UNIVERSITY OF STRATHCLYDE
DEPARTMENT OF MATHEMATICS AND STATISTICS

MM201 Linear Algebra and Differential Equations

Contents

1 Matrices and linear equations **3**

1.1 Introduction 3

1.2 Linear equations 3

1.3 Matrices 5

1.4 Systems of linear equations 8

1.5 Determinants 11

2 Vector spaces, algebras, linear algebra **15**

2.1 Introduction 15

2.2 Sets 15

2.3 The set of $m \times n$ matrices 17

2.4 Vector spaces 18

2.5 Vectors in n dimensions 18

2.6 Other examples of vector spaces 19

2.7 Euclidean n-space 21

2.8 Algebras 24

2.9 Transformations and linearity . . . 24

3 Subspaces, Bases & Dimension **26**

3.1 Subspaces . . . 26

3.2 Linear combinations and spanning sets . . . 27

3.3 Linear independence . . . 28

3.4 Basis and dimension . . . 29

3.5 Orthogonal bases . . . 31

3.6 The Gram-Schmidt orthogonalisation procedure . . . 32

4 Matrix-Related Subspaces **34**

4.1 Nullspace . . . 34

4.2 Row and column spaces; rank . . . 35

4.3 Summary of situation for $n \times n$ systems . . . 38

5 Eigenvalues, eigenvectors and diagonalisation **41**

5.1 Eigenvalues and eigenvectors . . . 41

5.2 Diagonalisation of matrices . . . 42

5.3 Results for symmetric matrices . . . 44

1 Matrices and linear equations

1.1 Introduction

Linear algebra is the study of **matrices**, their properties and many applications. Why the study of matrices is called **Linear Algebra** will become clear later; in mathematics all words, such as "linear" and "algebra", have very precise meanings which it takes time and effort to develop and explain.

Linear algebra is a very important branch of mathematics. Though no area of mathematics is thinkable without it, its main application areas are in operations research and optimisation, statistics and probability, numerical analysis and mathematical physics. Linear algebra is also an important prerequisite for learning about multivariate calculus, complex variables, integration, functional analysis and **differential equations**.

A **differential equation** is an equation which involves the derivatives of an unknown function as well as the function itself. Such equations play an extremely important and useful role in mathematical and statistical modelling of real-world phenomena in many different disciplines, such as engineering, physics, biology, medicine, epidemiology, economics and finance. Many of the mathematical ideas needed for understanding and solving differential equations come from **linear algebra**.

In this course we will meet some key ideas in linear algebra and differential equations and explore the links between them. This course also introduces concepts such as **vectors spaces** and **algebras**, which will help you develop abstract mathematical thinking.

We begin with linear algebra. First, we will look at the solution of systems of linear equations. This will motivate the introduction of matrices.

1.2 Linear equations

A **linear equation** is an equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b.$$

Here the unknowns are $x_1, \ldots, x_n$; the (known) numbers $a_1, a_2, \ldots, a_n$, (called the **coefficients** of the equation) and the right-hand side b will usually be real numbers. We will also call the unknowns $x_1, \ldots, x_n$ **variables**.

Note that each term in the equation is either a constant, or the product of a constant and a single variable (that is, there are no higher powers such as squares and cubes, and no products of variables such as x_1x_2 etc), which makes this equation **linear**, a term we will define rigorously later.

A **system** of equations is a set of equations to be solved for the same unknowns. We already know how to solve simple systems of linear equations involving two equations and two unknowns. For example, we can solve the system

$$\begin{aligned} x_1 + x_2 &= 2, \\ 2x_1 + x_2 &= 3, \end{aligned} \tag{1.1}$$

by multiplying the first equation by -2, and adding it to the second one to get the single equation $-x_2 = -1$, so that $x_2 = 1$. Substituting this value into either of the equations in (1.1), we can now solve for x_1. For example, using the first equation gives $x_1 + 1 = 2$ so that $x_1 = 1$. So we have found the **unique** solution of (1.1): $x_1 = x_2 = 1$.

Even with only two variables, however, things are not always this straightforward. For example, looking at the system

$$\begin{aligned} x_1 + x_2 &= 1, \\ 2x_1 + 2x_2 &= 2. \end{aligned}$$

We see that the second equation is just the first one multiplied by 2, so it does not contain any new information. Hence really we have only one equation,

$$x_1 + x_2 = 1\,, \tag{1.2}$$

for the two variables x_1 and x_2. We call such a system **underdetermined** as it has fewer equations than unknowns. To satisfy (1.2), we see that for any given value $x_1 \in \mathbb{R}$, we must have $x_2 = 1 - x_1$. So, this system has an **infinite** number of solutions. We can prescribe x_1 as we wish, e.g. by putting $x_1 = \lambda \in \mathbb{R}$ and then $x_2 = 1 - \lambda$. So if we choose $\lambda = 13$, we get a solution $(13, -12)$. Or if $\lambda = 0$, we get a solution $(0, 1)$ (here in the bracket the first number corresponds to a value of x_1 and the second, to a value of x_2). More formally, we can write the whole **solution set** $\mathcal{S}$ as

$$\mathcal{S} = \{(\lambda, 1 - \lambda) : \lambda \in \mathbb{R}\}.$$

We will talk more about this notation later.

Finally, we note that sometimes linear systems of equations simply cannot be solved. For example, the system

$$\begin{aligned} x_1 + x_2 &= 2, \\ x_1 + x_2 &= 3, \end{aligned}$$

gn clearly has no solution as $2 \neq 3$. Such systems of equations are said to be **inconsistent**.

In summary, there are three different possibilities:

1. the system has a unique solution;
2. the system has an infinite number of solutions;
3. the system has no solution.

Practical applications often involve hundreds or thousands of equations rather than just two, so we would like to be able to generalise the ideas we have just seen to much larger systems. We can write a general system of m linear algebraic equations involving n unknowns $x_1, x_2, \ldots, x_n$ (where m may or may not be equal to n) as

$$\begin{array}{ccccccccc} a_{11}\,x_1 & + & a_{12}\,x_2 & + & \cdots & + & a_{1n}\,x_n & = & b_1, \\ a_{21}\,x_1 & + & a_{22}\,x_2 & + & \cdots & + & a_{2n}\,x_n & = & b_2, \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}\,x_1 & + & a_{m2}\,x_2 & + & \cdots & + & a_{mn}\,x_n & = & b_m. \end{array} \tag{1.3}$$

To solve this system efficiently, we need some good notation, so we now introduce the idea of **matrices** to represent linear systems.

1.3 Matrices

In this section we review the key facts about matrices.

Definition 1.1 *A real (or complex) matrix is a rectangular array of real (complex) numbers. The general matrix A with m rows and n columns (written $m \times n$) is*

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad or \quad A = [a_{ij}]_{m\times n}.$$

Matrices are usually denoted by capital letters. The number a_{ij} is the entry in **row** i and **column** j of the matrix. We will call a_{ij} the (i, j)-th **element of** A. Note that in MM201, matrices will be **real** (unless stated otherwise). The set of all $m \times n$ real matrices will be denoted by $\mathbb{R}^{m\times n}$.

Observe that a real number is a 1×1 matrix. Before we discuss what can be done with matrices, let us introduce some important examples and classes of matrices.

Definition 1.2 *(a) An $m \times n$ matrix A is* ***square*** *if $m = n$.*

(b) The $m \times n$ ***zero*** *matrix $0_{m\times n}$ is the $m \times n$ matrix that contains only zero elements.*

(c) The $n \times n$ ***identity*** *(or* ***unit****) matrix is given by $I_n = [\delta_{ij}]_{n\times n}$, where δ_{ij} is the* **Kronecker delta** *defined by $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$*
(The subscript n is often omitted when the size is obvious).

(d) The **main diagonal** *of an $m \times n$ matrix A is the line of elements $a_{11}, a_{22}, \ldots, a_{kk}$, where $k = \min(m, n)$.*

(e) A square matrix is ***diagonal*** *if its only non-zero entries lie on the main diagonal, i.e. $a_{ij} = 0$ when $i \neq j$.*

(f) A square matrix is ***upper (lower) triangular*** *if all elements below (above) the main diagonal are zero, i.e. if A is lower triangular then $a_{ij} = 0$, $i < j$ and if A is upper triangular then $a_{ij} = 0$, $i > j$.*

What can one do with a singe ? We certainly can multiply a matrix by a number, but there i something new we can do: compute the **transpose** of a matrix.

Definition 1.3 *(a)* ***Multiplication by a scalar****: Let $A = [a_{ij}]_{m\times n}$ and let λ be a scalar. The product λA is the matrix $D = [d_{ij}]_{m\times n}$, where $d_{ij} = \lambda a_{ij}$ $\forall i, j : 1 \leq i \leq m,\ 1 \leq j \leq n$.*

(b) The ***transpose*** *of the $m \times n$ matrix A is the $n \times m$ matrix $A^T = [\hat{a}_{ij}]_{n\times m}$, where $\hat{a}_{ij} = a_{ji}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$.*

(c) A square matrix A is ***symmetric*** *if $A^T = A$ and* ***skew-symmetric*** *if $A^T = -A$.*

We say that two matrices A and B are **equal** is they have the same number of rows and columns and all their elements are equal, i.e. if $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{r\times s}$, then $A = B$ if and only if $m = r$, $n = s$ and $a_{ij} = b_{ij}$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

Sometimes we can also add and multiply matrices:

Definition 1.4 *(a)* ***Addition****: Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$. The sum $A+B$ is the $m \times n$ matrix $C = [c_{ij}]_{m\times n}$, where $c_{ij} = a_{ij} + b_{ij}$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.*

(b) ***Multiplication:*** *Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{r\times s}$. If $n = r$ then the product AB is the matrix $C = [c_{ij}]_{m\times s}$, where*

$$c_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj} \quad \forall i, j : 1 \le i \le m,\ 1 \le j \le s.$$

If $n \neq r$, then the product AB is not defined.

Examples 1A

There is no division of matrices in the sense that if A and B are matrices, the ratio A/B cannot be defined. However, consider this. Let a and b be two real numbers, $b \neq 0$. Then $\frac{a}{b} = a \cdot \frac{1}{b}$, that is, dividing a by b is the same as multiplying a by the *reciprocal* of b. Some square matrices have a "reciprocal" matrix, which is called the **inverse**; later we will discuss which square matrices have such an inverse and how to compute it.

Definition 1.5 *An $n\times n$ matrix A is **non-singular** (or **invertible**) if there exists an $n\times n$ matrix called the **inverse** of A, denoted by A^{-1}, such that $AA^{-1} = A^{-1}A = I_n$.*

Note that, if A^{-1} exists, then it is unique. If A^{-1} does not exist, A is said to be **singular**.

In physics, one often encounters non-singular matrices A such that $A^{-1} = A^T$. Such matrices are called **orthogonal**.

Another useful concept for square matrices is that of **trace**.

Definition 1.6 *The **trace** of a square ($n \times n$) matrix is the sum of the elements on the main diagonal, that is,*

$$\mathrm{tr}(A) = \sum_{k=1}^{n} a_{kk}.$$

Examples 1B

Finally, we list some standard results about matrices. These can easily be checked using the definitions of matrix operations given above.

Theorem 1.7 *(a) Assuming that the dimensions of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.*

(i) $A + B = B + A$ *(ii)* $A + (B + C) = (A + B) + C$
(iii) $A(BC) = (AB)C$ *(iv)* $A(B + C) = AB + AC$
(v) $(B + C)A = BA + CA$.

(b) *(i)* $(A+B)^T = A^T + B^T$ *(ii)* $(AB)^T = B^T A^T$ *(iii)* $(A^T)^T = A$.

(c) *If A, B are non-singular $n \times n$ matrices, then*

(i) $(AB)^{-1} = B^{-1} A^{-1}$ *(ii)* $(A^{-1})^{-1} = A$

(d) $I_n^{-1} = I_n$ *for each positive integer n.*

(e) *The matrix $A = [a_{ij}]_{2\times 2}$ is non-singular if $|A| = a_{11} a_{22} - a_{12} a_{21} \neq 0$. Here $|A|$ (or* $\det(A)$*) is called the* **determinant** *of A. In this case*

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{1.4}$$

Later we will discuss determinants and inverses of matrices larger than 2×2.

Examples 1C

1.4 Systems of linear equations

We now return to our general linear system (1.3). Using our new notation, it can be written in matrix form as

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{1.5}$$

where $A = [a_{ij}]_{m\times n}$ is an $m \times n$ matrix, $\boldsymbol{x} = [x_1\, x_2 \ldots x_n]^T$ is an $n \times 1$ matrix and $\boldsymbol{b} = [b_1\, b_2, \ldots b_m]^T$ is an $m \times 1$ matrix. (Note: matrices with just one row or one column are often written as small bold or underlined letters.)

If we know the inverse of A, then solving (1.5) to find $\boldsymbol{x}$ is easy. We can simply multiply both sides by the matrix A^{-1} as follows:

$$A\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b} \Leftrightarrow I\boldsymbol{x} = A^{-1}\boldsymbol{b} \Leftrightarrow \boldsymbol{x} = A^{-1}\boldsymbol{b}.$$

Sometimes, however, A^{-1} does not exist and, even when it does, it is almost always much too expensive in terms of computing time to calculate. **We therefore need a way of solving (1.5) which does not involve finding A^{-1} explicitly.**

To do this, we formalise the method of solving a linear system we used for two equations in §1.2 and extend it to deal with the general case. The method we will discuss involves using **elementary row operations** (EROs) to transform linear system (1.5) to an equivalent, but simpler, system with a coefficient matrix, B say, in so-called **echelon** form.

Definition 1.8 *We define three types of EROs as applied to an $m \times n$ matrix A:*

Type I	:	*Interchange rows r and s of A, with $r \neq s$;*
Type II	:	*Multiply row r of A by a non-zero scalar α;*
Type III	:	*Add $\alpha\cdot$(row s) to row r of A for a non-zero scalar α.*

Definition 1.9 *The $m \times n$ matrix B is a (row)* ***echelon matrix*** *if it has the following properties.*

(a) If a row does not consist entirely of zeros, then the first non-zero number in the row is 1: this is called a ***leading 1****.*

(b) If there are any rows that consist entirely of zeros then they are grouped together at the bottom of the matrix.

(c) In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row is further to the right than the leading 1 in the higher row.

If, also, each column that contains a leading 1 has zeros everywhere else, then B is a ***reduced (row) echelon matrix****.*

Notes: (i) When solving the resulting linear system, a variable associated with a **column** containing a leading 1 in the echelon matrix is called a **leading variable**: if there is no leading 1 in a column, the associated variable is called a **free variable**.

(ii) We will write $A \sim B$ to signify that A is **row equivalent** to B (i.e. B can be obtained by applying EROs to A).

(iii) The above operations could equally well be applied to the columns of a matrix rather than the rows (e.g. interchanging two columns etc). For solving linear systems, we use row operations only.

The technique for using EROs to solve a linear system is as follows. First, construct an **augmented matrix** $[A|\boldsymbol{b}]$, that is, append the right-hand side vector $\boldsymbol{b}$ to A as an extra column. Then use the following steps:

1. Start at the left of the augmented matrix $[A|\boldsymbol{b}]$ and make sure that the number at the top of the first column is nonzero (if it isn't, swap two rows).

2. Using EROs, make all the other numbers in the column underneath it zero.

3. Repeat this on the submatrix obtained by ignoring the first row and column, and keep going until you have an echelon matrix.

4. The entries of $\boldsymbol{x}$ can now be found by starting with the last equation and substituting it into the preceding ones (this is called **back substitution**).

Some practical hints:

(i) As a general rule, avoid fractions as long as you can. The easier the arithmetic, the less likely you are to make mistakes. (This is not, of course, true for a computer!).

(ii) Be careful when combining EROs. In particular, do not use rows before you have updated them.

(iii) Always check your answer by substituting the values of the variables back into the original equation!

Examples 1D

As we saw in §1.2, a linear system can have one, infinitely many or zero solutions. To identify these situations for larger systems, we can examine the bottom row of the final matrix from step 3 of the above algorithm. Suppose that, for an m by n system of equations, the last row of the final echelon matrix has entries

$$[0 \quad \ldots \quad 0 \quad \hat{a}_{mn} \quad | \quad \hat{b}_m].$$

This means that we have transformed the final equation from

$$a_{m1}\, x_1 + a_{m2}\, x_2 + \cdots + a_{mn}\, x_n = b_m$$

to the new equation

$$0\, x_1 + 0\, x_2 + \cdots + \hat{a}_{mn} x_n = \hat{b}_m. \tag{1.6}$$

The three possible situations (one, infinitely many, or no solutions) correspond to the three cases below:

1. $\hat{a}_{mn} \neq 0$. Here equation (1.6) has a unique solution given by $x_n = \hat{b}_m / \hat{a}_{mn}$, and the equations corresponding to the echelon matrix can be solved by back substitution using this value of x_n.

2. $\hat{a}_{mn} = 0$ **and** $\hat{b}_m = 0$. Here equation (1.6) reduces to “0=0”, that is, the equation for x_n has essentially disappeared, so the linear system is **underdetermined** and has infinitely many solutions. To find out what form these solutions take, we can let x_n take **any** value (as before, we use λ here to represent an arbitrary real number), then use back substitution to solve for x_1, x_2, etc in terms of λ. An additional parameter is required for each additional row of zeros in the echelon matrix.

3. $\hat{a}_{mn} = 0$ **and** $\hat{b}_m \neq 0$. Here equation (1.6) is $0 = \hat{b}_m \neq 0$, which has no solution so the linear system is **inconsistent**.

Examples 1E

1.5 Determinants

One important and useful quantity associated with a square matrix is its **determinant**. In this section we give a method for calculating the determinant of a given matrix, and discuss some of its properties.

First, we recall from (1.4) that the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is invertible if $a_{11}\,a_{22} - a_{12}\,a_{21} \neq 0$, with

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix},$$

where $\det(A) = a_{11}\,a_{22} - a_{12}\,a_{21}$ is the **determinant** of the 2×2 matrix A, and is sometimes also denoted by $|A|$. The aim now is to find the analogue of this formula for A^{-1} when $A = [a_{ij}]_{n \times n}$ for $n > 2$. To achieve this we first require the idea of a permutation.

Definition 1.10 *A* ***permutation*** *of the set of integers* $S = \{1, 2, \ldots, n\}$ *is an arrangement of these integers in some order without omissions or repetitions.*

A permutation is a one-to-one mapping $\sigma : S \to S$ where

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ j_1 & j_2 & j_3 & \cdots & j_n \end{pmatrix},$$

that is,

$$\sigma(i) = j_i \quad \text{for} \quad i = 1, 2, \ldots, n.$$

For example, if

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix},$$

then $\sigma(1) = 4$, $\sigma(2) = 2$, $\sigma(3) = 1$, $\sigma(4) = 3$.

The set of all possible permutations of S is denoted by S_n; it contains $n!$ permutations.

Definition 1.11 *An **inversion** is said to occur in a permutation* $(j_1, j_2, \ldots, j_n)$ *whenever a larger integer precedes a smaller one.*

The total number of inversions occurring in a permutation can be obtained as follows:

(i) Find the number of integers less than j_1 that follow j_1 in the permutation;

(ii) Find the number of integers less than j_2 that follow j_2 in the permutation;

(iii) Continue this process for $j_3, j_4, \ldots, j_{n-1}$.

The sum of all these numbers is the total number of inversions.

Definition 1.12 *(a) A permutation is **even** if the total number of inversions is an even integer and it is **odd** if the total number of inversions is an odd integer.*

*(b) The **sign** of a permutation* σ *is* $\operatorname{sgn}(\sigma) = (-1)^N$*, where* N *is the number of inversions. Thus* $\operatorname{sgn}(\sigma) = +1$ *or* -1 *depending on whether* N *is even or odd.*

If $A = [a_{ij}]_{n\times n}$ then the number $\det(A)$ is defined using the notation for permutations.

Definition 1.13 *If* $A = [a_{ij}]_{n\times n}$ *then the determinant of* A*,* $\det(A)$*, is defined by*

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)}\, a_{2\sigma(2)} \ldots a_{n\sigma(n)},$$

with the summation over all $n!$ *permutations in* S_n.

Examples 1F

For $n > 3$, this definition is extremely cumbersome. For example, if $n = 5$, there are $5! = 120$ terms in the sum. However, it is possible to show that $\det(A)$ can be evaluated by an alternative method — the **cofactor expansion**.

Definition 1.14 *Let* $A = [a_{ij}]_{n\times n}$*, and let* M_{ij} *be the* $(n-1)\times(n-1)$ *submatrix of* A *obtained by deleting row* i *and column* j *of* A*. Then* $\det(M_{ij})$ *is called the **minor** of* a_{ij} *and* $A_{ij} = (-1)^{i+j}\det(M_{ij})$ *is called the **cofactor** of* a_{ij}.

It can be shown that $\det(A)$ may be expressed in terms of the cofactors of each of the elements in any one row or column.

Theorem 1.15 *If* $A = [a_{ij}]_{n \times n}$*, then for any* $i = 1, 2, \ldots, n$*,*

$$\det(A) = a_{i1}\, A_{i1} + a_{i2}\, A_{i2} + \cdots + a_{in}\, A_{in},$$

(cofactor expansion along the i*th row) and for any* $j = 1, 2, \ldots, n$*,*

$$\det(A) = a_{1j}\, A_{1j} + a_{2j}\, A_{2j} + \cdots + a_{nj}\, A_{nj}$$

(cofactor expansion along the j*th column).*

To illustrate this theorem, consider the 3×3 case

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

The cofactor expansion along row 1 gives

$$\begin{aligned}
\det(A) &= a_{11}\, A_{11} + a_{12}\, A_{12} + a_{13}\, A_{13} \\
&= a_{11}(-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)^{1+2} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(-1)^{1+3} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}(a_{22}\, a_{33} - a_{23}\, a_{32}) - a_{12}(a_{21}\, a_{33} - a_{23}\, a_{31}) + a_{13}(a_{21}\, a_{32} - a_{22}\, a_{31}) \\
&= a_{11}\, a_{22}\, a_{33} - a_{11}\, a_{23}\, a_{32} - a_{12}\, a_{21}\, a_{33} + a_{12}\, a_{23}\, a_{31} + a_{13}\, a_{21}\, a_{32} - a_{13}\, a_{22}\, a_{31}.
\end{aligned}$$

Note that A_{ij} is either $\det(M_{ij})$ or $-\det(M_{ij})$, depending on the sign of $(-1)^{i+j}$. The sign pattern of $(-1)^{i+j}$ for a 3×3 matrix is $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$, and this is extended in the obvious way for $n > 3$.

Examples 1G

The following results (quoted without proof) are useful for simplifying the evaluation of $\det(A)$.

Theorem 1.16 *(a)* $\det(I_n) = 1$ *and* $\det(0_{n \times n}) = 0$.

(b) If A *has a zero row then* $\det(A) = 0$.

(c) $\det(A^T) = \det(A)$.

(d) If A *and* B *are square matrices of the same size, then* $\det(AB) = \det(A)\det(B)$.

(e) If A^{-1} *exists then* $\det(A^{-1}) = 1/\det(A)$ *(because* $\det(A)\det(A^{-1}) = \det(I_n) = 1$*).*

(f) *If A has two equal rows then* $\det(A) = 0$.

(g) *If B is formed by interchanging two rows of A then* $\det(B) = -\det(A)$.

(h) *If B is formed by multiplying any one row of A by α (i.e. performing the ERO $r_i' = \alpha r_i$) then* $\det(B) = \alpha \det(A)$.

(i) *If B is the matrix formed by adding $\alpha\times$ row j to row i with $i \neq j$ (i.e. performing the ERO $r_i' = r_i + \alpha r_j$) then* $\det(B) = \det(A)$.

Note: The last four statements also apply to columns and column operations.

Examples 1H

2 Vector spaces, algebras, linear algebra

2.1 Introduction

Now that we know something about matrices, we are going to make a serious mathematical move and "abstract" from the example of matrices some important mathematical concepts. The nature of the move is similar to looking at a chihuahua and an Alsatian and abstracting from these two examples the concept of a "dog". Once we have a concept of a dog, we can make statements about all dogs, not just chihuahuas or Alsatians.

Starting with matrices, we will end up with concepts such **vector spaces** and **algebras**. Then we will need to change our point of view of matrices to get to grips with the concept of **linear algebra**.

This is the most abstract chapter of the course, necessary to help you develop mathematical thinking and to prepare you for courses such as MM301, MM303, and MM403 in fourth year.

2.2 Sets

Clearly, we can talk about the **set** of $m \times n$ matrices. So let us discuss sets in general.

Definition 2.17 *(i) A **set** is a collection of objects, such that there is a well-defined way of determining whether a given object is included in, or excluded from the set.*

*(ii) An object in a set is called an **element** or **member** of that set. If the element x is a member of the set A, we write $x \in A$. Alternatively, if x is not a member of the set A, we write $x \notin A$.*

Some examples of sets include the set of students registered for MM201, the set of all Munros in Scotland, or the set of prime numbers. All of these examples have clear rules that allow someone with enough information to determine whether a given object is included in the collection or not. Important sets are the reals denoted by $\mathbb{R}$, the rationals $\mathbb{Q}$, the integers $\mathbb{Z}$, the natural numbers (non-negative integers) $\mathbb{N}$, and the complex numbers $\mathbb{C}$,

There are two common ways of describing sets:

1. by listing all of the elements of a set between curly brackets $\{\cdots\}$. For example, the set A of all integers x which satisfy $2 \leq x < 5$ can be written as $A = \{2, 3, 4\}$. Note

that the order of the elements in the set does not matter. Also, each entry is listed only once: repetitions are ignored.

2. by explicitly stating a defining property of the form

$$\{x \mid x \text{ satisfies property } P\}.$$

The $\mid$ symbol is read as 'such that' (sometimes I will write "s.t." for 'such that'); sometimes a colon (:) is used instead.

For example, the set A above can be written as

$$A = \{x \,|\, x \in \mathbb{Z} \text{ and } 2 \leq x < 5\}.$$

Two sets S and T are considered equal when they have exactly the same elements: we write this as $S = T$. If a set has a finite number of elements, it is called a **finite set**; otherwise, it is an **infinite set**. The number of elements in a finite set S is called the **order** of S, and written as $|S|$. The **empty set**, denoted by $\emptyset$, is the (unique) set which has no elements. That is, $\emptyset = \{\}$. Note that the empty set $\emptyset$ is NOT the same as $\{\emptyset\}$, which is 'the set which contains the empty set'.

Definition 2.18 *A set T is called a **subset** of a set S if every element of T is an element of S. We write this relationship as $T \subset S$. When T is NOT a subset of S, we write $T \not\subset S$.*

Note: $T \subseteq S$ used to represent a subset of S that can be also S itself, and $T \subset S$ is reserved for a so-called **proper subset** (so that $T \subseteq S$ and $T \neq S$).

Now let S and T be subsets of some set U..

- The **union** of S and T consists of the elements which are in either S or T or both. We write
$$S \cup T = \{x \in U \,|\, x \in S \text{ or } x \in T\}.$$
- The **intersection** of two subsets S and T consists of the elements which are in both S and T.
$$S \cap T = \{x \in U | x \in S \text{ and } x \in T\}.$$
- If S and T have no elements in common, they are said to be **disjoint** or **mutually exclusive**, that is, $S \cap T = \emptyset$.

- The **complement** of a subset S is the set of all elements of U which are NOT in S.

$$S' = \{x \in U | x \notin S\}.$$

- The **set difference** of S and T consists of the elements in S but NOT in T.

$$S - T = \{x | x \in S \text{ and } x \notin T\}.$$

  An alternative notation for $S - T$ is $S \backslash T$.

2.3 The set of $m \times n$ matrices

Consider first the set of all cities in Scotland. It certainly is a **set**, but there is not much you can do with it. It makes no sense to add Dundee to Aberdeen or to multiply Glasgow by -3. With $m \times n$ matrices it is not like that: there are allowable operations on that set. What are they?

Suppose we denote the set of $m \times n$ matrices by $\mathbb{R}^{m\times n}$ as before, let $A, B, C \in \mathbb{R}^{m\times n}$ and let $\alpha, \beta \in \mathbb{R}$ (scalars). Then we already know that:

1. If $A, B \in \mathbb{R}^{m\times n}$, then $A + B \in \mathbb{R}^{m\times n}$;
2. If $A \in \mathbb{R}^{m\times n}$, $\alpha \in \mathbb{R}$, then $\alpha A \in \mathbb{R}^{m\times n}$.

These two properties can be expressed in many different ways. We can say that **addition of elements is defined on** $\mathbb{R}^{m\times n}$ and that **multiplication by scalars is defined on** $\mathbb{R}^{m\times n}$, or, equivalently, we can say that $\mathbb{R}^{m\times n}$ is **closed under addition and multiplication by scalars**.

These operations, of addition and multiolication by scalars have many properties:

A1: $A + B = B + A$ (commutativity of addition)

A2: $A + (B + C) = (A + B) + C$ (associativity of addition)

A3: $A + O_{m\times n} = O_{m\times n} + A = A$ (existence of zero element)

A4: $A + (-A) = (-A) + A = O_{m\times n}$ (existence of additive inverse)

SM1: $\alpha(A + B) = \alpha A + \alpha B$ (distributivity of scalar multiplication)

SM2: $(\alpha + \beta)A = \alpha A + \beta A$

SM3: $(\alpha\beta)A = \alpha(\beta A)$ (associativity of scalar multiplication)

SM4: $1(A) = A$ (existence of multiplicative identity)

2.4 Vector spaces

This enumeration of properties of $\mathbb{R}^{m\times n}$ will now lead us to a general definition of a **vector space**. So without further ado:

Consider an arbitrary non-empty set V of elements, together with two operations called **addition** and **scalar multiplication**.

Definition 2.19 *If the following axioms are satisfied by all objects x, y, z in V and all scalars (real numbers) α and β, then we call V a **vector space**.*

A1: $x + y = y + x$ *(commutativity)*

A2: $x + (y + z) = (x + y) + z$ *(associativity)*

A3: $\exists\, e \in Z$ *s. t.* $\forall\, x \in V,\ x + e = e + x = x$ *(zero element)*

A4: $x + (-x) = (-x) + x = e$ *(additive inverse)*

SM1: $\alpha(x + y) = \alpha x + \alpha y$ *(distributivity)*

SM2: $(\alpha + \beta)x = \alpha x + \beta x$

SM3: $(\alpha\beta)x = \alpha(\beta x)$ *(associativity of multiplication)*

SM4 $1x = x$ *(multiplicative identity)*

This approach in mathematics, of starting with an example and then abstracting its properties, has the great advantage that once we set up such an abstract definition, any result we prove will be true for every single example of a vector space.

Note that for every vector space we **have** to say what addition and multiplication by scalars mean, and what is the zero element.

2.5 Vectors in n dimensions

The next very important example of a vector spaces is the space of **vectors in n dimensions**, where $n \in \mathbb{N}$, $n > 1$.

We can regard the n-tuple $(x_1, x_2, \ldots, x_n)$ of real numbers as a 'point' in n-dimensional space or as an n-dimensional vector (an n-vector).

Definition 2.20 *Let n be a positive integer.*

*(a) An **n-vector** is an **ordered** n-tuple of real numbers. $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, where $x_i \in \mathbb{R}$ for $i = 1, 2, \ldots, n$.*

(b) The set of all n-vectors is denoted by $\mathbb{R}^n$. For example, $(1, 0, 3, 4, -\sqrt{2}) \in \mathbb{R}^5$.

The definitions of addition and scalar multiplication in $\mathbb{R}^n$ are as follows:

Definition 2.21 *For two vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, the sum $\boldsymbol{x} + \boldsymbol{y}$ is defined by*

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1,\ x_2 + y_2, \ldots, x_n + y_n),$$

and if $\alpha \in \mathbb{R}$ then the scalar multiple $\alpha\boldsymbol{x}$ is defined by

$$\alpha\boldsymbol{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

Definition 2.22 *The **zero vector** in $\mathbb{R}^n$ is denoted by $\mathbf{0}$ and is defined by $\mathbf{0} = (0, 0, \ldots, 0)$.*

Check that these operations make $\mathbb{R}^n$ into a vector space.

Note: In Definition 2.19, we have specified that the scalars α and β are real numbers, giving a **real vector space**. However, it is also possible to define vector spaces where the scalars are complex numbers. In that case, the vector space will be said to be **complex**. All of the vector spaces we will deal with in this class will be real: the generalisation of this (and other) concepts to complex vector spaces are discussed in the Y3 classes MM301 and MM303. .

Other ways of writing vectors. If you think of it, a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ is "the same" as a $1 \times n$ matrix $[x_1 \ \ldots \ x_n]$ or as a $n \times 1$ matrix $[x_1 \ \ldots \ x_n]^T$. This will become useful later.

("The same" means that to each n-vector corresponds exactly one $1 \times n$ and exactly one $n \times 1$ matrix and vice versa.)

2.6 Other examples of vector spaces

In addition to $\mathbb{R}^{m \times n}$ and $\mathbb{R}^n$, there are some vector spaces you should know.

1. **Real numbers.** The system of real numbers is a real vector space. This is just the special case of $\mathbb{R}^n$ with $n = 1$.

2. **Complex numbers.** The set of complex numbers is defined by

$$\mathbb{C} = \{a + ib : \ a, b \in \mathbb{R}\},$$

where $i^2 = -1$. Complex numbers can be rigorously defined as points in the complex plane, that is, we can think of the complex number $x = a + bi$ () as the point in the Cartesian plane with coordinates (a, b).

Recall the standard addition operator for complex numbers (denoted by $+$): to add two complex numbers, $x = a + bi$ and $y = c + di$ say, we add real and imaginary parts separately:

$$x + y = (a + bi) + (c + di) = (a + c) + (b + d)i.$$

Similarly, to multiply a complex number by a real scalar α, we multiply both real and imaginary parts by the scalar:

$$\alpha x = \alpha(a + bi) = \alpha a + \alpha bi.$$

The set of complex numbers, together with these two operations, satisfies all of the properties in Definition 2.19, so the complex numbers are a real vector space.

3. **Real polynomials.** Let $n \in \mathbb{N}$ and let P_n be the set of all real polynomials of degree less than or equal to n, that is,

$$P_n := \{p = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \ \text{ where } a_i \in \mathbb{R} \text{ for } i = 1, \ldots, n\}.$$

If, for two polynomials $p, q \in P_n$, we define the operations

$$\begin{aligned} \textit{addition:} \quad p + q &= (a_0 + a_1 x + \ldots + a_n x^n) + (b_0 + b_1 x + \ldots + b_n x^n) \\ &= (a_0 + b_0) + (a_1 + b_1)x + \ldots + (a_n + b_n)x^n, \end{aligned}$$

$$\begin{aligned} \textit{multiplication by a scalar:} \quad \alpha p &= \alpha(a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n) \\ &= \alpha a_0 + \alpha a_1 x + \ldots + \alpha a_n x^n, \end{aligned}$$

then P_n is a real vector space.

4. **Real-valued continuous functions.** Let $a, b \in \mathbb{R}$ such that $a < b$, and define

$$C^0([a, b]) = \{f : [a, b] \to \mathbb{R} \ : \ f \text{ is continuous on } [a, b]\}.$$

For any two functions $f, g \in C^0[a, b]$, we can define their sum $h = f + g$ by

$$h(x) = f(x) + g(x),$$

and for every $f \in C^0([a, b])$ and scalar $\alpha \in \mathbb{R}$ we can define the product $k = \alpha f$ of f and α by

$$(\alpha f)(x) = \alpha f(x).$$

In this way we have defined addition and scalar multiplication in $C^0([a, b])$, so that $C^0[a, b]$ is a real vector space; you will see it again in MM303.

2.7 Euclidean n-space

There is something else we can do with the space of vectors $\mathbb{R}^n$.

In $\mathbb{R}^2$ and $\mathbb{R}^3$, we have already seen (in MM103) the idea of the **scalar product** (also called the dot or inner product) of two position vectors in 2 or 3 dimensional space. We now extend this notion to deal with pairs of vectors in $\mathbb{R}^n$, where n may exceed 3, thus generalising the concept of the 'angle between two vectors'.

Recall that in $\mathbb{R}^3$, the scalar product of $\boldsymbol{x} = (x_1, x_2, x_3)$ and $\boldsymbol{y} = (y_1, y_2, y_3)$ is defined by

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1\, y_1 + x_2\, y_2 + x_3\, y_3 = \sum_{i=1}^{3} x_i\, y_i.$$

Also, we have

$$\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\|\|\boldsymbol{y}\| \cos\theta,$$

where $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{x_1^2 + x_2^2 + x_3^2}$ is the length of $\boldsymbol{x}$ and θ is the angle between the 3D vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

We now generalise these definitions to vectors in $\mathbb{R}^n$.

Definition 2.23 *(a) The* ***Euclidean inner product*** *(or scalar product) of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is defined by*

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1\, y_1 + x_2\, y_2 + \cdots + x_n\, y_n = \sum_{i=1}^{n} x_i\, y_i$$

$\boldsymbol{x} \cdot \boldsymbol{y}$ is sometimes written as $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.

(b) When equipped with the Euclidean inner product, $\mathbb{R}^n$ is called ***Euclidean n-space****.*

(c) The ***Euclidean norm*** *(or length, or magnitude) of $\boldsymbol{x} \in \mathbb{R}^n$ is defined by*

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

If $\|\boldsymbol{x}\| = 1$, then $\boldsymbol{x}$ is called a ***unit vector*** *in $\mathbb{R}^n$.*

(d) The ***Euclidean distance*** *between $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is defined by*

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

(e) *The angle θ between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by*

$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \quad \textit{(provided } \boldsymbol{x} \neq \mathbf{0} \quad \textit{and} \quad \boldsymbol{y} \neq \mathbf{0})$$

The following are easily proved:

Note: (i) $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \mathbf{0}$.

(ii) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$

$(\boldsymbol{x} + \boldsymbol{y}) \cdot \boldsymbol{z} = \boldsymbol{x} \cdot \boldsymbol{z} + \boldsymbol{y} \cdot \boldsymbol{z}$

$(\alpha\boldsymbol{x}) \cdot \boldsymbol{y} = \alpha(\boldsymbol{x} \cdot \boldsymbol{y}) = \boldsymbol{x} \cdot (\alpha\boldsymbol{y}) \quad$ for $\alpha \in \mathbb{R}$

(iii) In $\mathbb{C}^n$, the inner product is defined for $\boldsymbol{w}, \boldsymbol{z} \in \mathbb{C}^n$ by

$$\boldsymbol{w} \cdot \boldsymbol{z} = \sum_{i=1}^{n} w_i\,\bar{z}_i \qquad (\bar{z}_i = \text{ complex conjugate of } z_i).$$

The norm of $\boldsymbol{z} \in \mathbb{C}^n$ is $\|\boldsymbol{z}\| = \sqrt{\boldsymbol{z} \cdot \boldsymbol{z}} = \left(\sum_{i=1}^{n} |z_i|^2\right)^{\frac{1}{2}}$.

Examples 2A

We now state one of the most important inequalities in mathematics:

Theorem 2.24 (Cauchy-Schwarz Inequality)

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\|\ \|\boldsymbol{y}\| \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

In component form:
$$\left|\sum_{i=1}^{n} x_i\, y_i\right| \leq \left(\sum_{i=1}^{n} (x_i)^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} (y_i)^2\right)^{\frac{1}{2}}$$

Proof 2.24 *If $\boldsymbol{x} = \mathbf{0}$ or $\boldsymbol{y} = \mathbf{0}$ (or both) the result is immediate. Suppose, therefore, that $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{y} \neq \mathbf{0}$. Then for any $\alpha \in \mathbb{R}$, define $R(\alpha) = \|\alpha\boldsymbol{x} + \boldsymbol{y}\|^2$ and note that $R(\alpha) \geq 0$. We now have*

$$\begin{aligned} R(\alpha) &= (\alpha\boldsymbol{x} + \boldsymbol{y}) \cdot (\alpha\boldsymbol{x} + \boldsymbol{y}) = \alpha^2\|\boldsymbol{x}\|^2 + 2\alpha(\boldsymbol{x} \cdot \boldsymbol{y}) + \|\boldsymbol{y}\|^2 \\ &= a\alpha^2 + 2b\alpha + c, \end{aligned} \tag{2.7}$$

say, where

$$a = \|\boldsymbol{x}\|^2 > 0;$$
$$b = \boldsymbol{x} \cdot \boldsymbol{y};$$
$$c = \|\boldsymbol{y}\|^2 > 0.$$

Since $R(\alpha) \geq 0 \quad \forall \alpha \in \mathbb{R}$, the discriminant of quadratic (2.7) must be non-positive (if the discriminant > 0 then $R(\alpha) < 0$ for a set in $\mathbb{R}$).

Hence $4b^2 - 4ac \leq 0 \;\Rightarrow\; b^2 \leq ac$

$$\Rightarrow \quad (\boldsymbol{x} \cdot \boldsymbol{y})^2 \leq \|\boldsymbol{x}\|^2 \, \|\boldsymbol{y}\|^2$$
$$\Rightarrow \quad |\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|.$$

Another useful inequality (which we have already seen in MM101) is given in the next theorem.

Theorem 2.25 (Triangle Inequality)

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

or in component form: $$\left[\sum_{i=1}^{n} (x_i + y_i)^2\right]^{\frac{1}{2}} \leq \left[\sum_{i=1}^{n} (x_i)^2\right]^{\frac{1}{2}} + \left[\sum_{i=1}^{n} (y_i)^2\right]^{\frac{1}{2}}$$

Proof 2.25

$$\begin{aligned}
\|\boldsymbol{x} + \boldsymbol{y}\|^2 &= (\boldsymbol{x} + \boldsymbol{y}) \cdot (\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{x} + 2(\boldsymbol{x} \cdot \boldsymbol{y}) + \boldsymbol{y} \cdot \boldsymbol{y} \\
&= \|\boldsymbol{x}\|^2 + 2(\boldsymbol{x} \cdot \boldsymbol{y}) + \|\boldsymbol{y}\|^2 \\
&\leq \|\boldsymbol{x}\|^2 + 2|\boldsymbol{x} \cdot \boldsymbol{y}| + \|\boldsymbol{y}\|^2 \qquad (\textit{since } \boldsymbol{x} \cdot \boldsymbol{y} \leq |\boldsymbol{x} \cdot \boldsymbol{y}|) \\
&\leq \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\| \, \|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \quad (\textit{by Cauchy-Schwarz}) \\
&= (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \\
\Rightarrow \quad \|\boldsymbol{x} + \boldsymbol{y}\| &\leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|.
\end{aligned}$$

In Euclidean $\mathbb{R}^2$ or $\mathbb{R}^3$ spaces two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be **orthogonal** (perpendicular) if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. This notion can also now be extended to general $\mathbb{R}^n$.

Definition 2.26 *(a) Two non-zero vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ are **orthogonal** if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. (This is sometimes written as $\boldsymbol{x} \perp \boldsymbol{y}$.)*

*If, also, $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$ then we say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthonormal***

*(b) The set of non-zero vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\} \subset \mathbb{R}^n$ is **orthogonal** if $\boldsymbol{x}_i \cdot \boldsymbol{x}_j = 0$ whenever $i \neq j$. If the vectors in this orthogonal set also satisfy $\|\boldsymbol{x}_i\| = 1$ for $i = 1, 2, \ldots, m$ then the set is said to be **orthonormal**. In this case we write*

$$\boldsymbol{x}_i \cdot \boldsymbol{x}_j = \delta_{ij}$$

*where δ_{ij} is the **Kronecker delta**.*

Examples 2B

2.8 Algebras

When you think about it, in saying that $m \times n$ matrices are a vector space, we forgot that some matrices can be multiplied. Let us now restrict ourselves to **square** $n \times n$ matrices.

Note that for any three $n \times n$ matrices A, B, and C and any scalar λ,

(M1) $(AB)C = A(BC)$ (associativity of multiplication)

(M2) $A(B + C) = AB + AC$ (distributivity of multiplication)

(M3) $(A + B)C = AC + BC$ (distributivity of multiplication)

(M4) $\lambda(AB) = (\lambda A)B = A(\lambda B)$.

And now we can abstract these properties and define an algebra.

Definition 2.27 *A vector space X for which one can ALSO define a multiplication operation with the above properties, i.e. for every three elements of X, $x.y, z$ and any scalar λ,*

(M1) $(xy)z = x(yz)$ *(associativity of multiplication)*

(M2) $x(y + z) = xy + xz$ *(distributivity of multiplication)*

(M3) $(x + y)z = xz + yz$ *(distributivity of multiplication)*

(M4) $\lambda(xy) = (\lambda x)y = x(\lambda y)$

is called an (associative) algebra.

So $n \times n$ algebras are an associative algebra. Another example of algebras is the vector space of continuous functions on $\mathbb{R}$ under function multiplication.[1]

2.9 Transformations and linearity

If you remember, we said that vectors in $\mathbb{R}^n$ can be identified with $n \times 1$ or $1 \times n$ matrices. But then you can think about $n \times n$ matrices in a completely new light. An $n \times n$ matrix times an $n \times 1$ matrix gives you an $n \times 1$ matrix. In other words, an $n \times n$ matrix times an n-vector is an n-vector: if A is an $n \times n$ matrix, and $\boldsymbol{x} \in \mathbb{R}^n$, then we can define a vector $\boldsymbol{y}$

[1] You might want to check that continuous functions on $\mathbb{R}$ under *composition* do not form an algebra.

by $\boldsymbol{y} = A\boldsymbol{x}$. Hence $n \times n$ matrices can be thought as a rule to make new n-vectors from old n-vectors. This is what we mean by a **transformation**. We say that $n \times n$ matrices define transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ ($m \times n$ matrices define transformations from $\mathbb{R}^n$ into $\mathbb{R}^m$).

Matrices multiplying vectors have the following property: if $\boldsymbol{u}$ and $\boldsymbol{v}$ are two n-vectors and α and β are scalars, then

$$A(\alpha\boldsymbol{u} + \beta\boldsymbol{v}) = \alpha A\boldsymbol{u} + \beta A\boldsymbol{v}$$

This property deserves a name of its own.

Definition 2.28 *If F is transformation from a vector space X into a vector space Y such that for all $u, v \in X$ and all scalars α, β we have that*

$$F(\alpha u + \beta v) = \alpha F(u) + \beta F(v),$$

we call F a **linear** *transformation.*

This is all I am going to say about transformations and linearity. You will hear a lot about these concepts in MM301 next year.

To summarise: the set of $n \times n$ matrices forms an algebra of linear transformations. **This** is the reason why the study of matrices goes under the name of **Linear Algebra**.

We will deal again with the product $A\boldsymbol{x}$ when we talk about eigenvalues and eigenvectors.

3 Subspaces, Bases & Dimension

It is clear from the last chapter that the concept of a vector space is very useful because it applies to many examples commonly used in mathematics. In this chapter we will study the properties of vector spaces in some detail. We begin with one of the most important ideas, that of a **subspace** of a vector space V. We will denote elements of a vector space by boldface lower-case letters.

3.1 Subspaces

Definition 3.1 *If U is a nonempty subset of a real vector space V then U is called a* ***subspace*** *of V if the following conditions hold:*

(a) U ***is closed under addition****: If $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of U, then $\boldsymbol{x}+\boldsymbol{y}$ is in U as well.*

(b) U ***is closed under scalar multiplication****:*

If α is any scalar and $\boldsymbol{x}$ is an element of U, then $\alpha\boldsymbol{x}$ is in U.

Let us prove the following result: If U is a subspace of V, the the zero element of V, $\mathbf{0}$ is in U. Here are two different proofs.

1. Let $\boldsymbol{x} \in U$. Such an $\boldsymbol{x}$ exists as U is nonempty. But then for every scalar α, $\alpha\boldsymbol{x} \in U$. Pick $\alpha = 0$. So $0\boldsymbol{x} \in U$. But $0\boldsymbol{x} = \mathbf{0}$. So $\mathbf{0} \in V$.

2. Let $\boldsymbol{x} \in U$. Now pick $\alpha = -1$. So by closure of U under scalar multiplication, $-\boldsymbol{x} \in U$. But then by closure of U under addition, $\boldsymbol{x} + (-\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{x} = \mathbf{0} \in U$, which is what we wanted to prove,

Hence we have

Zero test. If a subset U of V does **not** contain the zero vector, it is **not** a subspace of V.

Note that, with this definition, V is actually a subspace of V. Also, the subset U containing only the zero vector, $\mathbf{0}$, is a subspace of V. These are **trivial** subspaces of V, and all other subspaces are called **proper subspaces**.

It can be shown that any subspace U of a real vector space V is also a real vector space. This means that instead of having to verify all of the properties of vector spaces in Theorem 3.1 for

U, we can use a simpler test to check if U is a subspace of V. Then U will be automatically a vector space.

Note that the two properties, of closure under addition and scalar multiplication can be combined:

Subspace test. A subset U of a vector space V is a subspace of V if and only if

$$\boxed{\boldsymbol{z} = \alpha \boldsymbol{x} + \beta \boldsymbol{y} \in U, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in U \text{ and } \forall \alpha, \beta \in \mathbb{R}.} \tag{3.8}$$

Note that this test covers the two parts of Definition 3.1: using $\alpha = \beta = 1$ gives (a), $\beta = 0$ gives (b).

Examples 3A

3.2 Linear combinations and spanning sets

Definition 3.2 *Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\} \subseteq V$.*

(a) $\boldsymbol{x} \in V$ is a **linear combination** *of the vectors in S if $\boldsymbol{x}$ can be expressed in the form*

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_m \boldsymbol{v}_m = \sum_{i=1}^{m} \alpha_i \boldsymbol{v}_i, \tag{3.9}$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$.

(b) The ***span*** *of S, denoted by $sp(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ or $sp(S)$, is the set of all possible linear combinations of the elements in S.*

Theorem 3.3 *Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\} \subseteq V$. Then $sp(S)$ is a* ***subspace*** *of V.*

Proof 3.3 *Note first that $\mathbf{0} \in sp(S)$ (because we can get $\mathbf{0}$ by taking all of the coefficients in (3.9) to be zero). To show that $sp(S)$ is a subspace of V, we use test (3.8).*

Let $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in sp(S)$. We know that there exists $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$ and $\beta_1, \beta_2, \ldots, \beta_m \in \mathbb{R}$ such that

$$\boldsymbol{x} = \sum_{i=1}^{m} \alpha_i \boldsymbol{v}_i \quad \textit{and} \quad \boldsymbol{y} = \sum_{i=1}^{m} \beta_i \boldsymbol{v}_i.$$

So construct

$$\begin{aligned} \boldsymbol{z} &= \alpha \boldsymbol{x} + \beta \boldsymbol{y} \\ &= \alpha(\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_m \boldsymbol{v}_m) + \beta(\beta_1 \boldsymbol{v}_1 + \cdots + \beta_m \boldsymbol{v}_m) \\ &= (\alpha\alpha_1 + \beta\beta_1) \boldsymbol{v}_1 + (\alpha\alpha_2 + \beta\beta_2) \boldsymbol{v}_2 + \cdots + (\alpha\alpha_m + \beta\beta_m) \boldsymbol{v}_m \\ &= \gamma_1 \boldsymbol{v}_1 + \gamma_2 \boldsymbol{v}_2 + \cdots + \gamma_m \boldsymbol{v}_m, \end{aligned}$$

where $\gamma_i = \alpha\alpha_1 + beta\beta_1 \in \mathbb{R}$. Hence $\boldsymbol{z} \in sp(S)$ (as $\boldsymbol{z} = \alpha\boldsymbol{x} + \beta\boldsymbol{y}$ is a linear combination of the elements in S). This property holds $\forall\alpha, \beta \in \mathbb{R}$ and $\forall\boldsymbol{x}, \boldsymbol{y} \in sp(S)$ so it follows from (3.8) that $sp(S)$ is a subspace of V.

We say that $sp(S)$ is the subspace **generated** or **spanned** by $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$, and call S the **spanning set** for the subspace $sp(S)$.

Examples 3B

3.3 Linear independence

A subspace U of V has more than one spanning set. Given a spanning set S for a subspace U of V, it may be possible to reduce S to a smaller spanning set $S' \subset S$ by eliminating elements of S. For example, suppose that $S = sp\,\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ spans U. We can examine each $\boldsymbol{s}_i$ in turn: if $\boldsymbol{s}_i = \boldsymbol{0}$ or $\boldsymbol{s}_i$ is a linear combination of $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{i-1}$ then we can eliminate $\boldsymbol{s}_i$ to leave a smaller set which spans S. This reduction process leads to the following question: given a subspace U of V, is there a **minimum** number of elements needed in any spanning set? To answer this we introduce the idea of **linear independence**.

Definition 3.4 *Let $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ be a non-empty subset of V. Then the vector equation*

$$\alpha_1\, \boldsymbol{s}_1 + \alpha_2\, \boldsymbol{s}_2 + \cdots + \alpha_m\, \boldsymbol{s}_m = \boldsymbol{0} \tag{3.10}$$

with unknowns $\alpha_1, \alpha_2, \ldots, \alpha_m$ has at least one solution, namely

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

*If this is **the only** solution to (3.10), then $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m$ are said to be **linearly independent** elements (and S is a linearly independent set).*

*If there are other solutions to (3.10), then $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m$ are **linearly dependent** elements (and S is a linearly dependent set).*

In other words,

$$\begin{array}{llll} \displaystyle\sum_{k=1}^{m} \alpha_k\, \boldsymbol{s}_k = \boldsymbol{0} & \textit{meansthat} & \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0 & : \ S \textit{ is linearly independent} \\ \displaystyle\sum_{k=1}^{m} \alpha_k\, \boldsymbol{s}_k = \boldsymbol{0} & \text{with} & \alpha_k \neq 0 \textit{ for at least one } k & : \ S \textit{ is linearly dependent.} \end{array}$$

Notes:

(i) If $\mathbf{0} \in S$ then S is linearly dependent. For example, if $S = \{\mathbf{0}, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$, then $\alpha_1 \mathbf{0} + \alpha_2 \boldsymbol{s}_2 + \cdots + \alpha_m \boldsymbol{s}_m = \mathbf{0}$ $\forall \alpha_1 \in \mathbb{R}$ (that is, not only when $\alpha_1 = \alpha_2 = \alpha_3 = \cdots = 0$).

(ii) Let $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ be a set of non-zero elements in V ($m \geq 2$). Then

S is linearly dependent	$\Leftrightarrow$	at least one $\boldsymbol{s}_k$ is a linear combination of the other elements in S;
S is linearly independent	$\Leftrightarrow$	no $\boldsymbol{s}_k$ can be expressed as a linear combination of the other elements in S.

Examples 3C

3.4 Basis and dimension

We normally think of $\mathbb{R}$ as 1-dimensional, $\mathbb{R}^2$ as 2-dimensional and the space in which we live as 3-dimensional. Now we consider this notion of dimension for a general vector space V. First we introduce the idea of a basis.

Definition 3.5 *Let U be a subspace of V. The set $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\} \subseteq U$ is a* ***basis*** *for U if*

(a) B is a linearly independent set in U;

(b) B spans U.

Examples 3D

Theorem 3.6 *If $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$ is a basis for the subspace $U \subseteq V$ then $\boldsymbol{x} \in U$ can be expressed in the form*

$$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2 + \cdots + \alpha_m \boldsymbol{b}_m$$

in only one way. That is, each $\boldsymbol{x} \in U$ has a ***unique*** *representation as a linear combination of $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m$.*

Proof 3.6 *As B is a basis, B spans U so given $\boldsymbol{x} \in U$, $\exists \alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$ such that*

$$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2 + \cdots + \alpha_m \boldsymbol{b}_m.$$

Suppose there is another representation

$$\boldsymbol{x} = \beta_1 \boldsymbol{b}_1 + \beta_2 \boldsymbol{b}_2 + \cdots + \beta_m \boldsymbol{b}_m.$$

Then

$$\mathbf{0} = \boldsymbol{x} - \boldsymbol{x} = (\alpha_1 - \beta_1)\boldsymbol{b}_1 + (\alpha_2 - \beta_2)\boldsymbol{b}_2 + \cdots + (\alpha_m - \beta_m)\boldsymbol{b}_m.$$

But B is linearly independent, so this implies $\alpha_1 - \beta_1 = 0,\ \alpha_2 - \beta_2 = 0, \ldots, \alpha_m - \beta_m = 0$, that is, the representation of $\boldsymbol{x}$ as a linear combination of $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m$ is unique.

If $\boldsymbol{x} = \alpha_1\,\boldsymbol{b}_1 + \alpha_2\,\boldsymbol{b}_2 + \cdots + \alpha_m\,\boldsymbol{b}_m$ is the unique representation of $\boldsymbol{x} \in U$ in terms of the basis elements, then $\alpha_1, \alpha_2, \ldots, \alpha_m$ are said to be the **coordinates** (or **components**) of $\boldsymbol{x}$ with respect to the basis B.

Lemma 3.7 *If $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$ is a linearly independent set of elements in the subspace $U \subseteq V$ and if $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_d\}$ spans U, then $m \leq d$.*

Theorem 3.8 *Any two bases for a subspace $U \subseteq V$ contain the same number of elements.*

Proof 3.8 *Let B_1 and B_2 be two bases for U, containing n_1 and n_2 elements, respectively. Then B_1 is linearly independent and B_2 spans U so $n_1 \leq n_2$ (from the lemma above). But B_2 is also linearly independent and B_1 spans U so $n_2 \leq n_1$. Hence $n_1 = n_2$.*

Definition 3.9 *The **dimension** of a subspace $U \subseteq V$ is the number of elements in **any basis** for U. The dimension is denoted by dim U.*

Note that the trivial subspace that contains only the zero vector has dimension 0 (as it does not contain any linearly independent elements).

Examples 3E

Some useful results: Let U be a subspace of V with dim $U = d$.

(i) If $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_d\}$ is a set of d elements in U then S is a basis for U if **either** S is linearly independent, **or** S spans U.

(ii) Any set containing more than d elements from U cannot be linearly independent.

(iii) Any set containing less than d elements from U cannot span U.

(iv) If $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ is a linearly independent set in U and $m < d$, then S can be enlarged by adding additional elements to form a basis for U.

(v) If U is a subspace of V then dim $U \leq \dim V$; moreover, dim $U = \dim V \Leftrightarrow U = V$.

Examples 3F

3.5 Orthogonal bases

We have already met the idea of **orthogonality** in our discussion of $\mathbb{R}^n$ in §2.7. In this section we will assume that V is the real vector space $\mathbb{R}^n$ equipped with the **Euclidean inner product** and the **Euclidean norm** that we met in Chapter 2, that is, V is **Euclidean n-space**. Note, however, that the ideas discussed here apply to any vector space in which an inner product and norm can be defined.

Recall that a set $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\} \subset V$ is an **orthogonal** set if $\boldsymbol{s}_i \cdot \boldsymbol{s}_j = 0$ when $i \neq j$ and it is **orthonormal** if $\boldsymbol{s}_i \cdot \boldsymbol{s}_j = \delta_{ij}$ (see Definition 2.26). We now use the notion of orthogonality in constructing useful bases for vector spaces.

Theorem 3.10 *Let $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ ($\boldsymbol{s}_i \neq \mathbf{0}$ for $i = 1, 2, \ldots, m$) be an orthogonal set of elements in $\mathbb{R}^n$. Then S is linearly independent.*

Proof 3.10 *Suppose $\alpha_1 \boldsymbol{s}_1 + \alpha_2 \boldsymbol{s}_2 + \cdots + \alpha_m \boldsymbol{s}_m = \mathbf{0}$ and form the inner product of this with $\boldsymbol{s}_i$ (for any $i = 1, 2, \ldots, m$). This gives*

$$\begin{aligned} 0 = \mathbf{0} \cdot \boldsymbol{s}_i &= (\alpha_1 \boldsymbol{s}_1 + \alpha_2 \boldsymbol{s}_2 + \cdots + \alpha_m \boldsymbol{s}_m) \cdot \boldsymbol{s}_i \\ &= \alpha_1(\boldsymbol{s}_1 \cdot \boldsymbol{s}_i) + \alpha_2(\boldsymbol{s}_2 \cdot \boldsymbol{s}_i) + \cdots + \alpha_m(\boldsymbol{s}_m \cdot \boldsymbol{s}_i) \\ &= \alpha_i(\boldsymbol{s}_i \cdot \boldsymbol{s}_i) \quad (\textit{since } \boldsymbol{s}_k \cdot \boldsymbol{s}_i = 0 \text{ if } k \neq i) \\ &= \alpha_i \|\boldsymbol{s}_i\|^2 \\ \Rightarrow \quad \alpha_i = 0 \quad &(\textit{since } \|\boldsymbol{s}_i\|^2 \neq 0). \end{aligned}$$

This holds for any $i = 1, 2, \ldots, m$ so, as $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$, S is linearly independent.

Notes:

(i) If U is an m-dimensional subspace of V and $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\} \subset U$ is orthogonal, then $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m\}$ is a linearly independent set in U and so it must be a basis for U.

(ii) Every non-zero subspace U of V has an orthogonal basis, and any given basis for U can be converted to an orthogonal basis by applying the **Gram-Schmidt** process (see below). Of course, any orthogonal basis is readily converted to an orthonormal basis.

(iii) If $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$ is an **orthogonal** basis for the subspace U of V, then for each $\boldsymbol{x} \in U$, $\exists\ \alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$ such that

$$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2 + \cdots + \alpha_m \boldsymbol{b}_m.$$

So, for each $k = 1, 2, \ldots, m$ we have

$$\boldsymbol{x} \cdot \boldsymbol{b}_k = \alpha_k \|\boldsymbol{b}_k\|^2 \quad \text{(as in Proof 3.10)},$$

which gives

$$\alpha_k = \frac{\boldsymbol{x} \cdot \boldsymbol{b}_k}{\|\boldsymbol{b}_k\|^2}, \quad k = 1, 2, \ldots, m. \tag{3.11}$$

So

$$\boldsymbol{x} = \frac{\boldsymbol{x} \cdot \boldsymbol{b}_1}{\|\boldsymbol{b}_1\|^2}\boldsymbol{b}_1 + \frac{\boldsymbol{x} \cdot \boldsymbol{b}_2}{\|\boldsymbol{b}_2\|^2}\boldsymbol{b}_2 + \cdots + \frac{\boldsymbol{x} \cdot \boldsymbol{b}_m}{\|\boldsymbol{b}_m\|^2}\boldsymbol{b}_m,$$

that is, (3.11) gives a simple formula for calculating the coordinates of $\mathbf{x}$ with respect to the orthogonal basis. In addition, if B is an **orthonormal** basis for U then (3.11) reduces to

$$\alpha_k = \boldsymbol{x} \cdot \boldsymbol{b}_k, \quad k = 1, 2, \ldots, m,$$

that is, the coordinates of $\boldsymbol{x} \in U$ with respect to the orthonormal basis B are $\boldsymbol{x} \cdot \boldsymbol{b}_1, \boldsymbol{x} \cdot \boldsymbol{b}_2, \ldots, \boldsymbol{x} \cdot \boldsymbol{b}_m$.

Examples 3G

3.6 The Gram-Schmidt orthogonalisation procedure

The Gram-Schmidt orthogonalisation procedure allows us to take any basis and convert it into an orthogonal one. Thus let U be a non-zero subspace of V with basis $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$. We convert B to an **orthogonal** basis $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m\}$ as follows:

Step 1 Let $\boldsymbol{y}_1 = \boldsymbol{b}_1$.

Step 2 Let $\boldsymbol{y}_2 = \boldsymbol{b}_2 + \alpha_1^{(2)}\boldsymbol{y}_1$, where $\alpha_1^{(2)}$ is chosen so that $\boldsymbol{y}_2 \cdot \boldsymbol{y}_1 = 0$ (so $\boldsymbol{y}_2$ is orthogonal to $\boldsymbol{y}_1$).

Step 3 Let $\boldsymbol{y}_3 = \boldsymbol{b}_3 + \alpha_1^{(3)}\boldsymbol{y}_1 + \alpha_2^{(3)}\boldsymbol{y}_2$, where $\alpha_1^{(3)}$ and $\alpha_2^{(3)}$ are chosen so that

$$\boldsymbol{y}_3 \cdot \boldsymbol{y}_1 = 0 \quad \text{and} \quad \boldsymbol{y}_3 \cdot \boldsymbol{y}_2 = 0.$$

Step 4 In general, having found $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r$, let

$$\boldsymbol{y}_{r+1} = \boldsymbol{b}_{r+1} + \alpha_1^{(r+1)}\boldsymbol{y}_1 + \alpha_2^{(r+1)}\boldsymbol{y}_2 + \cdots + \alpha_r^{(r+1)}\boldsymbol{y}_r,$$

where $\alpha_1^{(r+1)}, \alpha_2^{(r+1)}, \ldots, \alpha_r^{(r+1)}$ are chosen such that

$$\boldsymbol{y}_{r+1} \cdot \boldsymbol{y}_1 = 0,\ \boldsymbol{y}_{r+1} \cdot \boldsymbol{y}_2 = 0, \ldots, \boldsymbol{y}_{r+1} \cdot \boldsymbol{y}_r = 0.$$

This process is continued until all the vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m$ have been obtained. These vectors belong to U (since each is a linear combination of vectors in B) and they are linearly independent (since they are orthogonal; see Theorem 3.10): they therefore form a basis for U (since $\dim U = m$).

If an **orthonormal** basis is required, we use $\left\{\dfrac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}, \dfrac{\boldsymbol{y}_2}{\|\boldsymbol{y}_2\|}, \cdots, \dfrac{\boldsymbol{y}_m}{\|\boldsymbol{y}_m\|}\right\}$.

4 Matrix-Related Subspaces

We saw in §2.6 that if $\mathbb{R}^{m\times n}$ is the set of $m \times n$ real matrices then, together with the addition and scalar multiplication operators as defined in Definition 1.4, it forms a vector space. In this section, we consider the relationships between some important subspaces of $\mathbb{R}^{m\times n}$, and their relevance to the solution of linear systems $A\boldsymbol{x} = \boldsymbol{b}$ for a real matrix $A \in \mathbb{R}^{m\times n}$ and vectors $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$.

4.1 Nullspace

The system $A\boldsymbol{x} = \boldsymbol{b}$ is called **homogeneous** if $\boldsymbol{b} = \mathbf{0}$, otherwise it is **inhomogeneous**. It is clear that $A\boldsymbol{x} = \mathbf{0}$ always has the trivial solution $\boldsymbol{x} = \mathbf{0} \in \mathbb{R}^n$, but we are usually interested in non-trivial solutions. For example,

$$\begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{has non-trivial solution} \quad \boldsymbol{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Theorem 4.1 *For a real $m \times n$ matrix A, the solution set of the homogeneous linear system*

$$A\boldsymbol{x} = \mathbf{0} \tag{4.12}$$

is a subspace of $\mathbb{R}^n$. This subspace is called the ***nullspace*** *of A, and denoted by $N(A)$.*

Proof 4.1 *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in $N(A)$ (so $A\boldsymbol{x} = \mathbf{0}$ and $A\boldsymbol{y} = \mathbf{0}$). Then, for a vector $\boldsymbol{z} = \alpha\boldsymbol{x} + \beta\boldsymbol{y}$, we have*

$$A\boldsymbol{z} = A(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha A\boldsymbol{x} + \beta A\boldsymbol{y} = \mathbf{0}$$

so $\boldsymbol{z} \in N(A)$. As this property is true for all $\boldsymbol{x}, \boldsymbol{y} \in N(A)$ and for all $\alpha, \beta \in \mathbb{R}$, $N(A)$ is a subspace of $\mathbb{R}^n$.

The **nullity** $n(A)$ of A is defined to be the dimension of the nullspace, that is,

$$n(A) = \dim N(A).$$

Note that the solution set of the inhomogeneous system $A\boldsymbol{x} = \boldsymbol{b}$ (where $\boldsymbol{b} \neq \mathbf{0}$) is **never** a subspace of $\mathbb{R}^n$ (as it does not contain the zero vector).

Theorem 4.2 *Let the system $A\boldsymbol{x} = \boldsymbol{b}$ be consistent and suppose $\boldsymbol{x}_1$ is a solution. Then $\boldsymbol{x}_2$ is also a solution if and only if $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{v}$, where $\boldsymbol{v}$ is a solution of the homogeneous system, that is, $A\boldsymbol{v} = \mathbf{0}$.*

Proof 4.2 *($\Rightarrow$) Let $A\boldsymbol{x}_1 = \boldsymbol{b}$ and $A\boldsymbol{x}_2 = \boldsymbol{b}$ and denote $\boldsymbol{v} = \boldsymbol{x}_2 - \boldsymbol{x}_1$.*

$$A\boldsymbol{v} = A(\boldsymbol{x}_2 - \boldsymbol{x}_1) = A\boldsymbol{x}_2 - A\boldsymbol{x}_1 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$$

Hence $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{v}$, where $A\boldsymbol{v} = \boldsymbol{0}$

($\Leftarrow$) Let $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{v}$ where $A\boldsymbol{x}_1 = \boldsymbol{b}$ and $A\boldsymbol{v} = \boldsymbol{0}$. Then

$$A\boldsymbol{x}_2 = A(\boldsymbol{x}_1 + \boldsymbol{v}) = A\boldsymbol{x}_1 + A\boldsymbol{v} = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}$$

Hence $\boldsymbol{x}_2$ is a solution.

Corollary 4.3 *$A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution if and only if $A\boldsymbol{x} = \boldsymbol{0}$ has only the trivial solution $\boldsymbol{x} = \boldsymbol{0}$ (i.e. $N(A) = \{\boldsymbol{0}\}$).*

Extending this idea, we say that the **general** solution of $A\boldsymbol{x} = \boldsymbol{b}$ has the form

$$\boldsymbol{x} = \boldsymbol{x}_1 + \lambda_1\boldsymbol{v}_1 + \lambda_2\boldsymbol{v}_2 + \ldots + \lambda_k\boldsymbol{v}_k$$

where $\boldsymbol{x}_1$ is a **particular** solution of $A\boldsymbol{x} = \boldsymbol{b}$, the vectors $\boldsymbol{v}_1,\ \boldsymbol{v}_1, \ldots \boldsymbol{v}_k$ are solutions of the homogeneous equation $A\boldsymbol{v} = \boldsymbol{0}$, and $\lambda_1,\ \lambda_2, \ldots \lambda_k \in \mathbb{R}$ (we will return to this idea in Theorem 4.10 below). Note that we have already seen solutions of linear equations written in this way when solving linear systems using EROs in Chapter 1: the number of parameters λ_i in the general solution corresponds to the number of free variables when solving by EROs.

4.2 Row and column spaces; rank

Definition 4.4 *Let A be a real $m \times n$ matrix.*

*(a) The **column space** of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A (regarded as vectors in $\mathbb{R}^m$). Its dimension, $c(A)$, is the **column rank** of A.*

*(b) The **row space** of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A. Its dimension, $\rho(A)$, is the **row rank** of A.*

We now make some observations about the relevance of these subspaces to the solution of a linear system $A\boldsymbol{x} = \boldsymbol{b}$.

Firstly, let $A = [a_{ij}]_{2\times 2}$ and $\boldsymbol{x} = (x_1,\ x_2)$. Then we see that

$$A\boldsymbol{x} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}\,x_1 & + & a_{21}\,x_2 \\ a_{21}\,x_1 & + & a_{22}\,x_2 \end{bmatrix} = x_1\begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + x_2\begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix},$$

that is, $A\boldsymbol{x}$ is a linear combination of the columns of A. It is readily shown that the same holds in the general $m \times n$ case: if $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n$ denote the columns of A and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ then

$$A\boldsymbol{x} = x_1\,\boldsymbol{c}_1 + x_2\,\boldsymbol{c}_2 + \cdots + x_n\,\boldsymbol{c}_n.$$

Thus, $A\boldsymbol{x} = \boldsymbol{b}$, where $A = [a_{ij}]_{m\times n}$, can be written as

$$x_1\,\boldsymbol{c}_1 + x_2\,\boldsymbol{c}_2 + \cdots + x_n\,\boldsymbol{c}_n = \boldsymbol{b}.$$

It therefore follows that $A\boldsymbol{x} = \boldsymbol{b}$ is consistent (has a solution) if and only if $\boldsymbol{b}$ is expressible as a linear combination of the columns of A, that is, if and only if $\boldsymbol{b}$ is in the column space of A.

Theorem 4.5 *If A is any matrix, then the row space and column space of A have the same dimension, that is, $\rho(A) = c(A)$.*

Definition 4.6 *The* **rank** $r(A)$ *of a matrix A is the common value of $\rho(A)$ and $c(A)$:*

$$r(A) = \rho(A) = c(A).$$

To find the rank of a matrix, we can use elementary row operations (EROs).

Theorem 4.7 *(a) EROs do not change the row space or the nullspace of a matrix. Thus, if B is derived from A by applying EROs, then row space of A = row space of B and $N(A) = N(B)$.*

(b) The non-zero rows in an echelon matrix B form a basis for the row space of B.

This theorem means that to calculate the rank of a matrix A we need only perform EROs to reduce A to an echelon matrix B: the required rank is then given by

$$r(A) = r(B) = \text{ number of non-zero rows in } B.$$

Theorem 4.8 ***Dimension Theorem for Matrices***

If A is a matrix with n columns then $r(A) + n(A) = n$.

Proof 4.8 *Since A has n columns, the linear system $A\boldsymbol{x} = \mathbf{0}$ has n unknowns. By applying suitable EROs to the augmented matrix $[A|\mathbf{0}]$, we can produce the equivalent augmented matrix $[B|\mathbf{0}]$ in which B is in echelon form. Recall that a variable x_k (kth component in $\boldsymbol{x}$) is called a* ***leading variable*** *if column k of B contains a* ***leading 1***, *otherwise x_k is a* ***free variable.***

$$\textit{number of leading variables} + \textit{number of free variables} = n.$$

But

$$\begin{aligned}\textit{number of leading variables} &= \textit{number of leading 1s} = \textit{number of non-zero rows of } B\\ &= r(B) = r(A).\end{aligned}$$

Also,

$$\begin{aligned}\textit{number of free variables} &= \textit{number of parameters in general solution of } A\boldsymbol{x} = \mathbf{0}\\ &= n(A).\end{aligned}$$

So $r(A) + n(A) = n$.

Note: $r(A)$ = number of leading variables that occur in solving $A\boldsymbol{x} = \mathbf{0}$;
$n(A)$ = number of free variables that occur in solving $A\boldsymbol{x} = \mathbf{0}$
= number of parameters in general solution of $A\boldsymbol{x} = \mathbf{0}$.

Examples 4A

Theorem 4.9 ***Consistency Theorem***
The linear system $A\boldsymbol{x} = \boldsymbol{b}$ *is consistent if and only if* $r(A) = r(A|\boldsymbol{b})$.

Proof 4.9 $A\boldsymbol{x} = \boldsymbol{b}$ *is consistent*

$\Leftrightarrow$ $\boldsymbol{b}$ *is in column space of* A

$\Leftrightarrow$ *column space of* A = *column space of* $A|\boldsymbol{b}$

$\Leftrightarrow$ $r(A) = r(A|\boldsymbol{b})$.

Theorem 4.10 *If* $\boldsymbol{x}_1$ *denotes any single solution of the consistent inhomogeneous linear system* $A\boldsymbol{x} = \boldsymbol{b}$, *and if* $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ *form a basis for* $N(A)$, *then* ***every*** *solution of* $A\boldsymbol{x} = \boldsymbol{b}$ *can be expressed in the form*

$$\boldsymbol{x} = \boldsymbol{x}_1 + \lambda_1 \boldsymbol{v}_1 + \lambda_2 \boldsymbol{v}_2 + \cdots + \lambda_k \boldsymbol{v}_k. \tag{4.13}$$

Conversely, for any choice of scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$, *the vector* $\boldsymbol{x}$ *given by (4.13) is a solution of* $A\boldsymbol{x} = \boldsymbol{b}$.

Proof 4.10 *Let* $\boldsymbol{x}_1$ *be a fixed solution of* $A\boldsymbol{x} = \boldsymbol{b}$ *and suppose* $\boldsymbol{x}$ *denotes any other solution. Then*

$$\begin{aligned}
&A(\boldsymbol{x} - \boldsymbol{x}_1) = A\boldsymbol{x} - A\boldsymbol{x}_1 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0} \\
&\Rightarrow \boldsymbol{x} - \boldsymbol{x}_1 \in N(A) \\
&\Rightarrow \boldsymbol{x} - \boldsymbol{x}_1 = \sum_{j=1}^{k} \lambda_j \, \boldsymbol{v}_j, \ \textit{since } \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \textit{ is a basis for } N(A) \\
&\Rightarrow \boldsymbol{x} = \boldsymbol{x}_1 + \sum_{j=1}^{k} \lambda_j \, \boldsymbol{v}_j.
\end{aligned}$$

Conversely, for any choice of $\lambda_1, \lambda_2, \ldots, \lambda_k$,

$$\begin{aligned}
A\boldsymbol{x} &= A(\boldsymbol{x}_1 + \lambda_1 \, \boldsymbol{v}_1 + \cdots + \lambda_k \, \boldsymbol{v}_k) = A\boldsymbol{x}_1 + \lambda_1(A\boldsymbol{v}_1) + \cdots + \lambda_k(A\boldsymbol{v}_k) \\
&= \boldsymbol{b} + \boldsymbol{0} + \boldsymbol{0} + \cdots + \boldsymbol{0} = \boldsymbol{b}
\end{aligned}$$

Note: In (4.13), we say

$$\begin{array}{rl}
\boldsymbol{x}_1 & \text{is a particular solution of } A\boldsymbol{x} = \boldsymbol{b} \\
\boldsymbol{x}_1 + \sum_{j=1}^{k} \lambda_j \, \boldsymbol{v}_j & \text{is the general solution of } A\boldsymbol{x} = \boldsymbol{b}\text{, and} \\
\sum_{j=1}^{k} \lambda_j \, \boldsymbol{v}_j & \text{is the general solution of } A\boldsymbol{x} = \boldsymbol{0}.
\end{array}$$

Hence the general solution of $A\boldsymbol{x} = \boldsymbol{b}$ can be written as any particular solution of $A\boldsymbol{x} = \boldsymbol{b}$ + the general solution of $A\boldsymbol{x} = \boldsymbol{0}$.

We will see a similar structure later when considering solutions of certain differential equations.

Theorem 4.11 *Let* A *be an* $m \times n$ *matrix of rank* r, *and let* $\boldsymbol{b} \in \mathbb{R}^m$ *be such that* $r(A) = r(A|\boldsymbol{b}) = r$. *Then the general solution of the consistent system* $A\boldsymbol{x} = \boldsymbol{b}$ *contains* $n - r$ *parameters.*

Proof 4.11 *If* k *is the number of parameters in the general solution, then*

$$k = n(A) = n - r(A) \quad \textit{(by Theorem 4.8)} \quad = \quad n - r.$$

4.3 Summary of situation for $n \times n$ systems

Theorem 4.12 *If* A *is an* $n \times n$ *matrix, then the following statements are equivalent.*

(a) A is non-singular (A^{-1} exists)

(b) $n(A) = 0$

(c) $r(A) = n$

(d) $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $\boldsymbol{b} \in \mathbb{R}^n$

(e) A is (row) equivalent to I_n.

Proof 4.12 *We will prove that (a) ⇒ (b), (b) ⇒ (c), (c) ⇒ (d), (d) ⇒ (e), (e) ⇒ (a).*

$$
\begin{aligned}
\text{(a)} \Rightarrow \text{(b)}: \quad A\boldsymbol{x}_0 = \boldsymbol{0} \;&\Rightarrow\; A^{-1}(A\boldsymbol{x}_0) = \boldsymbol{0} \Rightarrow (A^{-1}A)\boldsymbol{x}_0 = \boldsymbol{0} \Rightarrow \boldsymbol{x}_0 = \boldsymbol{0} \\
&\Rightarrow\; N(A) = \{\boldsymbol{0}\} \Rightarrow n(A) = 0.
\end{aligned}
$$

$$
\begin{aligned}
\text{(b)} \Rightarrow \text{(c)}: \quad n(A) = 0 \Rightarrow r(A) &= n - n(A) \quad \textit{(Thm 4.8)} \\
&= n - 0 = n.
\end{aligned}
$$

$$
\begin{aligned}
\text{(c)} \Rightarrow \text{(d)}: \quad r(A) = n \;&\Rightarrow\; \textit{column space of } A = \mathbb{R}^n \\
&\Rightarrow\; \textit{column space of } A|\boldsymbol{b} = \mathbb{R}^n \quad \forall \boldsymbol{b} \in \mathbb{R}^n \\
&\Rightarrow\; r(A|\boldsymbol{b}) = n \Rightarrow A\boldsymbol{x} = \boldsymbol{b} \text{ consistent.}
\end{aligned}
$$

$$
\begin{aligned}
\text{(d)} \Rightarrow \text{(e)}: \quad & A\boldsymbol{x} = \boldsymbol{b} \; \textit{has solution } \forall \boldsymbol{b} \in \mathbb{R}^n \\
&\Rightarrow \textit{column space of } A = \mathbb{R}^n \\
&\Rightarrow r(A) = n \\
&\Rightarrow A \sim I_n.
\end{aligned}
$$

$$
\begin{aligned}
\text{(e)} \Rightarrow \text{(a)}: \quad A \sim I_n \;&\Rightarrow\; \exists \textit{ matrices representing EROs } E_1, E_2, \ldots, E_k \textit{ s.t. } E_k E_{k-1} \ldots E_1\, A = I_n \\
&\Rightarrow\; A = E_1^{-1} E_2^{-1} \ldots E_k^{-1} I_n = E_1^{-1} E_2^{-1} \ldots E_k^{-1},
\end{aligned}
$$

and therefore A is non-singular and has inverse $A^{-1} = E_k\, E_{k-1} \ldots E_1$.

Examples 4B

We complete this section by returning to determinants, giving two theorems that relate $\det(A)$ to the existence of solutions of $A\boldsymbol{x} = \boldsymbol{b}$.

Theorem 4.13 *Let A and B be $n \times n$ matrices. Then A is singular* $\Leftrightarrow$ $\det(A) = 0$.

Proof 4.13 *($\Rightarrow$) Suppose A is singular. Then*

$$\begin{aligned}
&r(A) < n \quad \textit{(from Theorem 4.12)} \\
\Rightarrow\ & A \sim C \quad \textit{where } C \textit{ is an echelon matrix with at least one zero row} \\
\Rightarrow\ & A = PC \quad \textit{where } P \textit{ is an invertible } n \times n \textit{ matrix} \\
\Rightarrow\ & \det(A) = \det(P)\det(C) \quad \textit{by Theorem 1.16 (d)} \\
\Rightarrow\ & \det(A) = 0 \quad \textit{since } \det(C) = 0 \textit{ by Theorem 1.16 (b).}
\end{aligned}$$

($\Leftarrow$) Suppose $\det(A) = 0$, *but A is non-singular (i.e. A^{-1} exists). Then*

$$0 = \det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1 \quad \textit{(by Theorem 1.16 (a)).}$$

This is a contradiction, so A^{-1} cannot exist and A must be singular.

Note: If $\det(A) \neq 0 \Rightarrow A^{-1}$ exists and $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ for any given $\boldsymbol{b} \in \mathbb{R}^n$.

Theorem 4.14 *If A is an $n \times n$ matrix then $A\boldsymbol{x} = \mathbf{0}$ has a non-trivial solution $\boldsymbol{x}$ (i.e. $\boldsymbol{x} \neq \mathbf{0}$) if and only if* $\det(A) = 0$.

Proof 4.14

$$\begin{aligned}
&A\boldsymbol{x} = \mathbf{0} \quad \textit{for some } \boldsymbol{x} \neq \mathbf{0} \\
\Leftrightarrow\ & N(A) \quad \textit{contains at least one non-zero vector} \\
\Leftrightarrow\ & n(A) > 0 \\
\Leftrightarrow\ & A \quad \textit{is singular (by Theorem 4.12)} \\
\Leftrightarrow\ & \det(A) = 0 \quad \textit{(by Theorem 4.13).}
\end{aligned}$$

Examples 4C

5 Eigenvalues, eigenvectors and diagonalisation

5.1 Eigenvalues and eigenvectors

In subsection 2.9 we talked about taking a vector $\in \mathbb{R}^n$ and creating from it a new vector $\boldsymbol{y} \in \mathbb{R}^n$ by applying to $\boldsymbol{x}$ an $n \times n$ matrix A: $\boldsymbol{y} = A\boldsymbol{x}$. Sometimes it happens that the resulting $\boldsymbol{y}$ is "parallel" to $\boldsymbol{x}$...

Definition 5.1 *The number λ is an **eigenvalue** of the $n \times n$ matrix A if there exists a* **non-zero** *vector $\boldsymbol{x}$ such that*

$$A\boldsymbol{x} = \lambda\boldsymbol{x}.$$

Any **non-zero** $\boldsymbol{x}$ *satisfying this equation is called an **eigenvector** of A corresponding to the eigenvalue λ.*

Remarks: 1. Clearly, $\boldsymbol{x} = \boldsymbol{0}$ solves $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for every scalar λ, which is the reason that we are looking for the very particular λ (eigenvalues) for which $A\boldsymbol{x} = \lambda\boldsymbol{x}$ has a **non-zero** solution.

2. If $\boldsymbol{x}$ is an eigenvector of A corresponding to an eigenvalues λ, then so is $\alpha\boldsymbol{x}$ for any scalar α:

$$A(\alpha\boldsymbol{x}) = \alpha A\boldsymbol{x} = \alpha\lambda\boldsymbol{x} = \lambda(\alpha\boldsymbol{x}).$$

Eigenvalues and eigenvectors are of crucial importance in science as arise in the study of **stability** of mechanical and electrical systems, **stability** of numerical algorithms, and many other areas of application.

To find the eigenvalues of an $n \times n$ matrix A we rewrite $A\boldsymbol{x} = \lambda\boldsymbol{x}$ as

$$A\boldsymbol{x} = \lambda I_n \boldsymbol{x}, \qquad \text{or} \qquad (A - \lambda I_n)\boldsymbol{x} = \boldsymbol{0}.$$

By Theorem 4.14, this equation has a non-trivial solution $\boldsymbol{x}$ if and only if $\det(A - \lambda I_n) = 0$. The equation

$$\det(A - \lambda I_n) = 0$$

is called the **characteristic equation** of A, and the values of λ satisfying this equation are the eigenvalues. (Note that the characteristic equation may be equivalently written as $\det(\lambda I_n - A) = 0$). To find the eigenvector $\boldsymbol{x}_i$ corresponding to a particular eigenvalue λ_i, we can then solve the linear system

$$(A - \lambda_i I_n)\boldsymbol{x}_i = \boldsymbol{0}.$$

Notes:

(i) If A is $n \times n$, the characteristic equation $\det(A - \lambda I_n) = 0$ is a polynomial equation in λ of degree n. By the Fundamental Theorem of Algebra, this equation has n roots, some of which may be multiple roots and some of which may be complex. (If $\lambda \in \mathbb{C}$, then we have to seek an eigenvector in $\mathbb{C}^n$.

(ii) If λ is a root of $\det(A - \lambda I_n) = 0$ of multiplicity k, then k is called the **algebraic multiplicity** of the eigenvalue λ.

(iii) We have that $\boldsymbol{x}$ is an eigenvector of A corresponding to the eigenvalue λ if and only if $(A - \lambda I_n)\boldsymbol{x} = \mathbf{0}$, $\boldsymbol{x} \neq \mathbf{0}$, that is, if and only if $\boldsymbol{x}$ is in the nullspace of $A - \lambda I_n$. The nullspace $N(A - \lambda I_n)$ is called the **eigenspace** of A corresponding to λ. The dimension of $N(A - \lambda I_n)$ is called the **geometric multiplicity** of λ: it is the number of linearly independent eigenvectors corresponding to λ.

Examples 5A

5.2 Diagonalisation of matrices

In this section we will see how to transform A into simpler forms like diagonal matrices which are more convenient in many situations, for example, when solving systems of differential equations. This is not always possible!

The basic problem is this: given $A = [a_{ij}]_{n \times n}$, does there exist a non-singular matrix P such that $P^{-1}AP$ is a diagonal matrix? For any invertible P, A and $P^{-1}AP$ are called **similar** matrices. The problem, therefore, can be stated as determining whether or not A is similar to a diagonal matrix.

Definition 5.2 *An $n \times n$ matrix A is* ***diagonalisable*** *if there exists a non-singular matrix P such that $P^{-1}AP$ is a diagonal matrix. The matrix P is said to* ***diagonalise*** *A.*

We need the following result:

Lemma 5.3 *If $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a set of n linearly independent n-vectors, the $n \times n$ matrix*

$$P = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$$

having $\boldsymbol{x}_i$ as its i-th column is invertible.

To prove it, consider the solutions of $P\boldsymbol{y} = \boldsymbol{0}$. Now we can prove the following result:

Theorem 5.4 *An $n \times n$ matrix A is diagonalisable if and only if A has n linearly independent eigenvectors.*

Proof 5.4 *($\Rightarrow$) Suppose the linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ of A correspond to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively (the eigenvalues may not be distinct: each is repeated according to its algebraic multiplicity). Let $P = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ldots \boldsymbol{x}_n]$ be the $n \times n$ matrix with $\boldsymbol{x}_i$ as its ith column, for $i = 1, 2, \ldots, n$. Then*

$$
\begin{aligned}
AP &= A[\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ldots \boldsymbol{x}_n] = [\lambda_1 \boldsymbol{x}_1 \ \lambda_2 \boldsymbol{x}_2 \ldots \lambda_n \boldsymbol{x}_n] \\
&= [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ldots \boldsymbol{x}_n] \begin{bmatrix} \lambda_1 & & & O \\ & \lambda_2 & & \\ & O & \ddots & \\ & & & \lambda_n \end{bmatrix} \\
&= P \ \mathrm{diag}(\lambda_1, \ \lambda_2, \ldots, \lambda_n) \\
&= P\,D, \ \textit{say.}
\end{aligned}
$$

Note that by Lemma 5.3 P is invertible. Hence P^1 exists and $P^{-1}AP = D$, so that and A is diagonalisable.

($\Leftarrow$) Proof omitted.

Theorem 5.5 *If $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ $(m \leq n)$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, then $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ is a linearly independent set.*

Corollary 5.6 *It follows from Theorem 5.5 that if A has n distinct eigenvalues then A has n linearly independent eigenvectors, and A is diagonalisable.*

Important note: If an $n\times$ matrix A does not have n distinct eigenvalues, then it is diagonalisable if and only if the algebraic multiplicity and geometric multiplicity of each eigenvalue λ are equal, that is, if it has n linearly independent eigenvectors.

Examples 5B

5.3 Results for symmetric matrices

More can be said about symmetric matrices, $A = A^T$, which are very important in quantum mechanics.

Theorem 5.7 *Let A be a real $n \times n$ symmetric matrix. Then*

*(a) A has only **real** eigenvalues.*

(b) For each eigenvalue λ,

$$\textit{geometric multiplicity of } \lambda = \textit{algebraic multiplicity of } \lambda.$$

Hence A has n linearly independent eigenvectors and $\exists P$ s.t.

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

(c) If λ_i and λ_j are eigenvalues with $\lambda_i \neq \lambda_j$ then the corresponding eigenvectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are orthogonal, that is, $\boldsymbol{x}_i \cdot \boldsymbol{x}_j = 0$.

*(d) There exists an **orthogonal** matrix P such that*

$$P^T AP = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

(P is orthogonal if $PP^T = I$, or $P^{-1} = P^T$.)

Let us prove part (c). For that, check that if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and A is an $n \times n$ matrix, then

$$A\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x} \cdot A^T \boldsymbol{y}.$$

Now let $\boldsymbol{u}_1, \boldsymbol{u}_2$ be eigenvectors corresponding to two different eigenvalues $\lambda_1 \neq \lambda_2$ of A.

Then we have

$$\lambda_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2) = A\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = \boldsymbol{u}_1 \cdot A^T \boldsymbol{u}_2 =$$

$$\boldsymbol{u}_1 \cdot A\boldsymbol{u}_2 = \boldsymbol{u}_1 \cdot \lambda_2 \boldsymbol{u}_2 = \lambda_2(\boldsymbol{u}_1 \cdot \boldsymbol{u}_2),$$

which, since $\lambda_1 \neq \lambda_2$, is only possible if $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 0$.

Suppose A is an $n \times n$ symmetric matrix. To construct an orthogonal matrix P such that $P^T AP = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ we find an **orthonormal** set of eigenvectors of A, say $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n\}$: thus, $A\boldsymbol{z}_i = \lambda_i \boldsymbol{z}_i$, with $\boldsymbol{z}_i \cdot \boldsymbol{z}_j = \delta_{ij}$. The vectors $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n$ form the columns of P, so that

$$P = [\boldsymbol{z}_1 \; \boldsymbol{z}_2 \ldots \boldsymbol{z}_n].$$

The simplest case is when A has **distinct** eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with corresponding orthogonal eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ (Theorem 5.7(c)). We can then take $\boldsymbol{z}_i = \boldsymbol{x}_i/\|\boldsymbol{x}_i\|$ for $i = 1, 2, \ldots, n$.

Summary of strategy for symmetric matrices:

- Find eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and the corresponding linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$.
- If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct, take $\boldsymbol{z}_i = \boldsymbol{x}_i/\|\boldsymbol{x}_i\|,\ i = 1, \ldots, n$.
- If two eigenvalues are equal, $\lambda_i = \lambda_j$ say, apply Gram-Schmidt to the associated eigenvectors $\boldsymbol{x}_i, \boldsymbol{x}_j$ to produce orthonormal eigenvectors $\boldsymbol{z}_i, \boldsymbol{z}_j$. Let $\boldsymbol{z}_k = \boldsymbol{x}_k/\|\boldsymbol{x}_k\|$ for distinct eigenvalues to complete the set. (Use a similar approach if $\lambda_i = \lambda_j = \lambda_k$ etc).
- Take $P = [\boldsymbol{z}_1\ \boldsymbol{z}_2 \ldots \boldsymbol{z}_n]$, then $P^T AP = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

Examples 5C