 & -3/2 \end{array}\right).$$

Hence

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

For B we note

$$(B|I_4) \overset{\substack{A_{13}(-3)\\S_{24}}}{\longrightarrow} \left(\begin{array}{cccc|cccc} 1 & 3 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & -8 & 5 & 1 & -3 & 0 & 1 & 0 \\ 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \end{array}\right)$$

$$\overset{\substack{A_{23}(8)\\A_{24}(-2)\\A_{34}(1/5)}}{\longrightarrow} \left(\begin{array}{cccc|cccc} 1 & 3 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 45 & 25 & -3 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 & -3/5 & 1 & 1/5 & -2/5 \end{array}\right).$$

The left matrix is not yet in RRE form, but the presence of a zero row is sufficient to show that B is singular. ■

Remark 55 *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.*

Finally we define the *orthogonal matrices,* matrices important to the geometry of $\mathbb{R}^n$.

Definition 56 *An $n \times n$ matrix is* ***orthogonal*** *if $A^T = A^{-1}$.*

Proposition 57 *Let A and B be orthogonal $n \times n$ matrices. Then:*

(a) AB and A^{-1} are othogonal. Consequently the $n \times n$ orthogonal matrices form a group $O(n)$.

(b) A is orthogonal if and only if its columns (or rows) are n unit length, mutually perpendicular vectors.

(c) A preserves the dot product. If $\mathbf{x}$,$\mathbf{y} \in \mathbb{R}^n_{\text{col}}$ then $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

Proof. (a) We have that

$$(AB)^T = B^T A^T = B^{-1}A^{-1} = (AB)^{-1},$$

by the product rules for transposes and inverses, showing that AB is orthogonal. Similarly

$$\left(A^{-1}\right)^T = \left(A^T\right)^{-1} = \left(A^{-1}\right)^{-1},$$

showing that A^{-1} is orthogonal. ■

The reason that orthogonal matrices are important in geometry is that the orthogonal matrices are precisely those matrices that preserve the dot product.

Proposition 58 *Let A be an $n \times n$ matrix. Then A is orthogonal if and only if $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n_{\text{col}}$.*

Proof. Note that $\mathbf{x}^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$. So

$$\begin{aligned} & A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y} \\ \iff \quad & (A\mathbf{x})^T A\mathbf{y} = \mathbf{x}^T\mathbf{y} \\ \iff \quad & \mathbf{x}^T A^T A\mathbf{y} = \mathbf{x}^T\mathbf{y}. \end{aligned}$$

If $A^TA = I_n$ then the above is clearly true. Conversely, by choosing $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$, from the standard basis of $\mathbb{R}^n_{\text{col}}$ the above implies

$$\text{the } (i,j)\,\text{th entry of } A^TA = \delta_{ij} = \text{the } (i,j)\,\text{th entry of } I_n.$$

As this is true for all i, j then this implies $A^TA = I_n$. ■

3. VECTOR SPACES

Currently when you speak of vectors, you usually mean *coordinate vectors* represented either as a row vector in some $\mathbb{R}^n$ or as a column vector in some $\mathbb{R}^n_{\text{col}}$. But vectors exist without reference to coordinate systems. Wherever you are at the moment, look around you and choose some point near you and label it P, then pick a second point and label it Q. Then $\overrightarrow{PQ}$ is a vector. If you want to treat P as the origin then $\overrightarrow{PQ}$ is the position vector of Q. Or you might think of $\overrightarrow{PQ}$ as a movement and any parallel movement, with the same length and direction, equals the vector $\overrightarrow{PQ}$. Importantly though $\overrightarrow{PQ}$ has no coordinates, or at least doesn't until you make a choice of origin and axes. This is going to be an important aspect of the Linear Algebra I and II courses, namely choosing coordinates sensibly. This will also be an important aspect of the Geometry and Dynamics courses – in Geometry the change between two coordinate systems will need to be an isometry so that the lengths, areas, angles are measured to be the same; in Dynamics an inertial frame would be necessary for Newton's laws to hold and otherwise so-called 'fictitious forces' will arise.

But the vector spaces we will introduce are not just geometrical vectors like these coordinate or coordinateless vectors. A vector space's elements might contain functions, sequences, matrices, equations or, of course, vectors. Importantly, these more abstract vector spaces do have the same algebraic operations in common with the vectors familiar to you: namely, *addition* and *scalar multiplication*.

3.1 What is a vector space?

A real vector space is a non-empty set with operations of addition and scalar multiplication. Formally this means:

Definition 59 *A **real vector space** is a non-empty set V together with a binary operation $V \times V \to V$ given by $(u, v) \mapsto u + v$ (called **addition**) and a map $\mathbb{R} \times V \to V$ 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, $v \in V$, $\lambda \in \mathbb{R}$ (***distributivity*** *of scalar multiplication over vector addition);*
- $(\lambda+\mu)v = \lambda v + \mu v$ *for all* $v \in V$, λ, $\mu \in \mathbb{R}$ (***distributivity*** *of scalar multiplication over field addition);*
- $(\lambda\mu)v = \lambda(\mu v)$ *for all* $v \in V$, λ, $\mu \in \mathbb{R}$ *(scalar multiplication interacts well with field multiplication);*
- $1v = v$ *for all* $v \in V$ *(identity for scalar multiplication).*

$\mathbb{R}$ *is referred to as the* ***field of scalars*** *or* ***base field****. Elements of* V *are called* ***vectors*** *and elements of* $\mathbb{R}$ *are called* ***scalars****.*

Remark 60 *There are a lot of axioms on the above list, but the most important in practice are those requiring:*

- V *has a zero vector* 0_V*.*
- V *is closed under addition.*
- V *is closed under scalar multiplication.*

If these three axioms hold, and addition and scalar multiplication are defined naturally, then usually the remaining axioms will follow as a matter of routine checks.

The subsets of $\mathbb{R}^3$ *that are real vector spaces are the origin, lines through the origin, planes through the origin and all of* $\mathbb{R}^3$*. It's perhaps not surprising then that another term for a vector space is a 'linear space'.*

Example 61 *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:*

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_y) \tag{3.1}$$

$$\text{and } \lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n). \tag{3.2}$$

These satisfy the vector space axioms. The zero vector is $(0, 0, \ldots, 0)$ *and the additive inverse of* $(v_1, \ldots, v_n)$ *is* $(-v_1, \ldots, -v_n)$*.*

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* $\mathbb{R}$*.*

Notation 62 *I will often denote a single coordinate vector* $(v_1, v_2, \ldots, v_n)$ *as* $\mathbf{v}$*. I will use this bold notation for coordinate vectors, but vectors, as elements of a vector space, will not be written in bold.*

Example 63 *The field* $\mathbb{C}$ *is a real vector space, it is essentially the same as* $\mathbb{R}^2$ *as a vector space. (The technical term for 'essentially the same' is 'isomorphic'. More on this later.)*

Example 64 *For $m, n \geqslant 1$, the set $\mathcal{M}_{m\times n}(\mathbb{R})$ is a real vector space as previously stated in Remark 20.*

Example 65 *Let $V = \mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}$, the space of all real-valued functions on $\mathbb{R}$, with addition and scalar multiplication defined pointwise. That is,*

$$(f+g)(r) := f(r) + g(r), \qquad (\alpha f)(r) := \alpha f(r),$$

for $f, g \in V$ and $\alpha, r \in \mathbb{R}$.

Example 66 *Let $V = \{f : \mathbb{R} \to \mathbb{R}, f \text{ is differentiable}\}$, the space of all differentiable real-valued functions on $\mathbb{R}$, with addition and scalar multiplication defined pointwise. This is a vector space as, in particular,*

$$(f+g)' = f' + g', \qquad (\alpha f)' = \alpha f',$$

for $f, g \in V$ and $\alpha \in \mathbb{R}$, so that the sum of two differentiable functions and the scalar multiple of a differentiable function is differentiable. It's these facts that mean V is closed under addition and scalar multiplication.

Example 67 *Let $V = \{f : \mathbb{R} \to \mathbb{R}, f'' = f\}$. This is a vector space mainly as $f'' = f$ is a **linear** differential equation. That is, if f and g are solutions then so is $\alpha f + \beta g$ where α, β are real scalars. The general solution can be written as*

$$f(x) = A\cosh x + B\sinh x$$

or as

$$f(x) = Ae^x + Be^{-x}.$$

In expressing the general vector in this way note that in each case we are 'coordinatizing the space' and identifying V with $\mathbb{R}^2$. But note that the coordinate vector $(1, 0)$ corresponds to different vectors as we are using different choices of coordinates; it corresponds to the vector $\cosh x$ in the first case and to e^x in the second.

Example 68 *Let $V = \mathbb{R}^{\mathbb{N}} = \{(x_0, x_1, x_2, ...) : x_i \in \mathbb{R}\}$. This is the space of all real sequences with addition and scalar multiplication defined componentwise.*

Other important sequence spaces are

$$\begin{aligned} l^\infty &= \{(x_n)_{n=0}^\infty : (x_n) \ \textit{is bounded}\}. \\ c &= \{(x_n)_{n=0}^\infty : (x_n) \ \textit{converges}\}. \\ c_0 &= \{(x_n)_{n=0}^\infty : (x_n) \ \textit{converges to } 0\}. \end{aligned}$$

As an exercise, what theorems of analysis (concerning convergence) need to hold for these last three examples all to be vector spaces?

Our main focus in this course will be real vector spaces. However, vector spaces can be defined over any field, as can simultaneous equations be considered over any field. Formally a vector space V over a field $\mathbb{F}$ is a non-zero set V with addition $V \times V \to V$ and scalar multiplication $\mathbb{F} \times V \to V$ satisfying the vector space axioms in Definition 59. Common examples of other fields that we will encounter are:

- $\mathbb{C}$ – the field of complex numbers.
- $\mathbb{Q}$ – the field of rational numbers.
- $\mathbb{Z}_p$ – the field of integers modulo a prime number p.

The theory of vector spaces applies equally well for all fields. There can be some differences worth noting though depending on the choice of field.

- A non-zero real vector space is an infinite set. This need not be the case over a finite field like $\mathbb{Z}_p$.
- When we consider $\mathbb{C}$ as a vector space over $\mathbb{R}$, then every z can be uniquely written as $x1 + yi$ for two real scalars x and y. But when $\mathbb{C}$ is considered as a vector space over $\mathbb{C}$, then every z can be uniquely written as $z1$ for a single complex scalar z. (In due course we will appreciate that $\mathbb{C}$ is a 2-dimensional real vector space and a 1-dimensional complex vector space.

Lemma 69 *Let V be a vector space over $\mathbb{F}$. Then there is a unique additive identity element 0_V.*

Proof. Suppose that 0 and $0'$ are two elements that have the properties of 0_V. Then

$$\begin{aligned} 0 &= 0 + 0' \qquad [\text{as } 0' \text{ is a zero vector}] \\ &= 0' \qquad [\text{as } 0 \text{ is a zero vector}] \end{aligned}$$

and so $0 = 0'$, thus showing 0_V to be unique. ■

Remark 70 *Where it will not be ambiguous, we often write 0 for 0_V.*

Lemma 71 *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. With the notation introduced above we have

$$\begin{aligned} w_2 &= 0 + w_2 \qquad [\text{as } 0 \text{ is a zero vector}] \\ &= (w_1 + v) + w_2 \qquad [\text{by a hypothesis}] \\ &= w_1 + (v + w_2) \qquad [\text{by associativity}] \\ &= w_1 + 0 \qquad [\text{by a hypothesis}] \\ &= w_1. \qquad [\text{as } 0 \text{ is a zero vector}] \end{aligned}$$

■

Remark 72 *Using the notation of Lemma 71, we write $-v$ for the unique additive inverse of v.*

Proposition 73 *Let V be a vector space over a field $\mathbb{F}$. Take $v \in V$, $\lambda \in \mathbb{F}$. Then*
(a) $\lambda 0_V = 0_V$*;*
(b) $0v = 0_V$*;*
(c) $(-\lambda)v = -(\lambda v) = \lambda(-v)$*;*
(d) if $\lambda v = 0_V$ *then* $\lambda = 0$ *or* $v = 0_V$*.*
(e) $-v = (-1)v$*.*

Proof. (a) We have

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

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

$$0_V = \lambda 0_V.$$

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

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

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 $(-\lambda)v = -(\lambda v)$.
(d) Suppose that $\lambda v = 0_V$, and that $\lambda \neq 0$. Then λ^{-1} exists in $\mathbb{F}$, and

$$\lambda^{-1}(\lambda v) = \lambda^{-1} 0_V = 0_V \qquad \text{[by (a)]}.$$

So

$$(\lambda^{-1}\lambda)v = 0_V \qquad \text{[scalar } \cdot \text{ interacts well with field } \cdot\text{]},$$

showing

$$v = 1v = 0_V \qquad \text{[identity for scalar multiplication]}.$$

(e) Note that

$$\begin{aligned} v + (-1)v &= 1v + (-1)v && \text{[by a vector space axiom]} \\ &= (1 + (-1))\, v && \text{[distributivity]} \\ &= 0v && \text{[definition of additive inverse in field]} \\ &= 0_V. && \text{[by (a)]} \end{aligned}$$

Hence by the uniqueness of the additive inverse $(-1)v = -v$. ■

3.2 Subspaces

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

Definition 74 *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); (this usually involves showing $0_V \in U$).

(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).

Note that the operations of addition and scalar multiplication referred to are those of V, not some separate, different operations of U.

Definition 75 *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**.*

Proposition 76 ***(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 $u \in U$. Since U is closed under scalar multiplication, $0u = 0_V \in U$.

$\lambda u_1 + u_2 \in U$ 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 also 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: we note $0_V \in U$.

U is closed under addition: for u_1, $u_2 \in U$ have $u_1 + u_2 = 1u_1 + u_2 \in U$.

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 77 *If U is a subspace of the vector space V, then we write $U \leqslant V$.*

Proposition 78 *Let V be a vector space over $\mathbb{F}$, and let $U \leqslant V$. Then*

(a) U is a vector space over $\mathbb{F}$. In fact, the only subsets of V that are vector spaces over $\mathbb{F}$ are the subspaces;

(b) if $W \leqslant U$ then $W \leqslant V$ ("a subspace of a subspace is a subspace").

Proof. (a) 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 $U \times U \to U$. Likewise 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 shows that $-u = (-1)u \in U$.

The remaining four properties are all inherited from V. That is, they apply to general vectors of V and vectors in U are vectors in V.

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

Proposition 79 *Let V be a vector space. Take U, $W \leqslant V$. Then $U+W \leqslant V$ and $U \cap W \leqslant V$, where*

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

Indeed $U + W$ is the smallest subspace of V which contains U and W and $U \cap W$ is the largest subspace of V which is contained in both U and W.

Proof. (a) As $U \leqslant V$ and $W \leqslant V$ then $0_V \in U$ and $0_V \in W$ so that $0_V = 0_V + 0_V \in U + W$.

Say that $v_1, v_2 \in U + W$ and $\lambda \in \mathbb{F}$. By definition there exist $u_1, u_2 \in U$ and $w_1, w_2 \in W$ such that

$$v_1 = u_1 + w_1, \qquad v_2 = u_2 + w_2,$$

and then

$$\lambda v_1 + v_2 = \lambda(u_1 + w_1) + u_2 + w_2 = (\lambda u_1 + u_2) + (\lambda w_1 + w_2) \in U + W$$

as $\lambda u_1 + u_2 \in U$ and $\lambda w_1 + w_2 \in W$ because $U \leqslant V$ and $W \leqslant V$.

(b) The statements concerning the intersection are left as exercises. ■

3.3 Further examples

Example 80 *Consider a system of **homogeneous** linear equations with real coefficients a_{ij}:*

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

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

Let V be the set of real solutions of the this linear system. Then V is a real vector space. This becomes more apparent if we write the equations in matrix form. We see the system corresponds to $A\mathbf{x} = \mathbf{0}$, where $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is an $n \times 1$ column

vector of variables, and $\mathbf{0}$ *is shorthand for* $0_{n\times 1}$*. 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* $\mathbb{R}^n_{\text{col}}$*.*

Clearly V *is non-empty, because* $\mathbf{0} \in V$*.*

For $\mathbf{v}_1,\ \mathbf{v}_2 \in V$*, we have* $A\mathbf{v}_1 = \mathbf{0}$ *and* $A\mathbf{v}_2 = 0$*, so* $A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0}$*, so* $\mathbf{v}_1 + \mathbf{v}_2 \in V$*. So* V *is closed under addition.*

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

So $V \leqslant \mathbb{R}^n_{\text{col}}$*, and so* V *is a vector space.*

Example 81 *The set* $\mathbb{R}[x]$ *of all real polynomials in a variable* x *is a real vector space. We will show that it is a subspace of* $\mathbb{R}^{\mathbb{R}}$*. Addition and scalar multiplication are defined by*

$$\left(\sum a_n x^n\right) + \left(\sum b_n x^n\right) = \sum \left(a_n + b_n\right) x^n, \qquad \lambda \left(\sum a_n x^n\right) = \sum (\lambda a_n) x^n.$$

As the sums are finite, then the addition and scalar multiple are also finite and hence polynomials. Finally teh zero function is a polynomial.

Example 82 *Let* n *be a non-negative integer. The set of polynomials* $c_n x^n + \cdots + c_1 x + c_0$ *with* $c_0,\ c_1,\ \ldots,\ c_n \in \mathbb{R}$ *(that is, real polynomials with degree* $\leqslant n$*) is a real vector space, and a subspace of* $\mathbb{R}[x]$*.*

Example 83 *Let* X *be a set. Define* $\mathbb{R}^X := \{$*functions* f *with* $f\colon X \to \mathbb{R}\}$*, 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*

$$(f + g)(x) = f(x) + g(x) \qquad \text{and} \qquad (\lambda f)(x) = \lambda f(x).$$

Example 84 *We can study the solutions of a homogeneous linear second-order differential equation. These are twice-differentiable real functions* y *that satisfy an equation of the form*

$$y'' + a(x)y' + b(x)y = 0.$$

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* $\mathbb{R}^{\mathbb{R}}$*. Note* S *is clearly non-empty (the* 0 *function satisfies the differential equation), 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,* $S \leqslant \mathbb{R}^{\mathbb{R}}$*.*

This generalises to homogeneous linear differential equations of any order.

Example 85 *What are the subspaces of* $\mathbb{R}$*?*

Let $V = \mathbb{R}$*, let* U *be a non-trivial subspace of* V*. Then there exists* $u \in U$ *with* $u \neq 0$*. Take* $x \in \mathbb{R}$*. Let* $\lambda = \frac{x}{u}$*. Then* $x = \lambda u \in U$*, because* U *is closed under scalar multiplication. So* $U = V$*.*

So the only subspaces of $\mathbb{R}$ *are* $\{0\}$ *and* $\mathbb{R}$*.*

Example 86 *What are the subspaces of* $\mathbb{R}^2$*?*

Let $V = \mathbb{R}^2$*, let* U *be a non-trivial subspace of* V*. Then there exists* $\mathbf{u} \in U$ *with* $\mathbf{u} \neq (0,0)$*, say* $\mathbf{u} = (a,b)$*. We have* $\langle \mathbf{u} \rangle = \{\lambda \mathbf{u} : \lambda \in \mathbb{R}\} \subseteq U$*. (Such 'spans' will more generally be defined in the next chapter.)*

Case 1: $\langle \mathbf{u} \rangle = U$*.*

If $a \neq 0$*, then let* $m = \frac{b}{a}$*. Then* $\langle \mathbf{u} \rangle = \{(x,y) \in \mathbb{R}^2 \mid y = mx\}$*.*

If $a = 0$*, then* $\langle \mathbf{u} \rangle = \{(0,y) \mid y \in \mathbb{R}\}$*.*

So if U *is, geometrically, 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: $\langle \mathbf{u} \rangle \neq U$*.*

Then there is some $\mathbf{v} = (c,d) \in U \backslash \langle \mathbf{u} \rangle$*.*

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* $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$*. [Why?] So* U *contains the vectors* $(1,0)$ *and* $(0,1)$*, and hence* $U = \mathbb{R}^2$*. [Why?]*

So the subspaces of $\mathbb{R}^2$ *are* $\{\mathbf{0}\}$*, lines in* $\mathbb{R}^2$ *through the origin and* $\mathbb{R}^2$*.*

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 Spans and Spanning sets

Lemma 87 *Let V be a vector space over a field $\mathbb{F}$, and take $S = \{u_1, u_2, \ldots, u_m\} \subseteq V$. Define*

$$U := \{\alpha_1 u_1 + \cdots + \alpha_m u_m : \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}.$$

Then $U \leqslant V$.

Proof. Applying the subspace test, we note.

- $0_V \in U$: 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, $U \leqslant V$. ■

Definition 88 *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*

$$\langle u_1, \ldots, u_m \rangle := \{\alpha_1 u_1 + \cdots + \alpha_m u_m : \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}.$$

This is the smallest subspace of V that contains $u_1, \ldots, u_m$.

More generally, we can define the span of any set $S \subseteq V$ (even a potentially infinite set S) as

$$\langle S \rangle := \{\alpha_1 s_1 + \cdots + \alpha_m s_m : m \geqslant 0, s_1, \ldots, s_m \in S, \alpha_1, \ldots, \alpha_m \in \mathbb{F}\}.$$

Note that a linear combination only ever involves finitely many elements of S, even if S is infinite. There isn't enough structure in a vector space to be able to define infinite sums. By convention the span of the empty set in $\{0_V\}$.

Definition 89 *Let V be a vector space over $\mathbb{F}$. If $S \subseteq V$ is such that $V = \langle S \rangle$, then we say that S **spans** V, and that S is a **spanning set** for V.*

Example 90 $\{(1,1),(2,-1)\}$ *spans* $\mathbb{R}^2$ *as every* (x,y) *can be written*

$$(x,y) = \left(\frac{x+2y}{3}\right)(1,1) + \left(\frac{x-y}{3}\right)(2,-1).$$

Whilst the span of $\{(2,2),(-1,-1)\}$ *is the line* $y = x$ *in* $\mathbb{R}^2$.

Example 91 $\{(1,1,2),(2,-1,3)\}$ *spans the plane given parametrically as*

$$\mathbf{r} = \alpha(1,1,2) + \beta(2,-1,3) \qquad \alpha,\beta \in \mathbb{R}.$$

By eliminating α,β *from the expressions*

$$x = \alpha + 2\beta, \qquad y = \alpha - \beta, \qquad z = 2\alpha + 3\beta,$$

then we can see this is the plane with equation

$$5x + y - 3z = 0.$$

Example 92 *The three vectors* $\{(1,1,2),(2,-1,3),(3,0,5)\}$ *span the same plane* $5x+y-3z = 0$. *This is because*

$$(3,0,5) = (1,1,2) + (2,-1,3)$$

and so the third vector is itself a linear combination of the first two. Note that any point in the plane can be written in many different ways as a linear combination of the three vectors. For example

$$\begin{aligned}(0,3,1) &= 2(1,1,2) - 1(2,-1,3) + 0(3,0,5)\\ &= 1(1,1,2) - 2(2,-1,3) + 1(3,0,5)\\ &= 3(1,1,2) + 0(2,-1,3) - 1(3,0,5).\end{aligned}$$

This third vector means there is redundancy in the set. Any two of the three vectors are sufficient to span the plan. The issue here is that the three vectors are not linearly independent.

Definition 93 *Given a matrix, its* ***row space*** *is the span of its rows and its* ***column space*** *is the span of its column. For an* $m \times n$ *matrix* A, *we write* $\mathrm{Row}(A) \leqslant \mathbb{R}^n$ *for its row space and* $\mathrm{Col}(A) \leqslant \mathbb{R}^m_{\mathrm{col}}$ *for its column space.*

Example 94 *In Example 7 we met the matrix on the left below, and the matrix on the right is its RRE form.*

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 3 & 2 \\ 2 & -1 & 1 & 2 & 4 \\ 4 & -3 & 3 & 8 & 8 \end{array}\right), \qquad \left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 2 \\ 0 & 1 & -1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

A check will show that these two matrices have the same row space – we will see in Proposition 117 that EROs don't change row space. However it is clear that $(1,2,4)^T$ *is in the column space of the first matrix and not of the second – so EROs do change column space.*

4.2 Linear independence

Definition 95 *Let V be a vector space over $\mathbb{F}$. We say that $v_1, \ldots, v_m \in V$ are **linearly independent** if the only solution to the equation*

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0_V \qquad \text{where} \quad \alpha_1, \ldots, \alpha_m \in \mathbb{F}$$

is

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

Otherwise $v_1, \ldots, v_m$ are said to be linearly dependent, which means there is a non-trivial linear combination of $v_1, \ldots, v_m$ which adds to 0_V.

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

Example 96 *$\{(1,1,2),(2,-1,3)\} \subseteq \mathbb{R}^2$ is linearly independent. To check this, we see that comparing the x- and y-coordinates in*

$$\alpha(1,1,2) + \beta(2,-1,3) = (0,0,0),$$

implies

$$\alpha + \beta = 0, \qquad 2\alpha - \beta = 0.$$

These equations alone are enough to show $\alpha = \beta = 0$. Note though that these two vectors do not span $\mathbb{R}^3$.

Example 97 *$\{(1,1,2),(2,-1,3),(3,0,5)\}$ is linearly dependent. We previously noted that*

$$(1,1,2) + (2,-1,3) = (3,0,5)$$

so that

$$1(1,1,2) + 1(2,-1,3) + (-1)(3,0,5) = (0,0,0).$$

This is a non-trivial linear combination which adds up to $\mathbf{0}$.

Example 98 *Let V denote the vector space of differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Then the set*

$$S = \{\sin x, \cos x, \sin 2x\}$$

is linearly independent. Say that

$$\alpha \sin x + \beta \cos x + \gamma \sin 2x = 0_V,$$

noting 0_V denotes the zero function, so that the above is an identity of functions. If we set $x = 0$ then this gives $\beta = 0$. If we set $x = \pi/2$ then $\alpha = 0$. Hence $\gamma = 0$ also.

Proposition 99 *Let $S = \{v_1, \ldots, v_m\}$ be a linearly independent subset of a vector space V. Then*

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = \beta_1 v_1 + \cdots + \beta_m v_m$$

if and only if $\alpha_i = \beta_i$ for all $1 \leqslant i \leqslant m$. Hence we may 'compare coefficients'.

Proof. If $\alpha_i = \beta_i$ for all $1 \leqslant i \leqslant m$ then the result clearly follows. Conversely, we can rearrange the above equation as

$$(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_m - \beta_m)v_m = 0_V.$$

As S is linearly independent then $\alpha_i - \beta_i = 0$ for all i as required. ■

Example 100 *Let $V = \mathbb{C}$, considered as a real vector space. Then $\{1, i\}$ is linearly independent for if*

$$x + yi = 0_{\mathbb{C}}$$

then $x = \operatorname{Re} 0_{\mathbb{C}} = 0$ and $y = \operatorname{Im} 0_{\mathbb{C}} = 0$. Hence by the previous proposition 'comparing real and imaginary parts' is valid.

Example 101 *Let $V = \mathbb{R}[x]$, the vector space of polynomials with real coefficients. Then the set $S = \{1, x, x^2, \ldots\}$ is linearly independent. Recall that an infinite set is linearly independent if every finite subset is linearly independent. So say that*

$$a_0 1 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0_{\mathbb{R}[x]}$$

for some coefficients $a_0, a_1, a_2, \ldots, a_n$. Recall that the above is an identity of functions. We can see that $a_0 = 0$ by setting $x = 0$. We can then see that $a_1 = 0$ by differentiating and setting $x = 0$. In a similar fashion we can see that all the coefficients are zero and that S is linearly independent.

Lemma 102 *Let $v_1, \ldots, v_m$ be linearly independent elements of a vector space V. Let $v_{m+1} \in V$. Then $v_1, v_2, \ldots, v_m, v_{m+1}$ are linearly independent if and only if*

$$v_{m+1} \notin \langle v_1, \ldots, v_m \rangle.$$

Proof. ($\Longleftarrow$) Suppose that $v_{m+1} \notin \langle v_1, \ldots, v_m \rangle$. Take $\alpha_1, \ldots, \alpha_{m+1} \in \mathbb{F}$ such that

$$\alpha_1 v_1 + \cdots + \alpha_{m+1} v_{m+1} = 0_V.$$

We aim to show that the α_i are all 0. If $\alpha_{m+1} \neq 0$, then we have

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

which is a contradiction. So $\alpha_{m+1} = 0$ and hence $\alpha_1 v_1 + \cdots + \alpha_m v_m = 0_V$. But $v_1, \ldots, v_m$ are linearly independent, so this means that $\alpha_1 = \cdots = \alpha_m = 0$ as well.

($\Longrightarrow$) Conversely say that $v_1, v_2, \ldots, v_m, v_{m+1}$ are linearly independent. If $v_{m+1} \in \langle v_1, \ldots, v_m \rangle$ then there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that

$$v_{m+1} = \alpha_1 v_1 + \cdots + \alpha_m v_m$$

so that $\alpha_1 v_1 + \cdots + \alpha_m v_m - v_{m+1} = 0_V$ which contradicts the linear independence of $v_1, v_2, \ldots, v_m, v_{m+1}$. Hence $v_{m+1} \notin \langle v_1, \ldots, v_m \rangle$. ■

4.3 Bases

Definition 103 *Let V be a vector space. A **basis** of V is a linearly independent, spanning set. (The plural is 'bases', pronounced 'bay-seas'.)*

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

Remark 104 *It is important to note the language here. We can talk about 'a' basis of a vector space. Typically, vector spaces have many bases so we should not talk about 'the' basis. Some vector spaces have a 'standard' or 'canonical' basis though.*

Remark 105 *Not every vector space is finite-dimensional. For example, the space of real polynomials or the space of real sequences do not have finite bases. But in this course we'll generally study finite-dimensional vector spaces. The courses on Functional Analysis in Parts B and C (third and fourth year) explore the theory of infinite-dimensional vector spaces which have further analytical structure. Note in a vector space that only finite sums are well-defined. To meaningfully form an infinite sum, a notion of convergence is needed which is why further structure is needed.*

Where possible, we will work with general vector spaces, but sometimes we'll need to specialise to the finite-dimensional case.

Example 106 *In $\mathbb{R}^n$, for $1 \leqslant i \leqslant n$, let $\mathbf{e}_i$ be the row vector with coordinate 1 in the ith entry and 0 elsewhere. Then $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent: if*

$$\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n = \mathbf{0}$$

then by looking at the ith entry we see that $\alpha_i = 0$ for all i. Also, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ span $\mathbb{R}^n$, because

$$(a_1, \ldots, a_n) = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n.$$

*So $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of $\mathbb{R}^n$. We call it the **standard basis** or **canonical basis** of $\mathbb{R}^n$.*

Example 107 *Let $V = M_{m\times n}(\mathbb{F})$ denote the vector space of $m \times n$ matrices over a field $\mathbb{F}$. Then the standard basis of V is the set*

$$\{E_{ij} \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$$

which has entry of 1 for the (i, j)th entry, and all other entries are zero. Note that a matrix $A = (a_{ij})$ can be written

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}$$

and this is the unique expression of A as a linear combination of the standard basis.

Example 108 *Let $V = \{(x.y.z) \in \mathbb{R}^3 \mid x + 2y + z = 0\} \leqslant \mathbb{R}^3$. Then a basis for V is*

$$\{(1, 0, -1), (0, 2, -1)\}.$$

To see this note that x and y can be used to parameterize V and a general vector can be written uniquely as

$$(x, y, -x - 2y) = x(1, 0, -1) + y(0, 2, -1).$$

Example 109 *Let $V \leqslant \mathbb{R}^5$ be the space of vectors $(x_1, x_2, x_3, x_4, x_5)$ satisfying the three equations*

$$\begin{aligned} x_1 + x_2 - x_3 + x_5 &= 0; \\ x_1 + 2x_2 + x_4 + 3x_5 &= 0; \\ x_2 + x_3 + x_4 + 2x_5 &= 0. \end{aligned}$$

We can represent these equations as

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 1 \\ 1 & 2 & 0 & 1 & 3 \\ 0 & 1 & 1 & 1 & 2 \end{pmatrix} \overset{\text{RRE}}{\to} \begin{pmatrix} 1 & 0 & -2 & -1 & -1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

If we assign parameters to the last three columns (as there are no leading 1s in these columns) by setting $x_3 = \alpha, x_4 = \beta, x_5 = \gamma$ then

$$x_1 = 2\alpha + \beta + \gamma, \qquad x_2 = -\alpha - \beta - 2\gamma$$

and hence

$$\begin{aligned} (x_1, x_2, x_3, x_4, x_5) &= (2\alpha + \beta + \gamma, -\alpha - \beta - 2\gamma, \alpha, \beta, \gamma) \\ &= \alpha(2, -1, 1, 0, 0) + \beta(1, -1, 0, 1, 0) + \gamma(1, -2, 0, 0, 1). \end{aligned}$$

So a basis for V is

$$\{(2, -1, 1, 0, 0), (1, -1, 0, 1, 0), (1, -2, 0, 0, 1)\}.$$

Example 110 *The space $\mathbb{F}[x]$ of polynomials over a field $\mathbb{F}$ (that is, with coefficients from the field $\mathbb{F}$) has standard basis*

$$\left\{1, x, x^2, x^3, \ldots\right\}.$$

Every polynomial can be uniquely written as a finite linear combination of this basis.

Proposition 111 *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. ($\Rightarrow$) Let S be a basis of V. Take $v \in V$. Since S spans V, there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Further as S is linearly independent, then by Proposition 99 these scalars $\alpha_1, \ldots, \alpha_n$ are unique.

($\Leftarrow$) Conversely, 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. ■

Propostion 111 allows us to define:

Definition 112 *Given a basis $\{v_1, \ldots, v_n\}$ of V then every $v \in V$ can be uniquely written*

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

*and the scalars $\alpha_1, \ldots, \alpha_n$ are known as the **coordinates** of v with respect to the basis $\{v_1, \ldots, v_n\}$.*

Remark 113 *Thus choosing a basis $\{v_1, \ldots, v_n\}$ for a finite-dimensional vector space V identifies V with $\mathbb{R}^n$. To a vector v can be associated a coordinate vector $\mathbf{v} = (\alpha_1, \ldots, \alpha_n)$.*

A vector space has an origin, but no axes. Choosing a basis of V introduces α_i-axes into V and identifies a vector v with a coordinate vector $\mathbf{v}$. I will denote coordinate vectors in bold, or underline them when writing by hand. It is important to note that a coordinate vector is meaningless without the context of a basis as we can see in the following example.

Example 114 *Let $V = \{f : \mathbb{R} \to \mathbb{R}, f''(x) = 4f(x)\}$. Then the general solution of the differential equation can be written uniquely as*

$$f(x) = Ae^{2x} + Be^{-2x}$$

or as

$$f(x) = C \sinh 2x + D \cosh 2x.$$

So $\{e^{2x}, e^{-2x}\}$ is a basis of V as is $\{\sinh 2x, \cosh 2x\}$. Note that the same vector e^{2x} has coordinates $(A, B) = (1, 0)$ using the first basis and has coordinates $(C, D) = (1, 1)$ with respect to the second basis as

$$e^{2x} = \sinh 2x + \cosh 2x,$$

Similarly the same coordinate vector $(1, 0)$ represents the vector e^{2x} with respect to the first basis, but a different vector $\sinh 2x$ with respect to the second basis.

Remark 115 *The above, of course, raises the question of whether there is a best way to coordinatize a vector space – or equivalently a best way to choose a basis.*

Remark 116 *The question of whether all vector spaces have a basis is an important foundational one. Every vector space does have a basis provided we assume the so-called 'axiom of choice', which is not a standard axiom of set theory. However, it can be shown that a basis of a space like l^∞, the space of bounded real sequences, is necessarily uncountable. So the structure of vector spaces, solely, is not well suited to working with some infinite-dimensional vector spaces which explains why the topic of infinite-dimensional space is more one of 'functional analysis' where infinite linear combinations can be well-defined.*

We now turn to the question of how we determine whether a set of vectors is linearly independent or spanning. Recall that we write $\mathrm{Row}(M)$ for the row space of a matrix M, that is the span of the rows of M.

Proposition 117 *Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be a $k \times m$ matrix. Let $R = (r_{ij})$ be a matrix in RRE form which can be obtained by EROs from A.*

(a) The non-zero rows of R are independent.

(b) The rows of R are linear combinations of the rows of A.

(c) $\mathrm{Row}(BA)$ *is contained in* $\mathrm{Row}(A)$.

(d) If $k = m$ and B is invertible then $\mathrm{Row}(BA) = \mathrm{Row}(A)$.

(e) $\mathrm{Row}(R) = \mathrm{Row}(A)$.

Proof. (a) Denote the non-zero rows of R as $\mathbf{r}_1, \ldots, \mathbf{r}_r$ and suppose that $c_1\mathbf{r}_1 + \cdots + c_r\mathbf{r}_r = \mathbf{0}$. Say the leading 1 of $\mathbf{r}_1$ appears in the jth column. Then

$$c_1 + c_2 r_{2j} + c_3 r_{3j} + \cdots + c_r r_{rj} = 0.$$

But as R is in RRE form each of $r_{2j}, r_{3j}, \ldots, r_{rj}$ is zero, being entries under a leading 1. It follows that $c_1 = 0$. By focusing on the column which contains the leading 1 of $\mathbf{r}_2$ we can likewise show that $c_2 = 0$ and so on. As $c_i = 0$ for each i then the non-zero rows $\mathbf{r}_i$ are independent.

We shall prove (c) first and then (b) follows from it. Recall that

$$(i,j)\,\text{th entry of } BA = \sum_{s=1}^{m} b_{is}a_{sj} \qquad (1 \leqslant i \leqslant k,\ 1 \leqslant j \leqslant n).$$

Thus the ith row of BA is the row vector

$$\left(\sum_{s=1}^{m} b_{is}a_{s1}, \ldots, \sum_{s=1}^{m} b_{is}a_{sn}\right) \quad = \quad \sum_{s=1}^{m} b_{is}\underbrace{(a_{s1}, a_{s2}, \cdots, a_{sn})}_{s\text{th row of } A}, \tag{4.1}$$

which is a linear combination of the rows of A. So every row of BA is in $\mathrm{Row}(A)$. A vector in the row space $\mathrm{Row}(BA)$ is a linear combination of BA's rows which, in turn, are linear combinations of A's rows. Hence $\mathrm{Row}(BA)$ is contained in $\mathrm{Row}(A)$. Because $R = E_k \cdots E_1 A$ for some elementary matrices $E_1, E_2, \ldots, E_k$ then (b) follows from (c) with $B = E_k \cdots E_1$. Now (d) also follows from (c). We know $\mathrm{Row}(BA)$ is contained in $\mathrm{Row}(A)$ and likewise $\mathrm{Row}(A) = \mathrm{Row}(B^{-1}(BA))$ is contained in $\mathrm{Row}(BA)$. Finally (e) follows from (d) by taking $B = E_k \cdots E_1$ which is invertible as elementary matrices are invertible. ■

Corollary 118 *(**Test for Independence**) Let A be an $m \times n$ matrix. Then* $\mathrm{RRE}(A)$ *contains a zero row if and only if the rows of A are dependent.*

Proof. We have that $\mathrm{RRE}(A) = BA$ where B is a product of elementary matrices and so invertible. Say the ith row of BA is $\mathbf{0}$. By (4.1)

$$\mathbf{0} = \sum_{s=1}^{m} b_{is}(s\text{th row of } A).$$

Now b_{is} are the entries of the ith row of B which, as B is invertible, cannot all be zero. The above then shows the rows of A are linearly dependent.

Conversely suppose that the rows of A are linearly dependent. Let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ denote the rows of A and, without any loss of generality, assume that $\mathbf{r}_m = c_1\mathbf{r}_1 + \cdots + c_{m-1}\mathbf{r}_{m-1}$ for real numbers $c_1, \ldots, c_{m-1}$. By performing the EROs $A_{1m}(-c_1), \ldots, A_{(m-1)m}(-c_{m-1})$ we arrive at a matrix whose mth row is zero. We can continue to perform EROs on the top $m-1$ rows, leaving the bottom row untouched, until we arrive at a matrix in RRE form. Once we have shown RRE form is unique (to follow) then we have that RRE(A) has a zero row. ■

Corollary 119 *(**Test for a Spanning Set**) Let A be an $m \times n$ matrix. Then the rows of A span $\mathbb{R}^n$ if and only if*

$$\mathrm{RRE}\,(A) = \begin{pmatrix} I_n \\ 0_{(m-n)n} \end{pmatrix}.$$

Proof. Let $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ be the rows of A in $\mathbb{R}^n$ and suppose they span $\mathbb{R}^n$. Now row space is invariant under EROs. If it were the case that the ith column of RRE$\,(A)$ does not contain a leading 1 then $\mathbf{e}_i$ would not be in the row space. Consequently every column contains a leading 1 and so

$$\mathrm{RRE}\,(A) = \begin{pmatrix} I_n \\ 0_{(k-n)n} \end{pmatrix}.$$

Conversely if RRE$\,(A)$ has the above form then the rows of RRE$\,(A)$ are spanning and hence so are the original rows $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$. ■

Remark 120 *The above corollaries show that if vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent in $\mathbb{R}^n$ then $k \leqslant n$. (For if $k > n$ then the RRE form will necessarily have a zero row.) They further show that if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are spanning then there must be n leading 1s and hence we must have $k \geqslant n$. This then shows that a basis, any basis, of $\mathbb{R}^n$ contains n vectors.*

*The above takes a coordinate approach, and relies on some results we are yet to prove – especially uniqueness of the RRE form. We will shortly prove this result more formally, without making use of coordinates, but we will see that this is generally true of finite-dimensional vector spaces. This common cardinality of all bases is called the **dimension** of the vector space.*

Example 121 *(a) The vectors $\mathbf{v}_1 = (1, 2, -1, 0)$, $\mathbf{v}_2 = (2, 1, 0, 3)$, $\mathbf{v}_3 = (0, 1, 1, 1)$ in $\mathbb{R}^4$ are linearly independent. If we row reduce the matrix with rows $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ we get*

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 0 & 3 \\ 0 & 1 & 1 & 1 \end{pmatrix} \overset{\text{RRE}}{\to} \begin{pmatrix} 1 & 0 & 0 & 1.6 \\ 0 & 1 & 0 & -0.2 \\ 0 & 0 & 1 & 1.2 \end{pmatrix}$$

and hence the three vectors are independent because there is no zero row.

(b) A vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ if and only if $8x_1 + 6x_3 = x_2 + 5x_4$. One way to see this is to row reduce the matrix

$$\begin{pmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 0 & 3 \\ 0 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \end{pmatrix},$$

which reduces to

$$\begin{pmatrix} 1 & 0 & 0 & 1.6 \\ 0 & 1 & 0 & -0.2 \\ 0 & 0 & 1 & 1.2 \\ 0 & 0 & 0 & x_4 - 1.6x_1 + 0.2x_2 - 1.2x_3 \end{pmatrix}$$

which has a zero row if and only $8x_1 + 6x_3 = x_2 + 5x_4$.

4.4 Addendum

In this addendum we will show that the RRE form a matrix is unique. (Recall we've already shown existence.) The proof is somewhat technical and the proof (but not knowledge of the result) is off-syllabus, but is included for completeness. This also allows us to define *row rank*.

Theorem 122 ***(Uniqueness of RRE Form)*** *The reduced row echelon form of an $m \times n$ matrix A is unique.*

Proof. The proof below follows by fixing the number of rows m and arguing by induction on the number of columns n. The only $m \times 1$ matrices which are in RRE form are $\mathbf{0}$ and $\mathbf{e}_1^T$. The zero $m \times 1$ matrix will reduce to the former and non-zero $m \times 1$ matrices to the latter. In particular, the RRE form of an $m \times 1$ matrix is unique.

Suppose, as our inductive hypothesis, that all $m \times (n-1)$ matrices M have a unique reduced row echelon form RRE(M). Let A be an $m \times n$ matrix and let $\tilde{A}$ denote the $m \times (n-1)$ matrix comprising the first $n-1$ columns of A. Given *any* EROs which reduce A to RRE form, these EROs also reduce $\tilde{A}$ to RRE($\tilde{A}$) which is unique by hypothesis. Say RRE($\tilde{A}$) has r non-zero rows.

There are two cases to consider: (i) any RRE form of A has one more non-zero row than RRE($\tilde{A}$); (ii) any RRE form of A has the same number of non-zero rows as RRE($\tilde{A}$). These can be the only cases as the first $n-1$ columns of an RRE form of A are those of RRE($\tilde{A}$) and both matrices are in RRE form; note further that an extra non-zero row in any RRE form of A, if it exists, must equal $\mathbf{e}_n$. Case (i) occurs if $\mathbf{e}_n$ is in the row space of A and case (ii) if not. In particular, it is impossible that different sets of EROs might reduce a given A to both cases (i) and (ii).

So RRE(A) has one of the following two forms:

$$\text{(i)} \quad \begin{pmatrix} \boxed{\begin{matrix}\text{non-zero}\\ \text{RRE}(\tilde{A})\\ \text{rows}\end{matrix}} & \begin{matrix}0\\ \vdots\\ 0\end{matrix} \\ \begin{matrix}0 & \cdots & 0\end{matrix} & 1 \\ m-r-1 \text{ zero rows} & \end{pmatrix}, \qquad \text{(ii)} \quad \begin{pmatrix} \boxed{\begin{matrix}\text{non-zero}\\ \text{RRE}(\tilde{A})\\ \text{rows}\end{matrix}} & \begin{matrix}*\\ \vdots\\ *\end{matrix} \\ m-r \text{ zero rows} & \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1(R) \\ \vdots \\ \mathbf{r}_r(R) \\ m-r \text{ zero rows} \end{pmatrix}.$$

In case (i) the last column of any RRE form of A is $\mathbf{e}_{r+1}^T$ and so we see that RRE(A) is uniquely determined as we also know the first $n-1$ columns to be RRE($\tilde{A}$) by our inductive hypothesis. In case (ii), then any RRE form of A and RRE($\tilde{A}$) both have r non-zero rows. Let R_1 and R_2 be RRE forms of A. By hypothesis, their first $n-1$ columns agree and equal RRE($\tilde{A}$). By Proposition 117(e),

$$\text{Row}(R_1) = \text{Row}(A) = \text{Row}(R_2).$$

In particular, this means that the rows $\mathbf{r}_k(R_1)$ of R_1 are linear combinations of the rows $\mathbf{r}_k(R_2)$ of R_2. So, for any $1 \leqslant i \leqslant r$, there exist real numbers $\lambda_1, \ldots, \lambda_r$ such that

$$\mathbf{r}_i(R_1) = \textstyle\sum_{k=1}^{r} \lambda_k \mathbf{r}_k(R_2) \qquad \text{and hence} \qquad \mathbf{r}_i(\text{RRE}(\tilde{A})) = \textstyle\sum_{k=1}^{r} \lambda_k \mathbf{r}_k(\text{RRE}(\tilde{A}))$$

by focusing on the first $n-1$ columns. $\mathrm{RRE}(\tilde{A})$ is in RRE form and so its non-zero rows are independent; it follows that $\lambda_i = 1$ and $\lambda_j = 0$ for $j \neq i$. In particular $\mathbf{r}_i(R_1) = \mathbf{r}_i(R_2)$ for each i and hence $R_1 = R_2$ as required. ■

We may now define:

Definition 123 *The **row rank,** or simply **rank,** of a matrix A is the number of non-zero rows in* $\mathrm{RRE}(A)$. *We write this as* $\mathrm{rank}(A)$. *The uniqueness of* $\mathrm{RRE}(A)$ *means row rank is well-defined.*

Corollary 124 *Let $(A|\mathbf{b})$ be the matrix representing the linear system $A\mathbf{x} = \mathbf{b}$. Then the system is consistent (i.e. has at least one solution) if and only if* $\mathrm{rank}(A|\mathbf{b}) = \mathrm{rank}(A)$.

Proof. Note this result was already demonstrated for systems in RRE form during the proof of Proposition 44. Say that $\mathrm{RRE}(A) = PA$ where P is a product of elementary matrices that reduce A.

Now if E is an elementary matrix then $\mathrm{RRE}(EA) = \mathrm{RRE}(A)$ by the uniqueness of RRE form and so $\mathrm{rank}(EA) = \mathrm{rank}(A)$. We then have

$$\begin{array}{rcl} \mathrm{rank}(A|\mathbf{b}) = \mathrm{rank}(A) & \iff & \mathrm{rank}(PA|P\mathbf{b}) = \mathrm{rank}(PA) \\ & \iff & \text{the system } PA\mathbf{x} = P\mathbf{b} \text{ is consistent} \\ & \iff & \text{the system } A\mathbf{x} = \mathbf{b} \text{ is consistent} \end{array}$$

as the set of solutions to $A\mathbf{x} = \mathbf{b}$ is unaffected by EROs. ■

Proposition 125 *Let A be an $m \times n$ matrix and $\mathbf{b}$ in $\mathbb{R}^m_{\mathrm{col}}$.*

(a) the system $A\mathbf{x} = \mathbf{b}$ has no solutions if and only if $\left(\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & 1 \end{array}\right)$ *is in* $\mathrm{Row}(A|\mathbf{b})$. *If the system $A\mathbf{x} = \mathbf{b}$ is consistent then*

(b) there is a unique solution if and only if $\mathrm{rank}(A) = n$. *It follows that $m \geqslant n$.*

(c) there are infinitely many solutions if $\mathrm{rank}(A) < n$. *The set of solutions is an* $n - \mathrm{rank}(A)$ *parameter family.*

Proof. As we know that $(A|\mathbf{b})$ can be put into RRE form, and that EROs affect neither the row space nor the set of solutions, the above is just a rephrasing of Proposition 44. ■

Remark 126 *One might rightly guess that there is the equivalent notion of column rank. Namely the number of non-zero columns remaining when a matrix is similarly reduced using ECOs (elementary column operations). It is the case, in fact, that column rank and row rank are equal and we will prove this later. So we may refer to the rank of a matrix without ambiguity.* ■

5. DIMENSION

We are now in a position to define the *dimension* of a vector space with a basis, and to show that dimension is well-defined. Implicitly we have already seen this result in the tests for linear independent sets and for spanning sets. We showed in those tests that a linear independent subset of $\mathbb{R}^n$ cannot have more than n elements and that a spanning set of $\mathbb{R}^n$ cannot have fewer than n. The proof below has the merit of not relying on coordinates.

Theorem 127 (Steinitz Exchange Lemma) *Let V be a vector space over a field $\mathbb{F}$. Take $X = \{v_1, v_2, \ldots, v_n\} \subseteq V$. Suppose that $u \in \langle X \rangle$ but that $u \notin \langle X \backslash \{v_i\} \rangle$ for some i. Let*

$$Y = (X \backslash \{v_i\}) \cup \{u\}$$

(that is, we "exchange u for v_i"). Then $\langle Y \rangle = \langle X \rangle$.

Proof. Since $u \in \langle X \rangle$, there are $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$u = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

There is $v_i \in X$ such that $u \notin \langle X \backslash \{v_i\} \rangle$. Without loss of generality, we may assume that $i = n$. Since $u \notin \langle X \backslash \{v_n\} \rangle$, 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 $w \in \langle Y \rangle$ 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 $\langle Y \rangle \subseteq \langle X \rangle$. And if $w \in \langle X \rangle$ 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 $\langle Y \rangle \supseteq \langle X \rangle$. ■

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 fundamental to defining the dimension of a vector space.

Theorem 128 *Let V be a vector space. Let S, T be finite subsets of V. If S is linearly independent and T spans V, then $|S| \leqslant |T|$.*

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 $v_1, \ldots, v_n$. We will use the Steinitz Exchange Lemma to swap out the elements of T with those of S, one at a time, ultimately exhausting S.

Let $T_0 = \{v_1, \ldots, v_n\}$. Since $\langle T_0 \rangle = V$, then $u_1 \in \langle v_1, \ldots, v_i \rangle$ for some $1 \leqslant i \leqslant n$ and choose i to be minimal in this regard. Note then that $u_1 \in \langle v_1, \ldots, v_i \rangle$ but that $u_1 \notin \langle v_1, \ldots, v_{i-1} \rangle$. The Steinitz Exchange Lemma then shows that

$$\langle v_1, \ldots, v_i \rangle = \langle u_1, v_1, \ldots, v_{i-1} \rangle$$

and hence

$$\begin{aligned} V &= \langle v_1, \ldots, v_n \rangle \\ &= \langle v_1, \ldots, v_i \rangle + \langle v_{i+1}, \ldots, v_n \rangle \\ &= \langle u_1, v_1, \ldots, v_{i-1} \rangle + \langle v_{i+1}, \ldots, v_n \rangle \\ &= \langle u_1, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n \rangle. \end{aligned}$$

Now, by relabelling the elements of T, we can assume without loss of generality assume that u_1 has been exchanged for v_1 and we set

$$T_1 = \{u_1, v_2, \ldots, v_n\} \qquad \text{noting that} \qquad \langle T_1 \rangle = V.$$

We proceed inductively in this manner creating sets

$$T_k = \{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\} \qquad \text{such that} \qquad \langle T_k \rangle = V..$$

Note that at each stage $u_{k+1} \in \langle T_k \rangle$ but that $u_{k+1} \notin \langle u_1, \ldots, u_k \rangle$ as the set S is independent. Hence we can keep continuing to replace elements of T with elements of S. The process can only terminate when S is exhausted which means that $m \leqslant n$. ■

Corollary 129 *Let V be a finite-dimensional vector space. All bases of V are finite and of the same size.*

Proof. Since V is finite-dimensional then V has a finite basis B. By Theorem 128 any finite linearly independent subset of V has size at most $|B|$. Given another basis S of V, it is linearly independent, so every finite subset of S is linearly independent. So in fact S must be finite, and $|S| \leqslant |B|$. But B is linearly independent and S is spanning and so by Theorem 128 $|B| \leqslant |S|$. ■

Definition 130 *Let V be a finite-dimensional vector space. The **dimension** of V, written $\dim V$, is the size of any basis of V.*

Definition 131 *We can now redefine row rank using this notion of dimension. The **row rank** of a matrix is the dimension of its **row space.** When in RRE form, the non-zero rows of the matrix are linearly independent. Further EROs do not affect the row space. So the non-zero rows of a matrix in RRE form are a basis of the row space.*

5.1 Subspaces and Dimension

We include the following result here as it fits in naturally with some of the subsequent results; in what follows we will show:

- A spanning set contains a basis.
- A linearly independent set can be extended to a basis. (This result requires the notion of dimension.)
- A basis is a maximal linearly independent set.
- A basis is a minimal spanning set.

Proposition 132 *Let V be a vector space over $\mathbb{F}$ and let S be a finite spanning set. Then S contains a basis.*

Remark 133 *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 (that is, no linearly independent subset of S strictly contains T). Suppose, for a contradiction, that $\langle T\rangle \neq V$. Then, since $\langle S\rangle = V$, there must exist $v \in S\backslash\langle T\rangle$.

Now by Lemma 102 we see that $T \cup \{v\}$ is linearly independent, and $T \cup \{v\} \subseteq S$, and $|T \cup \{v\}| > |T|$, which contradicts the maximality of T. So T spans V, is linearly independent, and thus a basis. ■

Proposition 134 *Let U be a subspace of a finite-dimensional vector space V. Then U is finite-dimensional, and $\dim U \leqslant \dim V$. Further if $\dim U = \dim V$ then $U = V$.*

Proof. Let $n = \dim V$. Then, by Theorem 128, every linearly independent subset of V has size at most n. Let S be a largest linearly independent set contained in U (and so in V), so $|S| \leqslant n$.

Suppose, for a contradiction, that $\langle S\rangle \neq U$. Then there exists $u \in U\backslash\langle S\rangle$. Now by Lemma 102 $S \cup \{u\}$ is linearly independent, and $|S \cup \{u\}| > |S|$, which contradicts our choice of S. So $U = \langle S\rangle$ and S is linearly independent, so S is a basis of U, and as we noted earlier $|S| \leqslant n$.

Say now that $\dim U = \dim V$ and $U \neq V$. Then there exists $v \in V\backslash U$. This v may then be added to a basis of U to create a linearly independent subset of V with

$$\dim U + 1 = \dim V + 1$$

vectors, which is a contradiction. Hence $\dim U = \dim V$ implies $U = V$. ■

Proposition 135 *Let V be a finite-dimensional vector space over $\mathbb{F}$ and let S be a linearly independent set. Then there exists a basis B such that $S \subseteq B$.*

Proof. If $\langle S\rangle = V$ then we are done as S is linearly independent and spanning, and so a basis. If $\langle S\rangle \neq V$ then by Lemma 102 we can extend S to $S_1 = S \cup \{u_1\}$ where $u_1 \in U \backslash \langle S\rangle$ to create a larger linearly independent set. If $\langle S_1\rangle = V$, then we are done as S_1 is a basis. This process can continue and only terminates at some S_k if S_k is a basis. However this process must terminate as we know every linearly independent subset of V must contain at most $\dim V$ elements. ■

Corollary 136 *A maximal linearly independent subset of a finite-dimensional vector space is a basis.*

Proof. Let S be a maximal linearly independent subset of a finite-dimensional vector space V. If $\langle S\rangle \neq V$ then by Lemma 102 we can extend $S_1 = S \cup \{u_1\}$ which is still linearly independent, but contradicts the maximality of S. So $\langle S\rangle = V$. ■

Corollary 137 *A minimal spanning subset of a finite-dimensional vector space is a basis.*

Proof. Let S be a minimal spanning subset of a finite-dimensional vector space V. If S is not linearly independent, then there exists $v \in S$ which can be written as a linear combination of elements of $S\backslash\{v\}$. Then $S\backslash\{v\}$ is linearly independent, and as shown in Lemma 102 $S\backslash\{v\}$ is still spanning, which contradicts the minimality of S. ■

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

Example 138 *Let* $S = \{(0,1,2,3),(1,2,3,4),(2,3,4,5)\} \subseteq \mathbb{R}^4$. *Define*

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

So $\langle S\rangle = \mathrm{Row}(A)$. *Applying EROs to* A *does not change the row space. Now*

$$\mathrm{RRE}(A) = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

As has been commented before, the non-zero rows are a basis for the row space, or equivalently for $\langle S\rangle$.

5.2 The dimension formula

We previously saw that the sum $U+W$ and intersection $U \cap W$ of two subspaces are subspaces. We now prove a useful theorem connecting their dimensions. Recall that we can extend bases of subspaces to bases of larger spaces, but in general a basis of a vector space won't contain a basis of a subspace (or possibly even any elements from the subspace). Thus it makes sense to begin with $U \cap W$, the smallest of the relevant spaces.

The next result is particularly useful.

Theorem 139 (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 $U\cap W\leqslant U$ and $U\cap W\leqslant W$, so by Theorem 135 we can separately extend this set 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

$$\dim(U\cap W)=m, \qquad \dim U=m+p, \qquad \dim W=m+q.$$

We aim to show that $S=\{v_1, \ldots, v_m, u_1, \ldots, u_p, w_1, \ldots, w_q\}$ is a basis of $U+W$. It contains

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

elements. So if we can verify this aim then the result follows.

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

$$\begin{aligned} u &= \alpha_1 v_1+\cdots+\alpha_m v_m+\alpha'_1 u_1+\cdots+\alpha'_p u_p, \\ w &= \beta_1 v_1+\cdots+\beta_m v_m+\beta'_1 w_1+\cdots+\beta'_q w_q \end{aligned}$$

for some scalars α_i, α'_i, β_i, $\beta'_i\in\mathbb{F}$. Then

$$x=u+w=(\alpha_1+\beta_1)v_1+\cdots+(\alpha_m+\beta_m)v_m+\alpha'_1 u_1+\cdots+\alpha'_p u_p+\beta'_1 w_1+\cdots+\beta'_q w_q\in\langle S\rangle,$$

showing $\langle S\rangle=U+W$.

S is linearly independent: Take $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q\in\mathbb{F}$ such that

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

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 LHS is in U, and the vector on the RHS is in W. So they are both in $U\cap W$. As $v_1, \ldots, v_m$ form a basis of $U\cap W$, there are $\lambda_1, \ldots, \lambda_m\in\mathbb{F}$ such that

$$-(\gamma_1 w_1+\cdots+\gamma_q w_q)=\lambda_1 v_1+\cdots+\lambda_m v_m,$$

which rearranges to

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

But $\{v_1, \ldots, v_m, w_1, \ldots, w_q\}$ is linearly independent (it's a basis for W), and so each γ_i is 0. This then implies 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 (it's a basis for U), so each α_i and β_i equals 0. So S is linearly independent and the result follows. ■

Example 140 *Let V be a vector space of dimension 10. Let X, Y be subspaces of dimension 6. Then $X + Y \leqslant V$ so $\dim(X + Y) \leqslant \dim V = 10$. So, by the dimension formula,*

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

It is not hard to see that the possibilities

$$2 \leqslant \dim(X \cap Y) \leqslant 6,$$

are all possible. This is left as an exercise.

Definition 141 *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 $V = U \oplus W$.*

Proposition 142 *Let U, W be subspaces of a finite-dimensional vector space V. The following are equivalent:*

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

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

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

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

(e) 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: (a) $\Leftrightarrow$ (b) follows from the definition of direct sum.

Try using the dimension formula to prove that (a)/(b) are equivalent to (c)/(d)/(e). ■

Remark 143 *To conclude, two more general comments on direct sums.*

- *A vector space V is said to be the direct sum of subspaces $X_1, \ldots, X_k \leqslant V$ if every $v \in V$ can be uniquely written*

$$v = x_1 + x_2 + \cdots + x_k \qquad \text{where} \quad x_i \in X_i \quad \text{for all } i.$$

  Thus it is statement (b) in the proposition above which naturally generalizes.

- *Writing a vector space as a sum of subspaces is called an **internal direct sum**. Given vectors spaces $V_1, \ldots, V_k$ then the **external direct sum***

$$V_1 \oplus V_2 \oplus \cdots \oplus V_k$$

  has the Cartesian product $V_1 \times V_2 \times \cdots \times V_k$ as the underlying set, with addition and scalar multiplication defined componentwise. That is

$$(v_1, \ldots, v_k) + (w_1, \ldots, w_k) = (v_1 + w_1, \ldots, v_k + w_k);$$
$$\alpha.(v_1, \ldots, v_k) = (\alpha.v_1, \ldots, \alpha.v_k)$$

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 a Part C course.)

6.1 What is a linear transformation?

Definition 144 *Let V, W be vector spaces over $\mathbb{F}$. We say that a map $T\colon V \to W$ 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 $\lambda \in \mathbb{F}$ (preserves scalar multiplication).
*We call T a **linear transformation** or a **linear map**.*

Proposition 145 *Let V, W be vector spaces over $\mathbb{F}$, let $T\colon V \to W$.be a linear map. Then $T(0_V) = 0_W$.*

Proof. Note that $T(0_V) + T(0_V) = T(0_V + 0_V) = T(0_V)$, and hence $T(0_V) = 0_W$. ■

Proposition 146 *Let V, W be vector spaces over $\mathbb{F}$, let $T\colon V \to W$. The following are equivalent:*

(a) T is linear;

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

(c) for any $n \geqslant 1$, 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. ■

Example 147
- *Let V be a vector space. Then the **identity map** $id_V\colon V \to V$ given by $id_V(v) = v$ for all $v \in V$ is a linear map.*
- *Let V, W be vector spaces. The **zero map** $0\colon V \to W$ that sends every $v \in V$ to 0_W is a linear map.*

- *For m, $n \geqslant 1$, with $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. Then we define the left multiplication map*

$$L_A\colon \mathbb{R}^n_{\text{col}} \to \mathbb{R}^m_{\text{col}} \qquad \text{by} \qquad L_A(\mathbf{v}) = A\mathbf{v} \quad \text{for} \quad \mathbf{v} \in \mathbb{R}^n_{\text{col}}.$$

  This is a linear map. Similarly, we have a right multiplication map

$$R_A\colon \mathbb{R}^m \to \mathbb{R}^n \qquad \text{by} \qquad R_A(\mathbf{v}) = \mathbf{v}A \quad \text{for} \quad \mathbf{v} \in \mathbb{R}^m.$$

- *Take m, n, $p \geqslant 1$ with $A \in \mathcal{M}_{m\times n}(\mathbb{R})$. The left multiplication map $\mathcal{M}_{n\times p}(\mathbb{R}) \to \mathcal{M}_{m\times p}(\mathbb{R})$ 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 $v = u + w$. Define $P\colon V \to V$ by $P(v) = w$.*

  Proposition 148 *P is a linear map. P is called **projection of** V **onto** W **along** U.*

  Proof. *Take v_1, $v_2 \in V$ and α_1, $\alpha_2 \in \mathbb{F}$. 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 $\alpha_1 u_1 + \alpha_2 u_2 \in U$ and $\alpha_1 w_1 + \alpha_2 w_2 \in W$. So by uniqueness

$$P(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 w_1 + \alpha_2 w_2 = \alpha_1 P(v_1) + \alpha_2 P(v_2).$$

  ■

- *For $A = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{R})$, we define the **trace** of A to be*

$$\operatorname{trace}(A)\colon = a_{11} + a_{22} + \cdots + a_{nn},$$

  (the sum of the entries on the main diagonal of A). The map $\operatorname{trace}\colon \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$ *is a linear map.*

- *Let $\mathbb{R}_n[x]$ be the vector space of polynomials of degree at most n. Define $D\colon \mathbb{R}_n[x] \to \mathbb{R}_n[x]$ 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 $f\colon \mathbb{R} \to \mathbb{R}$. The differential operator $D\colon C^1(\mathbb{R}) \to \mathbb{R}^{\mathbb{R}}$ sending f to f' is a linear map.*

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

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

6.2 Combining linear transformations

We can add linear transformations (pointwise), and we can multiply a linear transformation by a scalar (pointwise).

Proposition 149 *Let V, W be vector spaces over a field $\mathbb{F}$. For S, $T\colon V \to W$ and $\lambda \in \mathbb{F}$, we may define linear maps $S+T$ and λS by*

$$\begin{aligned} S+T\colon V \to W \qquad & by \qquad (S+T)(v) = S(v)+T(v) \qquad for \qquad v \in V; \\ \lambda S\colon V \to W \qquad & by \qquad (\lambda S)(v) = \lambda S(v) \qquad for \qquad v \in V. \end{aligned}$$

With these operations (and the zero map $0\colon V \to W$), the set of linear transformations $V \to W$ forms a vector space denoted $\mathrm{Hom}(V,W)$.

Proof. Firstly $S+T$ is a linear map as

$$\begin{aligned} (S+T)(\alpha_1 v_1 + \alpha_2 v_2) &= S(\alpha_1 v_1 + \alpha_2 v_2) + T(\alpha_1 v_1 + \alpha_2 v_2) && \text{[by definition]} \\ &= \alpha_1 S(v_1) + \alpha_2 S(v_2) + \alpha_1 T(v_1) + \alpha_2 T(v_2) && \text{[by linearity]} \\ &= \alpha_1 (S(v_1) + T(v_1)) + \alpha_2 (S(v_2) + T(v_2)) && \text{[rearranging]} \\ &= \alpha_1 (S+T)(v_1) + \alpha_2 (S+T)(v_2) && \text{[by definition]} \end{aligned}$$

showing linearity. That λS is linear is left as an exercise. Verifying the vector space axioms for $\mathrm{Hom}(V,W)$ involves showing:

(i) $S+T = T+S$: this follows from commutativity of $+$ in W.
(ii) $S+(T+U) = (S+T)+U$: this follows from associativity of $+$ in W.
(iii) $S+0 = S$: this follows from properties of 0_W.
(iv) S has an additive inverse $(-S)(v) \stackrel{\text{def}}{=} -(S(v))$.
(v)

$$\lambda(S+T) = (\lambda S) + (\lambda T); \quad (\lambda+\mu)S = (\lambda S) + (\mu S); \quad (\lambda\mu)S = \lambda(\mu S); \quad 1S = S.$$

These properties follow from the same properties for vectors in W as addition and scalar multiplication of linear maps is defined pointwise. ■

We can also compose linear transformations.

Proposition 150 *Let U, V, W be vector spaces over $\mathbb{F}$. Let $S\colon U \to V$ and $T\colon V \to W$ be linear. Then $T \circ S\colon U \to W$ is linear.*

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

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

so $T \circ S$ is linear. ■

Remark 151 *We often write $T \circ S$ as TS. The notation $T \circ S$ removes any possible ambiguity about the order of the functions.*

Definition 152 *Let V, W be vector spaces and let $T\colon V \to W$ be linear. We say that T is **invertible** if there is a linear transformation $S\colon W \to V$ such that $ST = id_V$ and $TS = id_W$ (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}. An invertible linear map is called an **isomorphism.***

Remark 153 *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 154 *Let V, W be vector spaces. Let $T\colon V \to W$ 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 $S\colon W \to V$, but it remains to show that S is linear. Take w_1, $w_2 \in W$ and λ_1, $\lambda_2 \in \mathbb{F}$. Let $v_1 = S(w_1)$, $v_2 = S(w_2)$. Then

$$T(v_1) = TS(w_1) = w_1 \qquad \text{and} \qquad T(v_2) = TS(w_2) = w_2.$$

Now

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

So S is linear. ■

Proposition 155 *Let U, V, W be vector spaces. Let $S\colon U \to V$ and $T\colon V \to W$ be invertible linear transformations. Then $TS\colon U \to W$ is invertible, and $(TS)^{-1} = S^{-1}T^{-1}$.*

Proof. Exercise. ■

Proposition 156 *(a) Let V, W be vector spaces with V finite-dimensional. If there is an invertible linear map $T\colon V \to W$ then $\dim V = \dim W$.*

(b) Let V, W be finite-dimensional vector spaces with $\dim V = \dim W$. Then there is an invertible linear map $T\colon V \to W$.

Consequently V and W are isomorphic if and only if $\dim V = \dim W$.

Proof. (a) Let $v_1, \ldots, v_n$ be a basis for V. It is left as an exercise to show that $Tv_1, \ldots, Tv_n$ is a basis for W. Then $n = \dim V = \dim W$.

(b) Let $n = \dim V = \dim W$. Let $v_1, \ldots, v_n$ be a basis for V and $w_1, \ldots, w_n$ be a basis for W. It is left as an exercise to show that

$$T\colon V \to W \qquad \text{given by} \qquad T\left(\sum \alpha_i v_i\right) = \sum \alpha_i w_i$$

is a well-defined, invertible linear map. ■

Example 157 *Let $V = \mathbb{R}[x]$ denote the vector space of polynomials in x with real coefficients, and let W denote the vector space of real sequences $(a_n)_{n=0}^{\infty}$. Then V and W are both infinite-dimensional but are not isomorphic.*

Solution. The set $B = \{1, x, x^2, x^3, \ldots\}$ is a basis for V. That it is linearly independent shows that V is not finite-dimensional. The set of sequences $\{(\delta_{in})_{n=0}^{\infty} \mid i \geqslant 0\}$ is linearly independent and so W is also infinite-dimensional.

However the set $S = \{(t^n)_{n=0}^{\infty} \mid t \in \mathbb{R}\}$ is an uncountable linearly independent subset of W and hence W does not have a countable basis. We prove that S is linearly independent below. Suppose that

$$\alpha_1 (t_1^n) + \cdots + \alpha_k (t_k^n) = (0)$$

for real numbrs $\alpha_1, \ldots, \alpha_k, t_1, \ldots, t_k$ with the t_i distinct. Then for all $n \geqslant 0$ we have

$$\alpha_1 t_1^n + \cdots + \alpha_k t_k^n = 0.$$

These equations for $0 \leqslant n < k$ can be rewritten as the single matrix equation

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ t_1 & t_2 & t_3 & \cdots & t_k \\ t_1^2 & t_2^2 & t_3^2 & \cdots & t_k^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1^{k-1} & t_2^{k-1} & t_3^{k-1} & \cdots & t_k^k \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

As the t_i are distinct, then the above $k \times k$ matrix is invertible – this is proved in Linear Algebra II next term. Hence S is an uncountable linearly independent set. No such set exists in V and hence W is not isomorphic to V. ■

6.3 Rank and nullity

Definition 158 *Let V, W be vector spaces. Let $T : V \to W$ be linear. We define the **kernel** (or **null space**) of T to be*

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

*We define the **image** of T to be*

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

Here are some useful properties of kernels and images.

Proposition 159 *Let V, W be vector spaces over $\mathbb{F}$. Let $T : V \to W$ be linear. Then*

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

(b) T is injective if and only if $\ker T = \{0_V\}$.

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

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

Proof. (a) 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, $\ker T \leqslant V$.
$\underline{\operatorname{Im} T}$ 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, $\operatorname{Im} T \leqslant W$.

(b) Say that T is 1–1. If $v \in \ker T$ then $Tv = 0_W = T(0_V)$. By injectivity we have $v = 0_V$ and so $\ker T = \{0_V\}$.

Conversely say that $\ker T = \{0_V\}$ and that $Tv_1 = Tv_2$. Then $T(v_1 - v_2) = Tv_1 - Tv_2 = 0_W$. Then $v_1 - v_2 = 0_V$ and so T is 1–1.

(c) 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

$$w = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) \in \langle T(A) \rangle.$$

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

(d) Assume that V is finite-dimensional. Then $\ker T \leqslant V$ so $\ker T$ is finite-dimensional. Also, $\operatorname{Im} T$ is finite-dimensional by (iii). ■

Corollary 160 *Given a matrix A the image of L_A is* $\operatorname{Col}(A)$, *the column space of A and the image of R_A is* $\operatorname{Row}(A)$, *the row space of A.*

Proof. The canonical basis $\mathbf{e}_1, \ldots, \mathbf{e}_m$ spans $\mathbb{R}^m$ and hence the rows of A, $\mathbf{r}_1 = \mathbf{e}_1 A, \ldots, \mathbf{r}_m = \mathbf{e}_m A$ span $\operatorname{Im} R_A$. Hence $\operatorname{Im} R_A \subseteq \operatorname{Row}(A)$. Conversely if $\mathbf{v} \in \operatorname{Row}(A)$ then

$$\mathbf{v} = \alpha_1 \mathbf{r}_1 + \cdots + \alpha_m \mathbf{r}_m = (\alpha_1 \mathbf{e}_1 + \cdots + \alpha_m \mathbf{e}_m) A \in \operatorname{Im} R_A.$$

Likewise $\operatorname{Im} L_A = \operatorname{Col}(A)$. ■

Definition 161 *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 nullity$(T) := \dim(\ker T)$, and the* ***rank*** *of T to be rank$(T) := \dim(\operatorname{Im} T)$.*

Given the previous corollary, the rank of L_A equals the column rank of A and the rank of R_A equals the row rank of A.

The next theorem is very important!

Theorem 162 (Rank-Nullity Theorem) *Let V, W be vector spaces with V finite-dimensional. Let $T\colon V \to W$ be linear. Then*

$$\dim V = \operatorname{rank}(T) + \operatorname{nullity}(T).$$

Proof. Take a basis $v_1, \ldots, v_n$ for $\ker T$, where $n = \text{nullity}(T)$ and extend this to a basis $v_1, \ldots, v_n, v'_1, \ldots, v'_r$ of V so that $\dim V = n + r$. We claim that $B = \{Tv'_1, \ldots, Tv'_r\}$ is a basis for $\text{Im}\, T$.

B spans $\text{Im}\, T$:

By Proposition 159(c), $T(v_1), \ldots, T(v_n), T(v'_1), \ldots, T(v'_r)$ span $\text{Im}\, T$, being the images of a basis. But $v_1, \ldots, v_n \in \ker T$, meaning $T(v_1) = \cdots = T(v_n) = 0_W$, so in fact $Tv'_1, \ldots, Tv'_r$ span $\text{Im}\, T$.

B is linearly independent:

Take $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$ such that

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

As T is linear, we can rewrite this as $T(\alpha_1 v'_1 + \cdots + \alpha_r v'_r) = 0_W$. So $\alpha_1 v'_1 + \cdots + \alpha_r v'_r \in \ker T$. As $v_1, \ldots, v_n$ is a basis for $\ker T$, there are $\beta_1, \ldots, \beta_n \in \mathbb{F}$ such that

$$\alpha_1 v'_1 + \cdots + \alpha_r v'_r = \beta_1 v_1 + \cdots + \beta_n v_n,$$

But $v_1, \ldots, v_n, v'_1, \ldots, v'_r$ are linearly independent (being a basis for V), so

$$\beta_1 = \cdots = \beta_n = \alpha_1 = \cdots = \alpha_r = 0.$$

This shows $w_1, \ldots, w_r$ are linearly independent and the claim follows.

Now using the claim we have $\text{rank}T = r$, and so $\dim(V) = n + r = \text{nullity}(T) + \text{rank}(T)$. ■

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

Corollary 163 *Let V be a finite-dimensional vector space. Let $T\colon V \to V$ be linear. The following are equivalent:*

(a) T is invertible;

(b) $\text{rank}T = \dim V$;

(c) $\text{nullity}T = 0$.

Proof. (a) $\Rightarrow$ (b):

Assume that T is invertible. Then T is bijective, so is surjective, so $\text{Im}\, T = V$, meaning $\text{rank}T = \dim V$.

(b) $\Rightarrow$ (c):

Assume that $\text{rank}T = \dim V$. Then, by Rank-Nullity, $\text{nullity}T = 0$.

(c) $\Rightarrow$ (a):

Assume that $\text{nullity}T = 0$ so that $\ker T = \{0_V\}$. Then T is injective.

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

So T is bijective, so T is invertible ■

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

Corollary 164 *Let V be a finite-dimensional vector space. Let $T\colon V \to V$ 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 $S\colon V \to V$, so $T \circ S = id_V$. Since id_V is surjective, T is surjective, so $\text{rank}T = \dim V$.

So, by the previous corollary T is invertible, say with two-sided inverse S'.

Then $S' = S' \circ id_V = S' \circ (T \circ S) = (S' \circ T) \circ S = id_V \circ S = S$.

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

If instead we suppose that T has a left inverse $S\colon V \to V$, so $S \circ T = id_V$, then id_V is injective so that T is injective and hence $\text{nullity}T = 0$, and the argument is similar to the previous one. ■

Which, when reworded in terms of matrices, is the following result.

Corollary 165 *Let A, B be square matrices of the same size. If AB is invertible then A and B are invertible.*

Lemma 166 *Let V and W be vector spaces, with V finite-dimensional. Let $T\colon V \to W$ be linear and $U \leqslant V$. Then*

$$\dim U - \text{nullity}T \leqslant \dim T(U) \leqslant \dim U.$$

In particular, if T is injective then $\dim T(U) = \dim U$.

Proof. Let $S\colon U \to W$ be the restriction of T to U (that is, $S(u) = T(u)$ for all $u \in U$). Then S is linear, and $\ker S \leqslant \ker T$ so $\text{nullity}S \leqslant \text{nullity}T$. Also, $\text{Im}\, S = T(U)$. By Rank-Nullity,

$$\begin{aligned}\dim T(U) &= \dim \text{Im}\, S = \dim U - \text{nullity}S \leqslant \dim U;\\ \dim T(U) &= \dim U - \text{nullity}S \geqslant \dim U - \text{nullity}T.\end{aligned}$$

If T is injective, then $\text{nullity}T = 0$, so $\dim T(U) = \dim U$. ■

Remark 167 ***A take on the Rank-Nullity Theorem using matrices.***

Let A be an $m \times n$ matrix and consider its reduced form $\text{RRE}(A)$. *We know that there are as many leading* 1*s as there are non-zero rows and this is the row rank of A. We also know that the kernel of A can be parameterized by assigning parameters to every column which does not have a leading* 1. *Hence*

$$\begin{aligned}&\textit{row rank of } A = \textit{number of leading } 1\textit{s;}\\ &\textit{nullity of } A = \textit{number of columns without leading } 1\textit{s.}\end{aligned}$$

Hence

$$\begin{aligned}n &= \dim \mathbb{R}^n_{\text{col}}\\ &= (\textit{number of leading } 1\textit{s}) + (\textit{number of columns without leading } 1\textit{s})\\ &= (\textit{row rank of } A) + (\textit{nullity of } A)\,.\end{aligned}$$

Corollary 168 *(**Criteria for Invertibility**) Let A be an $n \times n$ matrix. The following statements are equivalent:*

(a) A is invertible.

(b) A has a left inverse.
(c) A has a right inverse.
(d) $\mathrm{Row}(A) = \mathbb{R}^n$.
(e) The columns of A are linearly independent.
(f) The rows of A are linearly independent.
(g) The only solution $\mathbf{x}$ in $\mathbb{R}^n_{\mathrm{col}}$ to the system $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.
(h) The row rank of A is n.
(i) $\mathrm{RRE}(A) = I_n$.

Proof. These are separately left as exercises. Some of the equivalencies have already been demonstrated. ■

7. LINEAR MAPS AND MATRICES

7.1 Representing linear maps with matrices

We saw examples of linear maps arising from multiplying by a matrix: for $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, we defined $L_A\colon \mathbb{R}^n_{\text{col}} \to \mathbb{R}^m_{\text{col}}$ by $L_A(\mathbf{v}) = A\mathbf{v}$, and we defined $R_A\colon \mathbb{R}^m \to \mathbb{R}^n$ by $R_A(\mathbf{v}) = \mathbf{v}A$.

We shall see that linear maps are to matrices, as vectors are to coordinate vectors. Importantly recall that a vector has different coordinates in different coordinate systems and that each choice of coordinates (or basis) associates coordinates with a vector. Similarly given a linear map $T\colon V \to W$ for each choice of coordinates (or bases) for V and W we will see that T is represented by a matrix; change your choice of bases and that matrix will change too!

Definition 169 *Let V be an n-dimensional vector space over $\mathbb{F}$ with an ordered basis $\mathcal{V}$ of vectors $v_1, \ldots, v_n$. Let W be an m-dimensional vector space over $\mathbb{F}$ with an ordered basis $\mathcal{W}$ of vectors $w_1, \ldots, w_m$. So every vector in V and W is represented by a coordinate vector in $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively.*

*Let $T\colon V \to W$ be a linear transformation. We define the **matrix for T with respect to the bases $\mathcal{V}$ and $\mathcal{W}$** to be the matrix which takes the coordinate vector of v to the coordinate vector of Tv.*

More explicitly, this is the $m \times n$ matrix $A = (a_{ij})$ where

$$T(v_i) = \sum_{k=1}^{m} a_{ki} w_k.$$

We will write ${}_{\mathcal{W}}T_{\mathcal{V}}$ for this matrix A.

Remark 170 *Firstly note that this matrix is well-defined. For each $1 \leqslant i \leqslant n$ then $T(v_i)$ can be uniquely expressed as a linear combination of $\mathcal{W}$.*

Remark 171 *Further $a_{1i}, \ldots, a_{mi}$ are the coordinates of $T(v_i)$. These are the entries in the ith column of A. The coordinate column vector of v_i is $\mathbf{e}_i^T$ and the ith column of A is $A\mathbf{e}_i^T$. So we can see that the entries of A are the coordinates of the images of the basis $\mathcal{V}$ as claimed.*

Note that this is what matrices normally do! Given an $m \times n$ matrix A then the first column of A equals $A\mathbf{e}_1$ where $\mathbf{e}_1 = (1, 0, \ldots, 0)$, and more generally the ith column is $A\mathbf{e}_i^T$.

Remark 172 *So if we use $\mathcal{V}$ and $\mathcal{W}$ to identify V and W with $\mathbb{F}^n_{\text{col}}$ and $\mathbb{F}^m_{\text{col}}$ then we identify T with L_A.*

Remark 173 *Importantly in this the bases are listed in an order. If the order of either basis changed then the matrix will change too.*

Remark 174 *If $V = W$ and we use the same ordered basis for both domain and codomain of $T\colon V \to V$, then we talk about the* ***matrix for T with respect to this basis****.*

Example 175 *Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(x, y, z) = (0, x, y)$. This is linear (check!). If we take*

$$\mathbf{i} = (1, 0, 0), \qquad \mathbf{j} = (0, 1, 0), \qquad \mathbf{k} = (0, 0, 1),$$

as the basis $\mathcal{E}$ for both the domain and codomain then we see that

$$T(1, 0, 0) = (0, 1, 0), \qquad T(0, 1, 0) = (0, 0, 1), \qquad T(0, 0, 1) = (0, 0, 0).$$

Hence

$$_{\mathcal{E}}T_{\mathcal{E}} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note that $({}_{\mathcal{E}}T_{\mathcal{E}})^2 = {}_{\mathcal{E}}T_{\mathcal{E}}^2$ and $({}_{\mathcal{E}}T_{\mathcal{E}})^3 = 0 = {}_{\mathcal{E}}T_{\mathcal{E}}^3$ (again check!).

Example 176 *Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(x, y, z) = (0, x, y)$ as in the previous example. Now let $\mathcal{F}$ be the ordered basis*

$$\mathbf{f}_1 = (1, 1, 1), \qquad \mathbf{f}_2 = (1, 1, 0), \qquad \mathbf{f}_3 = (0, 1, 1).$$

Find (i) ${}_{\mathcal{E}}T_{\mathcal{F}}$, (ii) ${}_{\mathcal{F}}T_{\mathcal{E}}$ and (iii) ${}_{\mathcal{F}}T_{\mathcal{F}}$.

Solution. (i) Note that

$$T(1, 1, 1) = (0, 1, 1), \qquad T(1, 1, 0) = (0, 1, 1), \qquad T(0, 1, 1) = (0, 0, 1).$$

Hence

$$_{\mathcal{E}}T_{\mathcal{F}} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

(ii) Note that

$$T(1, 0, 0) = (0, 1, 0) = -\mathbf{f}_1 + \mathbf{f}_2 + \mathbf{f}_3, \qquad T(0, 1, 0) = (0, 0, 1) = \mathbf{f}_1 - \mathbf{f}_2, \qquad T(0, 0, 1) = (0, 0, 0).$$

Hence

$$_{\mathcal{F}}T_{\mathcal{E}} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(iii) Note that

$$T(1, 1, 1) = (0, 1, 1) = \mathbf{f}_3, \qquad T(1, 1, 0) = (0, 1, 1) = \mathbf{f}_3, \qquad T(0, 1, 1) = (0, 0, 1) = \mathbf{f}_1 - \mathbf{f}_2.$$

Hence

$$_{\mathcal{F}}T_{\mathcal{F}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}.$$

■

Proposition 177 *Let V be an n-dimensional vector space over $\mathbb{F}$ with ordered basis $\mathcal{V}$. Let W be an m-dimensional vector space over $\mathbb{F}$ with ordered basis $\mathcal{W}$. Then*

(i) the matrix ${}_{\mathcal{W}}0_{\mathcal{V}}$ of the zero map is $0_{m\times n}$ for any choice of $\mathcal{V}$ and $\mathcal{W}$;

(ii) the matrix ${}_{\mathcal{V}}id_{\mathcal{V}}$ of identity map I_n for any choice of $\mathcal{V}$;

(iii) if $S\colon V \to W$, $T\colon V \to W$ are linear and α, $\beta \in \mathbb{F}$, then

$$ {}_{\mathcal{W}}(\alpha S + \beta T)_{\mathcal{V}} = \alpha\,({}_{\mathcal{W}}S_{\mathcal{V}}) + \beta\,({}_{\mathcal{W}}T_{\mathcal{V}})\,. $$

Proof. These are left as exercises. ■

Theorem 178 *Let U, V, W be finite-dimensional vector spaces over $\mathbb{F}$, of dimensions m, n, p, with ordered bases $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ respectively. Let $S\colon U \to V$ and $T\colon V \to W$ be linear. Let*

$$ A = {}_{\mathcal{V}}S_{\mathcal{U}} \quad \text{and} \quad B = {}_{\mathcal{W}}T_{\mathcal{V}}. $$

Then

$$ BA = {}_{\mathcal{W}}TS_{\mathcal{U}}. $$

Proof. Note that A is $n \times m$ and B is $p \times n$, so the product matrix BA is $p \times m$. Let $\mathcal{U}$ be $u_1, \ldots, u_m$, $\mathcal{V}$ be $v_1, \ldots, v_n$ and $\mathcal{W}$ be $w_1, \ldots, w_p$.

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

$$ S(u_i) = \sum_{j=1}^{n} a_{ji} v_j \text{ for } 1 \leqslant i \leqslant m; $$

$$ T(v_j) = \sum_{k=1}^{p} b_{kj} w_k \text{ for } 1 \leqslant j \leqslant n. $$

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

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

Thus, by definition, ${}_{\mathcal{W}}TS_{\mathcal{U}} = BA$. ■

Remark 179 *This is why we define multiplication of matrices in the way that we do!*

Remark 180 *As we are about to see, this gives a relatively clear and painless proof that matrix multiplication is associative as composition is associative.*

Corollary 181 *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

$$L_A \colon \mathbb{F}^n_{\text{col}} \to \mathbb{F}^m_{\text{col}} \quad \text{and} \quad L_B \colon \mathbb{F}^p_{\text{col}} \to \mathbb{F}^n_{\text{col}} \quad \text{and} \quad L_C \colon \mathbb{F}^q_{\text{col}} \to \mathbb{F}^p_{\text{col}}.$$

With respect to the standard bases of these spaces, the matrices of L_A, L_B, L_C are A, B, C respectively.Hence, by the previous theorem $A(BC)$ and $(AB)C$ are the matrices of

$$L_A \circ (L_B \circ L_C) \colon \mathbb{F}^q_{\text{col}} \to \mathbb{F}^m_{\text{col}}, \quad \text{and} \quad (L_A \circ L_B) \circ L_C \colon \mathbb{F}^q_{\text{col}} \to \mathbb{F}^m_{\text{col}}$$

respectively. But composition of functions is associative, so

$$L_A \circ (L_B \circ L_C) = (L_A \circ L_B) \circ L_C$$

and hence $A(BC) = (AB)C$. ■

Corollary 182 *Let V be a finite-dimensional vector space and let $T \colon V \to V$ be an invertible linear transformation. Let A be the matrix of T with respect to an ordered basis (for both domain and codomain). Then A is invertible, and A^{-1} is the matrix of T^{-1} with respect to the same basis.*

Proof. Exercise. ■

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 183 *Define $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x, y) = (2x + y, 3x - 2y)$. To find the matrix of T with respect to the standard ordered basis $\mathcal{E}$, note that*

$$T(1, 0) = (2, 3) \qquad \textit{and} \qquad T(0, 1) = (1, -2)$$

so the matrix for T with respect to this basis is

$$_{\mathcal{E}}T_{\mathcal{E}} = \begin{pmatrix} 2 & 1 \\ 3 & -2 \end{pmatrix}.$$

That is $T = L_A$. Let $\mathbf{f}_1 = (1, -2)$ and $\mathbf{f}_2 = (-2, 5)$. Then $\mathbf{f}_1$, $\mathbf{f}_2$ is an ordered basis of $\mathbb{R}^2$ which we will denote as $\mathcal{F}$. Note that

$$T(\mathbf{f}_1) = (0, 7) = 14\mathbf{f}_1 + 7\mathbf{f}_2$$
$$T(\mathbf{f}_2) = (1, -16) = -27\mathbf{f}_1 - 14\mathbf{f}_2$$

so the matrix for T with respect to this basis is

$$_{\mathcal{F}}T_{\mathcal{F}} = \begin{pmatrix} 14 & -27 \\ 7 & -14 \end{pmatrix}.$$

How are these two matrices related? Well, by Theorem 178, we can see that

$$_{\mathcal{F}}T_{\mathcal{F}} = \left(_{\mathcal{F}}I_{\mathcal{E}}\right)\left(_{\mathcal{E}}T_{\mathcal{E}}\right)\left(_{\mathcal{E}}I_{\mathcal{F}}\right).$$

*The matrix $_{\mathcal{E}}I_{\mathcal{F}}$ represents the identity transformation, so it is does not change vectors; however it changes the coordinate vector for a vector with respect to some basis $\mathcal{F}$ to the coordinate vector for the **same** vector with respect to a **different** basis $\mathcal{E}$. Note that the inverse of this matrix is $_{\mathcal{F}}I_{\mathcal{E}}$.*

We can take $\mathbf{f}_1$, $\mathbf{f}_2$ and write them with respect to $\mathbf{e}_1$, $\mathbf{e}_2$: we have

$$\mathbf{f}_1 = \mathbf{e}_1 - 2\mathbf{e}_2, \qquad \mathbf{f}_2 = -2\mathbf{e}_1 + 5\mathbf{e}_2$$

so we get a 'change of basis matrix'

$$_{\mathcal{E}}I_{\mathcal{F}} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}.$$

If this matrix is applied to $(1,0)^T$ then this coordinate vector represents $\mathbf{f}_1$. The image of the coordinate vector $(1,0)^T$ is $(1,-2)^T$ which represents $\mathbf{e}_1 - 2\mathbf{e}_2$. But this is of course the same vector! This vector just has different coordinates with respect to the bases $\mathcal{E}$ and $\mathcal{F}$.

It is then the case that

$$_{\mathcal{F}}I_{\mathcal{E}} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix},$$

which represents that

$$\mathbf{e}_1 = 5\mathbf{f}_1 + 2\mathbf{f}_2, \qquad \mathbf{e}_2 = 2\mathbf{f}_1 + \mathbf{f}_2.$$

And we can verify that

$$\begin{aligned} &\left(_{\mathcal{F}}I_{\mathcal{E}}\right)\left(_{\mathcal{E}}T_{\mathcal{E}}\right)\left(_{\mathcal{E}}I_{\mathcal{F}}\right) \\ &= \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 3 & -2 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 16 & 1 \\ 7 & 0 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 14 & -27 \\ 7 & -14 \end{pmatrix} = {}_{\mathcal{F}}T_{\mathcal{F}} \end{aligned}$$

as expected.

Corollary 184 *(**Change of basis theorem**) Let V be a finite-dimensional vector space over $\mathbb{F}$ with ordered bases $\mathcal{V}$, $\mathcal{V}'$. Let W be a finite-dimensional vector space over $\mathbb{F}$ with ordered bases $\mathcal{W}$, $\mathcal{W}'$. Let $T\colon V \to W$ be a linear map. Then*

$$_{\mathcal{W}'}T_{\mathcal{V}'} = \left(_{\mathcal{W}'}I_{\mathcal{W}}\right)\left(_{\mathcal{W}}T_{\mathcal{V}}\right)\left(_{\mathcal{V}}I_{\mathcal{V}'}\right).$$

Proof. This is an immediate corollary to Theorem 178. ■

Corollary 185 *(**Change of basis theorem version 2**) Let V be a finite-dimensional vector space over $\mathbb{F}$ with ordered bases $\mathcal{V}$, $\mathcal{V}'$ and let $T\colon V \to V$ be a linear map. Then*

$$_{\mathcal{V}'}T_{\mathcal{V}'} = \left(_{\mathcal{V}'}I_{\mathcal{V}}\right)\left(_{\mathcal{V}}T_{\mathcal{V}}\right)\left(_{\mathcal{V}}I_{\mathcal{V}'}\right).$$

If we set $A = {}_{\mathcal{V}'}T_{\mathcal{V}'}, B = {}_{\mathcal{V}}T_{\mathcal{V}}$ and $P = {}_{\mathcal{V}}I_{\mathcal{V}'}$ then note

$$A = P^{-1}BP.$$

Proof. This is a special case of the previous corollary. ■

Definition 186 *Take A, $B \in \mathcal{M}_{n\times n}(\mathbb{F})$. If there is an invertible $n \times n$ matrix P such that $A = P^{-1}BP$, then we say that A and B are **similar**. Similarity is then an equivalence relation.*

Remark 187 *So two matrices representing the same linear transformation from a finite-dimensional vector space to itself, but with respect to different bases, are similar.*

Remark 188 ***Properties of Linear Maps*** *As many different matrices can represent the **same** linear transformation $T\colon V \to V$ it would be disturbing if different conclusions about the properties of T could be determined by using different matrix representatives. For example, if we said a linear map T is invertible if a matrix representative of it is invertible, could T end up being invertible and not invertible? Reassuringly the answer is no.*

Let $A = {}_{\mathcal{V}}T_{\mathcal{V}}$ and $B = {}_{\mathcal{W}}T_{\mathcal{W}}$ be matrices representing T with respect to two bases, so that $A = P^{-1}BP$ for some invertible P. Then

- *A is invertible if and only if B is invertible.*
- *The trace of A equals the trace of B. [This follows from the identity* $\operatorname{trace}(MN) = \operatorname{trace}(NM)$*.]*
- *A functional identity satisfied by A, such as $A^2 = A$, is also satisfied by B.*
- *The determinant of A equals the determinant of B. [Determinants will be formally defined in Linear Algebra II next term.]*
- *The eigenvalues of A equal the eigenvalues of B. [Eigenvalues will be formally defined in Linear Algebra II next term.]*

Thus we may, in a well-defined fashion, refer to the invertibility, trace, determinant of a linear map.

Note we cannot, in a well-defined manner, refer to the transpose of a linear map. If $A = P^{-1}BP$ then it need not be the case that $A^T = P^{-1}B^TP$. Though you might note that this is true if P is orthogonal! (This is something that will be addressed when you meet adjoints in the second year.)

7.3 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. It makes sense to ask if rowrank(A) and colrank(A) related?

Remark 189 *From the definitions, we see that* $\mathrm{Col}(A) = \mathrm{Row}(A^T)$ *and so* $\mathrm{colrank}(A) = \mathrm{rowrank}(A^T)$. *Similarly,* $\mathrm{Row}(A) = \mathrm{Col}(A^T)$ *and so* $\mathrm{rowrank}(A) = \mathrm{colrank}(A^T)$.

We first prove the following:

Lemma 190 *The linear system* $(A|\mathbf{b})$ *is consistent if and only if* $\mathrm{Col}(A|\mathbf{b}) = \mathrm{Col}(A)$.

Proof. Say that A is $m \times n$ and denote the columns of A as $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$. Then

$$\begin{aligned}
(A|\mathbf{b}) \text{ is consistent} &\iff A\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{F}^n_{\text{col}} \\
&\iff x_1\mathbf{c}_1 + \cdots x_n\mathbf{c}_n = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{F}^n_{\text{col}} \\
&\iff \mathbf{b} \in \mathrm{Col}(A) \\
&\iff \mathrm{Col}(A|\mathbf{b}) = \mathrm{Col}(A).
\end{aligned}$$

■

Theorem 191 *The column rank of a matrix equals its row rank.*

Proof. We prove this by induction on the number of columns in the matrix. A non-zero $m \times 1$ matrix has column rank 1 and also row rank 1 as the matrix reduces to $\mathbf{e}_1^T$; the column rank and row rank of $0_{m\times 1}$ are both 0. So the $n = 1$ case is true. Suppose, as our inductive hypothesis, that column rank and row rank are equal for $m \times n$ matrices. Any $m \times (n+1)$ matrix $(A|\mathbf{b})$ can be considered as an $m \times n$ matrix A alongside $\mathbf{b}$ in $\mathbb{F}^m_{\text{col}}$. If the system $(A|\mathbf{b})$ is consistent then

$$\begin{aligned}
\mathrm{colrank}(A|\mathbf{b}) &= \mathrm{colrank}(A) && \text{[by previous lemma]} \\
&= \mathrm{rowrank}(A) && \text{[by inductive hypothesis]} \\
&= \mathrm{rowrank}(A|\mathbf{b}) && \text{[see Remark 45].}
\end{aligned}$$

On the other hand, if the system $(A|\mathbf{b})$ has no solutions then

$$\begin{aligned}
\mathrm{colrank}(A|\mathbf{b}) &= (\mathrm{colrank}A) + 1 && \text{as } \mathbf{b} \notin \mathrm{Col}(A) \\
&= (\mathrm{rowrank}A) + 1 && \text{[by inductive hypothesis]} \\
&= \mathrm{rowrank}(A|\mathbf{b}) && \text{[see Remark 45].}
\end{aligned}$$

So if the system is consistent the row rank and column rank maintain their common value. If inconsistent, then $\mathbf{b}$ adds a further dimension to the column space and $(0\,0\cdots 0\,|\,1)$ adds an extra dimension to the row space. Either way the column rank and row rank of $(A|\mathbf{b})$ still agree and the proof follows by induction. ■

We provide here a second proof of the result, as it takes a somewhat different approach.

Theorem 192 *Let $P = QR$ where P, Q, R are respectively $k \times l$, $k \times m$, $m \times l$ matrices over the same field.*
(a) Then the row rank of P is at most m.
(b) Let $p = \text{rowrank}P$. Then P may be written as the product of a $k \times p$ matrix and a $p \times l$ matrix.
(c) The row rank and column rank of a matrix are equal.

Solution. (a) By Proposition 117(c) we have

$$\text{Row}(P) = \text{Row}(QR) \leqslant \text{Row}(R).$$

Hence

$$\text{rowrank}(P) = \dim \text{Row}(P) \leqslant \dim \text{Row}(R) = \text{rowrank}(R) \leqslant m.$$

(b) There is a $k \times k$ invertible matrix E such that $EP = \text{RRE}(P)$ and hence

$$P = E^{-1}\text{RRE}(P).$$

Now let $\tilde{E}$ denote the first p columns of E^{-1} and $\tilde{P}$ denote the first p rows of $\text{RRE}(P)$. As the last $k - p$ rows of $\text{RRE}(P)$ are zero rows, we still have

$$P = \tilde{E}\tilde{P},$$

where $\tilde{E}$ is a $k \times p$ and $\tilde{P}$ is a $p \times l$ matrix.

(c) From (a) and (b) we know that the row rank of a $k \times l$ matrix P is the minimal value p such that P can be written as the product QR of a $k \times p$ matrix and a $p \times l$ matrix. Whenever $P = QR$ then

$$\underset{l \times k}{P^T} = \underset{l \times p}{R^T}\,\underset{p \times k}{Q^T}.$$

So the row rank of P^T is similarly p. But the $\text{rowrank}P^T = \text{colrank}P$ as required. ■

8. INNER PRODUCT SPACES

On first meeting vector spaces, it is quite natural to think of $\mathbb{R}^n$ as a typical example. However, as has already been commented, it is important to appreciate that $\mathbb{R}^n$ has a lot of structure beyond being just a real vector space. It has coordinates already assigned (and so a canonical basis) and distances and angles can be measured, for example using the dot (or scalar) product. Vector spaces, in general, have none of this extra structure.

The dot product is an example of an inner product; an inner product is a means of measuring distance and angles within a vector space. A vector space together with an inner product is called an *inner product space.* Initially we will consider inner products only on real vector spaces, but we will later discuss complex inner product spaces. Inner products appear in many areas of mathematics and they have particular importance in Fourier series and in quantum theory.

8.1 Bilinear forms

Definition 193 *Let V be a vector space over $\mathbb{F}$. A **bilinear form** B on V is a function $B\colon V \times V \to \mathbb{F}$, such that*

(a) $B(\alpha_1 v_1 + \alpha_2 v_2, v_3) = \alpha_1(v_1, v_3) + \alpha_2(v_2, v_3)$ *for all* v_1, v_2, $v_3 \in V$ *and* α_1, $\alpha_2 \in \mathbb{F}$;

(b) $B(v_1, \alpha_2 v_2 + \alpha_3 v_3) = \alpha_2(v_1, v_2) + \alpha_3(v_1, v_3)$ *for all* v_1, v_2, $v_3 \in V$ *and* α_2, $\alpha_3 \in \mathbb{F}$.

(a) says that B is linear in the first variable (when we fix the second variable), and (b) says the same for the second variable.

Example 194 *For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{F}^n$, we define*

$$B(\mathbf{x}, \mathbf{y}) = x_1 y_1 + \cdots + x_n y_n.$$

This gives a bilinear form. In $\mathbb{R}^n$ this is the familiar dot product, or scalar product, often written $\mathbf{x} \cdot \mathbf{y}$.

Example 195 *Take $A \in \mathcal{M}_{n\times n}(\mathbb{F})$. Note for $\mathbf{x}$, $\mathbf{y} \in \mathbb{F}^n$ that $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}A\mathbf{y}^T$ defines a bilinear form on $\mathbb{F}^n$.*

Note that the usual scalar product is an example of this in the special case that $A = I_n$, because $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}\mathbf{y}^T$. (Officially, $\mathbf{x}A\mathbf{y}^T$ is a 1×1 matrix, not an element of $\mathbb{F}$, but it is completely natural to identify 1×1 matrices with scalars).

And for $\mathbf{x}$, $\mathbf{y} \in \mathbb{F}^n_{\text{col}}$ then $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A\mathbf{y}$ defines a bilinear form on $\mathbb{F}^n_{\text{col}}$.

Definition 196 *Let V be a vector space over $\mathbb{F}$ and let B be a bilinear form on V. Take $v_1, \ldots, v_n \in V$. The **Gram matrix** of B with respect to $v_1, \ldots, v_n$ is the $n \times n$ matrix $(B(v_i, v_j)) \in \mathcal{M}_{n\times n}(\mathbb{F})$.*

Proposition 197 *Let V be a finite-dimensional vector space over $\mathbb{F}$ and let $v_1, \ldots, v_n$ be a basis for V. Let B be a bilinear form on V and let $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ be the associated Gram matrix. For $X, Y \in V$, let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{F}^n$ be the unique coordinate vectors such that*

$$X = x_1v_1 + \cdots + x_nv_n \qquad \text{and} \qquad Y = y_1v_1 + \cdots + y_nv_n.$$

Then $B(X, Y) = \mathbf{x}A\mathbf{y}^T$.

Remark 198 *Consequently the bilinear form in Example 195 essentially describes **all** bilinear forms on V. Note that if A is the Gram matrix of a bilinear form, then any other Gram matrix of B (that is, a Gram matrix with respect to a different basis) equals P^TAP where $P \in \mathcal{M}_{n\times n}(\mathbb{F})$ is invertible.*

Proof. We have

$$\begin{aligned}
B(X, Y) &= B\left(\sum_{i=1}^{n} x_iv_i, \sum_{j=1}^{n} y_jv_j\right) \\
&= \sum_{i=1}^{n} x_iB\left(v_i, \sum_{j=1}^{n} y_jv_j\right) && \text{[using linearity in the first entry]} \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_jB(v_i, v_j) && \text{[using linearity in the second entry]} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_ia_{ij}y_j \\
&= xAy^T.
\end{aligned}$$

■

Definition 199 *We say that a bilinear form $B\colon V \times V \to \mathbb{F}$ is **symmetric** if*

$$B(v_1, v_2) = B(v_2, v_1) \qquad \text{for all} \qquad v_1, v_2 \in V.$$

Note that a bilinear form is symmetric if and only if any Gram matrix of the bilinear form is symmetric.

8.2 Inner product spaces

Definition 200 *Let V be a real vector space. We say that a bilinear form $B\colon V \times V \to \mathbb{R}$ is **positive definite** if $B(v, v) \geqslant 0$ for all $v \in V$, with $B(v, v) = 0$ if and only if $v = 0$. N.B. we are defining **real** inner product spaces here; the requirement that $B(v, v) \geqslant 0$ does not make sense in a general field.*

Definition 201 *An **inner product** on a real vector space V is a positive definite, symmetric, bilinear form on V. Inner products are usually denoted $\langle x, y\rangle$ rather than $B(x, y)$.*

*We say that a real vector space is an **inner product space** if it is equipped with an inner product. Unless otherwise specified, $\langle -, -\rangle$ will denote an inner product, rather than a general bilinear form.*

Example 202 *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 $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\mathbf{x} \neq \mathbf{0}$, then*

$$\mathbf{x} \cdot \mathbf{x} = x_1^2 + \cdots + x_n^2 > 0,$$

*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 203 *Let $V = \mathbb{R}_n[x]$, the vector space of polynomials of degree $\leqslant n$. For f, $g \in V$, define*

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x$$

where $a < b$. Then $\langle -, -\rangle$ is bilinear – as integration is linear – and symmetric – as the integrand is symmetric in f and g.

If $f \in V$ and $f \neq 0$, then $f(x) = 0$ for only finitely many x in $[a, b]$, and $(f(x))^2 > 0$ at other x, and we find that

$$\langle f, f\rangle = \int_a^b f(x)^2\,\mathrm{d}x > 0.$$

So $\langle -, -\rangle$ is positive definite.

Hence $\langle -, -\rangle$ is an inner product on V. In fact, more generally, $\langle -, -\rangle$ defines an inner product on the space $C[a, b]$ of continuous real-valued functions on the interval $[a, b]$.

Importantly, inner products allow us to define length and angle, something which is not possible with the structure of a vector space alone.

Definition 204 *Let V be an inner product space. For $v \in V$, we define the **norm** (or **magnitude** or **length**) of v to be*

$$\|v\| := \sqrt{\langle v, v\rangle}.$$

The distance between two vectors v, $w \in V$ is defined to be

$$d(v, w) = \|v - w\|.$$

Proposition 205 *The norm $\| - \|$ has the following properties; for $v, w \in V$ and $\alpha \in \mathbb{R}$,*

(a) $\|v\| \geqslant 0$ and $\|v\| = 0$ if and only if $v = 0_V$.

(b) $\|\alpha v\| = |\alpha|\,\|v\|$.

*(c) $\|v + w\| \leqslant \|v\| + \|w\|$. This is known as the **triangle inequality**.*

Proof. (a) and (b) are straightforward. To prove (c) we will first prove the *Cauchy-Schwarz inequality.* ■

Proposition 206 ***(Cauchy-Schwarz Inequality)*** *For v, w in an inner product space V, then*

$$|\langle v, w\rangle| \leqslant \|v\| \, \|w\| \, .$$

Equality holds if and only if v and w are linearly dependent.

Proof. If $w = 0$ then the result is immediate, so assume that $w \neq 0$. For $t \in \mathbb{R}$, note that

$$\begin{aligned} 0 &\leqslant \|v + tw\|^2 \\ &= \langle v + tw, v + tw\rangle \\ &= \langle v, v\rangle + 2t\langle v, w\rangle + t^2\langle w, w\rangle \qquad \text{[by linearity and symmetry]} \\ &= \|v\|^2 + 2t\langle v, w\rangle + t^2 \|w\|^2 \, . \end{aligned}$$

As $\|w\| \neq 0$, the last line is a quadratic in t which is always non-negative. So it either has complex roots or a repeated real root, meaning its discriminant is non-positive. So

$$\text{discriminant} = 4\langle v, w\rangle^2 - 4 \|w\|^2 \|v\|^2 \leqslant 0$$

and the Cauchy-Schwarz inequality follows. For equality, the discriminant has to be zero which means there is a repeated real root $t = t_0$. But then $\|v + t_0 w\| = 0$ and hence $v + t_0 w = 0_V$ showing that v and w are linearly dependent. The converse is immediate. ■

Proof. Continuing the proof of Proposition 205 (c): we now see

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w\rangle \\ &= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &\leqslant \|v\|^2 + 2 \, |\langle v, w\rangle| + \|w\|^2 \\ &\leqslant \|v\|^2 + 2 \, \|v\| \, \|w\| + \|w\|^2 \qquad \text{[by the Cauchy-Schwarz inequality]} \\ &= (\|v\| + \|w\|)^2 \end{aligned}$$

and the triangle inequality follows. ■

Proposition 207 *The distance function $d(v, w) = \|v - w\|$ satisfies the following properties: for $u, v, w \in V$ we have*

(a) $d(v, w) \geqslant 0$ and $d(v, w) = 0$ if and only if $v = w$.
(b) $d(v, w) = d(w, v)$.
(c) $d(u, w) \leqslant d(u, v) + d(v, w)$.

Here (a), (b), (c) show d has the properties of a ***metric****.*

Proof. These properties follow straightforwardly from the properties of the norm. ■

You may recall that in $\mathbb{R}^2$ or $\mathbb{R}^3$ we have

$$\mathbf{x} \cdot \mathbf{y} = \|x\| \|y\| \cos\theta,$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. 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

$$\cos^{-1}\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right).$$

Note by the Cauchy-Schwarz inequality that this is well-defined as

$$\left|\frac{\langle x, y\rangle}{\|x\|\|y\|}\right| \leqslant 1$$

for any nonzero vectors x, y in an inner product space.

Example 208 *Let m, n be integers. Show that $\sin mx$ and $\cos nx$ are perpendicular in $C[-\pi, \pi]$ with the inner product from Example 203. Show also that $\cos mx$ perpendicular to $\cos nx$ when $m \neq n$ and find $\|\cos mx\|$.*

Solution. Note that

$$\langle \sin mx, \cos nx\rangle = \int_{-\pi}^{\pi} \sin mx \cos nx \,\mathrm{d}x = 0$$

as the integrand is odd. For the second part, recall the trigonometric identity

$$\cos mx \cos nx = \frac{1}{2}[\cos(m+n)x + \cos(m-n)x].$$

So if $m \neq n$ then

$$\begin{aligned}\langle \cos mx, \cos nx\rangle &= \frac{1}{2}\int_{-\pi}^{\pi} \cos(m+n)x + \cos(m-n)x \,\mathrm{d}x \\ &= \frac{1}{2}\left[\frac{\sin(m+n)x}{m+n} + \frac{\sin(m-n)x}{m-n}\right]_{-\pi}^{\pi} \\ &= 0.\end{aligned}$$

If $m = n \neq 0$ then

$$\|\cos mx\|^2 = \langle \cos mx, \cos mx\rangle = \frac{1}{2}\int_{-\pi}^{\pi} (\cos 2mx + 1)\ \mathrm{d}x = \pi,$$

and if $m = n = 0$ then

$$\|1\|^2 = \langle 1, 1\rangle = \int_{-\pi}^{\pi} \mathrm{d}x = 2\pi.$$

■

Remark 209 *The above orthogonality relations are crucial in the study of **Fourier series**. If we can represent a function on $-\pi < x < \pi$ as a Fourier series*

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx),$$

then, provided the integration and infinite sum can be interchanged, we would have

$$\begin{aligned}\int_{-\pi}^{\pi} f(x)\sin lx\,\mathrm{d}x &= \int_{-\pi}^{\pi}\left(\frac{1}{2}a_0\sin lx+\sum_{k=1}^{\infty}(a_k\cos kx\sin lx+b_k\sin kx\sin lx)\right)\mathrm{d}x\\ &= \left(\int_{-\pi}^{\pi}\frac{1}{2}a_0\sin lx\,\mathrm{d}x\right)+\sum_{k=1}^{\infty}a_k\left(\int_{-\pi}^{\pi}\cos kx\sin lx\,\mathrm{d}x\right)+\sum_{k=1}^{\infty}b_k\left(\int_{-\pi}^{\pi}\sin kx\sin lx\,\mathrm{d}x\right)\\ &= 0+\sum_{k=1}^{\infty}a_k\times 0+\sum_{k=1}^{\infty}b_k\pi\delta_{kl}\\ &= \pi b_l.\end{aligned}$$

Hence

$$b_l=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin lx\,\mathrm{d}x \qquad \textit{for} \qquad l\geqslant 1.$$

Similar calculations show that

$$a_l=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos lx\,\mathrm{d}x \qquad \textit{for} \qquad l\geqslant 0.$$

These are the Fourier coefficients of $f(x)$. Validating convergency issues and interchanging the integration and infinite sum are difficult matters of analysis.

8.3 Orthogonal Maps

Definition 210 *A linear map $\alpha\colon V\to V$ of an inner product space V is said to be **orthogonal** if*

$$\langle\alpha(v),\alpha(w)\rangle=\langle v,w\rangle$$

for all $v,w\in V$.

Proposition 211 *Let $\alpha\colon\mathbb{R}^n_{\mathrm{col}}\to\mathbb{R}^n_{\mathrm{col}}$ be a linear map and let A denote the matrix of α with respect to the standard basis. Then α is orthogonal with respect to the dot product if and only if A is an orthogonal matrix.*

Proof. Suppose that α is orthogonal with respect to the dot product. Denote the standard basis as $\mathbf{e}_1,\ldots,\mathbf{e}_n$. Then

$$\delta_{ij}=\mathbf{e}_i\cdot\mathbf{e}_j=\alpha(\mathbf{e}_i)\cdot\alpha(\mathbf{e}_j)=(A\mathbf{e}_i)^T(A\mathbf{e}_j)=\mathbf{e}_i^TA^TA\mathbf{e}_j.$$

Now $\mathbf{e}_i^TA^TA\mathbf{e}_j$ is the (i,j)th entry of A^TA, and this equals δ_{ij} which is the (i,j)th entry of I_n. Hence $A^TA=I_n$ as this is true for all i,j.

Reversing the implications of the above argument takes us from $A^TA = I_n$ to $\alpha(\mathbf{e}_i)\cdot\alpha(\mathbf{e}_j) = \delta_{ij}$. But then by linearity

$$\begin{aligned}\alpha\left(\sum_i u_i\mathbf{e}_i\right)\cdot\alpha\left(\sum_j v_j\mathbf{e}_j\right) &= \sum_i\sum_j u_iv_j\alpha(\mathbf{e}_i)\cdot\alpha(\mathbf{e}_j)\\ &= \sum_i\sum_j u_iv_j\delta_{ij}\\ &= \sum_i u_iv_i\\ &= \left(\sum_i u_i\mathbf{e}_i\right)\cdot\left(\sum_j v_j\mathbf{e}_j\right)\end{aligned}$$

and so α is orthogonal. ■

Proposition 212 *An orthogonal map is an isometry of an inner product space.*

Proof. Say α is orthogonal. Then, by linearity,

$$d(\alpha(v),\alpha(w))^2 = \|\alpha(v-w)\|^2 = \langle\alpha(v-w),\alpha(v-w)\rangle = \langle v-w, v-w\rangle = \|v-w\|^2 = d(v,w)^2.$$

Hence α is an isometry. In fact, it can be shown that any linear isometry of a finite-dimensional vector space is orthogonal. (This is proven in the Geometry course for $\mathbb{R}^n$.) ■

Definition 213 *Let V be an inner product space. We say that $\{v_1,\ldots,v_k\}\subseteq V$ is an **orthonormal set** if for all i, j we have*

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

So the vectors are of unit length and are mutually perpendicular.

Lemma 214 *In an inner product space V, an orthonormal set is linearly independent.*

Proof. Say $\{v_1,\ldots,v_k\}$ is orthonormal and $\alpha_1,\ldots,\alpha_k\in\mathbb{R}$ such that $\alpha_1v_1+\cdots+\alpha_kv_k = 0_V$. Then for $1\leqslant i\leqslant k$ we have

$$\begin{aligned}0 = \langle 0_V, v_i\rangle &= \langle\alpha_1v_1+\cdots+\alpha_kv_k, v_i\rangle\\ &= \alpha_1\langle v_1,v_i\rangle+\cdots+\alpha_k\langle v_k,v_i\rangle\\ &= \alpha_i\end{aligned}$$

so $\alpha_1 = \cdots = \alpha_k = 0$. ■

Remark 215 *Note then that n orthonormal vectors in an n-dimensional inner product space is an orthonormal basis. It is the case that every finite-dimensional inner product space has an orthonormal basis, but this result will be proved in Linear Algebra II.*

Recall 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 $X^{-1} = X^T$.

Proposition 216 *Take $X \in \mathcal{M}_{n\times n}(\mathbb{R})$. Consider $\mathbb{R}^n$ (or $\mathbb{R}^n_{\text{col}}$) equipped with the usual inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y}$. The following are equivalent:*

(a) $XX^T = I_n$;
(b) $X^TX = I_n$;
(c) the rows of X form an orthonormal basis of $\mathbb{R}^n$;
(d) the columns of X form an orthonormal basis of $\mathbb{R}^n_{\text{col}}$;
(e) for all $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n_{\text{col}}$, we have $X\mathbf{x}\cdot X\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$.

Proof. (a) $\Leftrightarrow$ (b): For any A, $B \in \mathcal{M}_{n\times n}(\mathbb{R})$, we have $AB = I_n$ if and only if $BA = I_n$.
(a) $\Leftrightarrow$ (c): Say the rows of X are $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Note that the (i,j)th entry of XX^T is $\mathbf{x}_i \cdot \mathbf{x}_j$. But $XX^T = I_n$ if and only if the (i,j) entry of XX^T is δ_{ij}, i.e. if and only if the rows are orthonormal. As there are n rows then the rows further form an orthonormal basis.
(b) $\Leftrightarrow$ (d): Say the columns of X are $\mathbf{y}_1, \ldots, \mathbf{y}_n$. We see that the (i,j)th entry of X^TX is $\mathbf{y}_i \cdot \mathbf{y}_j$. The remainder of the argument is as given above.
(b) $\Rightarrow$ (e): Recall that we can identify $\mathbf{x}\cdot\mathbf{y}$ with $\mathbf{x}^T\mathbf{y}$. Assume that $X^TX = I_n$ and take $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n_{\text{col}}$. Then

$$\begin{aligned}(X\mathbf{x})\cdot(X\mathbf{y}) &= (X\mathbf{x})^T(X\mathbf{y})\\ &= (\mathbf{x}^TX^T)(X\mathbf{y})\\ &= \mathbf{x}^T(X^TX)\mathbf{y}\\ &= \mathbf{x}^TI_n\mathbf{y}\\ &= \mathbf{x}^T\mathbf{y}\\ &= \mathbf{x}\cdot\mathbf{y}.\end{aligned}$$

(e) $\Rightarrow$ (d): Assume that $X\mathbf{x}\cdot X\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n_{\text{col}}$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n_{\text{col}}$. But then

$$\delta_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j = X\mathbf{e}_i \cdot X\mathbf{e}_j.$$

And so $X\mathbf{e}_1, \ldots, X\mathbf{e}_n$, which are the columns of X, form an orthonormal basis. ■

Remark 217 *Condition (e) says that the map $R_X\colon \mathbb{R}^n \to \mathbb{R}^n$ sending $\mathbf{x}$ to $\mathbf{x}X$ 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 the previous proposition says that X is orthogonal if and only if the map R_X is an isometry.*

8.4 Complex inner product spaces

Whilst the above theory of inner products applies very much to real vector spaces, rather than to vector spaces over a general field, the theory can be adapted and extended to vector

spaces over $\mathbb{C}$. However, the usual dot product on $\mathbb{C}^n$ isn't an inner product: it is bilinear and symmetric but we'd find in $\mathbb{C}^2$ that

$$\|(1,i)\|^2 = 1^2 + i^2 = 0$$

even though $(1,i) \neq (0,0)$. We can avoid this problem by defining the standard inner product on $\mathbb{C}^n$ to be

$$(z_1,\ldots,z_n)\cdot(w_1,\ldots,w_n) = \sum_{i=1}^{n} z_i\overline{w_i}.$$

We then have that

$$(z_1,\ldots,z_n)\cdot(z_1,\ldots,z_n) = \sum_{i=1}^{n} z_i\overline{z_i} = \sum_{i=1}^{n} |z_i|^2$$

which is non-negative and zero if and only if $(z_1,\ldots,z_n) = \mathbf{0}$.

Note that this form is linear in the first variable, positive definite, but "conjugate symmetric" in the sense that

$$\langle v_1, v_2\rangle = \overline{\langle v_2, v_1\rangle}.$$

Definition 218 *Let V be a complex vector space. A function $\langle -,-\rangle\colon V\times V\to\mathbb{C}$ is a **sesquilinear form** if*

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

(b) $\langle v_1, v_2\rangle = \overline{\langle v_2, v_1\rangle}$ 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 $\langle v, v\rangle \geqslant 0$ 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 219 *The prefix* sesqui- *relates to "$1\frac{1}{2}$ times"; for example a sesquicentennary is 150 years.*

Remark 220 *Positive definite sesquilinear forms are often called **Hermitian forms**, and complex inner product spaces are often called **Hermitian spaces**.*

Remark 221 *The equivalent of orthogonal maps for real inner product spaces are the **unitary** maps. That is, a linear map $U\colon V\to V$ of a complex inner product space V is unitary if $\langle Ux, Uy\rangle = \langle x, y\rangle$ for all $x, y \in V$. A unitary matrix is a square matrix such that $UU^* = I = U^*U$ where $U^* = \overline{U}^T$.*

Should you study quantum theory later, then you will see that the theory is generally set within complex inner product spaces. The wave function ψ of a particle is complex-valued and its norm-squared $\|\psi\|^2 = \psi\overline{\psi}$ is a probability density function.

You will explore inner product spaces further in Linear Algebra II and Part A Linear Algebra.