MTH4103: Geometry I

Dr John N. Bray,
Queen Mary, University of London

January–March 2014

Contents

Preface **iv**

1 Vectors **1**
- 1.1 Introduction . . . 1
- 1.2 Vectors . . . 2
- 1.3 The zero vector . . . 3
- 1.4 Vector negation . . . 3
- 1.5 Parallelograms . . . 4
- 1.6 Vector addition . . . 4
- 1.7 Some notation . . . 5
- 1.8 Rules for vector addition . . . 5
- 1.9 Vector subtraction . . . 8
- 1.10 Scalar multiplication . . . 9
- 1.11 Position vectors . . . 11

2 Cartesian Coördinates **14**
- 2.1 Sums and scalar multiples using coördinates . . . 16
- 2.2 Unit vectors . . . 16
- 2.3 Equations of lines . . . 17
  - 2.3.1 The line determined by two distinct points . . . 19

3 The Scalar Product **20**
- 3.1 The scalar product using coördinates . . . 21
- 3.2 Properties of the scalar product . . . 23
- 3.3 Equation of a plane . . . 24
- 3.4 The distance from a point to a plane . . . 25
- 3.5 The distance from a point to a line . . . 27

4 Intersections of Planes and Systems of Linear Equations **28**
- 4.1 Possible intersections of k planes in $\mathbb{R}^3$ . . . 29
  - 4.1.1 Empty intersections, unions, sums and products . . . 29
- 4.2 Some examples . . . 30
- 4.3 Notes . . . 34

5 Gaußian Elimination and Echelon Form **35**
5.1 Echelon form (Geometry I definition) 35
5.2 Gaußian elimination 36
5.2.1 Notes on Gaußian elimination 37
5.3 Solving systems of equations in echelon form 39
5.3.1 Back substitution 39
5.4 Summary 40
5.5 Intersection of a line and a plane 41
5.6 Intersection of two lines 42

6 The Vector Product **44**
6.1 Parallel vectors 44
6.2 Collinear vectors 45
6.3 Coplanar vectors 45
6.4 Right-handed and left-handed triples 47
6.4.1 Some finger exercises (for you to do) 48
6.5 The vector product 49
6.6 Properties of the vector product 50
6.6.1 The area of a parallelogram and triangle 51
6.6.2 The triple scalar product 52
6.6.3 The distributive laws for the vector product 54
6.7 The vector product in coördinates 54
6.8 Applications of the vector product 55
6.8.1 Distance from a point to a line 55
6.8.2 Distance between two lines 56
6.8.3 Equations of planes (revisited) 58
6.9 Is the cross product commutative or associative? 60
6.9.1 The triple vector product(s) 60

7 Matrices **62**
7.1 Addition of matrices 64
7.2 Rules for matrix addition 64
7.3 Scalar multiplication of matrices 65
7.4 Rules for scalar multiplication 65
7.5 Matrix multiplication 65
7.6 Rules for matrix multiplication 67
7.7 Some useful notation 69
7.8 Inverses of matrices 69
7.9 Matrices with no rows or no columns 71
7.10 Transposes of matrices 71

8 Determinants **73**
8.1 Inverses of 2×2 matrices, and determinants 73
8.2 Determinants of 3×3 matrices 76
8.3 Systems of linear equations as matrix equations 79
8.4 Determinants and inverses of $n \times n$ matrices 79
8.4.1 Determinants of $n \times n$ matrices 80
8.4.2 Adjugates and inverses of $n \times n$ matrices 81

9 Linear Transformations **82**
9.1 The vector space $\mathbb{R}^n$ 82
9.2 Linear transformations 83
9.3 Properties of linear transformations 84
9.4 Matrices and linear transformations 85
9.5 Composition of linear transformations and multiplication of matrices . . 86
9.6 Rotations and reflexions of the plane 87
9.7 Other linear transformations of the plane 89
9.7.1 Linear 'stretches' of axes and dilations 89
9.7.2 Shears (or transvections) 89
9.7.3 Singular transformations 90
9.7.4 Translations and affine maps 91

10 Eigenvectors and Eigenvalues **92**
10.1 Definitions 92
10.2 Eigenvectors corresponding to the eigenvalue 0 92
10.3 Finding all eigenvalues 93
10.4 Finding eigenvectors 94
10.5 Eigenvectors and eigenvalues for linear transformations of the plane . . . 96
10.6 Rotations and reflexions in $\mathbb{R}^3$ 97

Preface

These are notes for the course MTH4103: Geometry I that I gave (am giving) in Semester B of the 2013–14 academic year at Queen Mary, University of London. These notes are based on handwritten notes I inherited from Professor L. H. Soicher, who lectured this course in the academic years 2005–06 to 2009–10. The notes were typed up (using the LaTeX document preparation system) for the 2010–11 session by the two lecturers that year, namely Dr J. N. Bray (Chapters 1–6) and Prof. S. R. Bullett (Chapters 7–10). Naturally, several modifications were made to the notes in the process of typing them up, as one expects to happen when a new lecturer takes on a course. I have made many further revisions to the notes last year, and a few more this year, including some to take advantage of the new module MTH4110: Mathematical Structures. Since I have now typed up and/or edited the whole set of notes, the *culpa* for any errors, omissions or infelicities therein is entirely *mea*.

Despite the presence of these notes, one should still take (or have taken) notes in lectures. Certainly the examples given in lectures differ from those in these notes, and there is material I covered in the lectures that does not appear in these notes, and *vice versa*. Furthermore, the examinable material for the course is defined by what was covered in lectures (including the proofs). In these notes, I have seldom indicated which material is examinable.

My thanks go out to those colleagues who covered the several lectures I missed owing to illness.

Dr John N. Bray, 25th March 2014

Chapter 1

Vectors

1.1 Introduction

The word *geometry* derives from the Ancient Greek word γεωμετρία, with a rough meaning of *geometry, land-survey*, though I prefer *earth measurement*. There are two elements in this word: γῆ (or Γῆ), meaning *Earth* (among other related words), and either μετρέω, *to measure, to count*, or μέτρον, *a measure*. Given this etymology, one should have a fair idea of what geometry is about. (Consult the handouts on the web for a copy of the Greek alphabet; the letter names should give a [very] rough guide to the pronunciations of the letters themselves.)

In this module, we are interested in lines, planes, and other geometrical objects in 3-dimensional space (and maybe spaces of other dimensions).

We shall introduce standard notation for some number systems.

- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, the *natural numbers*. I always include 0 as a natural number; some people do not.

- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the *integers*. The notation comes from the German *Zahlen* meaning *numbers*.

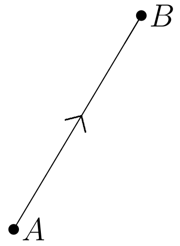
- $\mathbb{N}^+ = \mathbb{Z}^+ = \{1, 2, 3, \ldots\}$, the *positive integers*, and $\mathbb{N}_0 = \mathbb{Z}_{\geqslant 0} = \{0, 1, 2, 3, \ldots\}$, the *non-negative integers*. The word *positive* here means *strictly positive*, that is to say 0 is not considered to be a positive (or negative) number.

- $\mathbb{Q} = \{\, \frac{a}{b} : a, b \text{ are integers and } b \neq 0 \,\}$. The Q is the first letter of *quotient*.

- $\mathbb{R}$ denotes the set of 'real' numbers. Examples of real numbers are 2, $\frac{3}{5}$, $\sqrt{7}$ and π. Also, all decimal numbers, both terminating and not, are real numbers. In fact, each real number can be represented in decimal form, though this decimal is usually non-terminating and non-recurring. An actual definition of $\mathbb{R}$ is somewhat technical, and is deferred (for a long time). The set $\mathbb{R}$ is a very artificial construct and not very real at all.

1.2 Vectors

A *bound vector* is a bounded (and directed) line segment $\overrightarrow{AB}$ in 3-dimensional [real] space, which we shorten to 3-space and denote by $\mathbb{R}^3$, where A and B are points in this space. (There is no particular reason, except familiarity, to restrict ourselves to 3-space, and one often works in much higher dimensions, for example 196884-space.) We point out that the real world around us probably bears little resemblence to $\mathbb{R}^3$, despite the fact we are fondly imagining that it does.



Note that a bound vector $\overrightarrow{AB}$ is determined by (and determines) three things:

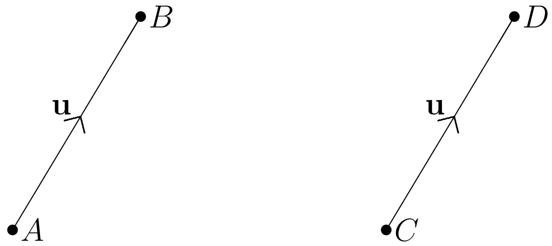
(i) its *length* or *magnitude*, denoted $|\overrightarrow{AB}|$,

(ii) its *direction* [provided $A \neq B$], and

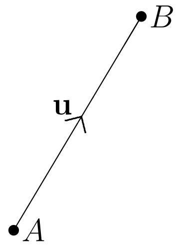
(iii) its *starting point*, which is A.

If $A = B$, then $\overrightarrow{AB} = \overrightarrow{AA}$ does not have a defined direction, and in this case $\overrightarrow{AB}$ is determined by its length, which is 0, and its starting point A.

If we ignore the starting point, and only care about the length and direction, we get the notion of a *free vector* (or simply *vector* in what follows). Thus $\overrightarrow{AB}$ and $\overrightarrow{CD}$ represent the same free vector if and only if they have the same length and direction.



We use $\mathbf{u}, \mathbf{v}, \mathbf{w}, \ldots$ to denote free vectors (these would be underlined when hand-written: thus $\underline{u}, \underline{v}, \underline{w}, \ldots$), and draw



to mean that the free vector $\mathbf{u}$ is represented by $\overrightarrow{AB}$; that is, the length and direction of $\mathbf{u}$ are those of the bound vector $\overrightarrow{AB}$. We denote the length of $\mathbf{u}$ by $|\mathbf{u}|$.

Note. We should *never* write something like $\overrightarrow{AB} = \mathbf{u}$, tempting though it may be. The reason is that two objects on each side of the equal sign are different types of object (a bound vector versus a free vector), and it is always inappropriate to relate different types of object using the equality sign.

Note. The module text uses $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}, \ldots$ for (free) vectors; this is perfectly standard in printed works. The previous lecturer used $\underline{u}, \underline{v}, \underline{w}, \ldots$ (see the 2010 exam, for example), which is decidedly non-standard in print. The book also uses $\mathbf{AB}$ for the free vector represented by $\overrightarrow{AB}$, which we shall never use. Better notations for the free vector represented by $\overrightarrow{AB}$ are free($\overrightarrow{AB}$) or $[\overrightarrow{AB}]$, but we shall hardly ever use these either.[1]

Note. Each bound vector represents a unique free vector. Also, for each free vector $\mathbf{u}$ and for each point A there is a unique point B such that $\overrightarrow{AB}$ represents $\mathbf{u}$. This is a consequence of a bound vector being determined by its length, direction and starting point, and a free vector being determined by its length and direction only. Of course, a suitable (and annoying) modification must be made to the above when the zero vector (see below) is involved. We leave such a modification to the reader.

1.3 The zero vector

The *zero vector* is the (free) vector with zero length. Its direction is undefined. We denote the zero vector by $\mathbf{0}$ (or $\underline{0}$ in handwriting). It is represented by $\overrightarrow{AA}$, where A can be any point. (It is also represented by $\overrightarrow{DD}$, where D can be any point, and so on.)

1.4 Vector negation

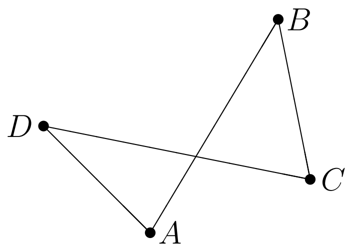
If $\mathbf{v}$ is a nonzero vector, then the negative of $\mathbf{v}$, denoted $-\mathbf{v}$, is the vector with the same length as $\mathbf{v}$ but opposite direction. We define $-\mathbf{0} := \mathbf{0}$. If $\overrightarrow{AB}$ represents $\mathbf{v}$ then $\overrightarrow{BA}$ represents $-\mathbf{v}$.

Note. Vector negation is a function from the set of free vectors to itself. It is therefore essential that it be defined for ***every*** element in the domain. That is, we must define the negative of every free vector. Note here the special treatment of the zero vector, which is not covered by the first sentence.

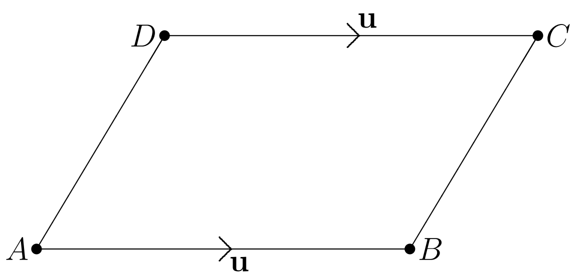
[1] Usually, $[\overrightarrow{AB}]$ would denote the equivalence class containing $\overrightarrow{AB}$. Here the relevant equivalence relation is that two bound vectors are equivalent if and only if they represent the same free vector. There is an obvious bijection between the set of these equivalence classes and the set of free vectors. See the module MTH4110: Mathematical Structures for definitions of equivalence relation and equivalence class.

1.5 Parallelograms

Suppose A, B, C, D are any points in 3-space. We obtain the figure $ABCD$ by joining A to B (by a [straight] line segment), B to C, C to D, and finally D to A. For example:



The figure $ABCD$ is called a *parallelogram* if $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vector.



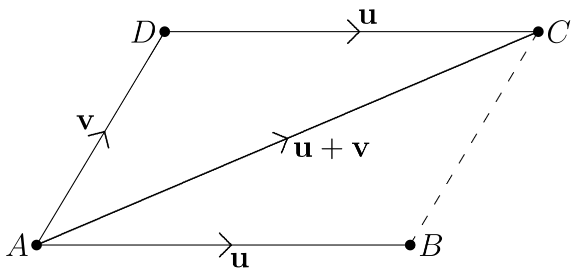
We note the following fact, which is really an axiom. We shall make some use of this later.

Fact (Parallelogram Axiom)**.** Note that $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vector ($\mathbf{u}$ say) if and only if $\overrightarrow{AD}$ and $\overrightarrow{BC}$ represent the same vector ($\mathbf{v}$ say). We can have $\mathbf{u} = \mathbf{v}$, though more usually we will have $\mathbf{u} \neq \mathbf{v}$.

Note. We now see the folly of writing expressions like $\overrightarrow{AD} = \mathbf{v}$, where one side is a bound vector, and one side is a free vector. For example, in the above parallelogram, we would notice that $\overrightarrow{AB} = \mathbf{u} = \overrightarrow{DC}$ and deduce, using a well-known property of equality, that $\overrightarrow{AB} = \overrightarrow{DC}$. But this is nonsense in the general case (when $A \neq D$), since the vectors $\overrightarrow{AB}$ and $\overrightarrow{DC}$ have different starting points and are therefore *not* equal.

1.6 Vector addition

Now suppose that $\mathbf{u}$ and $\mathbf{v}$ are any vectors. Choose a point A. Further, assume that points B and D are chosen so that $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{AD}$ represents $\mathbf{v}$. (The points B and D are unique.) We extend the A, B, D-configuration to a parallelogram by choosing a point C (which is unique) such that $\overrightarrow{DC}$ represents $\mathbf{u}$, as in the diagram below. (Note that $\overrightarrow{BC}$ represents $\mathbf{v}$ by the Parallelogram Axiom.)



The *sum* of $\mathbf{u}$ and $\mathbf{v}$, which we denote as $\mathbf{u} + \mathbf{v}$, is defined to be the vector represented by $\overrightarrow{AC}$.

Note that I have defined vector addition only for *free* vectors, *not* for bound vectors, so I do not wish to see you write things like $\overrightarrow{AB} + \overrightarrow{CD} = \cdots$.

1.7 Some notation

In lectures, you will often see abbreviations for various mathematical concepts. Some of these appear less often in printed texts. At least one of the symbols ($\forall$) was introduced around here. Most of these symbols can be negated.

- s.t. means 'such that'.
- $\forall$ means 'for all'.
- $\exists$ means 'there exists', while $\exists!$ means 'there exists unique'.
- $\nexists$ means 'there does not exist'.
- $a \in B$ means that the element a is a member of the set B.
- $a \notin B$ means that the element a is not a member of the set B.
- $A \subseteq B$ means that the set A is a subset of the set B (allows the case $A = B$).

1.8 Rules for vector addition

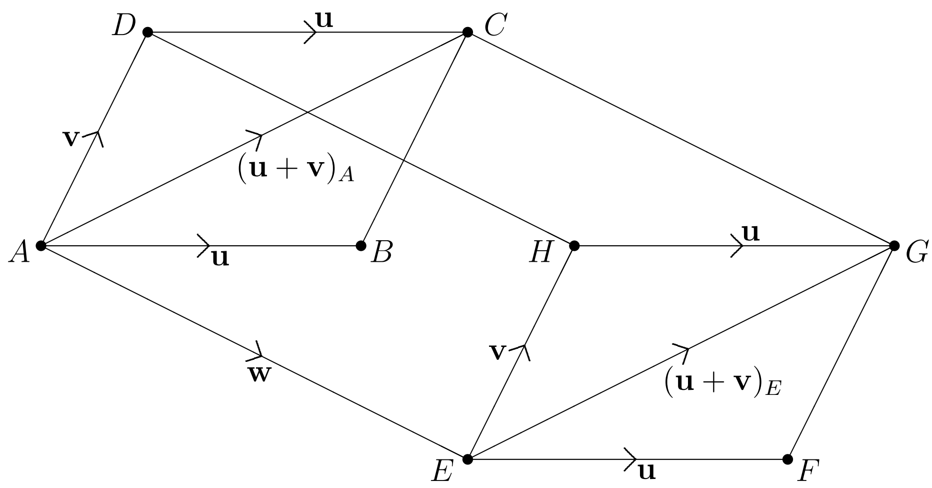
In the definition of $\mathbf{u} + \mathbf{v}$ you will notice the use of an arbitrary point A. When one encounters something like this, one is entitled to ask whether the definition depends on the point A or not. Mathematicians, being pedants, very often will ask such seemingly obvious questions. Temporarily, we shall use $(\mathbf{u} + \mathbf{v})_A$ to denote the value of $\mathbf{u} + \mathbf{v}$ obtained if the arbitrary point A was used in its definition. (There is no need to worry about the points B, C and D, since these are uniquely determined given $\mathbf{u}$, $\mathbf{v}$ and A.)

We also take the opportunity in the following couple of pages to introduce terms such as *commutative*, *associative*, *identity*, *inverse* and *distributive*. You should meet these terms many times in your mathematical career. In the lectures we proved Theorems 1.2

and 1.3 before Theorem 1.1. The box at the end of the proofs is the end-of-proof symbol. One can write things like ‘QED’ (*quod erat demonstrandum* meaning ‘which was to be shown’) instead.

Theorem 1.1. The definition of $\mathbf{u}+\mathbf{v}$ does not depend on the point A used to define it. In notation we have $(\mathbf{u}+\mathbf{v})_A = (\mathbf{u}+\mathbf{v})_E$ for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all points A and E.

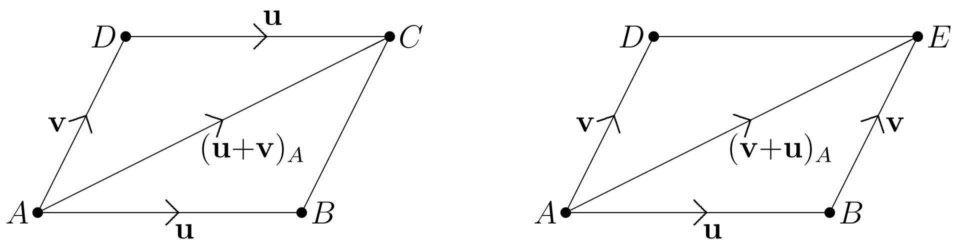
Proof. This proof involves three applications of the Paralellogram Axiom. Let $ABCD$ be the parallelogram obtained by using the Parallelogram Rule for vector addition to calculate $(\mathbf{u}+\mathbf{v})_A$, and let $EFGH$ be the parallelogram obtained by using the Parallelogram Rule for vector addition to calculate $(\mathbf{u}+\mathbf{v})_E$. Thus $\overrightarrow{AB}$, $\overrightarrow{DC}$, $\overrightarrow{EF}$ and $\overrightarrow{HG}$ all represent $\mathbf{u}$, while $\overrightarrow{AD}$ and $\overrightarrow{EH}$ both represent $\mathbf{v}$. Also $\overrightarrow{AC}$ represents $(\mathbf{u}+\mathbf{v})_A$ and $\overrightarrow{EG}$ represents $(\mathbf{u}+\mathbf{v})_E$. Finally, we define $\mathbf{w}$ to be the vector represented by $\overrightarrow{AE}$. All this information is in the diagram below.



Firstly, we examine the quadrilateral $AEHD$, and because two sides, $\overrightarrow{AD}$ and $\overrightarrow{EH}$ represent the same vector (namely $\mathbf{v}$), we conclude that the other two sides $\overrightarrow{AE}$ and $\overrightarrow{DH}$ represent the same vector, which is $\mathbf{w}$. We now turn our attention to the quadrilateral $DHGC$, and note that since $\overrightarrow{DC}$ and $\overrightarrow{HG}$ represent the same vector (namely $\mathbf{u}$), then so do $\overrightarrow{DH}$ and $\overrightarrow{CG}$, this common vector being $\mathbf{w}$. We have now shown that the sides $\overrightarrow{AE}$ and $\overrightarrow{CG}$ of the quadrilateral $AEGC$ represent $\mathbf{w}$, and so applying the Parallelgram Axiom for a third time, we see that $\overrightarrow{AC}$ and $\overrightarrow{EG}$ represent the same vector: that is, we have now shown that $(\mathbf{u}+\mathbf{v})_A = (\mathbf{u}+\mathbf{v})_E$. □

Theorem 1.2. For all vectors $\mathbf{u}$ and $\mathbf{v}$ we have $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$. This law is known as the *commutativity* of vector addition, and we say that vector addition is *commutative*. In fact, $(\mathbf{u}+\mathbf{v})_A = (\mathbf{v}+\mathbf{u})_A$ for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all points A.

Proof. The Parallelogram Rule for vector addition gives us the following parallelograms $ABCD$ and $ABED$ in which $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent $\mathbf{u}$; $\overrightarrow{AD}$ and $\overrightarrow{BE}$ represent $\mathbf{v}$; $\overrightarrow{AC}$ represents $(\mathbf{u}+\mathbf{v})_A$ and $\overrightarrow{AE}$ represents $(\mathbf{v}+\mathbf{u})_A$ (see the following diagram).

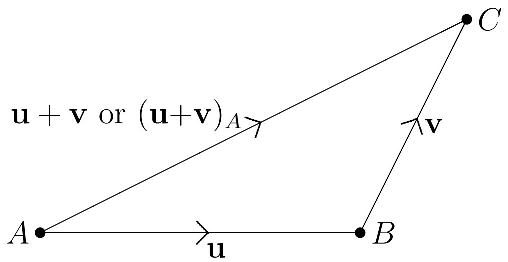


By the Parallelogram Axiom, $\overrightarrow{BC}$ represents $\mathbf{v}$, and by the uniqueness of a point X such that $\overrightarrow{BX}$ represents $\mathbf{v}$ we have that $C = E$. So now both $(\mathbf{u}+\mathbf{v})_A$ and $(\mathbf{v}+\mathbf{u})_A$ are represented by $\overrightarrow{AC}[=\overrightarrow{AE}]$ and therefore $(\mathbf{u}+\mathbf{v})_A = (\mathbf{v}+\mathbf{u})_A$. □

<u>*What is X?*</u> In maths, when one wants to refer to a quantity (so that we can describe some property satisfied by that quantity), we usually have its name, which is typically a letter of the alphabet, such as X or Y. It might be that there is no thing satisfying the properties required by X. For example, there is no real number X such that $X^2 + 1 = 0$. Or X need not be unique; for example there are precisely 3 real numbers X such that $X^3 + X^2 - 2X - 1 = 0$. [It does not matter what the actual values of X are, though in this case I can express them in other terms—one possible value of X is $2\cos\frac{2\pi}{7}$.]

The following gives an alternative way of defining vector addition, known as the *Triangle Rule.*

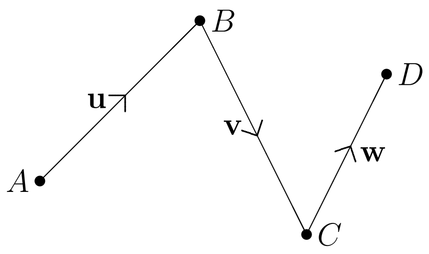
Theorem 1.3 (Triangle Rule)**.** Let A be a point, and let B and C be the unique points such that $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{BC}$ represents $\mathbf{v}$. Then $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$, or more accurately $(\mathbf{u}+\mathbf{v})_A$. The following diagram illustrates the Triangle Rule.



Proof. (For this proof the points A, B, C correspond to the points A, B, E (in that order) in the right-hand parallelogram of the picture in the proof of Theorem 1.2.) By the Parallelogram Rule we see that $\overrightarrow{AC}$ represents $(\mathbf{v}+\mathbf{u})_A$, and by the previous theorem, we have $(\mathbf{v}+\mathbf{u})_A = (\mathbf{u}+\mathbf{v})_A$, hence the result. □

Theorem 1.4. For all vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ we have $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$. This property is called *associativity.*

Proof. Pick a point A, and let B, C and D be the unique points such that $\overrightarrow{AB}$ represents $\mathbf{u}$, $\overrightarrow{BC}$ represents $\mathbf{v}$, and $\overrightarrow{CD}$ represents $\mathbf{w}$, see the diagram below.



By the Triangle Rule applied to triangle BCD, we find that $\overrightarrow{BD}$ represents $\mathbf{v}+\mathbf{w}$, and so by the Triangle Rule applied to triangle ABD we obtain that $\overrightarrow{AD}$ represents $\mathbf{u}+(\mathbf{v}+\mathbf{w})$. But the Triangle Rule applied to triangle ABC gives that $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$, and applying the Triangle Rule to triangle ACD shows that $\overrightarrow{AD}$ represents $(\mathbf{u}+\mathbf{v})+\mathbf{w}$. Since $\overrightarrow{AD}$ represents both $\mathbf{u}+(\mathbf{v}+\mathbf{w})$ and $(\mathbf{u}+\mathbf{v})+\mathbf{w}$, we conclude that they are equal. □

Theorem 1.5. For all vectors $\mathbf{u}$ we have $\mathbf{u}+\mathbf{0}=\mathbf{u}$. Thus $\mathbf{u}+\mathbf{0}=\mathbf{0}+\mathbf{u}=\mathbf{u}$. This asserts that $\mathbf{0}$ is an *identity* for vector addition.

Proof. Exercise (on exercise sheet). □

1.9 Vector subtraction

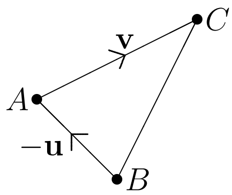
Definition. For vectors $\mathbf{u}$ and $\mathbf{v}$, we define $\mathbf{u}-\mathbf{v}$ by $\mathbf{u}-\mathbf{v}:=\mathbf{u}+(-\mathbf{v})$.

Theorem 1.6. For all vectors $\mathbf{u}$ we have $\mathbf{u}-\mathbf{u}=\mathbf{0}$. In other words, for each vector $\mathbf{u}$ we have $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$, and thus $\mathbf{u}+(-\mathbf{u})=(-\mathbf{u})+\mathbf{u}=\mathbf{0}$ by Theorem 1.2. This property means that $-\mathbf{u}$ is an *additive inverse* of $\mathbf{u}$.

Proof. Let $\overrightarrow{AB}$ represent $\mathbf{u}$. Then $\overrightarrow{BA}$ represents $-\mathbf{u}$, and thus, by the Triangle Rule, $\overrightarrow{AA}$ represents $\mathbf{u}+(-\mathbf{u})$. But $\overrightarrow{AA}$ (also) represents $\mathbf{0}$, and so $\mathbf{u}-\mathbf{u}=\mathbf{u}+(-\mathbf{u})=\mathbf{0}$, as required. □

Theorem 1.7. Suppose $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{AC}$ represents $\mathbf{v}$. Then $\overrightarrow{BC}$ represents $\mathbf{v}-\mathbf{u}$.

Proof. A diagram for this is as follows.



We have that $\overrightarrow{BA}$ represents $-\mathbf{u}$, and so, by the Triangle Rule, $\overrightarrow{BC}$ represents $(-\mathbf{u})+\mathbf{v}$. But $(-\mathbf{u})+\mathbf{v}=\mathbf{v}+(-\mathbf{u})$ by Theorem 1.2 (commutativity of vector addition), and $\mathbf{v}+(-\mathbf{u})=\mathbf{v}-\mathbf{u}$ by definition. □

1.10 Scalar multiplication

We now define how to multiply a real number α (a *scalar*) by a vector $\mathbf{v}$ to obtain another vector $\alpha\mathbf{v}$. We first specify that $\alpha\mathbf{v}$ has length

$$|\alpha||\mathbf{v}|,$$

where $|\alpha|$ is the absolute value of α and $|\mathbf{v}|$ is the length of $\mathbf{v}$. Thus if $\alpha = 0$ or $\mathbf{v} = \mathbf{0}$ then $|\alpha\mathbf{v}| = |\alpha||\mathbf{v}| = 0$, and so $\alpha\mathbf{v} = \mathbf{0}$. Otherwise (when $\alpha \neq 0$ and $\mathbf{v} \neq \mathbf{0}$), we have that $\alpha\mathbf{v}$ is nonzero, and we must specify the direction of $\alpha\mathbf{v}$. When $\alpha > 0$, we specify that $\alpha\mathbf{v}$ has the same direction as $\mathbf{v}$, and when $\alpha < 0$, we specify that $\alpha\mathbf{v}$ has the same direction as $-\mathbf{v}$ (and hence the opposite direction to $\mathbf{v}$).

Note. I have noticed that some students are writing $\mathbf{v}\alpha$ instead of $\alpha\mathbf{v}$. It is ugly, and I have not defined $\mathbf{v}\alpha$, so please do not use it. I may want to use that notation ($\mathbf{v}\alpha$) for something completely different.

From this definition of scalar multiplication, we observe the following elementary properties.

1. $0\mathbf{v} = \mathbf{0}$ for all vectors $\mathbf{v}$,
2. $\alpha\mathbf{0} = \mathbf{0}$ for all scalars α,
3. $1\mathbf{v} = \mathbf{v}$ for all vectors $\mathbf{v}$,
4. $(-1)\mathbf{v} = -\mathbf{v}$ for all vectors $\mathbf{v}$.

Further properties of scalar multiplication are given below as theorems. These are all harder to prove.

Theorem 1.8. For all vectors $\mathbf{v}$ and all scalars α, β, we have $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$.

Proof. We have $|\alpha(\beta\mathbf{v})| = |\alpha||\beta\mathbf{v}| = |\alpha|(|\beta||\mathbf{v}|) = (|\alpha||\beta|)|\mathbf{v}| = (|\alpha\beta|)|\mathbf{v}| = |(\alpha\beta)\mathbf{v}|$, and so $\alpha(\beta\mathbf{v})$ and $(\alpha\beta)\mathbf{v}$ have the same length.

If $\alpha = 0$, $\beta = 0$ or $\mathbf{v} = \mathbf{0}$ then $\alpha(\beta\mathbf{v}) = \mathbf{0} = (\alpha\beta)\mathbf{v}$ (easy exercise), so we now suppose that $\alpha \neq 0$, $\beta \neq 0$ and $\mathbf{v} \neq \mathbf{0}$, and show that $\alpha(\beta\mathbf{v})$ and $(\alpha\beta)\mathbf{v}$ have the same direction. The rest of the proof breaks into four cases, depending on the signs of α and β.

If $\alpha > 0$, $\beta > 0$ then $\beta\mathbf{v}$ and $(\alpha\beta)\mathbf{v}$ both have the same direction as $\mathbf{v}$, and $\alpha(\beta\mathbf{v})$ has the same direction as $\beta\mathbf{v}$, hence as $\mathbf{v}$. So $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$ in this case. If $\alpha < 0$, $\beta < 0$ then $\alpha\beta > 0$ and both $\alpha(\beta\mathbf{v})$ and $(\alpha\beta)\mathbf{v}$ have the same direction as $\mathbf{v}$, though multiplying by either one of α or β reverses direction.

If $\alpha < 0$, $\beta > 0$ or $\alpha > 0$, $\beta < 0$ then $\alpha(\beta\mathbf{v})$ and $(\alpha\beta)\mathbf{v}$ both have the same direction as $-\mathbf{v}$ (can you see why?). [This is an example of a hidden exercise in the text, and you should still try to do it, even though it will not appear on any exercise sheet.]

So we have now shown that $\alpha(\beta\mathbf{v})$ and $(\alpha\beta)\mathbf{v}$ have the same direction in all cases, completing the proof. □

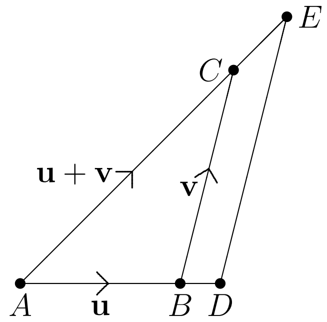
Theorem 1.9. For all vectors $\mathbf{v}$ and all scalars α, β, we have $(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v}+\beta\mathbf{v}$, where the correct bracketing on the right-hand side is $(\alpha\mathbf{v})+(\beta\mathbf{v})$. This is an example of a *distributive law.*

Proof. Exercise (not on the sheets). When evaluating $\alpha\mathbf{v}+\beta\mathbf{v}$ using the Parallelogram Rule, you may assume that the four (not necessarily distinct) vertices of the parallelogram all lie on a common (straight) line. $\square$

Theorem 1.10. For all vectors $\mathbf{u}$, $\mathbf{v}$ and all scalars α, we have $\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u}+\alpha\mathbf{v}$, where the correct bracketing on the right-hand side is $(\alpha\mathbf{u})+(\alpha\mathbf{v})$. This is also a distributive law.

Proof. This will be at best a sketch of a proof. It will really be an argument convincing you that the result is true using notions from Euclidean geometry. It is better to consider this 'theorem' as an axiom. Our demonstration below will only cover the case $\alpha > 0$, and the diagram is drawn with $\alpha > 1$.

Let $\overrightarrow{AB}$ represent $\mathbf{u}$ and $\overrightarrow{BC}$ represent $\mathbf{v}$, so that $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$ by the Triangle Rule. Extending the lines AB and AC as necessary, we let D be the point on the line AB such that $\overrightarrow{AD}$ represents $\alpha\mathbf{u}$, and let E be the point on the line AC such that $\overrightarrow{AE}$ represents $\alpha(\mathbf{u}+\mathbf{v})$.



(case $\alpha > 1$)

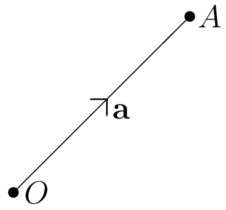
Then $|\overrightarrow{AD}| = \alpha|\overrightarrow{AB}|$ and $|\overrightarrow{AE}| = \alpha|\overrightarrow{AC}|$ and so triangles ABC and ADE are similar (that is one is a scaling of the other; here this scaling fixes the point A). Therefore $|\overrightarrow{DE}| = \alpha|\overrightarrow{BC}|$ and $\overrightarrow{BC}$ and $\overrightarrow{DE}$ have the same direction, and so $\overrightarrow{DE}$ represents $\alpha\mathbf{v}$. Now we use the Triangle Rule with triangle ADE to conclude that $\alpha\mathbf{u}+\alpha\mathbf{v} = \alpha(\mathbf{u}+\mathbf{v})$.

(Note that the notion of similarity in general allows translations, rotations and reflexions as well as scaling.) $\square$

Aside. In 2011, one student observed that Theorem 1.4 (among others) looked more like an axiom than a theorem. It is quite possible to choose an alternative axiomatisation of geometry in which several of the theorems here (including Theorem 1.4) are axioms. It is also very likely that the Parallelogram Axiom would no longer be an axiom in such an alternative system, but a theorem requiring proof.

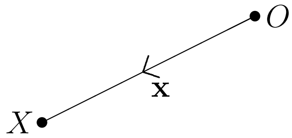
1.11 Position vectors

In order to talk about position vectors, we need to assume that we have fixed a point O as an *origin* in 3-space. Then if A is any point, the *position vector* of A is defined to be the free vector represented by the bound vector $\overrightarrow{OA}$.



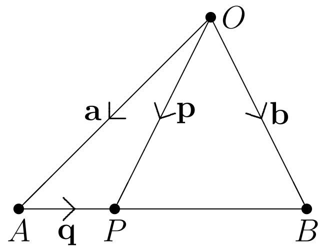
a is the position vector of the point A

Note that each vector $\mathbf{x}$ is the position vector of exactly one point X in 3-space. This point X has as distance and direction from the origin the length and direction of the vector $\mathbf{x}$.



Theorem 1.11. Let A and B be points with position vectors $\mathbf{a}$ and $\mathbf{b}$. Let P be a point on the line segment AB such that $|\overrightarrow{AP}| = \lambda|\overrightarrow{AB}|$. Then P has position vector $\mathbf{p} = (1-\lambda)\mathbf{a} + \lambda\mathbf{b}$.

Proof. A diagram for this situation is as follows.

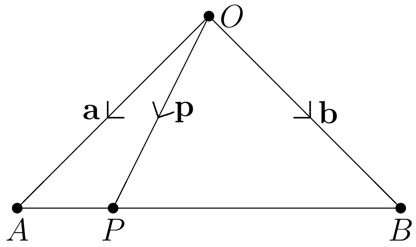


Let $\mathbf{c}$ be the vector represented by $\overrightarrow{AB}$, and let $\mathbf{q}$ be the vector represented by $\overrightarrow{AP}$. Then $\mathbf{q} = \lambda\mathbf{c}$. Now $\mathbf{a} + \mathbf{c} = \mathbf{b}$ (by the Triangle Rule), and adding $-\mathbf{a}$ to both sides we obtain $\mathbf{c} = \mathbf{b} - \mathbf{a}$. Therefore, using the Triangle Rule, we get $\mathbf{p} = \mathbf{a} + \mathbf{q} = \mathbf{a} + \lambda\mathbf{c}$, and using the various rules for vector addition and scalar multiplication, we get:

$$\begin{aligned}\mathbf{p} = \mathbf{a} + \lambda\mathbf{c} &= \mathbf{a} + \lambda(\mathbf{b} - \mathbf{a}) \\ &= \mathbf{a} + \lambda(\mathbf{b} + (-\mathbf{a})) \\ &= \mathbf{a} + \lambda\mathbf{b} + \lambda(-\mathbf{a}) \\ &= \mathbf{a} + \lambda\mathbf{b} + \lambda((-1)\mathbf{a}) \\ &= 1\mathbf{a} + \lambda\mathbf{b} + (-\lambda)\mathbf{a} \\ &= (1-\lambda)\mathbf{a} + \lambda\mathbf{b},\end{aligned}$$

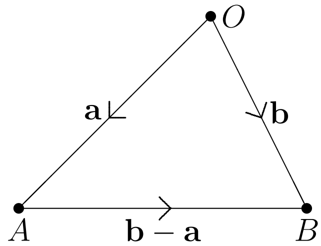
as required. □

Example. Suppose P is one quarter of the way from A along the line segment AB. Then $\mathbf{p} = (1 - \frac{1}{4})\mathbf{a} + \frac{1}{4}\mathbf{b} = \frac{3}{4}\mathbf{a} + \frac{1}{4}\mathbf{b}$ (see picture overleaf).



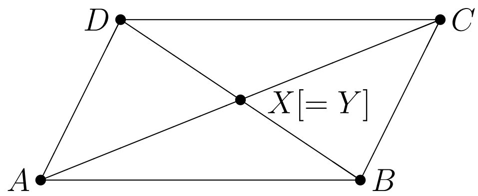
$$|\overrightarrow{AP}| = \frac{1}{4}|\overrightarrow{AB}|$$

Note. If A and B are any points, with position vectors $\mathbf{a}$ and $\mathbf{b}$ respectively, then the vector represented by $\overrightarrow{AB}$ is $\mathbf{b} - \mathbf{a}$ (see picture below).



Theorem 1.12 (Application of Theorem 1.11)**.** The diagonals of a parallelogram $ABCD$ meet each other in their midpoints.

Proof. This proof proceeds by determining the midpoints of the two diagonals and showing they are the same. The diagram of what we wish to prove is given below.



Let X be the midpoint of the diagonal AC, and let Y be the midpoint of the diagonal BD. Let A, B, C, D, X, Y have position vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{x}$, $\mathbf{y}$ respectively. Then, by Theorem 1.11, we have

$$\mathbf{x} = (1 - \tfrac{1}{2})\mathbf{a} + \tfrac{1}{2}\mathbf{c} = \tfrac{1}{2}\mathbf{a} + \tfrac{1}{2}\mathbf{c} = \tfrac{1}{2}(\mathbf{a} + \mathbf{c}) \quad \text{and} \quad \mathbf{y} = (1 - \tfrac{1}{2})\mathbf{b} + \tfrac{1}{2}\mathbf{d} = \tfrac{1}{2}(\mathbf{b} + \mathbf{d}).$$

Since $ABCD$ is a parallelogram, $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vector. Since $\overrightarrow{AB}$ represents $\mathbf{b} - \mathbf{a}$ and $\overrightarrow{DC}$ represents $\mathbf{c} - \mathbf{d}$, we have $\mathbf{b} - \mathbf{a} = \mathbf{c} - \mathbf{d}$. Adding $\mathbf{a} + \mathbf{d}$ to both sides (and using the rules of vector addition and subtraction) gives $\mathbf{b} + \mathbf{d} = \mathbf{c} + \mathbf{a} [= \mathbf{a} + \mathbf{c}]$. So now $\frac{1}{2}(\mathbf{a} + \mathbf{c}) = \frac{1}{2}(\mathbf{b} + \mathbf{d})$, which implies that $\mathbf{x} = \mathbf{y}$, whence $X = Y$. □

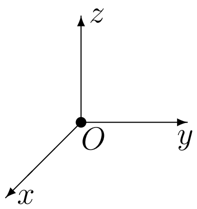
In lectures, the symbol $\Rightarrow$ often crops up, especially in proofs. It means 'implies' or 'implies that.' So $A \Rightarrow B$ means 'A implies B' or 'if A then B' or 'A only if B.' The symbol $\Leftarrow$ means 'is implied by,' so that $A \Leftarrow B$ means 'A is implied by B' or 'A if B.' The symbol $\Leftrightarrow$ means 'if and only if' so that $A \Leftrightarrow B$ means 'A if and only if B' or 'A is equivalent to B.' Throughout the above, A and B are statements.

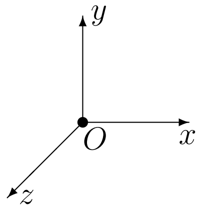
Chapter 2

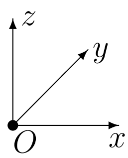
Cartesian Coördinates

The adjective *Cartesian* above refers to René Descartes (1596–1650), who was the first to coördinatise the plane as ordered pairs of real numbers, which provided the first systematic link between Euclidean geometry and algebra.

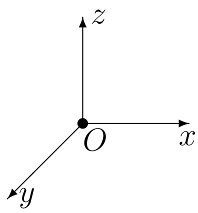
Choose an origin O in 3-space, and choose (three) mutually perpendicular axes through O, which we shall label as the x-, y- and z-axes. The x-axis, y-axis and z-axis form a right-handed system if they can be rotated to look like one of the following (which can all to rotated to look like the others).







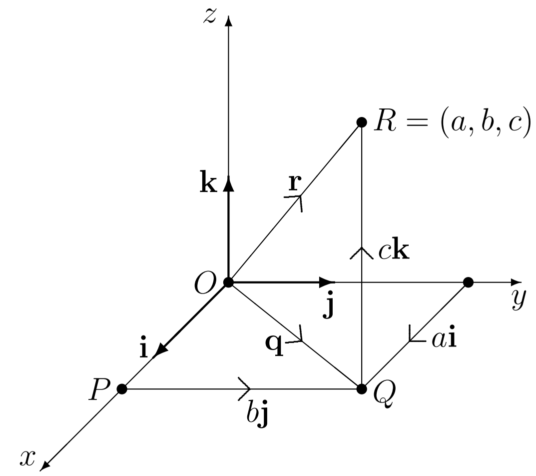
A left-handed system can be rotated to look like the following.



Swapping two axes or reversing the direction of one (or three) of the axes changes the handedness of the system. It is possible to make the shape of a right-handed system using our right-hand, with the thumb (pointing) along the x-axis, first [index] finger along the y-axis, and second [middle] finger along the z-axis. You should curl the other two fingers of your hand into you palm when you do this. (Unfortunately, it is possible, though much harder, to make the shape of a left-handed system using your right hand, but if you can make such a configuration it should be much more uncomfortable!) If you use your left hand, you should end up with a left-handed system.

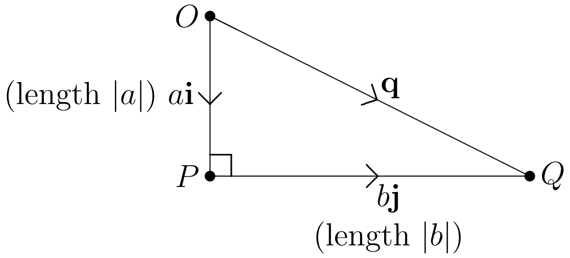
We let **i**, **j** and **k** denote vectors of length 1 in the directions of the x-, y- and z-axes respectively.

Let R be the point whose coördinates are (a, b, c), and let $\mathbf{r}$ be the position vector of R. Then $\mathbf{r} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$. See the diagram below.

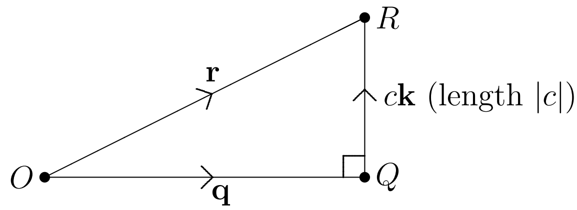


Let $\mathbf{q} = a\mathbf{i} + b\mathbf{j}$ be the position vector of the point Q. Then applying Pythagoras's Theorem to the right-angled triangle OPQ we get:

$$|\mathbf{q}| = |\overrightarrow{OQ}| = \sqrt{|\overrightarrow{OP}|^2 + |\overrightarrow{PQ}|^2} = \sqrt{|a|^2 + |b|^2} = \sqrt{a^2 + b^2}.$$



We also have the right-angled triangle OQR.



Applying Pythagoras's Theorem to triangle OQR gives:

$$|\mathbf{r}| = |\overrightarrow{OR}| = \sqrt{|\overrightarrow{OQ}|^2 + |\overrightarrow{QR}|^2} = \sqrt{|\mathbf{q}|^2 + |c|^2} = \sqrt{a^2 + b^2 + c^2}.$$

To summarise: If R is a point having coördinates (a, b, c), then the position vector of R is $\mathbf{r} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, which has length $\sqrt{a^2 + b^2 + c^2}$.

Notation. We write $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ for the vector $a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

2.1 Sums and scalar multiples using coördinates

Now let $\mathbf{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} d \\ e \\ f \end{pmatrix}$ and let α be a scalar. Then, using the rules for vector addition and scalar multiplication (sometimes multiple times per line) we get:

$$\begin{aligned}
\mathbf{u} + \mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} d \\ e \\ f \end{pmatrix} &= (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) + (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\
&= (a\mathbf{i} + d\mathbf{i}) + (b\mathbf{j} + e\mathbf{j}) + (c\mathbf{k} + f\mathbf{k}) \\
&= \quad (a+d)\mathbf{i} + (b+e)\mathbf{j} + (c+f)\mathbf{k} \quad = \begin{pmatrix} a+d \\ b+e \\ c+f \end{pmatrix},
\end{aligned}$$

along with

$$\begin{aligned}
\alpha\mathbf{u} = \alpha \begin{pmatrix} a \\ b \\ c \end{pmatrix} &= \alpha(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \\
&= \alpha(a\mathbf{i}) + \alpha(b\mathbf{j}) + \alpha(c\mathbf{k}) \\
&= (\alpha a)\mathbf{i} + (\alpha b)\mathbf{j} + (\alpha c)\mathbf{k} = \begin{pmatrix} \alpha a \\ \alpha b \\ \alpha c \end{pmatrix},
\end{aligned}$$

and

$$-\mathbf{u} = (-1)\mathbf{u} = (-1) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -a \\ -b \\ -c \end{pmatrix}.$$

Example. Let $\mathbf{u} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 3 \\ 5 \\ -1 \end{pmatrix}$. Then

$$3\mathbf{u} - 4\mathbf{v} = 3\mathbf{u} + (-4)\mathbf{v} = 3 \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + (-4) \begin{pmatrix} 3 \\ 5 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ -3 \\ 0 \end{pmatrix} + \begin{pmatrix} -12 \\ -20 \\ 4 \end{pmatrix} = \begin{pmatrix} -6 \\ -23 \\ 4 \end{pmatrix}.$$

2.2 Unit vectors

Definition 2.1. A *unit vector* is a vector of length 1.

For example, $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors. Let $\mathbf{r}$ be any nonzero vector, and define:

$$\hat{\mathbf{r}} := \left(\frac{1}{|\mathbf{r}|}\right) \mathbf{r}.$$

(Note that $|\mathbf{r}| > 0$, so that $\frac{1}{|\mathbf{r}|}$ and hence $\hat{\mathbf{r}}$ exist.) But we also have $\frac{1}{|\mathbf{r}|} > 0$, and thus $|\hat{\mathbf{r}}| = |\frac{1}{|\mathbf{r}|}||\mathbf{r}| = \frac{1}{|\mathbf{r}|}|\mathbf{r}| = 1$, so that $\hat{\mathbf{r}}$ is the unit vector in the same direction as $\mathbf{r}$. (There is only one unit vector in the same direction as $\mathbf{r}$.)

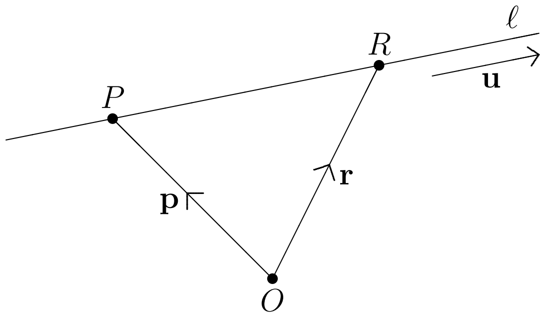
Example. Let $\mathbf{r} = \begin{pmatrix} -1 \\ 5 \\ -4 \end{pmatrix}$. Then $|\mathbf{r}| = \sqrt{(-1)^2 + 5^2 + (-4)^2} = \sqrt{42}$. Therefore

$$\hat{\mathbf{r}} = \left(\frac{1}{|\mathbf{r}|}\right)\mathbf{r} = \frac{1}{\sqrt{42}}\begin{pmatrix} -1 \\ 5 \\ -4 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{42} \\ 5/\sqrt{42} \\ -4/\sqrt{42} \end{pmatrix}.$$

If we want the unit vector in the *opposite* direction to $\mathbf{r}$, this is simply $-\hat{\mathbf{r}}$, and if we want the vector of length 7 in the opposite direction to $\mathbf{r}$, this is $-7\hat{\mathbf{r}}$.

2.3 Equations of lines

Let ℓ be the line through the point P in the direction of the *nonzero* vector $\mathbf{u}$.



Now a point R is on the line ℓ if and only if $\overrightarrow{PR}$ represents a scalar multiple of $\mathbf{u}$. Let $\mathbf{p}$ and $\mathbf{r}$ be the position vectors of P and R respectively. Then $\overrightarrow{PR}$ represents $\mathbf{r} - \mathbf{p}$, and so R is on ℓ if and only if $\mathbf{r} - \mathbf{p} = \lambda\mathbf{u}$ for some real number λ; equivalently $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$ for some real number λ. We thus get the *vector equation* of the line ℓ:

$$\mathbf{r} = \mathbf{p} + \lambda\mathbf{u} \qquad (\lambda \in \mathbb{R}),$$

where $\mathbf{p}$ and $\mathbf{u}$ are constants and $\mathbf{r}$ is a variable (depending on λ) which denotes the position of a general point on ℓ.

Moving to coördinates, we let $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ and $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$. Then:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{r} = \mathbf{p} + \lambda\mathbf{u} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \lambda\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} p_1 + \lambda u_1 \\ p_2 + \lambda u_2 \\ p_3 + \lambda u_3 \end{pmatrix},$$

from which we get the *parametric equations* of the line ℓ, namely:

$$\left.\begin{array}{l} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{array}\right\}.$$

Assuming that $u_1, u_2, u_3 \neq 0$, we may eliminate λ to get:

$$\frac{x - p_1}{u_1} = \frac{y - p_2}{u_2} = \frac{z - p_3}{u_3},$$

which are the *Cartesian equations* of the line ℓ. (Each of these fractions is equal to λ.) The following should tell you how to get the Cartesian equations of ℓ when one or two of the u_i are zero (they cannot all be zero). If $u_1 = 0$ and $u_2, u_3 \neq 0$ then the Cartesian equations are

$$x = p_1,\ \frac{y - p_2}{u_2} = \frac{z - p_3}{u_3},$$

and if $u_1 = u_2 = 0$, $u_3 \neq 0$, the Cartesian equations are $x = p_1$, $y = p_2$ (with no mention of z anywhere).

Example. As an example, we determine the vector, parametric and Cartesian equations of the line ℓ through the point $(3, -1, 2)$ in the direction of the vector $\begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$.

The vector equation is:

$$\mathbf{r} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + \lambda \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}.$$

The parametric equations are:

$$\left.\begin{array}{rl} x = & 3 - 2\lambda \\ y = & -1 + \lambda \\ z = & 2 + 4\lambda \end{array}\right\}.$$

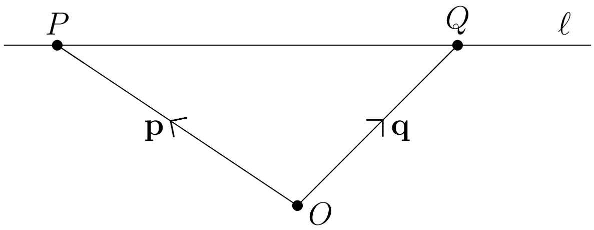
And the Cartesian equations are:

$$\frac{x - 3}{-2} = \frac{y + 1}{1} = \frac{z - 2}{4}.$$

Is the point $(7, -3, -6)$ on ℓ? Yes, since the vector equation is satisfied with $\lambda = -2$ (this value of λ can be determined from the parametric equations). What about the point $(1, 1, 1)$? No, because the Cartesian equations are not satisfied: $1 = \frac{1-3}{-2} \neq \frac{1+1}{1} = 2$. Alternatively, we can look at the parametric equations: the first would give $\lambda = 1$, and the second would give $\lambda = 2$, an inconsistency.

2.3.1 The line determined by two distinct points

Suppose we are given two points P and Q on a line ℓ, with $P \neq Q$, and we want to determine (say) a vector equation for ℓ. Suppose P has position vector $\mathbf{p}$ and Q has position vector $\mathbf{q}$.



Then ℓ is a line through P in the direction of $\overrightarrow{PQ}$, and thus in the direction of $\mathbf{q} - \mathbf{p}$ (the vector that $\overrightarrow{PQ}$ represents). Therefore, a vector equation for ℓ is:

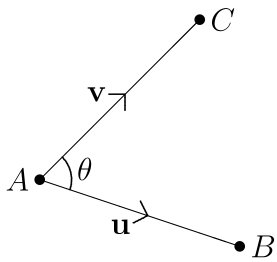
$$\mathbf{r} = \mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}).$$

Parametric and Cartesian equations for ℓ can be derived from this vector equation in the usual way.

Chapter 3

The Scalar Product

The scalar product is a way of multiplying two vectors to produce a scalar (real number). Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors represented by $\overrightarrow{AB}$ and $\overrightarrow{AC}$.



We define the angle between $\mathbf{u}$ and $\mathbf{v}$ to be the angle θ (in radians) between $\overrightarrow{AB}$ and $\overrightarrow{AC}$, with $0 \leqslant \theta \leqslant \pi$. A handy chart for converting between degrees and radians is given below.

radians	0	$\frac{\pi}{180}$	$\frac{\pi}{12}$	$\frac{\pi}{10}$	$\frac{\pi}{6}$	$\frac{\pi}{5}$	$\frac{\pi}{4}$	1	$\frac{\pi}{3}$	$\frac{2\pi}{5}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	π	2π
degrees	0	1	15	18	30	36	45	$\frac{180}{\pi} \approx 57.3$	60	72	90	120	135	180	360

Definition 3.1. The *scalar product* (or *dot product*) of $\mathbf{u}$ and $\mathbf{v}$ is denoted $\mathbf{u}\cdot\mathbf{v}$, and is defined to be 0 if either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$. If both $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$, we define $\mathbf{u}\cdot\mathbf{v}$ by

$$\mathbf{u}\cdot\mathbf{v} := |\mathbf{u}||\mathbf{v}| \cos\theta,$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. (Note that I have had to specify what θ is in the definition itself; you **must** do the same.) We say that $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\mathbf{u}\cdot\mathbf{v} = 0$.

Note that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or the angle between $\mathbf{u}$ and $\mathbf{v}$ is $\frac{\pi}{2}$. (This includes the case $\mathbf{u} = \mathbf{v} = \mathbf{0}$.)

Note. Despite the notation concealing this fact somewhat, the scalar product is a ***function***. Its codomain (and range) is $\mathbb{R}$, and its domain is the set of ordered pairs of (free) vectors. As usual, we must make sure that the function is defined (in a unique manner) for ***all*** elements of the domain, and this includes those pairs having the zero vector in one or both positions.

3.1 The scalar product using coördinates

Theorem 3.2. Let $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$. Then $\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$.

Proof. If $\mathbf{u} = \mathbf{0}$ (in which case $u_1 = u_2 = u_3 = 0$) or $\mathbf{v} = \mathbf{0}$ (in which case $v_1 = v_2 = v_3 = 0$) we have $\mathbf{u}\cdot\mathbf{v} = 0 = u_1v_1 + u_2v_2 + u_3v_3$, as required.

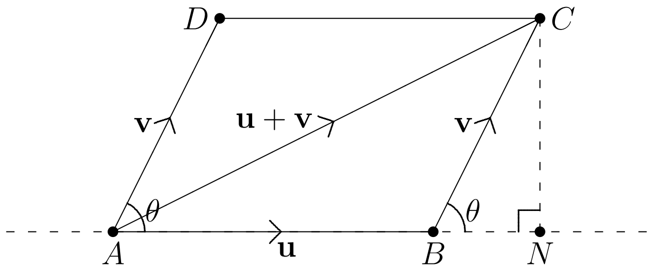
Now suppose that $\mathbf{u}, \mathbf{v} \neq 0$, and let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. We calculate $|\mathbf{u} + \mathbf{v}|^2$ in two different ways. Firstly we use coördinates.

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 = \left| \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{pmatrix} \right|^2 &= (u_1 + v_1)^2 + (u_2 + v_2)^2 + (u_3 + v_3)^2 \\ &= u_1^2 + 2u_1v_1 + v_1^2 + u_2^2 + 2u_2v_2 + v_2^2 + u_3^2 + 2u_3v_3 + v_3^2 \\ &= |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2(u_1v_1 + u_2v_2 + u_3v_3), \end{aligned}$$

that is:

$$|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2(u_1v_1 + u_2v_2 + u_3v_3). \tag{3.1}$$

Our second way to do this is geometrical. Pick a point A, and consider the parallelogram $ABCD$, where $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{AD}$ represents $\mathbf{v}$. Thus $\overrightarrow{BC}$ represents $\mathbf{v}$, and so $\overrightarrow{AC}$ represents $\mathbf{u} + \mathbf{v}$ by the Triangle Rule. Drop a perpendicular from C to the line through A and B, meeting the said line at N, and let M be an arbitrary point on the line through A and B strictly to 'right' of B (i.e. when traversing the line through A and B in a certain direction we encounter the points in the order A, B, M). Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$ (i.e. θ is the size of angle BAD). A result from Euclidean geometry states that the angle MBC also has size θ. The following diagram has all this information.



(Note that this diagram is drawn with $0 < \theta < \frac{\pi}{2}$. If $\theta = \frac{\pi}{2}$ then $N = B$, and if $\theta > \frac{\pi}{2}$ then N lies to the 'left' of B, probably between A and B, but possibly even to the 'left' of A.) We have $|\overrightarrow{AN}| = |(|\mathbf{u}| + |\mathbf{v}|\cos\theta)|$, even when $\theta \geqslant \frac{\pi}{2}$, and even when N is to the 'left' of A. We also have that $|\overrightarrow{CN}| = |\mathbf{v}||\sin\theta|$. Applying Pythagoras (which is fine here even when $\theta \geqslant \frac{\pi}{2}$), and using the fact that $|a|^2 = a^2$ whenever $a \in \mathbb{R}$, we obtain:

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 = |\overrightarrow{AC}|^2 = |\overrightarrow{AN}|^2 + |\overrightarrow{CN}|^2 &= (|\mathbf{u}| + |\mathbf{v}|\cos\theta)^2 + (|\mathbf{v}|\sin\theta)^2 \\ &= |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}|\cos\theta + |\mathbf{v}|^2(\cos\theta)^2 + |\mathbf{v}|^2(\sin\theta)^2 \\ &= |\mathbf{u}|^2 + |\mathbf{v}|^2(\cos^2\theta + \sin^2\theta) + 2\mathbf{u}\cdot\mathbf{v}. \end{aligned}$$

Here $\cos^2\theta$ means $(\cos\theta)^2$ and $\sin^2\theta$ means $(\sin\theta)^2$. Using the standard identity that $\cos^2\theta + \sin^2\theta = 1$ for all θ, we obtain:

$$|\mathbf{u}+\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2\mathbf{u}\cdot\mathbf{v}. \tag{3.2}$$

Comparing Equations 3.1 and 3.2 gives us the result. □

Note that if $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ then $\mathbf{u}\cdot\mathbf{u} = u_1^2 + u_2^2 + u_3^2 = |\mathbf{u}|^2$ (even when $\mathbf{u} = \mathbf{0}$).

Example. We determine $\cos\theta$, where θ is the angle between $\mathbf{u} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$. We have $|\mathbf{u}| = \sqrt{2^2 + (-1)^2 + 1^2} = \sqrt{6}$ and $|\mathbf{v}| = \sqrt{1^2 + 2^2 + (-3)^2} = \sqrt{14}$, along with:

$$\mathbf{u}\cdot\mathbf{v} = 2 \times 1 + (-1) \times 2 + 1 \times (-3) = 2 - 2 - 3 = -3.$$

The formula $\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}||\mathbf{v}|\cos\theta$ gives $-3 = \sqrt{6}\sqrt{14}\cos\theta = 2\sqrt{21}\cos\theta$, the last step being since $\sqrt{6}\sqrt{14} = \sqrt{2}\sqrt{3}\sqrt{2}\sqrt{7} = 2\sqrt{21}$. Thus we get:

$$\cos\theta = \frac{-3}{2\sqrt{21}} = -\frac{1}{2}\sqrt{\frac{3}{7}}.$$

(The last equality was obtained by cancelling a factor of $\sqrt{3}$ from the numerator and denominator. There is no need to do this if it does not make the fraction 'neater', and here I do not think it does.)

Note. The following is an example of totally unacceptable working when calculating a dot product.

$$\begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \times 3 \\ (-1) \times (-2) \\ (-2) \times 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ -2 \end{pmatrix} = 3 + 2 + (-2) = 3.$$

This is because the first and third so-called equalities are nothing of the sort. The first is trying to equate a scalar (LHS) with a vector (RHS), while the third tries to equate a vector with a scalar. The above has ***TWO*** errors, and we shall simply mark such stuff as being wrong.

3.2 Properties of the scalar product

Let $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ and let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. From the definition $\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}||\mathbf{v}|\cos\theta$ of the scalar product we observe that:

- if $0 \leqslant \theta < \frac{\pi}{2}$ then $\mathbf{u}\cdot\mathbf{v} > 0$;
- if $\theta = \frac{\pi}{2}$ then $\mathbf{u}\cdot\mathbf{v} = 0$; and
- if $\frac{\pi}{2} < \theta \leqslant \pi$ then $\mathbf{u}\cdot\mathbf{v} < 0$.

Moreover,

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|}.$$

Now let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be any vectors. Then:

1. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$;
2. $\mathbf{u}\cdot(\mathbf{v} + \mathbf{w}) = (\mathbf{u}\cdot\mathbf{v}) + (\mathbf{u}\cdot\mathbf{w})$;
3. $(\mathbf{u} + \mathbf{v})\cdot\mathbf{w} = (\mathbf{u}\cdot\mathbf{w}) + (\mathbf{v}\cdot\mathbf{w})$;
4. $\mathbf{u}\cdot(\alpha\mathbf{v}) = \alpha(\mathbf{u}\cdot\mathbf{v}) = (\alpha\mathbf{u})\cdot\mathbf{v}$ for all scalars α;
5. $\mathbf{u}\cdot(-\mathbf{v}) = (-\mathbf{u})\cdot\mathbf{v} = -(\mathbf{u}\cdot\mathbf{v})$; and
6. $(-\mathbf{u})\cdot(-\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}$.

There is however no (non-vacuous) associative law for the dot product. This is because neither of the quantities $(\mathbf{u}\cdot\mathbf{v})\cdot\mathbf{w}$ and $\mathbf{u}\cdot(\mathbf{v}\cdot\mathbf{w})$ is defined. (In both cases, we are trying to form the dot product of a vector and a scalar in some order, and in neither order does such a product exist.)

Each of the above equalities can be proved by using Theorem 3.2, which expresses the dot product in terms of coördinates. To prove (1) we observe that:

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3 = v_1u_1 + v_2u_2 + v_3u_3 = \mathbf{v}\cdot\mathbf{u}.$$

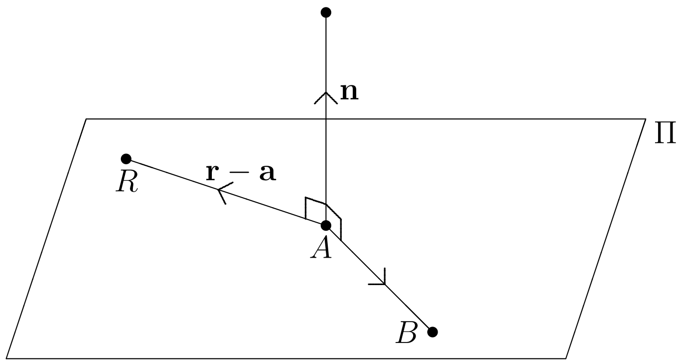
In order to prove the equality $\mathbf{u}\cdot(\alpha\mathbf{v}) = \alpha(\mathbf{u}\cdot\mathbf{v})$ of (4) we observe the following.

$$\begin{aligned}\mathbf{u}\cdot(\alpha\mathbf{v}) = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \cdot \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \alpha v_3 \end{pmatrix} &= u_1(\alpha v_1) + u_2(\alpha v_2) + u_3(\alpha v_3) \\ &= \alpha(u_1v_1) + \alpha(u_2v_2) + \alpha(u_3v_3) \\ &= \alpha(u_1v_1 + u_2v_2 + u_3v_3) = \alpha(\mathbf{u}\cdot\mathbf{v}).\end{aligned}$$

The proofs of the rest of these equalities are left as exercises.

3.3 Equation of a plane

Let $\mathbf{n}$ be a vector and Π be a plane. We say that $\mathbf{n}$ is *orthogonal* to Π (or Π is orthogonal to $\mathbf{n}$) if for all points A, B on Π, we have that $\mathbf{n}$ is orthogonal to the vector represented by $\overrightarrow{AB}$. We also say that $\mathbf{n}$ is a *normal* (or normal vector) to Π, hence the notation $\mathbf{n}$.



Suppose that $\mathbf{n} \neq \mathbf{0}$, A is a point, and we wish to determine an equation of the (unique) plane Π that is orthogonal to $\mathbf{n}$ and contains A. Now a point R, with position vector $\mathbf{r}$, is on Π exactly when $\overrightarrow{AR}$ represents a vector orthogonal to $\mathbf{n}$, that is when $(\mathbf{r}-\mathbf{a})\cdot\mathbf{n} = 0$, where $\mathbf{a}$ is the position vector of $\mathbf{a}$. Equivalently, we have $\mathbf{r}\cdot\mathbf{n} - \mathbf{a}\cdot\mathbf{n} = 0$, which gives:

$$\mathbf{r}\cdot\mathbf{n} = \mathbf{a}\cdot\mathbf{n},$$

a *vector equation* of the plane Π, where $\mathbf{r}$ is the position vector of an arbitary point on Π, $\mathbf{a}$ is the position vector of a fixed point on Π, and $\mathbf{n}$ is a nonzero vector orthogonal to Π. In coördinates, we let

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} \quad \text{and} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Then the point (x, y, z) is on Π exactly when:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \cdot \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$$

that is, when

$$n_1 x + n_2 y + n_3 z = d,$$

where $d = a_1 n_1 + a_2 n_2 + a_3 n_3$. This is a *Cartesian equation* of the plane Π.

Example. We find a Cartesian equation for the plane through $A = (2, -1, 3)$ and orthogonal to $\mathbf{n} = \begin{pmatrix} -2 \\ 3 \\ 5 \end{pmatrix}$. A vector equation is $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} -2 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} -2 \\ 3 \\ 5 \end{pmatrix}$, which gives rise to the Cartesian equation $-2x + 3y + 5z = 8$.

Example. The equation $2x-y+3z = 6$ specifies the plane Π orthogonal to $\mathbf{n} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$ and containing the point $(1, -1, 1)$. This is because we can write the equation as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix},$$

which has the form $\mathbf{r}\cdot\mathbf{n} = \mathbf{a}\cdot\mathbf{n}$ for suitable vectors $\mathbf{r}$ and $\mathbf{a}$. The point $(1, 2, 3)$ is not on Π since $2 \times 1 + (-1) \times 2 + 3 \times 3 = 2 - 2 + 9 = 9 \neq 6$. The point $(1, 2, 2)$ is on Π since $2 \times 1 + (-1) \times 2 + 3 \times 2 = 2 - 2 + 6 = 6$.

Note that the coördinates of $\mathbf{n}$ can always be taken to be the coefficients of x, y, z in the Cartesian equation. (It is valid to multiply such an $\mathbf{n}$ by any nonzero scalar, but must ensure we do the corresponding operations to the right-hand sides of any equations we use. Thus both $2x - y + 3x = 6$ and $-4x + 2y - 6x = -12$ are Cartesian equations of the plane Π in the second example above.) Finding a point on Π is harder. A sensible strategy is to set two of x, y, z to be zero (where the coefficient of the third is nonzero). Here setting $x = y = 0$ gives $3z = 6$, whence $z = 2$, so that $(0, 0, 2)$ is on Π. Setting $x = z = 0$ gives $y = -6$, so that $(0, -6, 0)$ is on Π, and setting $y = z = 0$ gives $x = 3$, so that $(3, 0, 0)$ is on Π. (This sensible strategy does not find the point $(1, -1, 1)$ that is on Π.)

In the case of the plane Π' with equation $x + y = 1$, setting $x = z = 0$ gives the point $(0, 1, 0)$ on Π', while setting $y = z = 0$ gives the point $(1, 0, 0)$ on Π'. But if we set $x = y = 0$, we end up with the equation $0 = 1$, which has no solutions for z, so we do not find a point here.

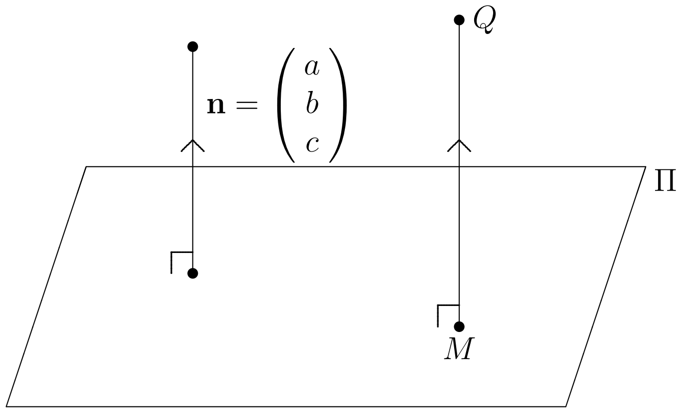
Note. [Not lectured.] Another form of a vector equation for a plane, corresponding to the vector equation for a line is as follows. Take any 3 points A, B, C on Π such that A, B, C are not on the same line. Let A, B, C have position vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ respectively. Then a vector equation for Π is:

$$\mathbf{r} = \mathbf{a} + \lambda(\mathbf{b} - \mathbf{a}) + \mu(\mathbf{c} - \mathbf{a}),$$

where λ and μ range (independently) over the whole of $\mathbb{R}$.

3.4 The distance from a point to a plane

Let Π be the plane having equation $ax + by + cz = d$, so that Π is orthogonal to $\mathbf{n} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \neq \mathbf{0}$. Let Q be a point, and let M be the point on Π closest to Q.



Let $\mathbf{m}$, $\mathbf{q}$ be the position vectors of M, Q. Then the vector $\mathbf{q}-\mathbf{m}$ represented by $\overrightarrow{MQ}$ is orthogonal to Π, and so (since we are in just 3 dimensions) $\mathbf{q}-\mathbf{m}$ is a scalar multiple of $\mathbf{n}$, that is $\mathbf{q}-\mathbf{m} = \lambda\mathbf{n}$ for some scalar $\lambda \in \mathbb{R}$. Therefore:

$$(\mathbf{q}-\mathbf{m})\cdot\mathbf{n} = (\lambda\mathbf{n})\cdot\mathbf{n} = \lambda(\mathbf{n}\cdot\mathbf{n}) = \lambda|\mathbf{n}|^2,$$

and so $\mathbf{q}\cdot\mathbf{n}-\mathbf{m}\cdot\mathbf{n} = \lambda|\mathbf{n}|^2$. But $\mathbf{m}$ is on Π, which has equation $\mathbf{r}\cdot\mathbf{n} = d$, and so $\mathbf{m}\cdot\mathbf{n} = d$. Therefore $\mathbf{q}\cdot\mathbf{n} - d = \lambda|\mathbf{n}|^2$, and thus:

$$\lambda = \frac{\mathbf{q}\cdot\mathbf{n} - d}{|\mathbf{n}|^2}$$

But the distance from Q to Π (which is the distance from Q to M, where M is the closest point on Π to Q) is in fact $|\overrightarrow{MQ}| = |\lambda\mathbf{n}| = |\lambda||\mathbf{n}|$, that is:

$$\left|\frac{\mathbf{q}\cdot\mathbf{n} - d}{|\mathbf{n}|^2}\right| |\mathbf{n}| = \frac{|\mathbf{q}\cdot\mathbf{n} - d|}{|\mathbf{n}|}.$$

If one looks at other sources one may see a superficially dissimilar formula for this distance. To obtain this, we let P be *any* point on Π, and let $\mathbf{p}$ be the position vector of $\mathbf{p}$, so that $\mathbf{p}\cdot\mathbf{n} = d$. Thus $\mathbf{q}\cdot\mathbf{n} - d = \mathbf{q}\cdot\mathbf{n} - \mathbf{p}\cdot\mathbf{n} = (\mathbf{q}-\mathbf{p})\cdot\mathbf{n}$. Therefore the distance can also be expressed as:

$$\frac{|(\mathbf{q}-\mathbf{p})\cdot\mathbf{n}|}{|\mathbf{n}|} = \left|(\mathbf{q}-\mathbf{p})\cdot\frac{\mathbf{n}}{|\mathbf{n}|}\right| = \left|\text{free}(\overrightarrow{PQ})\cdot\frac{\mathbf{n}}{|\mathbf{n}|}\right| = |(\mathbf{q}-\mathbf{p})\cdot\hat{\mathbf{n}}|,$$

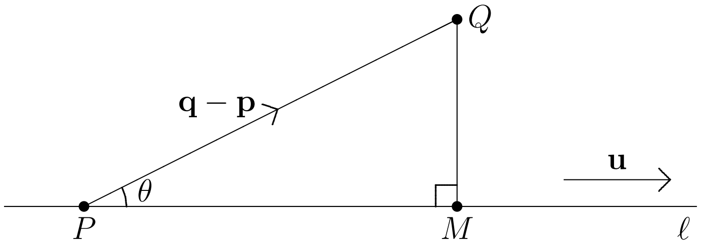
where $\text{free}(\overrightarrow{PQ}) = \mathbf{q}-\mathbf{p}$ is the (free) vector represented by $\overrightarrow{PQ}$, and $\hat{\mathbf{n}}$ is the unit vector in the direction of $\mathbf{n}$.

Example. We find the distance of $(3,-2,4)$ from the plane defined by $2x+3y-5z = 7$. With our notation we have $\mathbf{n} = \begin{pmatrix} 2 \\ 3 \\ -5 \end{pmatrix}$, $d = 7$, $\mathbf{q} = \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix}$, and so the distance is

$$\frac{|\mathbf{q}\cdot\mathbf{n} - d|}{|\mathbf{n}|} = \frac{|(6-6-20)-7|}{\sqrt{2^2+3^2+(-5)^2}} = \frac{27}{\sqrt{38}}.$$

3.5 The distance from a point to a line

Let ℓ be the line with (vector) equation $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$, where $\mathbf{u} \neq \mathbf{0}$, and let Q be a point with position vector $\mathbf{q}$. If $Q = P$ (where P has position vector P), then Q is on ℓ, and the distance between Q and ℓ is 0. Else we drop a normal from Q to ℓ meeting ℓ at the point M.



The distance from Q to ℓ is $|\overrightarrow{MQ}|$, that is $|\mathbf{q} - \mathbf{m}|$, is easily seen from the diagram to be $|\overrightarrow{PQ}| \sin\theta$, where θ is the angle between $\mathbf{q} - \mathbf{p}$ (the vector that $\overrightarrow{PQ}$ represents) and the vector $\mathbf{u}$ (which is the direction [up to opposite] of the line ℓ). [Pedants should recall that $\mathbf{u} \neq \mathbf{0}$, and also note that $\sin\theta \geqslant 0$, since $0 \leqslant \theta \leqslant \pi$.] For now we content ourselves with noting that this distance is $|\mathbf{q} - \mathbf{p}| \sin\theta$ (which 'morally' applies even when $\mathbf{q} = \mathbf{p}$). When we encounter the cross product, we shall be able to express this distance as $|(\mathbf{q} - \mathbf{p}) \times \mathbf{u}|/|\mathbf{u}|$. Note that $\sin\theta$ can be calculated using dot products, since $\cos\theta$ can be so calculated, and we have $\sin\theta = \sqrt{1 - \cos^2\theta}$. On calculating $|\mathbf{q} - \mathbf{p}| \sin\theta$, we find that the distance from Q to ℓ is

$$\frac{\sqrt{|\mathbf{q} - \mathbf{p}|^2|\mathbf{u}|^2 - ((\mathbf{q} - \mathbf{p})\cdot\mathbf{u})^2}}{|\mathbf{u}|},$$

a formula which applies even when $\mathbf{q} = \mathbf{p}$. In the case when $|\mathbf{u}| = 1$ the above formula simplifies to $\sqrt{|\mathbf{q} - \mathbf{p}|^2 - ((\mathbf{q} - \mathbf{p})\cdot\mathbf{u})^2}$.

Exercise. Use methods from Calculus I to minimise the distance from R to Q, where R, the typical point on ℓ, has position vector $\mathbf{r}$ with $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$. Show that this minimum agrees with the distance from Q to ℓ given above. Hint: The quantity $|\mathbf{r} - \mathbf{q}|$ is always at least 0. So $|\mathbf{r} - \mathbf{q}|$ is minimal precisely when $|\mathbf{r} - \mathbf{q}|^2$ is minimal. But $|\mathbf{r} - \mathbf{q}|^2 = (\mathbf{r} - \mathbf{q})\cdot(\mathbf{r} - \mathbf{q})$, and this is easier to deal with.

Chapter 4

Intersections of Planes and Systems of Linear Equations

Using coördinates, a plane Π is defined by an equation:

$$ax + by + cz = d,$$

where a, b, c, d are real numbers, and at least one of a, b, c is nonzero. The set of points on Π consists precisely of the points (p, q, r) with $ap + bq + cr = d$. In set-theoretic notation, this set is:

$$\{\, (p, q, r) : p, q, r \in \mathbb{R} \mid ap + bq + cr = d \,\}.$$

Suppose we wish to determine the intersection (as a set of points) of k given planes $\Pi_1, \Pi_2, \ldots, \Pi_k$ given by the respective equations:

$$\left.\begin{array}{c} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \\ \vdots \\ a_kx + b_ky + c_kz = d_k \end{array}\right\}. \tag{4.1}$$

Now a point (a, b, c) is in this intersection precisely when it is on each of the planes $\Pi_1, \Pi_2, \ldots, \Pi_k$, which is the case precisely when $x = p$, $y = q$, $z = r$ is a solution to each of the equations $a_ix + b_iy + c_iz$ of (4.1), where $1 \leqslant i \leqslant k$. Thus, to determine the intersection, we need to determine the solutions to the system of k linear equations (4.1) in 3 unknowns. The technique we shall use to do this, called *Gaußian elimination* or *reduction to echelon form* can be applied to determine the solutions to a system of k linear equations in any number of unknowns.

Note. Gaußian elimination is named after Carl Friedrich Gauß (1777–1855). The symbol ß is a special German symbol called *Eszett* or *scharfes S*, and is pronounced like the English word-initial S, and is often rendered into English as ss; thus Gauß is often written as Gauss in English.

4.1 Possible intersections of k planes in $\mathbb{R}^3$

For $k = 0, 1, 2, 3, 4, \ldots$ we detail below various configurations of k planes in $\mathbb{R}^3$, and what their intersections are. (The cases $k = 0, 1$ did not appear in lectures.)

- The intersection of 0 planes of $\mathbb{R}^3$ is the whole of $\mathbb{R}^3$. (See below for why.)
- The intersection of 1 plane(s) Π_1 of $\mathbb{R}^3$ is simply Π_1.
- The intersection of 2 planes Π_1, Π_2 of $\mathbb{R}^3$ is usually a line. The only exceptions occur when Π_1 and Π_2 are parallel. In such a case, if $\Pi_1 \neq \Pi_2$, then Π_1 and Π_2 intersect nowhere, whereas if $\Pi_1 = \Pi_2$, then Π_1 and Π_2 intersect in the plane Π_1. Example 3 below is a case when Π_1 and Π_2 are parallel but not equal.
- In general, 3 planes Π_1, Π_2, Π_3 intersect at precisely one point (Example 1 below is like this). Exceptional situations arise when two (or all) of the planes are parallel. Assuming that no two of Π_1, Π_2, Π_3 are parallel, exceptional situations arise only when the intersections of Π_1 with Π_2, Π_2 with Π_3 and Π_3 with Π_1 are parallel lines. These lines either coincide, in which case Π_1, Π_2, Π_3 intersect in this line (Example 2 below is like this), or the three lines are distinct, in which case Π_1, Π_2, Π_3 have empty intersection (Example 2′ below is like this).
- In general, 4 or more planes intersect at no points whatsoever. Another way of saying this is that their intersection is $\varnothing$, the empty set. Non-empty intersections are possible in exceptional cases.

4.1.1 Empty intersections, unions, sums and products

This was not done in lectures. Empty products and so on are somewhat subtle, and cause a lot of confusion and stress. Take the following as definitions.

- If I intersect 0 sets, each of which is presumed to belong to some "universal" set, then their intersection is that "universal" set. In the case above, the "universal" set was $\mathbb{R}^3$. A *"universal" set* is a set that contains (as elements) all the entities one wishes to consider in a given situation. If no "universal" set is understood (or exists) in the context in which you happen to be working, then the intersection of 0 sets is undefined. Taking the complement of a set is only defined when a "universal" set is around.
- The union of 0 sets is the empty set $\varnothing$. (There is no need to assume the existence a "universal" set here.)
- The sum of 0 real (or rational or complex) numbers is 0, and the sum of 0 vectors is $\mathbf{0}$. [In general, the sum of 0 things is the additive identity of the object these things are taken to belong to, when such a thing exists and is unique.]
- The product of 0 real (or rational or complex) numbers is 1. [In general, the product of 0 things is the multiplicative identity of the object these things are taken to belong to, when such a thing exists and is unique.]

4.2 Some examples

Before we formalise the notions of *linear equation*, *Gaußian elimination*, *echelon form* and *back substitution* in the next chapter, we give some examples of solving systems of linear equations using these methods.

Example 1. We determine all solutions to the system of equations:

$$\left.\begin{array}{rcl} x + y + z &=& 1 \\ -2x + 2y + z &=& -1 \\ 3x + y + 5z &=& 7 \end{array}\right\}. \tag{4.2}$$

We use the first equation to eliminate x in the second and third equations. We do this by adding twice the first equation to the second, and -3 times the first equation to the third, to get:

$$\left.\begin{array}{rcl} x + y + z &=& 1 \\ 4y + 3z &=& 1 \\ -2y + 2z &=& 4 \end{array}\right\}.$$

We now use the second equation to eliminate y in third by adding $\frac{1}{2}$ times the second equation to the third, which gives:

$$\left.\begin{array}{rcl} x + y + z &=& 1 \\ 4y + 3z &=& 1 \\ \frac{7}{2}z &=& \frac{9}{2} \end{array}\right\}.$$

We have now reduced the system to something called *echelon form*, and this is easy to solve by a process known as *back substitution*. The third equation gives $z = \frac{9/2}{7/2} = \frac{9}{7}$. Then the second equation gives $4y + 3(\frac{9}{7}) = 1$, and so $4y = -\frac{20}{7}$, whence $y = -\frac{5}{7}$. Then the first equation gives $x - \frac{5}{7} + \frac{9}{7} = 1$, whence $x = \frac{3}{7}$.

We conclude that the only solution to the system of equations (4.2) is $x = \frac{3}{7}$, $y = -\frac{5}{7}$, $z = \frac{9}{7}$. Thus the three planes defined by the equations of (4.2) intersect in the single point $(\frac{3}{7}, -\frac{5}{7}, \frac{9}{7})$. Recall that, in general, three planes intersect in precisely one point.

It is always good practice to check that any solution you get satisfies the original equations. You will probably pick up most mistakes this way. If your 'solution' does not satisfy the original equations then you have certainly made a mistake. If the original equations are satisfied, then you have possibly made a mistake and got lucky, and you could still have overlooked some solution(s) other than the one(s) you found. Naturally, the check works fine here.

Example 2. We determine all solutions to the system of equations:

$$\left.\begin{array}{rcl} -y - 3z &=& -7 \\ 2x - y + 2z &=& 4 \\ -4x + 3y - 2z &=& -1 \end{array}\right\}. \tag{4.3}$$

We want a nonzero x-term (if possible) in the first equation, so we interchange the first two equations, to get:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -7 \\ -4x + 3y - z &= -1 \end{aligned}\right\}.$$

We now use the first equation to eliminate the x-term from the other equations. To do this we add twice the first equation to the third equation; we leave the second equation alone, since its x-term is already zero. We now have:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -7 \\ y + 3z &= 7 \end{aligned}\right\}.$$

We now use the second equation to eliminate y from the third equation. We do this by adding the second equation to the third equation, which gives:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -7 \\ 0 &= 0 \quad (!) \end{aligned}\right\}. \tag{4.4}$$

This system of equations is in echelon form, but has the rather interesting equation $0 = 0$. This prompts the following definition.

Definition. An equation $ax + by + cz = d$ is called *degenerate* if $a = b = c = 0$ (NB: we do allow $d \neq 0$ as well as $d = 0$). Otherwise it is *non-degenerate*.

There are two types of degenerate equations.

1. The degenerate equation $0 = 0$ (in 3 variables x, y, z) has as solutions $x = p$, $y = q$, $z = r$, for *all* real numbers p, q, r.

2. The degenerate equation $0 = d$, with $d \neq 0$, has *no* solutions. Note that the $=$ sign is being used in two different senses in the previous sentence: the first use relates two sides of an equation, and the second use is as equality. This may be confusing, but I am afraid you are going to have to get used to it.

Since the equation $0 = 0$ yields no restrictions whatsoever, we may discard it from the system of equations (4.4) to obtain:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -7 \end{aligned}\right\}. \tag{4.5}$$

This system of equations is in echelon form, and has no degenerate equations, and we solve this system of equations using the process of back substitution. The variable z can be any real number t, since z is not a leading variable in any of the equations of (4.5), where we define the term *leading variable* in the next chapter. Then the second equation gives $-y-3t=-7$, and so $y=7-3t$. Then the first equation gives $2x-(7-3t)+2t=4$, and so $2x=11-5t$, and hence $x=\frac{11}{2}-\frac{5}{2}t$.

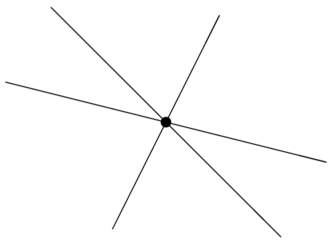
We have that $x=\frac{11}{2}-\frac{5}{2}t$, $y=7-3t$, $z=t$ is a solution for all real numbers t (this is infinitely[1] many solutions). Therefore, the intersection of the three planes defined by (4.3) is

$$\{\,(\tfrac{11}{2}-\tfrac{5}{2}t, 7-3t, t) : t\in\mathbb{R}\,\}.$$

This intersection is a line, with parametric equations:

$$\left.\begin{aligned} x &= \tfrac{11}{2}-\tfrac{5}{2}\lambda \\ y &= \phantom{\tfrac{1}{2}}7-3\lambda \\ z &= \phantom{\tfrac{11}{2}-\tfrac{5}{2}}\lambda \end{aligned}\right\}.$$

When we take a cross-section through the configuration of planes defined by the original equations, we get a diagram like the following, where all the planes are perpendicular to the page.



The sceptic will wonder whether we have lost any information during the working of this example. We shall discover that the method we use preserves all the information contained in the original equations. Nevertheless, it is still prudent to check, for all real numbers t, that $(x,y,z)=(\frac{11}{2}-\frac{5}{2}t, 7-3t, t)$ is a solution to all of the original equations (4.3).

Example 2′. The equations here have the same left-hand sides as those in Example 2. However, their right-hand sides are different (I may have made a different alteration in lectures).

$$\left.\begin{aligned} -\ \ y-3z &= -5 \\ 2x-\ \ y+2z &= 4 \\ -4x+3y-\ \ z &= -1 \end{aligned}\right\}. \tag{4.6}$$

The steps one must perform to bring the equations into echelon form are the same as in Example 2. In all cases, the left-hand sides should all be the same. However, the

[1]In 2010, the BBC broadcast a Horizon programme about Infinity, including contributions from Professor P. J. Cameron of our department. One thing we learned in the programme is that there are different sizes of infinite set. The sets $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ all have the same size, denoted $\aleph_0$ (the countably infinite cardinality), while $\mathbb{R}$ has a strictly bigger size, denoted $2^{\aleph_0}$. The symbol $\aleph$ is the first letter of the Hebrew alphabet, is called 'aleph, and traditionally stands for a glottal stop.

right-hand sides will differ. Echelonisation proceeds as follows. First swap the first and second equations:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -5 \\ -4x + 3y - z &= -1 \end{aligned}\right\} .$$

Add twice the first equation to the third:

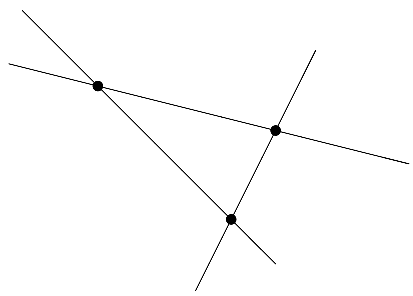
$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -5 \\ y + 3z &= 7 \end{aligned}\right\} .$$

Add the second equation to the third:

$$\left.\begin{aligned} 2x - y + 2z &= 4 \\ - y - 3z &= -5 \\ 0 &= 2 \end{aligned}\right\} .$$

But $0 = 2$ is a degenerate equation with *no* solutions, so the whole system of equations has no solutions, and thus the original system of equations has no solutions. There is no need to engage in back substitution for this example.

When we take a cross-section through the configuration of planes defined by the original equations, we get a diagram like the following, where all the planes are perpendicular to the page.



Checking all the solutions we obtain is vacuous in this example. A better bet is to follow through the echelonisation to try to figure out how to obtain the equation $0 = 2$ from the original equations (4.6). In this case, we find that we obtain $0 = 2$ by adding the first and third equations of (4.6) to twice the second equation of (4.6).

Example 3. We determine the intersection of the two planes defined by:

$$\left.\begin{aligned} x + 2y - z &= 2 \\ -2x - 4y + 2z &= 1 \end{aligned}\right\} . \tag{4.7}$$

Add twice the first equation to the second to get:

$$\left.\begin{aligned} x + 2y - z &= 2 \\ 0 &= 5 \end{aligned}\right\} .$$

But $0 = 5$ is a degenerate equation with *no* solutions, so the whole system of equations has no solutions, and thus the original system of equations has no solutions. So the intersection of the planes is $\varnothing = \{\}$, the empty set, which is the *only* set that has *no* elements. In this case, the original two planes were parallel, but not equal.

Example 4. [Not done in lectures.] If we have a system of $k = 0$ equations in unknowns x, y, z, then the solutions to this system of equations is $x = r$, $y = s$, $z = t$, where r, s, t can be any real numbers. The solution set is thus:

$$\{ (r, s, t) : r, s, t \in \mathbb{R} \} = \mathbb{R}^3,$$

which corresponds to my earlier assertion that the intersection of 0 planes in $\mathbb{R}^3$ is the whole of $\mathbb{R}^3$.

4.3 Notes

We have been solving systems of linear equations by employing two basic types of operations on these equations to bring them into an easy-to-solve form called *echelon form*. These operations are:

(A) adding a multiple of one equation to another;

(I) interchanging two equations; and

(M) [not used by us] multiplying an equation by a *nonzero* number.

These are called *elementary operations* on the system of linear equations, and the corresponding operations on matrices (we define matrices later) are called *elementary row operations*.

These elementary operations (including (M)) are all invertible, and as a consequence *never* change the set of solutions of a system of linear equations (see Coursework 4). It is for this reason that we kept writing down equations we had seen previously, and not just the new equations we had found. The whole system of linear equations is important, and if we did not keep track of the whole system this then we might lose some information on the way and inadvertently deduce more solutions to our equations than the original equations had.

One should be careful how one annotates row operations. Please bear in mind that we are operating on systems of equations, which should thus be linked by a brace (you will lose marks for forgetting this). Writing $R_1 + 2R_2$ does not tell me the row operation you have performed. Does this mean add 2 copies of Row 2 to Row 1 (an operation you would never use)? In that case, you could write $R_1 \mapsto R_1 + 2R_2$ or $R_1' = R_1 + 2R_2$. Or does $R_1 + 2R_2$ replace Row 2? (This is not one of our basic operations, but I have still seen it in work I had to mark.) During Gaußian elimination, failure to indicate the row operations used, or indicating them ambiguously, is also liable to lose marks.

Chapter 5

Gaußian Elimination and Echelon Form

A *linear equation* in variables $x_1, x_2, \ldots, x_n$ is of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = d,$$

where $a_1, a_2, \ldots, a_n, d$ are scalars. The x_1-term a_1x_1 is the first term, The x_2-term a_2x_2 is the second term, and in general the x_i-term a_ix_i is the i^{th} term. In the case when $n = 3$, we generally use a, b, c, x, y, z instead of $a_1, a_2, a_3, x_1, x_2, x_3$; thus $a_1x_1 + a_2x_2 + a_3x_3 = d$ becomes $ax + by + cz = d$.

A linear equation $a_1x_1+a_2x_2+\cdots+a_nx_n = d$ is *degenerate* if $a_1 = a_2 = \cdots = a_n = 0$; otherwise it is *non-degenerate*. The equation can be degenerate even if $d \neq 0$.

5.1 Echelon form (Geometry I definition)

Before defining echelon form we make some comments. Firstly, echelon form is properly a concept for matrices; we give an equivalent definition for systems of linear equations. Secondly, there are 3 different definitions of echelon form, of varying strengths, and it is possible that you may have encountered a different one previously. We use the weakest of the definitions in this course. **For this course, it is important that you understand and use the definition of echelon form given below.**

Definition 5.1. A system of linear equations in $x_1, x_2, \ldots, x_n$ is in *echelon form* if every non-degenerate equation begins with strictly fewer zero terms than each equation below it and any degenerate equation occurs after (below) the non-degenerate equations.

Note that any system of linear equations in echelon form has at most n non-degenerate equations (if any), but can have arbitrarily many degenerate equations (if any). A system with no non-degenerate equations (or no equations whatsoever) is automatically in echelon form. The right-hand sides of a system of linear equations play *no* rôle in determining whether that system is in echelon form. We *do not* insist that the first nonzero term of each non-degenerate equation be x_i for some i. (This extra condition is currently required in the MTH5112: Linear Algebra I definition of echelon form.)

Example. As linear equations in the variables x, y, z, we have that:

- $x+2y=8$ begins with 0 zero terms (the zero z-term does not begin the equation);
- $3y-4z=-5$ begins with 1 zero term;
- $-5z=2$ begins with 2 zero terms;
- $0=0$ and $0=5$ both begin with 3 zero terms (the right-hand sides are irrelevant).

Thus the system of equations

$$\left.\begin{aligned} x+2y \qquad &= \ \ 8 \\ 3y-4z &= -5 \\ -5z &= \ \ 2 \\ 0 &= \ \ 0 \\ 0 &= \ \ 5 \end{aligned}\right\}$$

is in echelon form.

Also, the following systems of equations are in echelon form:

$$\left.\begin{aligned} x+2y \qquad &= \ \ 3 \\ 3y-4z &= -5 \\ -5z &= \ \ 4 \end{aligned}\right\} \qquad \left.\begin{aligned} x+2y \qquad &= 3 \\ -5z &= 4 \\ 0 &= 0 \end{aligned}\right\} \qquad \left.\begin{aligned} x+2y \qquad &= \ \ 3 \\ -5z &= \ \ 4 \\ 0 &= -5 \end{aligned}\right\},$$

and these are in echelon form too:

$$\left.\begin{aligned} 3y-4z &= -5 \\ -5z &= \ \ 7 \end{aligned}\right\} \qquad \left.\begin{aligned} 3y-4z &= -5 \\ 0 &= \ \ 7 \end{aligned}\right\} \qquad \left.\begin{aligned} 3y-4z &= -5 \\ 0 &= \ \ 0 \end{aligned}\right\} \qquad 3y-4z=-5 \ \}.$$

The following systems of equations are *not* in echelon form.

$$\left.\begin{aligned} x+2y \qquad &= \ \ 7 \\ -3z &= \ \ 4 \\ 2y-5z &= -3 \end{aligned}\right\} \qquad \left.\begin{aligned} x+4y-\ \ z &= -1 \\ 0 &= \ \ 4 \\ 3y-4z &= -5 \end{aligned}\right\} \qquad \left.\begin{aligned} x+3y \qquad &= \ \ 1 \\ y+4z &= -2 \\ y-5z &= \ \ 3 \end{aligned}\right\}.$$

In the last case, the last two equations commence with the same number (one) of zero terms but are non-degenerate.

5.2 Gaußian elimination

We now describe the process of *Gaußian elimination*, which is used to bring a system of linear equations (in $x_1, x_2, \ldots, x_n$ or x, y, z, etc.) into echelon form. **You <u>must</u> use exactly the algorithm described here, even though other ways of reducing to (a possibly different) echelon form may be mathematically valid.**[1] In particular, our algorithm does not use the M-operations of Section 4.3; these are required in general for reducing to the stronger versions of echelon form.

[1] The course MTH5112: Linear Algebra I uses a strictly stronger definition of echelon form than we do, and consequently uses a slightly different version of Gaußian elimination than we do in MTH4103: Geometry I. This modified version does require the M-operations of Section 4.3. You **<u>must not</u>** use the Linear Algebra I version of Gaußian elimination in this course.

Step 1. If the system is in echelon form then stop. (There is nothing to do.)

If each equation has zero x_1-term then go to Step 2. (There are no x_1-terms to eliminate.)

If the first equation has a zero x_1-term then interchange it with the first equation that has nonzero x_1-term. (So now the first equation has nonzero x_1-term.)

Add appropriate multiples of the first equation to the others to eliminate their x_1-terms.

Step 2. (At this point, all equations except perhaps the first should have zero x_1-term.)

If the system is in echelon form then stop.

If each equation with zero x_1-term also has zero x_2-term then go to Step 3.

If the first equation with zero x_1-term has a zero x_2-term then interchange it with the first equation that has zero x_1-term and nonzero x_2-term. (So now the first equation with zero x_1-term has nonzero x_2-term.)

Add appropriate multiples of the first equation with zero x_1-term to the equations below it to eliminate their x_2-terms.

Step m for $3 \leqslant m \leqslant n$. (At this point, all equations except perhaps the first up to $m-1$ should have zero x_1-, x_2-, …, x_{m-1}-terms. Each of these first $\leqslant m-1$ equations should begin with strictly more zero terms than their predecessors.)

If the system is in echelon form then stop.

If each equation with zero x_1-, x_2-, …, x_{m-1}-terms also has zero x_m-term then go to Step $m+1$. (This situation should only arise in the case when $m < n$; if $m = n$ this case should only 'arise' if you are already in echelon form [in which case you should have stopped already].)

If the first equation with zero x_1-, x_2-, …, x_{m-1}-terms also has zero x_m-term, then interchange it with the first equation having zero x_1-, x_2-, …, x_{m-1}-terms and nonzero x_m-term.

Add appropriate multiples of the first equation with zero x_1-, x_2-, …, x_{m-1}-terms to the equations below it to eliminate their x_m-terms.

If $m = n$ then the system of equations should now be in echelon form, so you can stop.

5.2.1 Notes on Gaußian elimination

1. If a degenerate equation of the form $0 = 0$ is created at any stage, it can be discarded, and the Gaußian elimination continued without it.
2. If a degenerate equation of the form $0 = d$, with $d \neq 0$, is created at any stage then the system of equations has no solutions, and the Gaußian elimination can be stopped.
3. The operations required to bring a system of equations into echelon form are independent of the right-hand sides of the equations. (Compare Examples 2 and $2'$ in Section 4.2.)

Example. We perform Gaußian elimination on the following system of equations.

$$\left.\begin{array}{r} x + 2y + z = 2 \\ x + 2y + 3z = -8 \\ 3x + 5y + 2z = 6 \\ -2x - 2y + z = 0 \end{array}\right\}. \tag{5.1}$$

Step 1. To eliminate the x-term in all equations but the first, we add -1 times the first equation to the second, -3 times the first equation to the third, and 2 times the first equation to the fourth. This gives:

$$\left.\begin{array}{r} x + 2y + z = 2 \\ 2z = -10 \\ - y - z = 0 \\ 2y + 3z = 4 \end{array}\right\}.$$

Step 2. The first equation having zero x-term (the second) also has zero y-term. The first equation having zero x-term and nonzero y-term is the third, so interchange the second and third equations, to get:

$$\left.\begin{array}{r} x + 2y + z = 2 \\ - y - z = 0 \\ 2z = -10 \\ 2y + 3z = 4 \end{array}\right\}.$$

Now that the second equation has zero x-term and nonzero y-term, use this to eliminate y from the equations below it. To do this we need only add twice the second equation to the fourth, to obtain:

$$\left.\begin{array}{r} x + 2y + z = 2 \\ - y - z = 0 \\ 2z = -10 \\ z = 4 \end{array}\right\}.$$

Step 3. We are still not in echelon form, and the first equation with zero x-term and y-term has nonzero z-term. (This is the third equation.) We use the third equation to eliminate z from all equations after the third (by adding $-\frac{1}{2}$ times the third equation to the fourth). This gives:

$$\left.\begin{array}{r} x + 2y + z = 2 \\ - y - z = 0 \\ 2z = -10 \\ 0 = 9 \end{array}\right\}.$$

This system of equations is now in echelon form, but has no solutions (as $0 = 9$ has no solutions). Hence the original system (5.1) has no solutions. Geometrically, this reflects the fact that four planes of $\mathbb{R}^3$ generally have empty intersection.

5.3 Solving systems of equations in echelon form

We now show how to find all solutions to a system of linear equations in echelon form. If the system contains an equation an equation of the form $0 = d$ with $d \neq 0$, then the system has no solutions, and we have nothing more to do. If the system has any equations of the form $0 = 0$, then throw these out, since they contribute no restriction on the solution set whatsoever. Thus we may now only consider systems of linear equations in echelon form which have no degenerate equations.

Definition. In a non-degenerate linear equation (written in standard form), the first variable, reading left to right, in a nonzero term is called the *leading variable* of the equation.

Example. In these examples, we assume that we have three variables x, y, z, in that order. In the first three cases, the linear equation is in standard form; in the latter two cases it is not.

- The leading variable of $-x + y = 3$ is x.
- The leading variable of $3y - 2z = 7$ is y.
- The leading variable of $-4z = 8$ is z.
- The standard form of $z - y + 2x - y = 4$ is $2x - 2y + z = 4$, so the leading variable is x.
- The equation $x + (-1)x - 3 = 4$ is degenerate (it is equivalent to $-3 = 4$, or $0 = 7$), and so has no leading variable defined.

5.3.1 Back substitution

Back substitution is an algorithm to determine *all* the solutions to a system of non-degenerate equations in echelon form. It proceeds as follows.

Step 1. Variables which are not leading variables of any of the equations in the system can take arbitrary (real) values. Assign a symbolic value to each such non-leading variable.[2]

Step 2. Given symbolic values for the non-leading variables, solve for the leading variables, starting from the bottom and working up.

[2] In general, we shall want to apply this procedure when the equations are defined over fields F other than $\mathbb{R}$. In this case, the non-leading variables should take arbitrary values in F. You should see the concept of *field* defined in MTH4104: Introduction to Algebra. (MTH4104 is compulsory for about half of you this year, and strongly recommended for the rest of you next year.) Examples of fields are $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ (but not $\mathbb{N}$ or $\mathbb{Z}$). You should have met all these sets in MTH4110: Mathematical Structures, and gained some extra familiarity with $\mathbb{C}$ in MTH4101: Calculus II.

Example 1. We apply back substitution to the following system of equations, which is in echelon form:

$$\left.\begin{aligned} x + 3y - z &= 7 \\ z &= 0 \end{aligned}\right\}.$$

There is just one non-leading variable, namely y. Thus y can take any real value, say $y = t$. We now solve for the leading variables x and z, starting with z.

The last equation gives $z = 0$. Therefore, the first equation gives $x + 3t + 0 = 7$, and so $x = 7 - 3t$. Thus, the solutions of the system are $x = 7 - 3t$, $y = t$, $z = 0$, where t can be any real number.

We remark that the intersection of the planes defined by these equations is thus $\{\,(7-3t, t, 0) : t \in \mathbb{R}\,\}$, which is the set of points on the line having parametric equations:

$$\left.\begin{aligned} x &= 7 - 3\lambda \\ y &= \quad\;\; \lambda \\ z &= 0 \end{aligned}\right\},$$

and Cartesian equations $\frac{x-7}{-3} = y$, $z = 0$.

Example 2. We apply back substitution to the following system of equations, which is in echelon form:

$$\left.\begin{aligned} x + 3y + 5z &= 9 \\ 2y + 4z &= 6 \\ 3z &= 3 \end{aligned}\right\}.$$

All the variables are leading in one of the equations. So we start immediately on Step 2. The (third) equation $3z = 3$ gives $z = 1$. Then the second equation becomes $2y + 4 = 6$, whence we get that $y = 1$. Then the first equation becomes $x + 3 + 5 = 9$, and so $x = 1$. Therefore the only solution of this system of equations is $x = y = z = 1$.

Example 3. We apply back substitution to the following system of just one equation, which is in echelon form:

$$2x - y + 3z = 5 \;\Big\}.$$

The non-leading variables y and z can take arbitrary real values, say $y = s$ and $z = t$. We now solve for the only leading variable, x, using the only equation of the system. We have $2x - s + 3t = 5$, so $2x = 5 + s - 3t$, and hence $x = \frac{5}{2} + \frac{s}{2} - \frac{3t}{2}$. Thus the solutions of the system are $x = \frac{5}{2} + \frac{s}{2} - \frac{3t}{2}$, $y = s$, $z = t$, where s and t can be any real numbers.

We remark that is follows that $\{\,(\frac{5}{2} + \frac{s}{2} - \frac{3t}{2}, s, t) : s, t \in \mathbb{R}\,\}$ is the set of points on the plane defined by $2x - y + 3z = 5$.

5.4 Summary

To solve a system of linear equations, first use Gaußian elimination to bring the system into echelon form, and then use back substitution to the system in echelon form, remembering to deal appropriately with degenerate equations along the way. You should review your notes, starting at the beginning of Chapter 4, to see many examples of this.

5.5 Intersection of a line and a plane

Consider a line ℓ defined by the parametric equations:

$$\left.\begin{array}{l} x = p_1 + u_1\lambda \\ y = p_2 + u_2\lambda \\ z = p_3 + u_3\lambda \end{array}\right\},$$

and a plane Π defined by

$$ax + by + cz = d.$$

We find the intersection of ℓ and Π by determining the λ for which

$$a(p_1 + u_1\lambda) + b(p_2 + u_2\lambda) + c(p_3 + u_3\lambda) = d. \tag{5.2}$$

This gives one linear equation in one unknown λ. The solution is unique, except in the degenerate case when $au_1 + bu_2 + cu_3 = 0$, in which case ℓ is parallel to Π. In this case, there are 0 or infinitely many solutions, depending on whether ℓ is in Π or not (which is if and only if $ap_1 + bp_2 + cp_3 = d$ or not). Note that rearranging Equation 5.2 to standard form for a linear equation in λ yields

$$(au_1 + bu_2 + cu_3)\lambda = d - (ap_1 + bp_2 + cp_3). \tag{5.3}$$

So if Π and ℓ have vector equations $\mathbf{r}\cdot\mathbf{n} = d$ and $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$ respectively, we find that

$$(\mathbf{n}\cdot\mathbf{u})\lambda = d - \mathbf{n}\cdot\mathbf{p}. \tag{5.4}$$

(You should be able to work out what $\mathbf{n}$, $\mathbf{p}$ and $\mathbf{u}$ are here.) Thus if $\mathbf{n}\cdot\mathbf{u} \neq 0$ we get the unique solution $\lambda = (d - \mathbf{n}\cdot\mathbf{p})/(\mathbf{n}\cdot\mathbf{u})$.

Example. For example, let ℓ be the line with parametric equations:

$$\left.\begin{array}{rl} x = & 1 + 2\lambda \\ y = & 2 + 3\lambda \\ z = & -1 - 4\lambda \end{array}\right\},$$

and let Π be the plane defined by $x - y + 2z = 3$. To determine the intersection $\Pi \cap \ell$ of Π and ℓ we solve

$$(1 + 2\lambda) - (2 + 3\lambda) + 2(-1 - 4\lambda) = 3.$$

This gives $-3 - 9\lambda = 3$, or $-9\lambda = 6$, with unique solution $\lambda = -\frac{2}{3}$. (If we wish to use the formula below Equation 5.4 to calculate λ we note that $d = 3$, $\mathbf{n}\cdot\mathbf{p} = -3$ and $\mathbf{n}\cdot\mathbf{u} = -9$.) Thus ℓ and Π intersect in the single point (x_0, y_0, z_0), with

$$\begin{aligned} x_0 &= 1 + 2(-\tfrac{2}{3}) = -\tfrac{1}{3}, \\ y_0 &= 2 + 3(-\tfrac{2}{3}) = 0, \\ z_0 &= -1 - 4(-\tfrac{2}{3}) = \tfrac{5}{3}. \end{aligned}$$

As a set of points, the intersection $\ell \cap \Pi$ of ℓ and Π is $\{(-\frac{1}{3}, 0, \frac{5}{3})\}$.

Notation. Let A and B be sets. Then the *intersection* of A and B is denoted $A \cap B$, and is the set of elements that are in both A and B. The *union* of A and B is denoted $A \cup B$, and is the set of elements that are in A or B (or both). As a mnemonic, the symbol $\cap$ resembles a lower case N, which is the second letter of intersection, and the symbol $\cup$ resembles a lower case U, the first letter of union.

The following properties hold for $\cap$ and $\cup$, for *all* sets A, B, C. The last pair of properties only makes sense in the presence of a 'universal' set $\mathscr{E}$.

- $A \cap B = B \cap A$ and $A \cup B = B \cup A$.
- $A \cap (B \cap C) = (A \cap B) \cap C$ and $A \cup (B \cup C) = (A \cup B) \cup C$.
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- $A \cap A = A \cup A = A$.
- $A \cap \varnothing = \varnothing$ and $A \cup \varnothing = A$.
- $A \cap \mathscr{E} = A$ and $A \cup \mathscr{E} = \mathscr{E}$.

5.6 Intersection of two lines

Consider lines ℓ and m defined (respectively) by the parametric equations:

$$\left.\begin{array}{r}x = p_1 + u_1\lambda\\ y = p_2 + u_2\lambda\\ z = p_3 + u_3\lambda\end{array}\right\}\qquad\text{and}\qquad\left.\begin{array}{r}x = q_1 + v_1\mu\\ y = q_2 + v_2\mu\\ z = q_3 + v_3\mu\end{array}\right\}.$$

We find the intersection of ℓ and m by determining the λ and μ for which:

$$\begin{aligned}p_1 + u_1\lambda &= q_1 + v_1\mu,\\ p_2 + u_2\lambda &= q_2 + v_2\mu,\\ p_3 + u_3\lambda &= q_3 + v_3\mu.\end{aligned}$$

This is equivalent to solving the following system of linear equations:

$$\left.\begin{array}{r}u_1\lambda - v_1\mu = q_1 - p_1\\ u_2\lambda - v_2\mu = q_2 - p_2\\ u_3\lambda - v_3\mu = q_3 - p_3\end{array}\right\}.$$

Note that, in general, we expect this system of equations to have no solution.

Example. In this example, the lines ℓ and m are defined (respectively) by the parametric equations:

$$\left.\begin{array}{r}x = 1 + \lambda\\ y = 2 + 3\lambda\\ z = 1 - 4\lambda\end{array}\right\}\qquad\text{and}\qquad\left.\begin{array}{r}x = 2 + 3\mu\\ y = 1 - \mu\\ z = 3 + 2\mu\end{array}\right\}.$$

Thus we must solve (for λ and μ) the following equations:

$$\left.\begin{aligned} 1+\lambda \;\; &= 2+3\mu, \\ 2+3\lambda &= 1-\mu, \\ 1-4\lambda &= 3+2\mu. \end{aligned}\right\}$$

This is equivalent to the following system of linear equations:

$$\left.\begin{aligned} \lambda - 3\mu &= 1 \\ 3\lambda + \mu &= -1 \\ -4\lambda - 2\mu &= 2 \end{aligned}\right\}.$$

We now apply Gaußian elimination. Adding -3 times the first equation to the second and 4 times the first equation to the third gives:

$$\left.\begin{aligned} \lambda - 3\mu &= 1 \\ 10\mu &= -4 \\ -\,14\mu &= 6 \end{aligned}\right\}.$$

We now need to eliminate μ in the third equation. To do this add $\frac{7}{5}[= -(\frac{-14}{10})]$ times the second equation to the third, to get:

$$\left.\begin{aligned} \lambda - 3\mu &= 1 \\ 10\mu &= -4 \\ 0 &= \tfrac{2}{5} \end{aligned}\right\}.$$

There is no solution to this system (because of the equation $0 = \frac{2}{5}$). Thus we conclude that the lines ℓ and m do not meet. As a set of points, the intersection of ℓ and m is $\varnothing$, the empty set.

Here, the lines ℓ and m are *skew*, that is, they do not meet in any point, but they are not parallel either (since their direction vectors are not scalar multiples of each other).

Chapter 6

The Vector Product

6.1 Parallel vectors

Suppose that $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors. We say that $\mathbf{u}$ and $\mathbf{v}$ are *parallel*, and write $\mathbf{u} \parallel \mathbf{v}$, if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$ (which will also force $\mathbf{v}$ to be a scalar multiple of $\mathbf{u}$). Note that $\mathbf{u}$ and $\mathbf{v}$ are parallel if and only if they have the same or opposite directions, which happens exactly when $\mathbf{u}$ and $\mathbf{v}$ are at an angle of 0 or π.

Example. The vectors $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \\ -6 \end{pmatrix}$ are parallel, but $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ -4 \\ -6 \end{pmatrix}$ are not parallel.

The relation of parallelism has the following properties, where $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are *nonzero* vectors.

1. $\mathbf{u} \parallel \mathbf{u}$ for all vectors $\mathbf{u}$. [This property is called *reflexivity*.]
2. $\mathbf{u} \parallel \mathbf{v}$ implies that $\mathbf{v} \parallel \mathbf{u}$. [This property is called *symmetry*.]
3. $\mathbf{u} \parallel \mathbf{v}$ and $\mathbf{v} \parallel \mathbf{w}$ implies that $\mathbf{u} \parallel \mathbf{w}$. [This property is called *transitivity*.]

A relation that is reflexive, symmetric and transitive is called an *equivalence relation*. This concept was introduced in MTH4110: Mathematical Structures. Thus, parallelism is a equivalence relation on the set of nonzero vectors

Note. A wording like "$\mathbf{u}$ and $\mathbf{v}$ are parallel if ..." presumes that the property of parallelism is symmetric between $\mathbf{u}$ and $\mathbf{v}$. This may be obvious from the definition that follows, or else a proof would need to be supplied. If you want to define a concept that is asymmetric (or not obviously symmetric), a wording like "$\mathbf{u}$ is parallel to $\mathbf{v}$ if ..." is more appropriate. [Can you prove that the relations of parallelism and collinearity (see below) are symmetric? Be careful of the zero vector for the latter relation.]

Also, when one provides more than one definition of a concept, one should prove the equivalence of the definitions. Can you prove the various definitions of collinear to be equivalent?

6.2 Collinear vectors

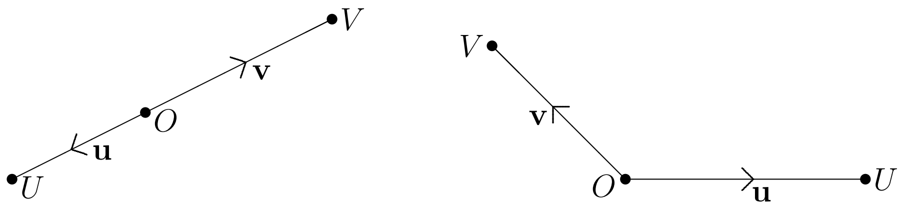
It is useful to extend the notion of parallelism to pairs of vectors involving the zero vector. However, we shall give this notion a different name, and call it collinearity. We say that two vectors $\mathbf{u}$ and $\mathbf{v}$ are *collinear* if $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$ and $\mathbf{v}$ are parallel (this includes the case $\mathbf{u} = \mathbf{v} = \mathbf{0}$).

An equivalent definition of collinearity is that $\mathbf{u}$ and $\mathbf{v}$ are collinear if there exist [real] numbers (scalars) α and β not both zero such that

$$\alpha\mathbf{u} + \beta\mathbf{v} = \mathbf{0}. \tag{6.1}$$

(This covers the case when one or both of $\mathbf{u}$ and $\mathbf{v}$ is $\mathbf{0}$, as well as the general case when $\mathbf{u}$ and $\mathbf{v}$ are parallel.) The relation of collinearity is reflexive and symmetric (in all dimensions), but is not transitive in dimension at least 2. However, it is nearly transitive [not a technical term]: if $\mathbf{u}$ and $\mathbf{v}$ are collinear, and $\mathbf{v}$ and $\mathbf{w}$ are collinear, then either $\mathbf{u}$ and $\mathbf{w}$ are collinear or $\mathbf{v} = \mathbf{0}$ (or both).

Another definition of collinearity is that $\mathbf{u}$ and $\mathbf{v}$ are collinear if O, U and V all lie on some [straight] line, where $\overrightarrow{OU}$ represents $\mathbf{u}$ and $\overrightarrow{OV}$ represents $\mathbf{v}$. This line is unique except in the case $\mathbf{u} = \mathbf{v} = \mathbf{0}$, that is $O = U = V$. Two diagrams illustrating this concept are given below. In the left-hand one $\mathbf{u}$ and $\mathbf{v}$ are collinear, and in the right-hand one they are not.



Note. The word collinear is logically made up of two parts: the prefix *co-*, and the stem *linear*. Thus it would seem that the logical spelling of *collinear* should be as *colinear*. However, *co-* is a variant of *com-* or *con-*, roughly meaning 'with' or 'together', and this assimilates to *col-* and *cor-* before words beginning with L and R respectively. Thus, in contrast to words like *coplanar*, we get words like *collinear* and *correlation* since the plain *co-* prefix is rare before words beginning with L and R. (The phenomenon of assimilation also affects the prefix Latinate prefix *in-*, giving us words like *impossible*, *immaterial*, *illegal* and *irrational*. The Germanic prefix *un-* is unaffected by this phenomenon.)

6.3 Coplanar vectors

We say that vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are *coplanar* if the points O, U, V, W all lie on some plane, where $\overrightarrow{OU}$, $\overrightarrow{OV}$, $\overrightarrow{OW}$ represent the free vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ respectively. (Note that this definition is symmetric in $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.) Thus we see that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are coplanar if and only if either $\mathbf{u}$ and $\mathbf{v}$ are collinear or there exist scalars λ and μ such that $\mathbf{w} = \lambda\mathbf{u} + \mu\mathbf{v}$.

From the above, we get the following symmetrical algebraic formulation of coplanarity: $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are *coplanar* if there exist [real] numbers (scalars) α, β, γ such that

$$\alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w} = \mathbf{0} \quad \text{and} \quad (\alpha, \beta, \gamma) \neq (0, 0, 0). \tag{6.2}$$

Note that to determine whether a particular triple $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is coplanar, we can use the above equation $\alpha\mathbf{u} + \beta\mathbf{v} + \gamma\mathbf{w} = \mathbf{0}$, which induces 3 linear equations in the unknowns α, β, γ (or n equations if $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ rather than $\mathbb{R}^3$). These equations always have the solution $\alpha = \beta = \gamma = 0$, and $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are coplanar if and only if the equations have a solution *other than* $\alpha = \beta = \gamma = 0$.

Example 1. If $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$ then $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are coplanar. For example, if $\mathbf{u} = \mathbf{0}$, we can take $\alpha = 1$, $\beta = \gamma = 0$ in Equation 6.2. Geometrically, the points O, U, V, W consist of at most 3 distinct points, and any three points (in $\mathbb{R}^3$) lie on at least one plane.

Example 2. Suppose that $\mathbf{u}$ and $\mathbf{v}$ are (nonzero and) parallel. Then $\mathbf{v} = \lambda\mathbf{u}$ for some scalar λ. Therefore we have $\lambda\mathbf{u} - \mathbf{v} = \mathbf{0}$, and so in Equation 6.2 we can take $\alpha = \lambda$, $\beta = -1$, $\gamma = 0$. Geometrically, O, U, V lie on one line (which is unique since $U, V \neq O$), and so O, U, V, W lie on at least one plane (which is unique, except when W lies on the line determined by O, U, V).

Example 3. We let $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$. Then Equation 6.2 yields the following linear equations in α, β, γ:

$$\left.\begin{aligned} \alpha + 4\beta + 7\gamma &= 0 \\ 2\alpha + 5\beta + 8\gamma &= 0 \\ 3\alpha + 6\beta + 9\gamma &= 0 \end{aligned}\right\}.$$

Echelonisation gives the following system of equations:

$$\left.\begin{aligned} \alpha + 4\beta + 7\gamma &= 0 \\ -3\beta - 6\gamma &= 0 \\ 0 &= 0 \end{aligned}\right\}.$$

Back substitution then gives the solution $\alpha = t$, $\beta = -2t$, $\gamma = t$, where t can be any real number. Setting $t = 1$ (any nonzero value will do), we see that $\mathbf{u} - 2\mathbf{v} + \mathbf{w} = \mathbf{0}$, so that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are coplanar.

Example 4. We let $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$. Then Equation 6.2 yields the following linear equations in α, β, γ:

$$\left.\begin{aligned} \alpha + 2\beta + 3\gamma &= 0 \\ 2\alpha + 3\beta + \gamma &= 0 \\ 3\alpha + \beta + 2\gamma &= 0 \end{aligned}\right\}.$$

Echelonisation gives the following system of equations:

$$\left.\begin{aligned} \alpha + 2\beta + 3\gamma &= 0 \\ -\beta - 5\gamma &= 0 \\ 18\gamma &= 0 \end{aligned}\right\}.$$

Thus we see that the only solution to this system of equations is $\alpha = \beta = \gamma = 0$. Therefore **u**, **v**, **w** are *not* coplanar.

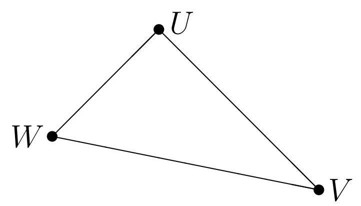
6.4 Right-handed and left-handed triples

Suppose now that **u**, **v**, **w** are not coplanar. Pick an origin O in 3-space, and define U, V, W by the condition that $\overrightarrow{OU}$, $\overrightarrow{OV}$, $\overrightarrow{OW}$ represent **u**, **v**, **w** respectively. We now give three methods to determine whether we have a right-handed or left-handed triple. You should try to convince yourselves of the equivalence of these methods. Some of the results stated under the "finger exercises" (Section 6.4.1) are easier to prove with one of these methods than another.

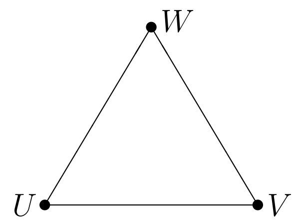
Method 1. We let the plane determined by U, V, W be the page of these notes. (Note that U, V, W are not all on the same line, so that this plane is unique.) We now orient ourselves so that O is in front of the page, which is the same side as us. (Note that O cannot be on the page.) We now follow the vertices of the triangle UVW around *clockwise*. If they occur in the order U, V, W (or V, W, U or W, U, V) then the triple is right-handed. Else, for the orders U, W, V and V, U, W and W, V, U, the triple is left-handed.

If O is behind the page (the other side to us), then the *anticlockwise* orders U, V, W and V, W, U and W, U, V give us a right-handed triple, while the anticlockwise orders U, W, V and V, U, W and W, V, U give us a left-handed triple.

The following diagrams illustrate the situations that arise, where the page is the plane containing U, V and W. Note that O being on the page would make **u**, **v**, **w** coplanar, so this does not happen here.



right-handed if O is **in front of** the page
left-handed if O is **behind** the page



right-handed if O is **behind** the page
left-handed if O is **in front of** the page

Method 2. We let the plane determined by O, U, V be the page of these notes. We then look at the plane from the side such the angle from $\mathbf{u}$ to $\mathbf{v}$ *proceeding anticlockwise* is between 0 and π. (If the anticlockwise angle lies between π and 2π then look at the plane from the other side.) If $\mathbf{w}$ points 'towards' you, that is W is the same side of the O, U, V-plane as you are, then $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is right-handed. If $\mathbf{w}$ points 'away from' you then $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is left-handed.

Method 3. Using your right hand put your thumb in the direction of $\mathbf{u}$, and your first [index] finger in the direction of $\mathbf{v}$. If W lies on the side of the plane through O, U, V indicated by your second [middle] finger, we call $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ a *right-handed triple*; otherwise it is a *left-handed triple*.

For any triple $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ of vectors, precisely one of the following properties holds: it is coplanar, it is right-handed, or it is left-handed.

Examples. $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ is a right-handed triple. $\mathbf{i}$, $\mathbf{j}$, $-\mathbf{k}$ and $\mathbf{k}$, $\mathbf{j}$, $\mathbf{i}$ are both left-handed triples.

Note. The vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ forming a right-handed or left-handed triple need not be mutually orthogonal, but they *must not* be coplanar.

6.4.1 Some finger exercises (for you to do)

If $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are coplanar, then so is any triple that is a permutation of $\pm\mathbf{u}$, $\pm\mathbf{v}$, $\pm\mathbf{w}$, for any of the 8 possibilities for sign. From now on, we assume that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are not coplanar. The basic operations we can do to a triple of vectors are:

1. multiply one the vectors by a constant $\lambda > 0$;
2. negate one of them; or
3. swap two of them.

The first operation preserves the handedness of a triple, but the latter two operations send right-handed triples to left-handed triples and vice versa. (Each of these operations sends coplanar triples to coplanar triples.) Combining these operations, we see that negating an even number of the vectors preserves handedness, while negating an odd number of the vectors changes the handedness. One can convert $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ to each of $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$ and $\mathbf{w}$, $\mathbf{u}$, $\mathbf{v}$ using two swaps. Thus one sees that (in 3 dimensions) cycling $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ does not change the handedness of the system. We have the following.

- If $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a right-handed triple, then the following triples are also right-handed: $\mathbf{u}, \mathbf{v}, \mathbf{w}$; $\mathbf{v}, \mathbf{w}, \mathbf{u}$; $\mathbf{w}, \mathbf{u}, \mathbf{v}$; $\mathbf{u}, -\mathbf{v}, -\mathbf{w}$; $-\mathbf{u}, \mathbf{v}, -\mathbf{w}$; $-\mathbf{u}, -\mathbf{v}, \mathbf{w}$; $\mathbf{u}, -\mathbf{w}, \mathbf{v}$; $\mathbf{u}, \mathbf{w}, -\mathbf{v}$; $-\mathbf{u}, \mathbf{w}, \mathbf{v}$; $\mathbf{w}, \mathbf{v}, -\mathbf{u}$; $\mathbf{w}, -\mathbf{v}, \mathbf{u}$; $-\mathbf{w}, \mathbf{v}, \mathbf{u}$; $-\mathbf{w}, -\mathbf{v}, -\mathbf{u}$; and so on.
- If $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a right-handed triple, then the following triples are left-handed: $\mathbf{u}, \mathbf{w}, \mathbf{v}$; $\mathbf{w}, \mathbf{v}, \mathbf{u}$; $\mathbf{v}, \mathbf{u}, \mathbf{w}$; $-\mathbf{u}, \mathbf{v}, \mathbf{w}$; $\mathbf{u}, -\mathbf{v}, \mathbf{w}$; $\mathbf{u}, \mathbf{v}, -\mathbf{w}$; $-\mathbf{u}, -\mathbf{v}, -\mathbf{w}$; $-\mathbf{v}, -\mathbf{w}, -\mathbf{u}$; $-\mathbf{w}, -\mathbf{u}, -\mathbf{v}$; $\mathbf{w}, -\mathbf{v}, -\mathbf{u}$; $-\mathbf{w}, \mathbf{v}, -\mathbf{u}$; $-\mathbf{w}, -\mathbf{v}, \mathbf{u}$; and so on.

6.5 The vector product

We now describe a method of multiplying two vectors to obtain another vector.

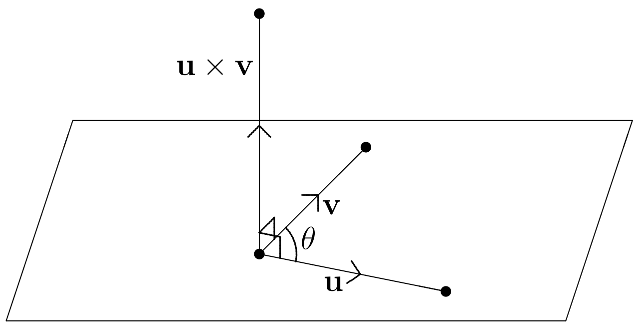
Definition 6.1. Suppose that $\mathbf{u}$, $\mathbf{v}$ are nonzero non-parallel vectors (of $\mathbb{R}^3$) at angle θ. Then the *vector product* or *cross product* $\mathbf{u} \times \mathbf{v}$ is defined to be the vector satisfying:

(i) $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}| \sin\theta$ (note that $\sin\theta > 0$ here, since $0 < \theta < \pi$);

(ii) $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$; and

(iii) $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} \times \mathbf{v}$ is a right-handed triple.

If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$, $\mathbf{v}$ are parallel, then the vector product $\mathbf{u} \times \mathbf{v}$ is defined to be $\mathbf{0}$. Thus $|\mathbf{u} \times \mathbf{v}| = 0 = |\mathbf{u}||\mathbf{v}| \sin\theta$ also holds when $\mathbf{u}$ and $\mathbf{v}$ are (nonzero and) parallel. A picture illustrating a typical cross product appears below.



Note. Occasionally, you will see the cross product $\mathbf{u} \times \mathbf{v}$ written as $\mathbf{u} \wedge \mathbf{v}$. However, this wedge product properly means something else, so you should not use such a notation.

We note that if $\mathbf{u}$, $\mathbf{v}$ are nonzero and non-parallel, then $|\mathbf{u} \times \mathbf{v}| > 0$ (so that $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$); otherwise $|\mathbf{u} \times \mathbf{v}| = 0$ (so that $\mathbf{u} \times \mathbf{v} = \mathbf{0}$). In particular, for all vectors $\mathbf{u}$, we have $\mathbf{u} \times \mathbf{u} = \mathbf{0}$. Further, we note that $\mathbf{i} \times \mathbf{j} = \mathbf{k}$, since:

(i) $|\mathbf{k}| = 1 = 1 \times 1 \times 1 = |\mathbf{i}||\mathbf{j}| \sin\frac{\pi}{2}$;

(ii) $\mathbf{k}$ is orthogonal to both $\mathbf{i}$ and $\mathbf{j}$; and

(iii) $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ is a right-handed triple.

Similarly, we get the following table of cross products:

	$\mathbf{i}$	$\mathbf{j}$	$\mathbf{k}$
$\mathbf{i}$	$\mathbf{0}$	$\mathbf{k}$	$-\mathbf{j}$
$\mathbf{j}$	$-\mathbf{k}$	$\mathbf{0}$	$\mathbf{i}$
$\mathbf{k}$	$\mathbf{j}$	$-\mathbf{i}$	$\mathbf{0}$

where the entry in the $\mathbf{u}$ row and $\mathbf{v}$ column is $\mathbf{u} \times \mathbf{v}$.

6.6 Properties of the vector product

For all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and for all scalars α, the following properties hold

1. $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$;
2. $(\alpha\mathbf{u}) \times \mathbf{v} = \alpha(\mathbf{u} \times \mathbf{v})$ and $\mathbf{u} \times (\alpha\mathbf{v}) = \alpha(\mathbf{u} \times \mathbf{v})$;
3. $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = \mathbf{v}\cdot(\mathbf{w} \times \mathbf{u}) = \mathbf{w}\cdot(\mathbf{u} \times \mathbf{v})$ and
   $\mathbf{u}\cdot(\mathbf{w} \times \mathbf{v}) = \mathbf{w}\cdot(\mathbf{v} \times \mathbf{u}) = \mathbf{v}\cdot(\mathbf{u} \times \mathbf{w}) = -[\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})]$;
4. $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v})\cdot\mathbf{w}$;
5. $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{w})$ and $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{v} \times \mathbf{w})$.

The proofs of these will be done below, with some other necessary results interspersed. We also note the result that the vector product is not associative, that is in general $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) \neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$, even though both sides of this are always defined. We leave it as an exercise to find an example of this non-equality. There are, however, triples $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ of vectors such that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$; for example, we have $\mathbf{i} \times (\mathbf{j} \times \mathbf{k}) = \mathbf{i} \times \mathbf{i} = \mathbf{0} = \mathbf{k} \times \mathbf{k} = (\mathbf{i} \times \mathbf{j}) \times \mathbf{k}$. Nor is the vector product commutative.

In the proofs below, we shall make use of certain facts without noting them explictly. One such fact is that collinearity is a symmetric relation. That is, $\mathbf{u}$ and $\mathbf{v}$ are collinear if and only if $\mathbf{v}$ and $\mathbf{u}$ are collinear.

Theorem 6.2. For all vectors $\mathbf{u}$ and $\mathbf{v}$ we have $\mathbf{v} \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$. This property is called *anti-commutativity.*

Note. The fact that the vector product is anti-commutative ***does not*** prove that it fails to be commutative. It could be that $\mathbf{u}\times\mathbf{v} = \mathbf{0}$ for all $\mathbf{u}$ and $\mathbf{v}$, or that $-(\mathbf{u}\times\mathbf{v}) = \mathbf{u}\times\mathbf{v}$ even when $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$. You must exhibit an ***explicit example*** to show that the vector product is not commutative, for example $\mathbf{i} \times \mathbf{j} = \mathbf{k} \neq -\mathbf{k} = \mathbf{j} \times \mathbf{i}$. (In some situations there may be more subtle ways to show a property does not always hold, but the explicit example still seems to be the best approach.)

Proof. If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$ and $\mathbf{v}$ are parallel (that is if $\mathbf{u}$ and $\mathbf{v}$ are collinear) then $\mathbf{v} \times \mathbf{u} = \mathbf{0} = -\mathbf{0} = -(\mathbf{u} \times \mathbf{v})$. Otherwise, we let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$, and let $\mathbf{w} = \mathbf{u} \times \mathbf{v}$. Now $|-\mathbf{w}| = |\mathbf{w}| = |\mathbf{u}||\mathbf{v}|\sin\theta = |\mathbf{v}||\mathbf{u}|\sin\theta$, $-\mathbf{w}$ is orthogonal to $\mathbf{u}$, $\mathbf{v}$ (since $\mathbf{w}$ is), and $\mathbf{v}$, $\mathbf{u}$, $-\mathbf{w}$ is a right-handed triple. Thus $\mathbf{v} \times \mathbf{u} = -\mathbf{w} = -(\mathbf{u} \times \mathbf{v})$. □

Theorem 6.3. For all vectors $\mathbf{u}$ and $\mathbf{v}$ and all scalars α we have $(\alpha\mathbf{u})\times\mathbf{v} = \alpha(\mathbf{u}\times\mathbf{v}) = \mathbf{u} \times (\alpha\mathbf{v})$.

Proof. We prove the first of these of equalities only. The proof of the other is similar and is left as an exercise (on Coursework 5). Alternatively, we can use properties we will have already established, for we have

$$\mathbf{u} \times (\alpha\mathbf{v}) = -((\alpha\mathbf{v}) \times \mathbf{u}) = -(\alpha(\mathbf{v} \times \mathbf{u})) = -(\alpha(-(\mathbf{u} \times \mathbf{v}))) = \alpha(\mathbf{u} \times \mathbf{v}).$$

If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$, $\mathbf{v}$ are parallel or $\alpha = 0$ then $(\alpha\mathbf{u}) \times \mathbf{v} = \mathbf{0} = \alpha(\mathbf{u} \times \mathbf{v})$. Otherwise (when $\alpha \neq 0$ and $\mathbf{u}, \mathbf{v}$ not collinear), we let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$, and let $\mathbf{w} = \mathbf{u} \times \mathbf{v}$. The angle between $\alpha\mathbf{u}$ and $\mathbf{v}$ is θ if $\alpha > 0$ and $\pi - \theta$ if $\alpha < 0$, and we have

$$|\alpha\mathbf{w}| = |\alpha||\mathbf{w}| = |\alpha||\mathbf{u}||\mathbf{v}|\sin\theta = |\alpha\mathbf{u}||\mathbf{v}|\sin\theta = |\alpha\mathbf{u}||\mathbf{v}|\sin(\pi - \theta).$$

Moreover, $\alpha\mathbf{w}$ is orthogonal to $\alpha\mathbf{u}$ and $\mathbf{v}$ (since $\mathbf{w}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$), and $\alpha\mathbf{u}$, $\mathbf{v}$, $\alpha\mathbf{w}$ is a right-handed triple (whether $\alpha > 0$ or $\alpha < 0$). Thus $(\alpha\mathbf{u}) \times \mathbf{v} = \alpha(\mathbf{u} \times \mathbf{v})$. □

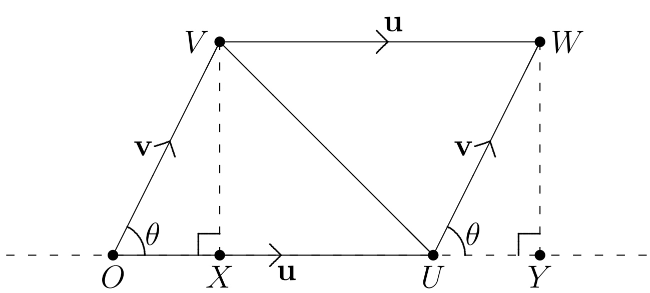
6.6.1 The area of a parallelogram and triangle

In this subsection, we prove formulae involving the vector product for areas of parallelograms and triangles having sides that represent the vectors $\mathbf{u}$ and $\mathbf{v}$.

Definition. Two geometric figures are said to be *congruent* if one can be obtained from the other by a combination of rotations, reflexions and translations. A figure is always congruent to itself, and two congruent figures always have the same area (if indeed they have an area at all: some really really weird bounded figures do not have an area). Two geometric figures are said to be *similar* if one can be obtained from the other by a combination of rotations, reflexions, translations, and scalings by nonzero amounts. Both congruence and similarity are equivalence relations on the set of geometric figures.

Theorem 6.4. Let $\overrightarrow{OU}$ and $\overrightarrow{OV}$ represent $\mathbf{u}$ and $\mathbf{v}$ respectively, and let W be the point making $OUWV$ into a parallelogram. Then the parallelogram $OUWV$ has area $|\mathbf{u} \times \mathbf{v}|$ and the triangle OUV has area $\frac{1}{2}|\mathbf{u} \times \mathbf{v}|$.

Proof. If $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$ and $\mathbf{v}$ are parallel then $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ and O, U, V, W all lie on the same line, so that the (degenerate) parallelogram $OUWV$ and triangle OUV both have area $0 = |\mathbf{u} \times \mathbf{v}| = \frac{1}{2}|\mathbf{u} \times \mathbf{v}|$. For the rest of the proof, we assume that $\mathbf{u}$ and $\mathbf{v}$ are not collinear, and for clarity, we also refer to the diagram below.



Otherwise, let ℓ be the line through O and U, and let X and Y be the points on ℓ with VX and WY perpendicular to ℓ. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Now triangles OXV and UYW are congruent, and thus have the same area, and so $OUWV$ has the same area as the rectangle $XYWV$, which is:

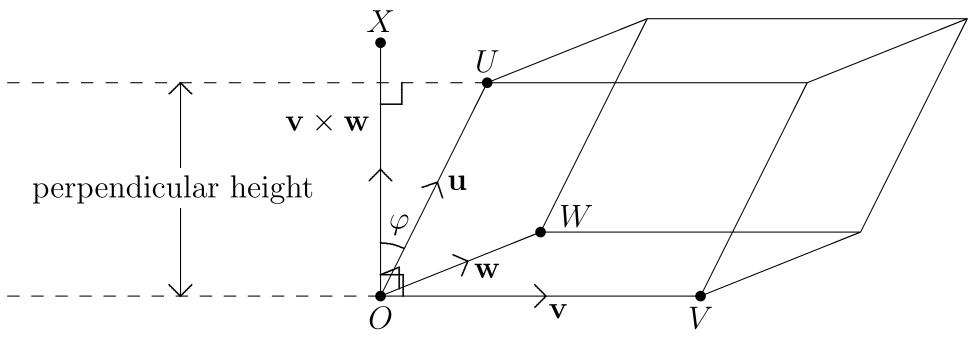
$$|\overrightarrow{VW}||\overrightarrow{VX}| = |\mathbf{u}|(|\mathbf{v}|\sin\theta) = |\mathbf{u} \times \mathbf{v}|.$$

The triangles OUV and WUV are congruent triangles covering the whole of the parallelogram $OUWV$ without overlap (except on the line UV of area 0). Thus OUV has half the area of $OUWV$, that is $\frac{1}{2}|\mathbf{u} \times \mathbf{v}|$. □

6.6.2 The triple scalar product

Definition. The *triple scalar product* of the ordered triple of vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$. Note that we have not defined the cross product of a scalar and a vector (either way round), and so $\mathbf{u} \times (\mathbf{v}\cdot\mathbf{w})$ and $(\mathbf{u}\cdot\mathbf{v}) \times \mathbf{w}$ do not exist.

Theorem 6.5. The volume of a *parallelepiped* with sides corresponding to $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ (as per the diagram below) is $|\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})|$.



Proof. Each face of a parallelepiped is a parallelogram, and here the base is a parallelogram with sides corresponding to $\mathbf{v}$ and $\mathbf{w}$. Thus the parallelepiped has volume

$$V = \text{area of base} \times \text{perpendicular height} = |\mathbf{v} \times \mathbf{w}| \times \text{perpendicular height},$$

by Theorem 6.4. If $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ or $\mathbf{u} = \mathbf{0}$ then $V = 0 = |\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})|$. Otherwise, the perpendicular height is $||\mathbf{u}| \cos\varphi| = |\mathbf{u}||\cos\varphi|$ where φ is the angle between $\mathbf{u}$ and $\mathbf{v}\times\mathbf{w}$, since $\mathbf{v} \times \mathbf{w}$ is orthogonal to the base. Thus

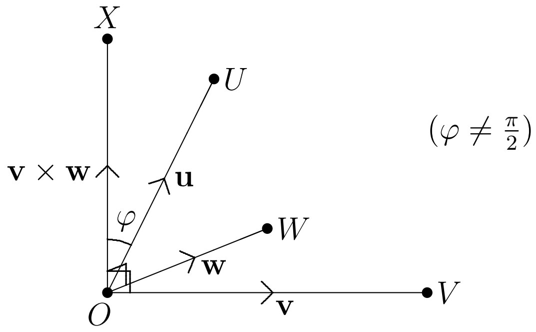
$$V = |\mathbf{v} \times \mathbf{w}||\mathbf{u}||\cos\varphi| = ||\mathbf{u}||\mathbf{v} \times \mathbf{w}| \cos\varphi| = |\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})|,$$

by the definition of the scalar product. □

Remark. The volume of the tetrahedron determined by $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ is $\frac{1}{6}|\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})|$. The volume is calculated as $\frac{1}{3} \times$ base area $\times$ perpendicular height. The tetrahedron has four vertices, namely O, U, V and W, and four triangular faces, which are OUV, OVW, OWU and UVW (see diagram above).

When is a triple scalar product 0? positive? negative?

We have $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = 0$ if and only if the volume of a parallelepiped with sides corresponding to $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is 0, which happens exactly when $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are coplanar. Otherwise, we consider the following diagram (on the next page), where $\varphi \neq \frac{\pi}{2}$, since we do not wish to have U in the plane determined by O, V and W.



Now $\mathbf{v} \times \mathbf{w}$ is orthogonal to the plane Π through O, V and W, and $\mathbf{v}$, $\mathbf{w}$, $\mathbf{v} \times \mathbf{w}$ is a right-handed triple. Also, $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = |\mathbf{u}||\mathbf{v} \times \mathbf{w}| \cos\varphi$, where φ is the angle between $\mathbf{u}$ and $\mathbf{v} \times \mathbf{w}$.

If $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) > 0$ then $0 \leqslant \varphi < \frac{\pi}{2}$, and so U is on the same side of Π as X. This implies that $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$ is a right-handed triple, and thus so is $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$.

If $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) < 0$ then $\frac{\pi}{2} < \varphi \leqslant \pi$, and so U is on the other side of Π to X. This implies that $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$ is a left-handed triple, and thus so is $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$.
Thus we conclude the following.

1. $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = 0$ if and only if $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are coplanar.

2. $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) > 0$ if and only if $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a right-handed triple.

3. $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) < 0$ if and only if $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a left-handed triple.

Theorem 6.6. For all vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ we have $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) = \mathbf{v}\cdot(\mathbf{w}\times\mathbf{u}) = \mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})$. Thus $\mathbf{u}\cdot(\mathbf{w} \times \mathbf{v}) = \mathbf{w}\cdot(\mathbf{v} \times \mathbf{u}) = \mathbf{v}\cdot(\mathbf{u} \times \mathbf{w}) = -[\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})]$ for all vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.

Proof. We prove the first line. The second line follows from the first, together with anti-commutativity of the cross product and various properties of the dot product.

In absolute terms, each of these triple products ($\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$, $\mathbf{v}\cdot(\mathbf{w}\times\mathbf{u})$ and $\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})$) gives the volume V of a parallelepiped with sides corresponding to $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.

If $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are coplanar then each triple product gives $V = 0$.

If $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a right-handed triple, then so are $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$ and $\mathbf{w}$, $\mathbf{u}$, $\mathbf{v}$, and each triple product gives $+V$.

If $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a left-handed triple, then so are $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$ and $\mathbf{w}$, $\mathbf{u}$, $\mathbf{v}$, and each triple product gives $-V$. □

Theorem 6.7. For all vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ we have $\mathbf{u}\cdot(\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v})\cdot\mathbf{w}$.

Proof. We have $(\mathbf{u} \times \mathbf{v})\cdot\mathbf{w} = \mathbf{w}\cdot(\mathbf{u} \times \mathbf{v}) = \mathbf{u}\cdot(\mathbf{v} \times \mathbf{w})$, using commutativity of the dot product followed by the previous theorem. □

6.6.3 The distributive laws for the vector product

Theorem 6.8. For all vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ we have $(\mathbf{u}+\mathbf{v})\times\mathbf{w} = (\mathbf{u}\times\mathbf{w}) + (\mathbf{v}\times\mathbf{w})$ and $\mathbf{u}\times(\mathbf{v}+\mathbf{w}) = (\mathbf{u}\times\mathbf{v}) + (\mathbf{u}\times\mathbf{w})$.

Proof. We prove just the first of these. Let $\mathbf{t}$ be any vector. Then we have

$$\begin{aligned}
\mathbf{t}\cdot((\mathbf{u}+\mathbf{v})\times\mathbf{w}) &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{w}\times\mathbf{t}) && \text{(by Theorem 6.6)}\\
&= \mathbf{u}\cdot(\mathbf{w}\times\mathbf{t}) + \mathbf{v}\cdot(\mathbf{w}\times\mathbf{t}) && \text{(by Distributive Law for } \cdot)\\
&= \mathbf{t}\cdot(\mathbf{u}\times\mathbf{w}) + \mathbf{t}\cdot(\mathbf{v}\times\mathbf{w}) && \text{(by Theorem 6.6)}
\end{aligned}$$

Therefore we have

$$\begin{aligned}
0 &= \mathbf{t}\cdot((\mathbf{u}+\mathbf{v})\times\mathbf{w}) - \mathbf{t}\cdot(\mathbf{u}\times\mathbf{w}) - \mathbf{t}\cdot(\mathbf{v}\times\mathbf{w})\\
&= \mathbf{t}\cdot(((\mathbf{u}+\mathbf{v})\times\mathbf{w}) - (\mathbf{u}\times\mathbf{w}) - (\mathbf{v}\times\mathbf{w}))
\end{aligned}$$

Let $\mathbf{s} = ((\mathbf{u}+\mathbf{v})\times\mathbf{w}) - (\mathbf{u}\times\mathbf{w}) - (\mathbf{v}\times\mathbf{w})$. Since $\mathbf{t}$ can be any vector we have $\mathbf{s}\cdot\mathbf{t} = \mathbf{t}\cdot\mathbf{s} = 0$ for every vector $\mathbf{t}$, and so by the Feedback Question (Part (c)) of Coursework 2, we must have $\mathbf{s} = \mathbf{0}$. Thus $(\mathbf{u}+\mathbf{v})\times\mathbf{w} = (\mathbf{u}\times\mathbf{w}) + (\mathbf{v}\times\mathbf{w})$, as required.

The proof of the second equality has a similar proof to the first, or it can be deduced from the first using the anti-commutative property of the vector product. □

6.7 The vector product in coördinates

We now use the rules for the vector product we have proved to find a formula for the vector product of two vectors given in coördinates.

Theorem 6.9. Let $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$. Then we have:

$$\mathbf{u}\times\mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

Proof. We have $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$ and $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$. Therefore, the distributive and scalar multiplication laws (Theorems 6.3 and 6.8) give us:

$$\begin{aligned}
\mathbf{u}\times\mathbf{v} &= (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}) \times (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k})\\
&= (u_1\mathbf{i}\times v_1\mathbf{i}) + (u_1\mathbf{i}\times v_2\mathbf{j}) + (u_1\mathbf{i}\times v_3\mathbf{k}) + (u_2\mathbf{j}\times v_1\mathbf{i})\\
&\qquad + (u_2\mathbf{j}\times v_2\mathbf{j}) + (u_2\mathbf{j}\times v_3\mathbf{k}) + (u_3\mathbf{k}\times v_1\mathbf{i}) + (u_3\mathbf{k}\times v_2\mathbf{j}) + (u_3\mathbf{k}\times v_3\mathbf{k})\\
&= u_1v_1(\mathbf{i}\times\mathbf{i}) + u_1v_2(\mathbf{i}\times\mathbf{j}) + u_1v_3(\mathbf{i}\times\mathbf{k}) + u_2v_1(\mathbf{j}\times\mathbf{i})\\
&\qquad + u_2v_2(\mathbf{j}\times\mathbf{j}) + u_2v_3(\mathbf{j}\times\mathbf{k}) + u_3v_1(\mathbf{k}\times\mathbf{i}) + u_3v_2(\mathbf{k}\times\mathbf{j}) + u_3v_3(\mathbf{k}\times\mathbf{k})\\
&= \mathbf{0} + u_1v_2\mathbf{k} + u_1v_3(-\mathbf{j}) + u_2v_1(-\mathbf{k}) + \mathbf{0} + u_2v_3\mathbf{i} + u_3v_1\mathbf{j} + u_3v_2(-\mathbf{i}) + \mathbf{0}\\
&= (u_2v_3 - u_3v_2)\mathbf{i} + (-u_1v_3 + u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}\\
&= (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k},
\end{aligned}$$

which is the result we wanted. □

There is a more useful way to remember this formula for $\mathbf{u} \times \mathbf{v}$. We define:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc,$$

which is the *determinant* of the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We have $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$. Then we have:

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} \mathbf{i} + \begin{vmatrix} u_3 & v_3 \\ u_1 & v_1 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \mathbf{k} = \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \mathbf{k}.$$

Example. Let $\mathbf{u} = \begin{pmatrix} 2 \\ -5 \\ 7 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} -3 \\ -1 \\ 4 \end{pmatrix}$. Then

$$\begin{aligned} \mathbf{u} \times \mathbf{v} &= \begin{vmatrix} -5 & -1 \\ 7 & 4 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 2 & -3 \\ 7 & 4 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 2 & -3 \\ -5 & -1 \end{vmatrix} \mathbf{k} \\ &= (-20 - (-7))\mathbf{i} - (8 - (-21))\mathbf{j} + (-2 - 15)\mathbf{k} = -13\mathbf{i} - 29\mathbf{j} - 17\mathbf{k}. \end{aligned}$$

In order to *check* (not *prove*) that we have calculated $\mathbf{u} \times \mathbf{v}$ correctly, we evaluate the scalar products $\mathbf{u} \cdot \mathbf{w}$ and $\mathbf{v} \cdot \mathbf{w}$, where we have calculated $\mathbf{u} \times \mathbf{v}$ to be $\mathbf{w}$. Both scalar products should be 0 since $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$.

Example. We find the volume of a parallelepiped with sides corresponding to the vectors $\mathbf{u} = \begin{pmatrix} 3 \\ 5 \\ -1 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} -5 \\ 3 \\ 2 \end{pmatrix}$. This volume is $|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|$. We have

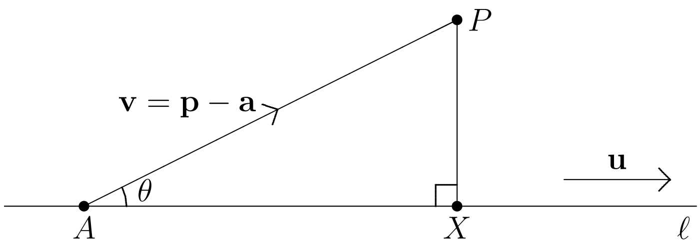
$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} 5 & 3 \\ 1 & 2 \end{vmatrix} \mathbf{i} - \begin{vmatrix} -1 & -5 \\ 1 & 2 \end{vmatrix} \mathbf{j} + \begin{vmatrix} -1 & -5 \\ 5 & 3 \end{vmatrix} \mathbf{k} = 7\mathbf{i} - 3\mathbf{j} + 22\mathbf{k}.$$

Thus $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = 3(7) + 5(-3) + (-1)(22) = 21 - 15 - 22 = -16$, and so the volume is $|-16| = 16$. Note that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is a left-handed triple, since $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) < 0$.

6.8 Applications of the vector product

6.8.1 Distance from a point to a line

Let P be a point, and let ℓ be a line with vector equation $\mathbf{r} = \mathbf{a} + \lambda \mathbf{u}$ (so that $\mathbf{u} \neq \mathbf{0}$), where $\mathbf{a}$ is the position vector of the point A lying on ℓ (see diagram).



Let $\mathbf{v}$ be the vector represented by $\overrightarrow{AP}$, so that $\mathbf{v} = \mathbf{p} - \mathbf{a}$, where $\mathbf{p}$ is the position vector of P. Suppose that P is not on ℓ, and let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. If X is a [or rather the] point on ℓ nearest to P then angle AXP is $\frac{\pi}{2}$, and the distance from P to ℓ is:

$$|\overrightarrow{PX}| = |\overrightarrow{AP}| \sin\theta = |\mathbf{v}| \sin\theta = \frac{|\mathbf{u} \times \mathbf{v}|}{|\mathbf{u}|} = \frac{|\mathbf{v} \times \mathbf{u}|}{|\mathbf{u}|} = \frac{|(\mathbf{p} - \mathbf{a}) \times \mathbf{u}|}{|\mathbf{u}|}.$$

Note that when P is on ℓ then $\mathbf{v}$ is a scalar multiple of $\mathbf{u}$, and $|\mathbf{u} \times \mathbf{v}|/|\mathbf{u}| = 0$, which is the correct distance from P to ℓ in this case also.

You should compare this to the result obtained in Section 3.5, taking due account of the different labelling used in that section. The result that $|\mathbf{a} \times \mathbf{b}| = \sqrt{|\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a}\cdot\mathbf{b})^2}$ for all vectors $\mathbf{a}$ and $\mathbf{b}$ (proof exercise) should also prove to be useful here.

Example. Find the distance from $P = (-3, 7, 4)$ to the line ℓ with vector equation $\mathbf{r} = \begin{pmatrix} 2 \\ -2 \\ -3 \end{pmatrix} + \lambda \begin{pmatrix} 4 \\ -5 \\ 3 \end{pmatrix}$. Here $\mathbf{a} = \begin{pmatrix} 2 \\ -2 \\ -3 \end{pmatrix}$, $A = (2, -2, -3)$ and $\mathbf{u} = \begin{pmatrix} 4 \\ -5 \\ 3 \end{pmatrix}$. So $\overrightarrow{AP}$ represents $\mathbf{v} = \mathbf{p} - \mathbf{a} = \begin{pmatrix} -3 \\ 7 \\ 4 \end{pmatrix} - \begin{pmatrix} 2 \\ -2 \\ -3 \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \\ 7 \end{pmatrix}$. Now

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} -5 & 9 \\ 3 & 7 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 4 & -5 \\ 3 & 7 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 4 & -5 \\ -5 & 9 \end{vmatrix} \mathbf{k} = \begin{pmatrix} -62 \\ -43 \\ 11 \end{pmatrix}.$$

Therefore $|\mathbf{u} \times \mathbf{v}| = \sqrt{(-62)^2 + (-43)^2 + 11^2} = \sqrt{3844 + 1849 + 121} = \sqrt{5814} = 3\sqrt{646}$ and $|\mathbf{u}| = \sqrt{4^2 + (-5)^2 + 3^2} = \sqrt{50} = 5\sqrt{2}$, and thus we conclude that the distance is

$$\frac{|\mathbf{u} \times \mathbf{v}|}{|\mathbf{u}|} = \frac{3\sqrt{646}}{5\sqrt{2}} = \frac{3\sqrt{323}}{5}.$$

(Note that we have $\mathbf{v} \times \mathbf{u} = (\mathbf{p} - \mathbf{a}) \times \mathbf{u} = -(\mathbf{u} \times \mathbf{v})$, and so $|\mathbf{v} \times \mathbf{u}| = |(\mathbf{p} - \mathbf{a}) \times \mathbf{u}| = |\mathbf{u} \times \mathbf{v}|$.)

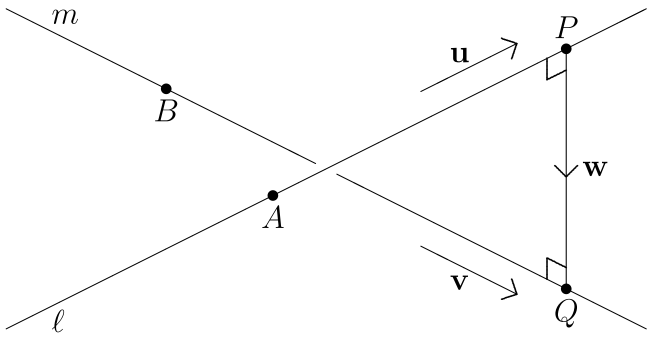
6.8.2 Distance between two lines

We derive a formula for the distance between lines ℓ and m having vector equations

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{u} \quad \text{and} \quad \mathbf{r} = \mathbf{b} + \mu\mathbf{v}$$

respectively. (By distance between ℓ and m, we mean the shortest distance from a point on ℓ to a point on m.) For the vector equations above to be valid, we require that $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$. If $\mathbf{u}$ and $\mathbf{v}$ are parallel, then this distance is the distance from P to m, where P is *any* point on ℓ.

So from now on, we assume that $\mathbf{u}$ and $\mathbf{v}$ are not parallel (and are nonzero). Let $\overrightarrow{PQ}$ be a shortest directed line segment from a point on ℓ to a point on m. This situation is shown in the diagram below (on the next page).



Then the vector $\mathbf{w}$ represented by $\overrightarrow{PQ}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, and so must be a scalar multiple of $\mathbf{u} \times \mathbf{v}$. Thus

$$\mathbf{w} = \mathbf{q} - \mathbf{p} = \alpha(\mathbf{u} \times \mathbf{v}),$$

for some scalar α, where $\mathbf{p}$ and $\mathbf{q}$ are (respectively) the position vectors of P and Q. Now P is on ℓ and so $\mathbf{p} = \mathbf{a} + \lambda\mathbf{u}$ for some λ, and Q is on m and so $\mathbf{q} = \mathbf{b} + \mu\mathbf{v}$ for some μ. Therefore, $\mathbf{q} - \mathbf{p} = \alpha(\mathbf{u} \times \mathbf{v}) = \mathbf{b} - \mathbf{a} + \mu\mathbf{v} - \lambda\mathbf{u}$, and thus:

$$\begin{aligned}
\alpha|\mathbf{u} \times \mathbf{v}|^2 &= \alpha(\mathbf{u} \times \mathbf{v})\cdot(\mathbf{u} \times \mathbf{v}) \\
&= (\mathbf{b} - \mathbf{a} + \mu\mathbf{v} - \lambda\mathbf{u})\cdot(\mathbf{u} \times \mathbf{v}) \\
&= (\mathbf{b} - \mathbf{a})\cdot(\mathbf{u} \times \mathbf{v}) + \mu(\mathbf{v}\cdot(\mathbf{u} \times \mathbf{v})) - \lambda(\mathbf{u}\cdot(\mathbf{u} \times \mathbf{v})) \\
&= (\mathbf{b} - \mathbf{a})\cdot(\mathbf{u} \times \mathbf{v}),
\end{aligned}$$

where the last step holds since $\mathbf{u}\cdot(\mathbf{u} \times \mathbf{v}) = \mathbf{v}\cdot(\mathbf{u} \times \mathbf{v}) = 0$. So we obtain

$$\alpha = \frac{(\mathbf{b} - \mathbf{a})\cdot(\mathbf{u} \times \mathbf{v})}{|\mathbf{u} \times \mathbf{v}|^2}.$$

Thus the length of $\overrightarrow{PQ}$ is

$$|\overrightarrow{PQ}| = |\mathbf{w}| = |\alpha||\mathbf{u} \times \mathbf{v}| = \frac{|(\mathbf{b} - \mathbf{a})\cdot(\mathbf{u} \times \mathbf{v})|}{|\mathbf{u} \times \mathbf{v}|},$$

and this is the distance between ℓ and m. This formula does not apply when $\mathbf{u}$ and $\mathbf{v}$ are parallel (it becomes $\frac{0}{0}$, which is not helpful).

Example. We calculate the distance between the lines ℓ and m having vector equations $\mathbf{r} = \mathbf{a} + \lambda\mathbf{u}$ and $\mathbf{r} = \mathbf{b} + \mu\mathbf{v}$ respectively, where

$$\mathbf{a} = \begin{pmatrix} 0 \\ 4 \\ -1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} 1 \\ -3 \\ -2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}.$$

We have $\mathbf{b} - \mathbf{a} = 2\mathbf{i} - 5\mathbf{j} + \mathbf{k}$ and

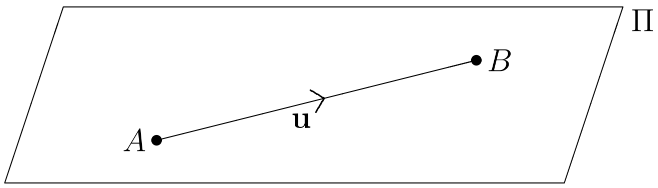
$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} 1 \\ -3 \\ -2 \end{pmatrix} \times \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix} = \begin{vmatrix} -3 & 1 \\ -2 & 2 \end{vmatrix}\mathbf{i} - \begin{vmatrix} 1 & -3 \\ -2 & 2 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 1 & -3 \\ -3 & 1 \end{vmatrix}\mathbf{k} = \begin{pmatrix} -4 \\ 4 \\ -8 \end{pmatrix}.$$

Thus we get $(\mathbf{b}-\mathbf{a})\cdot(\mathbf{u}\times\mathbf{v}) = -8-20-8 = -36$ and $|\mathbf{u}\times\mathbf{v}| = \sqrt{(-4)^2+4^2+(-8)^2} = \sqrt{16+16+64} = \sqrt{96} = 4\sqrt{6}$. Therefore the distance from ℓ to m is

$$\frac{|(\mathbf{b}-\mathbf{a})\cdot(\mathbf{u}\times\mathbf{v})|}{|\mathbf{u}\times\mathbf{v}|} = \frac{|-36|}{\sqrt{96}} = \frac{36}{4\sqrt{6}} = \frac{9}{\sqrt{6}} \left[= \frac{3\sqrt{3}}{\sqrt{2}} = \frac{3\sqrt{6}}{2}\right].$$

6.8.3 Equations of planes (revisited)

Let Π be a plane, and let $\mathbf{u}$ be a nonzero vector. We say that Π is *parallel* to $\mathbf{u}$ (or $\mathbf{u}$ is *parallel* to Π) if there are points A and B on Π such that $\overrightarrow{AB}$ represents $\mathbf{u}$.



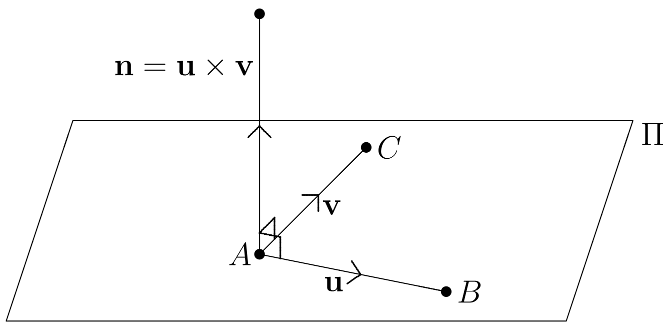
Two planes Π and Π' are *parallel* if every nonzero vector parallel to one is parallel to the other. Now suppose that $\mathbf{u}$ and $\mathbf{v}$ are nonzero non-parallel vectors. Then $\mathbf{n} := \mathbf{u}\times\mathbf{v}$ is a nonzero vector orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Thus if Π is a plane through the point P and parallel to both $\mathbf{u}$ and $\mathbf{v}$ then $\mathbf{n}[\neq \mathbf{0}]$ is orthogonal to Π and a vector equation for Π is

$$\mathbf{r}\cdot\mathbf{n} = \mathbf{p}\cdot\mathbf{n},$$

where $\mathbf{p}$ is the position vector of P. (In the diagram below, P can be taken to correspond to A.)

We are also interested in the equation of a plane determined by three points which do not all lie on the same line. So let A, B and C be three points which do not all lie on the same line, let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be their position vectors, and let Π be the unique plane containing A, B and C. One possible vector equation of Π is $\mathbf{r} = \mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})+\mu(\mathbf{c}-\mathbf{a})$, and this was given in Section 3.3. But we really want an equation of the form $\mathbf{r}\cdot\mathbf{n} = \mathbf{p}\cdot\mathbf{n}$, where $\mathbf{p}$ is the position vector of a point in Π and $\mathbf{n}$ is orthogonal to Π. So we can take $\mathbf{p} = \mathbf{a}$. But how can we determine a suitable normal vector $\mathbf{n}$?

We note that Π is parallel to the vectors $\mathbf{u} := \mathbf{b}-\mathbf{a}$ and $\mathbf{v} := \mathbf{c}-\mathbf{a}$ represented by $\overrightarrow{AB}$ and $\overrightarrow{AC}$ respectively. Then $\mathbf{u}$ and $\mathbf{v}$ are nonzero non-parallel vectors. Therefore $\mathbf{u}\times\mathbf{v}$ is a nonzero (proof exercise) vector perpendicular to $\mathbf{u}$ and $\mathbf{v}$, and thus orthogonal to Π, and so we can take $\mathbf{n} = \mathbf{u}\times\mathbf{v}$.



Finally, we find the plane determined by two points and a vector. So let A and B be points having position vectors $\mathbf{a}$ and $\mathbf{b}$ respectively. We now examine the plane Π containing points A and B and parallel to $\mathbf{v}$, where $\mathbf{u} := \mathbf{b} - \mathbf{a}$ and $\mathbf{v}$ are not collinear (which in particular means that $\mathbf{v} \neq \mathbf{0}$ and $A \neq B$). Then Π has vector equation $\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$ where $\mathbf{n} = \mathbf{u} \times \mathbf{v} = (\mathbf{b} - \mathbf{a}) \times \mathbf{v}$. (The diagram is relevant to this case too.)

Example 1. We determine a Cartesian equation for the plane Π through the point $A = (3, 4, 1)$ and parallel to $\mathbf{u} = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$. We have

$$\mathbf{n} = \mathbf{u} \times \mathbf{v} = \begin{vmatrix} 3 & 1 \\ -1 & 3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 1 & 2 \\ -1 & 3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \mathbf{k} = 10\mathbf{i} - 5\mathbf{j} - 5\mathbf{k} = \begin{pmatrix} 10 \\ -5 \\ -5 \end{pmatrix}.$$

A vector equation for Π is

$$\mathbf{r} \cdot \begin{pmatrix} 10 \\ -5 \\ -5 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 10 \\ -5 \\ -5 \end{pmatrix},$$

and so a Cartesian equation is $10x - 5y - 5z = 5$, which is equivalent to $2x - y - z = 1$ (after dividing through by 5).

Example 2. We determine a Cartesian equation for the plane Π through the points $A = (1, 2, 4)$, $B = (2, 4, 1)$ and $C = (4, 1, 2)$. Then Π is parallel to:

$$\mathbf{u} = \mathbf{b} - \mathbf{a} = \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} \text{ and } \mathbf{v} = \mathbf{c} - \mathbf{a} = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -2 \end{pmatrix},$$

and so Π is orthogonal to $\mathbf{n} = \mathbf{u} \times \mathbf{v} = -7\mathbf{i} - 7\mathbf{j} - 7\mathbf{k}$. Since A is also on Π a Cartesian equation for Π is $-7x - 7y - 7z = \mathbf{a} \cdot \mathbf{n} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \cdot \begin{pmatrix} -7 \\ -7 \\ -7 \end{pmatrix} = -49$, or equivalently $x + y + z = 7$. Since we can take $\mathbf{n}$ to be a nonzero scalar multiple of $\mathbf{u} \times \mathbf{v}$, we could have taken $\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$, which would have led directly to the Cartesian equation $x + y + z = 7$. (As a check, we can verify that A, B and C lie on the plane defined by this equation.)

6.9 Is the cross product commutative or associative?

We have already indicated that it is neither. For example, $\mathbf{i} \times \mathbf{j} = \mathbf{k} \neq -\mathbf{k} = \mathbf{j} \times \mathbf{i}$ and $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = -\mathbf{j} \neq \mathbf{0} = (\mathbf{i} \times \mathbf{i}) \times \mathbf{j}$. It is important to give *explicit* examples when the equalities fail to hold (or to somehow establish by stealth that such examples exist). In particular, we should ***not*** use the anti-commutative law (Theorem 6.2) to prove that the vector product is not commutative. This is because:

(i) It could be the case that $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ for all $\mathbf{u}$ and $\mathbf{v}$, in which case the vector product would be both commutative and anti-commutative. (It may seem obvious that the vector product is not always zero; the point is that this property must still be explicitly *checked* [once].)

(ii) Even if $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$, it could still be the case that $\mathbf{u} \times \mathbf{v}$ is equal to a vector $\mathbf{w}$ having the properties that $\mathbf{w} = -\mathbf{w}$ and $\mathbf{w} \neq \mathbf{0}$. (There are mathematical systems in which this can happen. However, $\mathbb{R}^3$ is not one of them, see Lemma 6.10 below.)

We shall classify all pairs $\mathbf{u}, \mathbf{v}$ of vectors such that $\mathbf{u} \times \mathbf{v} = \mathbf{v} \times \mathbf{u}$ and all triples $\mathbf{u}, \mathbf{v}, \mathbf{w}$ of vectors such that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$.

Lemma 6.10. Let $\mathbf{w}$ be a vector (in $\mathbb{R}^3$, or even $\mathbb{R}^n$) such that $\mathbf{w} = -\mathbf{w}$. Then $\mathbf{w} = \mathbf{0}$.

Proof. Adding $\mathbf{w}$ to both sides of $\mathbf{w} = -\mathbf{w}$ gives $2\mathbf{w} = \mathbf{0}$, and then multiplying both sides by $\frac{1}{2}$ gives $\mathbf{w} = \mathbf{0}$, as required. □

Theorem 6.11. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors. Then $\mathbf{u} \times \mathbf{v} = \mathbf{v} \times \mathbf{u}$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are collinear, which is if and only if $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.

Proof. If $\mathbf{u}$ and $\mathbf{v}$ are collinear then $\mathbf{u} \times \mathbf{v} = \mathbf{v} \times \mathbf{u} = \mathbf{0}$. If $\mathbf{u}$ and $\mathbf{v}$ are not collinear $\mathbf{u} \neq \mathbf{0}$, $\mathbf{v} \neq \mathbf{0}$, and the angle θ between $\mathbf{u}$ and $\mathbf{v}$ satisfies $0 < \theta < \pi$, and so $\sin\theta > 0$. Thus, from the definition of $\mathbf{u} \times \mathbf{v}$, we get $|\mathbf{u} \times \mathbf{v}| = |\mathbf{u}||\mathbf{v}| \sin\theta > 0$, whence $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$. Then the previous lemma (Lemma 6.10) gives us that $\mathbf{u} \times \mathbf{v} \neq -(\mathbf{u} \times \mathbf{v}) = \mathbf{v} \times \mathbf{u}$. □

6.9.1 The triple vector product(s)

Definition. The *triple vector products* of the ordered triple of vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are defined to be $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ and $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$.

Theorem 6.12. For all vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ we have $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u}\cdot\mathbf{w})\mathbf{v} - (\mathbf{u}\cdot\mathbf{v})\mathbf{w}$ and $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u}\cdot\mathbf{w})\mathbf{v} - (\mathbf{v}\cdot\mathbf{w})\mathbf{u}$

Proof. A somewhat tedious calculation using coördinates can prove this result, and this is left as an exercise for the reader. □

Some aspects of this result can be proved geometrically. We let $\mathbf{p} = \mathbf{v} \times \mathbf{w}$ and $\mathbf{q} = \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$, and suppose that $\mathbf{v}$ and $\mathbf{w}$ are not collinear. (It is annoying that the case when $\mathbf{v}$, $\mathbf{w}$ are collinear must be treated separately. But the first equality is fairly easy to prove in this case.) Since $\mathbf{p} = \mathbf{v} \times \mathbf{w}$ is orthogonal to $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{q} = \mathbf{u} \times \mathbf{p}$, we see that the point with position vector $\mathbf{q}$ must be in the plane determined by O, $\mathbf{v}$ and $\mathbf{w}$. That is $\mathbf{q} = \alpha\mathbf{v} + \beta\mathbf{w}$ for some scalars α and β. The requirement that $\mathbf{q}$ be orthogonal to $\mathbf{u}$ determines the ratio between α and β.

Theorem 6.13. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors. Then $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ if and only $(\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{v}\cdot\mathbf{w})\mathbf{u}$. The latter condition holds if and only if one (or both) of the following hold:

1. $\mathbf{u}$ and $\mathbf{w}$ are collinear; or
2. $\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{w}$.

Proof. We take the formulae for $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ and $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ given in Theorem 6.12. Thus $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ if and only if

$$(\mathbf{u}\cdot\mathbf{w})\mathbf{v} - (\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{u}\cdot\mathbf{w})\mathbf{v} - (\mathbf{v}\cdot\mathbf{w})\mathbf{u}. \tag{6.3}$$

We now perform the reversible operations of subtracting $(\mathbf{u}\cdot\mathbf{w})\mathbf{v}$ from both sides of (6.3) followed by negating both sides, to get that $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$ if and only if $(\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{v}\cdot\mathbf{w})\mathbf{u}$. If $\mathbf{u}$ and $\mathbf{w}$ are collinear, then $\mathbf{u} = \mathbf{0}$ or $\mathbf{w} = \mathbf{0}$ or $\mathbf{u} = \lambda\mathbf{w}$ for some λ. In the first two cases we get $(\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{v}\cdot\mathbf{w})\mathbf{u} = 0$, and in the last case we get $(\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{v}\cdot\mathbf{w})\mathbf{u} = (\lambda(\mathbf{u}\cdot\mathbf{v}))\mathbf{u}$. If $\mathbf{u}$ and $\mathbf{w}$ are not collinear, then $(\mathbf{u}\cdot\mathbf{v})\mathbf{w} = (\mathbf{v}\cdot\mathbf{w})\mathbf{u}$ if and only if $-(\mathbf{v}\cdot\mathbf{w})\mathbf{u} + (\mathbf{u}\cdot\mathbf{v})\mathbf{w} = \mathbf{0}$, which happens if and only if $\mathbf{v}\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{v} = 0$ by the definition of collinear, see Equation 6.1 of Section 6.2. But $\mathbf{v}\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{v} = 0$ if and only if $\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{w}$. □

Chapter 7

Matrices

Definition. An $m \times n$ *matrix* is an array of numbers set out in m rows and n columns, where $m, n \in \mathbb{N} = \{0, 1, 2, 3, \ldots\}$. (See Section 7.9 for the case $m = 0$ or $n = 0$.)

Examples. 1. $\begin{pmatrix} 1 & -1 & 5 \\ 2 & 0 & 6 \end{pmatrix}$ has 2 rows and 3 columns, and so it is a 2×3 matrix.

2. $\begin{pmatrix} 1 & 0 \\ 7 & -1 \\ \sqrt{2} & 3 \\ 3 & 1 \end{pmatrix}$ is a 4×2 matrix.

3. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ is a 3×3 matrix.

4. A vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is a 3×1 matrix.

The word *matrix* was introduced into mathematics in 1850 by James Joseph Sylvester (1814–1897), though the idea of writing out coefficients of equations as rectangular arrays of numbers dates back to antiquity (there are Chinese examples from about 200 BC), and Carl Friedrich Gauß (1777–1855) used this notation in his work on simultaneous equations. Sylvester and Arthur Cayley (1821–1895) developed the theory of matrices we use today, and which we shall investigate in this chapter.

We write $A = (a_{ij})_{m \times n}$ to mean that A is an $m \times n$ matrix whose (i, j)-entry is a_{ij}, that is, a_{ij} is in the i-th row and j-th column of A. For an $m \times n$ matrix A, we write A_{ij} or $A_{i,j}$ for the (i, j)-entry of A; thus if $A = (a_{ij})_{m \times n}$ then $A_{ij} = a_{ij}$. If $A = (a_{ij})_{m \times n}$ then

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

and we say that A has *size* $m \times n$. An $n \times n$ matrix is called a *square* matrix.

Example. Let
$$A = (a_{ij})_{2\times 2} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 4 & 3 \end{pmatrix}.$$
Then A is a square matrix of size 2×2. The $(1,2)$-entry of A is $a_{12} = -1$, and the $(2,2)$-entry of A is $a_{22} = 3$.

Example. We write out in full $A = (a_{ij})_{3\times 2}$ with $a_{ij} = i(i+j)$.
$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix},$$
with
$$\begin{array}{ll} a_{11} = 1(1+1) = 2, & a_{12} = 1(1+2) = 3, \\ a_{21} = 2(2+1) = 6, & a_{22} = 2(2+2) = 8, \\ a_{31} = 3(3+1) = 12, & a_{32} = 3(3+2) = 15, \end{array}$$
and so
$$A = \begin{pmatrix} 2 & 3 \\ 6 & 8 \\ 12 & 15 \end{pmatrix}.$$

Definition. Matrices A and B are *equal* if they have the same size and the same (i,j)-entry for every possible value of i and j. That is, if $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{p\times q}$, then A and B are equal if and only if $p = m$, $q = n$ and $a_{ij} = b_{ij}$ for all i, j with $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$.

Notation. 0_{mn} is the $m \times n$ matrix with every entry 0. The matrix 0_{mn} may also be denoted by $0_{m,n}$ or $0_{m\times n}$, especially in cases of ambiguity. (For example, does 0_{234} mean $0_{23\times 4}$ or $0_{2\times 34}$?) We call $0_{mn} = 0_{m,n} = 0_{m\times n}$ the *zero* $m \times n$ matrix.
For example, $0_{23} = 0_{2,3} = 0_{2\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is the zero 2×3 matrix.

Notation. I_n is the $n\times n$ matrix with $(1,1)$-entry $=$ $(2,2)$-entry $= \cdots =$ (n,n)-entry $= 1$ and all other entries 0. We call I_n the *identity* $n \times n$ matrix. Note that I_n is always a square matrix. For example, we have:
$$\mathrm{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathrm{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Another way to express this definition is that I_n is the unique $n \times n$ matrix having all its (top-left to bottom-right) diagonal entries equal to 1 and all other entries 0. Note that by the *diagonal* of a square matrix, we always mean its top-left to bottom-right diagonal; we are not generally interested in the bottom-left to top-right diagonal.

7.1 Addition of matrices

If $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$ then $A + B$ is defined to be the $m \times n$ matrix whose (i, j)-entry is $a_{ij} + b_{ij}$.

Note. We can only add matrices of the same size.

If $A = (a_{ij})_{m\times n}$, the *negative* of A, written $-A$, is defined to be the $m \times n$ matrix whose (i, j)-entry is $-a_{ij}$. Thus $-A = D$, where $D = (d_{ij})_{m\times n}$ with $d_{ij} = -a_{ij}$.

Examples. We have

$$\begin{pmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 2 \\ -2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 3 & -1 \\ 2 & -4 & 9 \end{pmatrix}$$

and

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

7.2 Rules for matrix addition

The next theorem states that matrix addition is associative and commutative. Furthermore, the is an identity element for matrix addition, and each matrix has an inverse under the operation of matrix addition.

Theorem 7.1. Let A, B and C be $m \times n$ matrices. Then

(i) $(A + B) + C = A + (B + C)$,

(ii) $A + B = B + A$,

(iii) $A + 0_{mn} = 0_{mn} + A = A$, and

(iv) $A + (-A) = (-A) + A = 0_{mn}[= 0_{m\times n}]$.

Proof. (i): Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{m\times n}$ and $C = (c_{ij})_{m\times n}$. Then $A + B$ is an $m \times n$ matrix whose (i, j)-entry is $a_{ij} + b_{ij}$ and so $(A + B) + C$ is an $m \times n$ matrix whose (i, j)-entry is $(a_{ij} + b_{ij}) + c_{ij}$.

Similarly, $B + C$ is an $m \times n$ matrix whose (i, j)-entry is $b_{ij} + c_{ij}$ and so $A + (B + C)$ is an $m \times n$ matrix whose (i, j)-entry is $a_{ij} + (b_{ij} + c_{ij})$. The result follows, since

$$(a_{ij} + b_{ij}) + c_{ij} \;=\; a_{ij} + (b_{ij} + c_{ij})$$

by the associative law for addition of real numbers.

Proofs of (ii), (iii) and (iv): exercises for you. □

7.3 Scalar multiplication of matrices

Let $A = (a_{ij})_{m\times n}$ and let α be a scalar (that is, a real number). Then αA is defined to be the $m \times n$ matrix whose (i, j)-entry is αa_{ij}. Thus $\alpha A := B$ where $B = (b_{ij})_{m\times n}$ with $b_{ij} = \alpha a_{ij}$ for all i and j. Note that $(-1)A = -A$.

Examples. We have

$$3\begin{pmatrix} -1 & 3 \\ 2 & 5 \\ -7 & 6 \end{pmatrix} = \begin{pmatrix} -3 & 9 \\ 6 & 15 \\ -21 & 18 \end{pmatrix}$$

and

$$(-2)\begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix} + 4\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 4 \\ -6 & -8 \end{pmatrix} + \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 8 \\ -2 & -8 \end{pmatrix}.$$

7.4 Rules for scalar multiplication

Theorem 7.2. Let A and B be $m \times n$ matrices and let α and β be scalars. Then:

(i) $\alpha(A + B) = \alpha A + \alpha B$,

(ii) $(\alpha + \beta)A = \alpha A + \beta A$,

(iii) $\alpha(\beta A) = (\alpha\beta)A$,

(iv) $1A = A$, and

(v) $0A = 0_{mn}[= 0_{m\times n}]$.

Proof. (i): Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{m\times n}$. Then $A + B$ is an $m \times n$ matrix with (i, j)-entry $a_{ij} + b_{ij}$ and $\alpha(A+B)$ is an $m\times n$ matrix with (i, j)-entry $\alpha(a_{ij} + b_{ij}) = \alpha a_{ij} + \alpha b_{ij}$. But αA is an $m \times n$ matrix with (i, j)-entry αa_{ij} and αB is an $m \times n$ matrix with (i, j)-entry αb_{ij}, and so $\alpha A + \alpha B$ is an $m \times n$ matrix with (i, j)-entry $\alpha a_{ij} + \alpha b_{ij}$.

(iii): Let $A = (a_{ij})_{m\times n}$. Then βA is an $m \times n$ matrix with (i, j)-entry βa_{ij} and so $\alpha(\beta A)$ is an $m \times n$ matrix with (i, j)-entry $\alpha(\beta a_{ij}) = (\alpha\beta)a_{ij}$. But $(\alpha\beta)A$ is also an $m \times n$ matrix with (i, j)-entry $(\alpha\beta)a_{ij}$.

Proofs of (ii), (iv), (v): exercises for you. □

7.5 Matrix multiplication

This is rather more interesting and more complicated than scalar multiplication. It will become clearer later why we define matrix multiplication in the way that we do.

Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$. Then the *product* AB is the $m \times p$ matrix with (i, j)-entry

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

Thus $AB = C$, where $C = (c_{ij})_{m\times p}$ with

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

Observe that to get the (i,j)-entry of AB, we focus on the i-th row of A and the j-th column of B:

$$\left(\boxed{a_{i1} \quad a_{i2} \quad \ldots \quad a_{in}}\right)\left(\boxed{\begin{matrix} b_{ij} \\ b_{2j} \\ \vdots \\ b_{nj} \end{matrix}}\right),$$

and we form their dot product, that is the sum $a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{1n}b_{nj}$.

Note. The matrix product AB of matrices A and B is defined if and only if the number of columns of A is the same as the number of rows of B. In the case when $A = (a_{ij})_{m\times n}$ and $B = (a_{ij})_{n\times p}$ we note that A has n columns and B has n rows.

Example 1. Let

$$A = \begin{pmatrix} 1 & -1 \\ 3 & 7 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 9 & 5 \\ 6 & 4 & 8 \end{pmatrix}.$$

Now A is a 2×2 matrix and B is a 2×3 matrix. They can be multiplied since the number of columns of A equals the number of rows of B. The product AB will be a 2×3 matrix as it will have the same number of rows as A and the same number of columns as B. Note that BA is *not* defined since the number of columns (3) of B is not equal to the number of rows (2) of A.

To get the $(1,3)$-entry of AB for example, we pick out the 1st row of A and the 3rd column of B:

$$\begin{pmatrix} \boxed{1 \quad -1} \\ * \quad\quad * \end{pmatrix}\begin{pmatrix} * & * & \boxed{\begin{matrix} 5 \\ 8 \end{matrix}} \\ * & * & \end{pmatrix},$$

and we form their dot product, the sum $1\times 5 + (-1)\times 8 = -3$. In full:

$$AB = \begin{pmatrix} 1\times 2 + (-1)\times 6 & 1\times 9 + (-1)\times 4 & 1\times 5 + (-1)\times 8 \\ 3\times 2 + 7\times 6 & 3\times 9 + 7\times 4 & 3\times 5 + 7\times 8 \end{pmatrix} = \begin{pmatrix} -4 & 5 & -3 \\ 48 & 55 & 71 \end{pmatrix}.$$

Example 2. We have

$$\begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1\times 1 + 0\times(-2) + (-2)\times 3 \\ 2\times 1 + 1\times(-2) + 4\times 3 \end{pmatrix} = \begin{pmatrix} -5 \\ 12 \end{pmatrix}.$$

Just as in the first example, the product of the matrices in the reverse order is not defined.

Example 3. We have

$$\begin{pmatrix} -6 & 4 \\ -9 & 6 \end{pmatrix} \begin{pmatrix} 6 & -4 \\ 9 & -6 \end{pmatrix} = \begin{pmatrix} 6 & -4 \\ 9 & -6 \end{pmatrix} \begin{pmatrix} -6 & 4 \\ -9 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0_{22} = 0_{2\times 2}.$$

Example 4. Let $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then

$$AB = \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} \quad \text{and} \quad BA = \begin{pmatrix} 3 & 2 \\ 7 & 4 \end{pmatrix},$$

and so $AB \neq BA$.

Note. In general, the case $AB \neq BA$ is more common than the case $AB = BA$. In any case, matrix multiplication is not commutative. Also AB can be defined without BA being defined, and vice-versa. Moreover, AB and BA need not have the same size, for if we let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times m}$ then AB is an $m \times m$ matrix while BA is an $n \times n$ matrix, so that AB and BA have different sizes when $m \neq n$.

7.6 Rules for matrix multiplication

Theorem 7.3. Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{m\times n}$ $X = (x_{ij})_{n\times p}$ and $Y = (y_{ij})_{n\times p}$. Then:

(i) $(A + B)X = AX + BX$ and $A(X + Y) = AX + AY$; and

(ii) $\alpha(AX) = (\alpha A)X = A(\alpha X)$ for every scalar α.

Now let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{n\times p}$ and $C = (c_{ij})_{p\times q}$. Then:

(iii) $(AB)C = A(BC)$; and

(iv) $\mathrm{I}_m A = A\mathrm{I}_n = A$.

Proof. (i): Let $C = A + B$. Then $C = (c_{ij})_{m\times n}$ with $c_{ij} = a_{ij} + b_{ij}$. We have that $(A + B)X = CX$ is an $m \times p$ matrix, with (i, j)-entry

$$\begin{aligned} \textstyle\sum_{k=1}^{n} c_{ik}x_{kj} &= c_{i1}x_{1j} + c_{i2}x_{2j} + \cdots + c_{in}x_{nj} \\ &= (a_{i1} + b_{i1})x_{1j} + (a_{i2} + b_{i2})x_{2j} + \cdots + (a_{in} + b_{in})x_{nj} \\ &= (a_{i1}x_{1j} + a_{i2}x_{2j} + \cdots + a_{in}x_{nj}) + (b_{i1}x_{1j} + b_{i2}x_{2j} + \cdots b_{in}x_{nj}) \\ &= ((i, j)\text{-entry of } AX) + ((i, j)\text{-entry of } BX) \end{aligned}$$

Since, in addition, $AX + BX$ has the same size, $m \times p$, as $(A + B)X$, we conclude that $(A + B)X = AX + BX$. The proof that $A(X + Y) = AX + AY$ is similar, and is left as an exercise, as is the proof of (ii).

(iii): Let $AB = X$, where $X = (x_{ij})_{m\times p}$, and let $BC = Y$, where $Y = (y_{ij})_{n\times q}$. Then $(AB)C = XC$ is an $m \times q$ matrix with (i,j)-entry

$$x_{i1}c_{1j} + x_{i2}c_{2j} + \cdots + x_{ip}c_{pj}.$$

But

$$\begin{aligned} x_{i1} &= a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1},\\ x_{i2} &= a_{i1}b_{12} + a_{i2}b_{22} + \cdots + a_{in}b_{n2},\\ &\cdots\\ x_{ip} &= a_{i1}b_{1p} + a_{i2}b_{2p} + \cdots + a_{in}b_{np}. \end{aligned}$$

Thus the (i,j)-entry of $(AB)C$ is

$$(a_{i1}b_{11} + \cdots + a_{in}b_{n1})c_{1j} + (a_{i1}b_{12} + \cdots + a_{in}b_{n2})c_{2j} + \cdots + (a_{i1}b_{1p} + \cdots + a_{in}b_{np})c_{pj}.$$

Multiplying out all brackets, we get that the (i,j)-entry of $(AB)C$ is

$$\sum_{s=1}^{p}\sum_{r=1}^{n}(a_{ir}b_{rs})c_{sj},$$

which is to say the sum of all terms $(a_{ir}b_{rs})c_{sj}$ as r varies over $1, 2, \ldots, n$ and s varies over $1, 2, \ldots, p$. (There are no issues with how the terms are ordered or grouped since addition of real numbers is both associative and commutative.)
A similar calculation shows that the (i,j)-entry of the $m \times q$ matrix $AY = A(BC)$ is

$$\sum_{r=1}^{n}\sum_{s=1}^{p}a_{ir}(b_{rs}c_{sj}).$$

But by the associative law for multiplying real numbers we have

$$(a_{ir}b_{rs})c_{sj} = a_{ir}(b_{rs}c_{sj})$$

for all i, j, r and s. Thus $(AB)C$ and $A(BC)$ are both $m \times q$ matrices and their (i,j)-entries are the same for all possible i and j. Hence $(AB)C = A(BC)$.

(iv): We prove that $\mathrm{I}_m A = A$, and leave the proof that $A\mathrm{I}_n = A$ as an exercise. Now I_m is an $m \times m$ matrix and so $\mathrm{I}_m A$ is an $m \times n$ matrix. The i-th row of I_m has zero entries everywhere except for the (i,i)-entry, which is 1. Thus $\mathrm{I}_m A$ has (i,j)-entry

$$0a_{1j} + \cdots + 0a_{i-1,j} + 1a_{ij} + 0a_{i+1,j} + \cdots + 0a_{mj} = a_{ij},$$

which is also the (i,j)-entry of A. □

Note. Since $(AB)C = A(BC)$ we can just write ABC for this product. In fact, matrix multiplication is associative in the strong sense that if one of $(AB)C$ and $A(BC)$ exists, then so does the other, and they are equal. But matrix multiplication is not commutative: we have seen examples where AB exists but BA does not, and also examples where both AB and BA exist and $AB = BA$ (Example 3) and $AB \neq BA$ (Example 4). There is also no anti-commutativity property for matrix multiplication.

Example. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} -1 & -1 \\ 2 & 3 \end{pmatrix}.$$

We compute $(AB)C$ and $A(BC)$ and check that they are the same:

$$(AB)C = \left(\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\right)\begin{pmatrix} -1 & -1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix};$$

and

$$A(BC) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ 2 & 3 \end{pmatrix}\right) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -3 & -4 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}.$$

Note. There is a different matrix product, known as the *Hadamard product* or *pointwise product*, which takes two $m \times n$ matrices A and B and produces the $m \times n$ matrix $A \circ B$, where $(A \circ B)_{ij} = A_{ij}B_{ij}$ for all i and j. While this alternative matrix product has some mathematical utility, it is nowhere near as useful as the standard (and more complicated) matrix product. I leave it you to discern what properties hold for the operation $\circ$.

There also yet another type of matrix product, known as the *Kronecker product* or *tensor product*, which can be applied to any two matrices. I am not going to tell you how to define this product, but the interested reader can easily find this information on the World Wide Web.

7.7 Some useful notation

For an $n \times n$ matrix A we define $A^0 = \mathrm{I}_n$, $A^1 = A$, $A^2 = AA$, $A^3 = AAA$, and so on. As a consequence of the (generalised) associative law for matrix multiplication we have $A^pA^q = A^{p+q}$ and $(A^p)^q = A^{pq}$ for all $p, q \in \mathbb{N} = \{0, 1, 2, \ldots\}$. Note that A^pA^q means $(A^p)(A^q)$. For $m \times n$ matrices B and C we write $B - C$ for $B + (-C)$.

7.8 Inverses of matrices

Every nonzero real number α has a multiplicative inverse, that is, if $0 \neq \alpha \in \mathbb{R}$ there exists $\beta \in \mathbb{R}$ such that $\alpha\beta = \beta\alpha = 1$. What about matrices?

Definition 7.4. An $n \times n$ matrix A is said to be *invertible* if there is some $n \times n$ matrix B such that $AB = BA = \mathrm{I}_n$.

Example 1. We have

$$\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{I}_2$$

and

$$\begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{I}_2.$$

Thus $A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ is invertible (and so is $B = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix}$ by the same calculation).

Example 2. The matrix $\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ is not invertible, because no matter what entries we take for the matrix $B = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ we get

$$\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} B = \begin{pmatrix} p+2r & q+2s \\ 0 & 0 \end{pmatrix} \neq \mathrm{I}_2 \quad \text{and} \quad B\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} p & 2p \\ r & 2r \end{pmatrix} \neq \mathrm{I}_2.$$

Theorem 7.5. Let A be an $n \times n$ matrix and let B and C be $n \times n$ matrices such that $AB = BA = \mathrm{I}_n$ and $AC = CA = \mathrm{I}_n$. Then $B = C$.

Proof. We have $B = B\mathrm{I}_n = B(AC) = (BA)C = \mathrm{I}_n C = C$. □

Definition. If A is an invertible $n \times n$ matrix then the unique matrix B such that $AB = BA = \mathrm{I}_n$ is called the *inverse* of A. We denote the inverse of A by A^{-1}. For $n \in \mathbb{N}^+$ we define $A^{-n} := (A^{-1})^n$, the n-th power of the inverse of A. (This definition is consistent for $n = 1$.)

Note. If A is invertible, then $AA^{-1} = A^{-1}A = \mathrm{I}_n$, and so by the definitions of invertible and inverse, we see that A^{-1} is invertible and $(A^{-1})^{-1} = A$. If A is invertible then $A^pA^q = A^{p+q}$ and $(A^p)^q = A^{pq}$ for all $p, q \in \mathbb{Z}$. In particular, $(A^m)^{-1} = A^{-m}$ for all $m \in \mathbb{Z}$. Also, I_n is always invertible (we have ${\mathrm{I}_n}^{-1} = \mathrm{I}_n$), as is any nonzero scalar multiple of I_n. We have ${\mathrm{I}_n}^m = \mathrm{I}_n$ for all $n \in \mathbb{N}$ and $m \in \mathbb{Z}$.

Fact. If A and B are $n \times n$ matrices with $AB = \mathrm{I}_n$, then $BA = \mathrm{I}_n$. Thus if a square matrix has a 1-sided "inverse" it is actually invertible. Now let $m > n$ (the case $m < n$ is similar), and let A be an $m \times n$ matrix and B an $n \times m$ matrix. Then AB is an $m \times m$ matrix, and can never be I_m, and BA is an $n \times n$ matrix, and sometimes we can have $BA = \mathrm{I}_n$.

Theorem 7.6. If A, B are invertible $n \times n$ matrices then AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. Using the associative law of matrix multiplication implicitly throughout, we have:

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A\mathrm{I}_nA^{-1} = AA^{-1} = \mathrm{I}_n$$

and

$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}\mathrm{I}_nB = B^{-1}B = \mathrm{I}_n.$$ □

7.9 Matrices with no rows or no columns

Somewhat surprisingly, it makes sense mathematically to consider matrices having no rows or no columns (or both). Since such matrices have no entries then (vacuously) all entries in such matrices are equal to 0. Thus such matrices are $0_{0\times m}$ or $0_{m\times 0}$ for some $m \in \mathbb{N} = \{0, 1, 2, \ldots, \}$. We also have $0_{0\times 0} = \mathrm{I}_0$. Among these matrices, only I_0 is invertible, with ${\mathrm{I}_0}^{-1} = \mathrm{I}_0$. The following relations hold (which are the only possible products involving these matrices):

1. $0_{m\times 0}0_{0\times n} = 0_{m\times n}$, with mn entries;
2. $0_{0\times m}0_{m\times 0} = 0_{0\times 0} = \mathrm{I}_0$;
3. $0_{0\times m}A = 0_{0\times n}$, where $A = (a_{ij})_{m\times n}$; and
4. $B0_{n\times 0} = 0_{m\times 0}$, where $B = (b_{ij})_{m\times n}$.

Note that in the case $0_{m\times 0}0_{0\times n} = 0_{m\times n}$, with $m, n \geqslant 1$, we have taken the product of two matrices having no entries to produce a matrix with a nonzero (mn) number of entries (all of which are 0).

7.10 Transposes of matrices

Let A be an $m \times n$ matrix. The *tranpose* of A is the $n \times m$ matrix denoted A^T, where $A^\mathsf{T} = (a^\mathsf{T}_{ij})_{n\times m}$ with $a^\mathsf{T}_{ij} = a_{ji}$ for all i and j where $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$. Thus we see that transposition interchanges the rôles of rows and columns in a matrix. To transpose a matrix A, we read off the rows of A and write them down as the columns of A^T.

Examples. We have ${\mathrm{I}_n}^\mathsf{T} = \mathrm{I}_n$ for all n. Also, we have

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^\mathsf{T} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 3 & 5 \\ 7 & 11 & 13 \\ 17 & 19 & 23 \end{pmatrix}^\mathsf{T} = \begin{pmatrix} 2 & 7 & 17 \\ 3 & 11 & 19 \\ 5 & 13 & 23 \end{pmatrix}.$$

The properties of the transposition operator are summarised below.

Theorem 7.7. Let $A = (a_{ij})_{m\times n}$, $B = (a_{ij})_{m\times n}$ and $C = (c_{ij})_{n\times p}$. Then:

(i) $(A^\mathsf{T})^\mathsf{T} = A$;

(ii) $(A + B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}$;

(iii) $(-A)^\mathsf{T} = -(A^\mathsf{T})$;

(iv) $(\lambda A)^\mathsf{T} = \lambda(A^\mathsf{T})$ for all scalars λ;

(v) $(AC)^\mathsf{T} = C^\mathsf{T} A^\mathsf{T}$; and

(vi) if A is an invertible $n \times n$ matrix, then so is A^T, and $(A^\mathsf{T})^{-1} = (A^{-1})^\mathsf{T}$.

Proof. (i), (ii), (iii) and (iv): Exercises for the reader.

(v): Firstly, we note that A, C, AC, $(AC)^\mathsf{T}$, C^T, A^T and $C^\mathsf{T} A^\mathsf{T}$ have sizes $m \times n$, $n \times p$, $m \times p$, $p \times m$, $p \times n$, $n \times m$ and $p \times m$ respectively, so that $(AC)^\mathsf{T}$ and $C^\mathsf{T} A^\mathsf{T}$ both have the same size, namely $p \times m$. Now we calculate, for $1 \leqslant i \leqslant p$ and $1 \leqslant j \leqslant m$, that:

$$(i,j)\text{-entry of } (AC)^\mathsf{T} = (j,i)\text{-entry of } AC = \sum_{k=1}^{n} a_{jk} c_{ki}.$$

Moreover, we have

$$(i,j)\text{-entry of } C^\mathsf{T} A^\mathsf{T} = \sum_{k=1}^{n} c^\mathsf{T}_{ik} a^\mathsf{T}_{kj} = \sum_{k=1}^{n} c_{ki} a_{jk} = \sum_{k=1}^{n} a_{jk} c_{ki}.$$

Since $(AC)^\mathsf{T}$ and $C^\mathsf{T} A^\mathsf{T}$ have the same size and identical (i,j)-entries for all i and j they are equal.

(vi): Using Part (v), we see that $(A^{-1})^\mathsf{T} A^\mathsf{T} = (AA^{-1})^\mathsf{T} = \mathrm{I}_n{}^\mathsf{T} = \mathrm{I}_n$ and $A^\mathsf{T}(A^{-1})^\mathsf{T} = (A^{-1}A)^\mathsf{T} = \mathrm{I}_n{}^\mathsf{T} = \mathrm{I}_n$. Thus A^T is invertible, with inverse $(A^{-1})^\mathsf{T}$. □

Note. In view of the above theorem, we write $-A^\mathsf{T}$ instead of $(-A)^\mathsf{T}$ or $-(A^\mathsf{T})$; λA^T instead of $(\lambda A)^\mathsf{T}$ or $\lambda(A^\mathsf{T})$; and $A^{-\mathsf{T}}$ instead of $(A^\mathsf{T})^{-1}$ or $(A^{-1})^\mathsf{T}$. Also, if A and B are invertible $n \times n$ matrices then $(AB)^{-\mathsf{T}} = A^{-\mathsf{T}} B^{-\mathsf{T}}$.

Chapter 8

Determinants

The founder of the theory of determinants is usually taken to be Gottfried Wilhelm Leibniz (1646–1716), who also shares the credit for inventing calculus with Sir Isaac Newton (1643–1727)[1]. But the idea of 2×2 determinants goes back at least to the Chinese around 200 BC. The word *determinant* itself was first used in its present sense in 1812 by Augustin-Louis Cauchy (1789–1857), and he developed much of the general theory we know today.

8.1 Inverses of 2×2 matrices, and determinants

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix. Recall that the determinant of A is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = ad - cb.$$

We also denote this determinant by $\det(A)$ or $\det A$ or $|A|$. (Just as with other operators, such as cos, we avoid using the brackets except to resolve ambiguity. Thus we prefer $\cos\theta$ and $\det A$ in preference to $\cos(\theta)$ and $\det(A)$, and $\det A^{\mathsf{T}}$ means $\det(A^{\mathsf{T}})$.)

Examples. Let $A = \begin{pmatrix} 2 & 3 \\ 2 & -1 \end{pmatrix}$. Then $\det A = -2 - 6 = -8$. Also, we have

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = -2, \quad \det \mathrm{I}_2 = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \quad \text{and} \quad \begin{vmatrix} 2 & -1 \\ -4 & 2 \end{vmatrix} = 4 - 4 = 0.$$

Lemma 8.1. Let A be any 2×2 matrix. Then $\det A^{\mathsf{T}} = \det A$.

Proof. If $A^{\mathsf{T}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $A^{\mathsf{T}} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$, and so $\det A^{\mathsf{T}} = ad - cb = \det A$. □

[1]His dates by the Gregorian calendar are 4th January 1643 – 31st March 1727. But the Julian calendar was in use in England at the time, and his Julian dates are 25th December 1642 – 20th March 1727. At the time, the year in England started on 25th March, so that legally he died in 1726.

Now let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and let $\Delta := \det A = ad - bc$. We note that if $\Delta \neq 0$ and

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

then

$$AB = \begin{pmatrix} (ad-bc)/\Delta & (-ab+ba)/\Delta \\ (cd-dc)/\Delta & (-cb+da)/\Delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$BA = \begin{pmatrix} (da-bc)/\Delta & (db-bd)/\Delta \\ (-ca+ac)/\Delta & (-cb+ad)/\Delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so that A is invertible and $A^{-1} = B$. Thus if A is a 2×2 matrix and $\det A \neq 0$, then A is invertible. What if $\det A = 0$? We shall show that in this case A is *not* invertible, but first we show the following.

Theorem 8.2. If A and B are 2×2 matrices then $\det(AB) = \det(A)\det(B)$.

Notation. Our conventions for bracketing the determinant operator are similar to those used for trigonometric functions. Thus $\det AB$ means $\det(AB)$, even though $(\det A)B$ makes sense, and $\det A \det B$ means $\det(A)\det(B) = (\det A)(\det B)$, though $\det(A \det B)$ makes less sense since we write the scalar first when notating scalar multiplication of matrices. Similarly, if λ is a scalar then $\det \lambda A$ means $\det(\lambda A)$. Note that while notations like $\det 3A$ and $\det A \det B$ are often used, we should *never* write things like $\det -A$ or $\det -2A$ or $\det A + B$. Also, $\det AB$ is seldom seen.

Proof. We prove this by direct computation. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $B = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$. Then $AB = \begin{pmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{pmatrix}$ and

$$\begin{aligned} \det(AB) &= (ap+br)(cq+ds) - (cp+dr)(aq+bs) \\ &= apcq + apds + brcq + brds - cpaq - cpbs - draq - drbs \\ &= apds + brcq - cpbs - draq \\ &= (ad-cb)(ps-rq) \\ &= \det A \det B. \end{aligned}$$

□

Example. Let $A = \begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix}$. Then

$$\det(AB) = \begin{vmatrix} -6 & 5 \\ 14 & 0 \end{vmatrix} = 0 - 70 = -70,$$

which is the same as $\det A \det B = (3-(-2))(-2-12) = 5 \times (-14) = -70$.

Theorem 8.3. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2-matrix. Then A is invertible if and only if $\det A \neq 0$, and when $\Delta := \det A \neq 0$, then $A^{-1} = \begin{pmatrix} d/\Delta & -b/\Delta \\ -c/\Delta & a/\Delta \end{pmatrix}$.

Proof. We have already seen that when $\Delta = \det A \neq 0$ then A is invertible with inverse as above. It therefore just remains to show that when $\Delta = 0$ then A is not invertible. So suppose that $\det A = 0$. Then $\det(AB) = \det A \det B = 0$ for every 2×2 matrix B. But since $\det \mathrm{I}_2 = 1$ this means there cannot exist a matrix B with $AB = \mathrm{I}_2$. Hence A is not invertible. $\square$

Examples. 1. Let $A = \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}$. Then $\det A = 0$, so A is not invertible.

2. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then $\det A = -2 \neq 0$, so A is invertible, and

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}.$$

Let us check this:

$$AA^{-1} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{I}_2,$$

$$A^{-1}A = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{I}_2.$$

We conclude this section by introducing the *adjugate* of a 2×2 matrix. This might not seem to be a terribly exciting or useful concept for 2×2 matrices, but is far more useful for 3×3 matrices, and in general for $n \times n$ matrices ($n \in \mathbb{N}^+$). The adjugate of a square matrix A is denoted $\operatorname{adj} A$ (or $\operatorname{adj}(A)$), and for 2×2 matrices is defined by

$$\operatorname{adj} \begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Direct calculation shows that we have $A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)\mathrm{I}_2$ for all 2×2 matrices A. Thus if $\det A \neq 0$ (precisely the conditions needed for A to be invertible) we have $A^{-1} = \frac{1}{\det A}(\operatorname{adj} A)$. These relations generalise to $n \times n$ matrices for all $n \geqslant 1$.

We emphasise that the adjugate of a matrix is not the same as its adjoint, and you must be careful not to confuse the two terms. (Unfortunately, what we now call the adjugate has in the past been termed the adjoint!) The *adjoint* of a matrix over $\mathbb{C}$ is defined to the transpose of its complex conjugate, and is denoted A^* or $A^\dagger$, though sometimes A^* simply means the complex conjugate of A. Thus the adjoint of matrix over $\mathbb{R}$ is the same as its transpose. (If you have not yet met complex numbers and the complex conjugate, you will have done so by the time you take your first year exams.)

8.2 Determinants of 3×3 matrices

Consider the 3×3 matrix

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}.$$

We define the *determinant* of A by:

$$\det A = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} := \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$$

where

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad \text{and} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

Thus

$$\begin{aligned} \det A &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \cdot \left(\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \times \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \right) \\ &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \cdot \left(\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \mathbf{k} \right) \\ &= a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1). \end{aligned}$$

Example. Let

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

Then

$$\begin{aligned} \det A &= 3 \begin{vmatrix} 0 & -3 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} - 2 \begin{vmatrix} 2 & -1 \\ 0 & -3 \end{vmatrix} \\ &= 3(3) - 2(3) - 2(-6) = 15. \end{aligned}$$

Remark. Notice that the right-hand side of our equation

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1)$$

can be rearranged to give $a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3)$. Thus we can also write:

$$\begin{aligned} \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} &= a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ &= a_1(b_2c_3 - b_3c_2) - b_1(a_2c_3 - a_3c_2) + c_1(a_2b_3 - a_3b_2) \\ &= a_1(b_2c_3 - c_2b_3) - b_1(a_2c_3 - c_2a_3) + c_1(a_2b_3 - b_2a_3). \end{aligned}$$

This formula ("expanding along the first row") is often used to define a 3×3 determinant (even by me!), and is more frequently seen than the triple scalar product definition ("expanding along the first column") used in this course. The reason for our seemingly strange approach is that we can easily prove many properties of 3×3 determinants using results we have already established for triple scalar products (see Theorem 8.5). Note that it also follows from this rearrangement that

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

In other words, the determinant of a 3×3 matrix A is equal to the determinant of its transpose A^{T}, a result we record as Theorem 8.4.

Another helpful rearrangement of the terms in $\det A$ is to group those terms with positive and negative coefficients together, to get

$$\det A = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1.$$

The products having positive coefficient correspond to diagonals going in the direction top-left to bottom-right in the diagram below.

$$\begin{array}{cc|ccc|cc} & & a_1 & a_2 & a_3 & & \\ & b_3 & b_1 & b_2 & b_3 & b_1 & \\ c_2 & c_3 & c_1 & c_2 & c_3 & c_1 & c_2 \end{array}$$

The top-right to bottom-left 'diagonals' in the above diagram correspond to those products with negative coefficient.

Theorem 8.4. For all 3×3 matrices A we have $\det A^{\mathsf{T}} = \det A$.

Proof. See above. □

Theorem 8.5. Let $A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$. Then

(i) Interchanging two columns of A negates $\det A$. For example

$$\begin{vmatrix} b_1 & a_1 & c_1 \\ b_2 & a_2 & c_2 \\ b_3 & a_3 & c_3 \end{vmatrix} = - \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = -\det A.$$

(ii) Multiplying any column of A by λ multiplies $\det A$ by λ.

(iii) Adding a scalar multiple of any column of A to any other column does not change the determinant.

Proof. These all follow very easily from properties we have already proved for the scalar triple product.

(i): This follows at once from Theorem 6.6, which gives $\det A = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = -\mathbf{a}\cdot(\mathbf{c}\times\mathbf{b}) = -\mathbf{b}\cdot(\mathbf{a}\times\mathbf{c}) = -\mathbf{c}\cdot(\mathbf{b}\times\mathbf{a})$.

(ii): We have $\lambda\det A = \lambda(\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})) = (\lambda\mathbf{a})\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{a}\cdot((\lambda\mathbf{b})\times\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times(\lambda\mathbf{c}))$, by Theorem 6.3 and Section 3.2.

(iii): Let B be the matrix obtained by adding λ times the second column to the first column of A. Then standard results on the dot and cross product, such as $\mathbf{b}\cdot(\mathbf{b}\times\mathbf{c}) = 0$, give us $\det B = (\mathbf{a}+\lambda\mathbf{b})\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) + \lambda\mathbf{b}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \det A$. The other cases are similar. □

Example. We can use these properties to compute determinants quickly.

$$\begin{vmatrix} 6 & 10 & 13 \\ 2 & 20 & 4 \\ -1 & -20 & -2 \end{vmatrix} = 10\begin{vmatrix} 6 & 1 & 13 \\ 2 & 2 & 4 \\ -1 & -2 & -2 \end{vmatrix} = 10\begin{vmatrix} 6 & 1 & 1 \\ 2 & 2 & 0 \\ -1 & -2 & 0 \end{vmatrix} = 10\begin{vmatrix} 0 & 1 & 1 \\ 2 & 2 & 0 \\ -1 & -2 & 0 \end{vmatrix}$$
$$= 10\left(0\begin{vmatrix} 2 & 0 \\ -2 & 0 \end{vmatrix} - 2\begin{vmatrix} 1 & 1 \\ -2 & 0 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ 2 & 0 \end{vmatrix}\right) = 10(0(0) - 2(2) - 1(-2)) = -20.$$

Here to get the first equality we have applied Part (ii) of the theorem to take out a factor of 10 from the second column, to get the second equality we have applied Part (iii) of the theorem to subtract twice the first column from the third column, and for the 3rd equality we have applied Part (iii) to subtract 6 times the third column from the first column. (Note that we can use the theorem to simplify the matrix even more, so that we end up computing $\det \mathrm{I}_3$, which is 1.)

Remark. Since $\det A^{\mathsf{T}} = \det A$ for all 3×3 matrices A, Theorem 8.5 holds with the word 'column' replaced by 'row' throughout. Thus we can just as well use row operations as column operations when simplifying determinants.

The rest of this section is presented without proof. In particular, the next theorem is rather difficult to prove. (The one after that is not too hard.)

Theorem 8.6. For all 3×3 matrices A and B we have $\det(AB) = \det A \det B$.

Proof. In theory, we can expand $\det(AB)$, which has $3!\times 3^3 = 6\times 27 = 162$ terms, perform all the cancellations to be left with $(3!)^2 = 6^2 = 36$ terms, and observe that this is $\det A\det B$. This is not very satisfactory, and new ideas are needed, which will also extend to general $n\times n$ determinants. However, this is beyond the scope of this course, though I may write up the proof as an appendix in the hope that some of you might understand it after having completed MTH4104: Introduction to Algebra. □

In view of the above theorem, it is evident that the 3×3 matrix A is not invertible if $\det A = 0$, since $\det \mathrm{I}_3 = 1$. It turns out that A is invertible whenever $\det A \neq 0$. But first, we define the adjugate, adj A, of a 3×3 matrix A. Let A_{ij} be the 2×2 matrix obtained by removing the i-th row and j-th column from A. (Note that this notation conflicts with what we used earlier.) Then

$$\operatorname{adj} A := \begin{pmatrix} \det A_{11} & -\det A_{21} & \det A_{31} \\ -\det A_{12} & \det A_{22} & -\det A_{32} \\ \det A_{13} & -\det A_{23} & \det A_{33} \end{pmatrix}.$$

Theorem 8.7. For all 3×3 matrices A we have $A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)\mathrm{I}_3$. Thus if $\det A \neq 0$ then A is invertible, and we have $A^{-1} = \frac{1}{\det A}(\operatorname{adj} A)$.

8.3 Systems of linear equations as matrix equations

We can write any system of m simultaneous linear equations in n unknowns $x_1, x_2, \ldots, x_n$ as a matrix equation

$$A\mathbf{x} = \mathbf{d} \tag{8.1}$$

where A is an $m \times n$ matrix, $\mathbf{x}$ is an $n \times 1$ matrix and $\mathbf{d}$ is an $m \times 1$ matrix, with the entries of $A = (a_{ij})_{m \times n}$ and $\mathbf{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_m \end{pmatrix}$ being known and the entries of $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ being the unknowns. The linear equations are $a_{j1}x_1 + \cdots + a_{jn}x_n = d_j$ for $1 \leqslant j \leqslant m$.

I commented previously that properly echelon form is a property of matrices. We define the matrix A to be *echelon form* if the corresponding system $A\mathbf{x} = \mathbf{d}$ of linear equations is in echelon form. So A is in echelon form (Geometry I version) if each nonzero row of A commences with strictly fewer 0s than those below it. (This forces all zero rows to occur at the 'bottom' of A.) We do not insist (for Geometry I) that the first nonzero entry in a nonzero row be 1.

If A is a square $n \times n$ matrix (so now $m = n$) such that $\det A \neq 0$ then the matrix A has a (unique) inverse A^{-1}, and by multiplying both sides of the matrix equation (8.1) by A^{-1} we see that

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{d}.$$

Since $A^{-1}A\mathbf{x} = \mathrm{I}_n\mathbf{x} = \mathbf{x}$ we deduce that

$$\mathbf{x} = A^{-1}\mathbf{d}$$

is a solution of the simultaneous equations, and indeed that it is the *unique* solution.

8.4 Determinants and inverses of $n \times n$ matrices

Notation. Throughout this section, A_{ij} will denote the $(m-1)\times(n-1)$ matrix obtained from $m \times n$ matrix A by deleting the i-th row and j-column. The notation A_{ij} only makes sense when $m, n \geqslant 1$. The (i, j)-entries of the matrices A, $\tilde{A}$ and B shall be denoted a_{ij}, $\tilde{a}_{ij}$ and b_{ij} respectively.

8.4.1 Determinants of $n \times n$ matrices

In general, if $A = (a_{ij})_{n\times n}$ is an $n \times n$ matrix the determinant of A is a sum of $n! = 1\cdot 2\cdot \cdots \cdot n$ numbers (summands), where each summand is the product of n entries of A multiplied by a sign ($+1$ or -1). Also, in each summand, the n entries of A used in the product come from distinct rows and columns. `<UndefinedConcepts>` Using concepts you will meet in MTH4104: Introduction to Algebra, we have

$$\det A := \sum_{\sigma\in \mathrm{S}_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}, \tag{8.2}$$

where S_n denotes the set (group) of all $n!$ permutations of $\{1, 2, \ldots, n\}$ and $\operatorname{sgn}\sigma$, the *sign* or *parity* of σ, is $+1$ if this is even and -1 if this is odd. `</UndefinedConcepts>` From the above formula, we see that the 0×0 matrix I_0 has determinant 1, and the 1×1 matrix $\begin{pmatrix} a \end{pmatrix}$ has determinant a. (The notation $\begin{vmatrix} a \end{vmatrix}$ should not be used for the determinant of the $\begin{pmatrix} a \end{pmatrix}$ because of its confusion with the absolute value function $|a|$.) The following table summarises $n \times n$ determinants for small n.

n	#terms	determinant of A
0	1	1
1	1	a_{11}
2	2	$a_{11}a_{22} - a_{12}a_{21}$
3	6	$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$
4	24	$a_{11}a_{22}a_{33}a_{44} + \cdots$
5	120	$a_{11}a_{22}a_{33}a_{44}a_{55} + \cdots$

For $n \in \mathbb{N}$ we can define the $n \times n$ determinant recursively (that is in terms of smaller determinants) as follows. Firstly, to gives us a foundation, we define the 0×0 determinant to be 1. Then for $n \geqslant 1$ we can calculate (or even define) $\det A$ as

$$\begin{aligned}\det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + a_{13}\det A_{13} - \cdots + (-1)^{n+1} a_{1n}\det A_{1n}\\ &= \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det A_{1k}.\end{aligned}$$

Note that this definition agrees with the formulae above for $n = 1$, 2 and 3 (as it should). Instead of expanding along the first row (as above), we can expand along (down) the first column to get

$$\det A = \sum_{k=1}^{n} (-1)^{k+1} a_{k1} \det A_{k1}.$$

In fact, we can expand along arbitrary rows and columns. For all i and j, we have

$$\det A = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det A_{ik} = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A_{kj}.$$

It is possible to use the formula of (8.2) in order to show the following.

Theorem 8.8. For all $n \in \mathbb{N}$ and all $n \times n$ matrices A, we have $\det A^{\mathsf{T}} = \det A$.

Theorem 8.9. For all $n \in \mathbb{N}$, all $n \times n$ matrices A and all scalars λ, the following hold.

1. If B is obtained from A by swapping two of its rows, then $\det B = -\det A$.
2. If B is obtained from A by multiplying one of its rows by λ, then $\det B = \lambda \det A$. Thus $\det \lambda A = \lambda^n \det A$.
3. If B is obtained from A by adding λ times one of its rows to another row, then $\det B = \det A$.

Theorem 8.10. Let A, $\tilde{A}$ and B be $n \times n$ matrices differing only in the j-th row, such that $b_{jk} = a_{jk} + \tilde{a}_{jk}$ for all k. Then $\det B = \det A + \det \tilde{A}$.

In view of Theorem 8.8, Theorem 8.9 holds with the word 'row' replaced by 'column', as does Theorem 8.10 with the understanding that now $b_{kj} = a_{kj} + \tilde{a}_{kj}$ for all k. Note that if $n \geqslant 2$ and A and B are $n \times n$ matrices then $\det(A+B) \neq \det A + \det B$ in general (but not universally). Another consequence of the above theorems is that if two rows (or columns) of A are equal, or even scalar multiples of each other, then $\det A = 0$. The following is far trickier to prove.

Theorem 8.11. For all $n \in \mathbb{N}$ and all $n \times n$ matrices A and B, we have $\det(AB) = \det A \det B$.

8.4.2 Adjugates and inverses of $n \times n$ matrices

The 0×0 matrix I_0 has inverse I_0. From now on we let $n \geqslant 1$. If $A = \begin{pmatrix} a \end{pmatrix}$ then $\operatorname{adj} A = \begin{pmatrix} 1 \end{pmatrix}$, and if $a = \det A \neq 0$ then A is invertible and $A^{-1} = \begin{pmatrix} \frac{1}{a} \end{pmatrix}$, and if $a = \det A = 0$ then A is not invertible. We have already seen the adjugate and inverse of a 2×2 and a 3×3 matrix.

In general, we let $A = (a_{ij})_{n \times n}$ for some $n \geqslant 1$, and let A_{ij} be the result of removing the i-th row and j-th column from A. The (i, j)-*minor* of A is $m_{ij} := \det A_{ij}$. The *cofactor* matrix of A is $C = (c_{ij})_{n \times n}$ where $c_{ij} := (-1)^{i+j} m_{ij} = (-1)^{i+j} \det A_{ij}$. The *adjugate* of A is the transpose of the cofactor matrix. That is $\operatorname{adj} A := C^{\mathsf{T}}$, and so $\operatorname{adj} A$ has (i, j)-entry $(-1)^{i+j} \det A_{ji}$.

The following two theorems hold, and show one way (not always very efficient) to compute A^{-1} in the case when A is invertible.

Theorem 8.12. For all $n \geqslant 1$ we have $A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)\mathrm{I}_n$ for all $n \times n$ matrices A.

Theorem 8.13. For all $n \geqslant 1$, the $n \times n$ matrix A is invertible if and only if $\det A \neq 0$, and if $\det A \neq 0$ we have $A^{-1} = \frac{1}{\det A}(\operatorname{adj} A)$.

Chapter 9

Linear Transformations

9.1 The vector space $\mathbb{R}^n$

As we saw in Chapter 2, once we have chosen an origin and unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, we can assign a *position vector* $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ to each point in 3-space. From now on we shall *represent* this position vector by the *column vector* of coefficients of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, that is to say the column vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, though we shall continue to represent the *points* of 3-space by triples (*row* vectors) (x, y, z).

For any given $n \in \mathbb{N}$, we let $\mathbb{R}^n$ denote the set of all column n-vectors ($n \times 1$ matrices):

$$\mathbb{R}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} : a_1, a_2, \ldots, a_n \in \mathbb{R} \right\}.$$

The same notation $\mathbb{R}^n$ is also often used to denote the set of all n-tuples (*row* vectors) $(a_1, a_2, \ldots, a_n)$, but as most of our computations from now on will involve column vectors, for the rest of this module we shall reserve the notation $\mathbb{R}^n$ for these. For the special case when $n = 0$, we note that $\mathbb{R}^0$ has just a single vector (which is necessarily the zero vector). We denote by $\mathbf{0}_n$ the column vector which has n entries, all of which are 0; this is the *zero vector* of $\mathbb{R}^n$.

Addition and scalar multiplication of vectors in $\mathbb{R}^n$ is componentwise, entirely analogous to the situation for $\mathbb{R}^3$ as detailed in Section 2.1, and also for matrices, as given in Sections 7.1 and 7.3. For all $\lambda \in \mathbb{R}$ we have

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} \quad \text{and} \quad \lambda \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

From the properties of addition and multiplication by scalars that we have already proved for matrices, we observe that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbb{R}$ we have the following ten properties.

(a.0) $\mathbf{u} + \mathbf{v} \in \mathbb{R}^n$.
(a.1) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
(a.2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
(a.3) There exists $\mathbf{e} \in \mathbb{R}^n$ such that $\mathbf{v} + \mathbf{e} = \mathbf{v} = \mathbf{e} + \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$. [$\mathbf{e} = \mathbf{0}_n$.]
(a.4) For every $\mathbf{v} \in \mathbb{R}^n$ there is a $-\mathbf{v} \in \mathbb{R}^n$ such that $(-\mathbf{v}) + \mathbf{v} = \mathbf{v} + (-\mathbf{v}) = \mathbf{e}$.
(m.0) $\alpha\mathbf{v} \in \mathbb{R}^n$.
(m.1) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$.
(m.2) $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$.
(m.3) $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$.
(m.4) $1\mathbf{v} = \mathbf{v}$.

This shows that $\mathbb{R}^n$ satisfies the rules to be an algebraic structure called a *vector space* (over $\mathbb{R}$). These are studied in detail in the module MTH5112: Linear Algebra I. You will come across many other examples of vector spaces, for example the set of all $m \times n$ matrices for a given m and n, as well as some more exotic examples such as the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

Properties (a.0)–(a.4) are for vector addition, while properties (m.0)–(m.4) are those for scalar multiplication. We call (a.0) and (m.0) *closure* properties, with the axioms of a vector space regarded as being (a.1)–(a.4) and (m.1)–(m.4).[1] I have told you the names of Properties (a.1)–(a.4), (m.1) and (m.2) before. Can you remember them? The symbol $\mathbf{e}$ in (a.3) and (a.4) is usually wriiten as $\mathbf{0}$ in this context.[2] Note also that the element $\mathbf{e}$ of (a.3) is unique, and that for any $\mathbf{v}$, the element $-\mathbf{v}$ of (a.4) is unique.

Vector subtraction is defined as you might expect (see earlier chapters), namely that $\mathbf{u} - \mathbf{v}$ is defined to be $\mathbf{u} + (-\mathbf{v})$. It is also possible to define the division of a vector by a nonzero scalar by $\mathbf{v}/\lambda := (1/\lambda)\mathbf{v}$ where $\mathbf{v}$ is a vector and $\lambda \neq 0$ is a scalar. However, this notation is rather ugly and should be avoided.

9.2 Linear transformations

Definition 9.1. A function $t : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$ we have:

(i) $t(\mathbf{u} + \mathbf{v}) = t(\mathbf{u}) + t(\mathbf{v})$, and
(ii) $t(\alpha\mathbf{u}) = \alpha t(\mathbf{u})$.

If $m = n$ we call t a *linear transformation of* $\mathbb{R}^n$. Linear transformations are also called *linear maps*.

---

[1] Any vector space must by definition also satisfy (a.0) and (m.0), but the definition of a vector space is usually stated in such a way that they are not regarded as axioms.

[2] Apparently the letter 'e' comes from a German word, but nobody seems to be quite sure which one. Possibilities include *Eins* meaning *1 (the digit)*, *Einheit* meaning *unit(y)*, or a compound word such as *Einselement* or *Einheitselement*.

Example 1. Let $t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function

$$t\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ -b \end{pmatrix}.$$

If $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, and $\alpha \in \mathbb{R}$, we have

$$t(\mathbf{u}+\mathbf{v}) = t\begin{pmatrix} u_1+v_1 \\ u_2+v_2 \end{pmatrix} = \begin{pmatrix} u_1+v_1 \\ -u_2-v_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ -u_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ -v_2 \end{pmatrix} = t(\mathbf{u}) + t(\mathbf{v})$$

and

$$t(\alpha\mathbf{u}) = t\begin{pmatrix} \alpha u_1 \\ u\alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha u_1 \\ -\alpha u_2 \end{pmatrix} = \alpha\begin{pmatrix} u_1 \\ -u_2 \end{pmatrix} = \alpha t(\mathbf{u}).$$

For each point (a, b) in the plane, t maps its position vector $\begin{pmatrix} a \\ b \end{pmatrix}$ to $\begin{pmatrix} a \\ -b \end{pmatrix}$, the position vector of $(a, -b)$. Geometrically, t is a reflexion in the x-axis.

Example 2. The function $t : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

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

is *not* a linear transformation, since, for example,

$$t\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = t\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

but

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

9.3 Properties of linear transformations

Theorem 9.2. Let $t : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. then for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for all scalars α, β we have:

(i) $t(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha t(\mathbf{u}) + \beta t(\mathbf{v})$;

(ii) $t(\mathbf{0}_n) = \mathbf{0}_m$;

(iii) $t(-\mathbf{u}) = -t(\mathbf{u})$.

Proof. (i). $t(\alpha\mathbf{u}+\beta\mathbf{v}) = t(\alpha\mathbf{u})+t(\beta\mathbf{v}) = \alpha t(\mathbf{u})+\beta t(\mathbf{v})$ (since t is a linear transformation).
(ii). $t(\mathbf{0}_n) = t(\mathbf{0}_n + \mathbf{0}_n) = t(\mathbf{0}_n) + t(\mathbf{0}_n)$. Adding $-t(\mathbf{0}_n)$ to both sides gives $\mathbf{0}_m = t(\mathbf{0}_n)$.
(iii). $t(-\mathbf{u}) = t((-1)\mathbf{u}) = (-1)t(\mathbf{u}) = -t(\mathbf{u})$. □

Notice that Part (ii) of this theorem gives us an alternative proof that the map t in Example 2 (above) is not a linear transformation, since it tells us that every linear transformation sends the origin in $\mathbb{R}^n$ to the origin in $\mathbb{R}^m$. Notice also that Part (i) of the theorem tells us that if t is a linear transformation then t maps every straight line $\{\, \lambda\mathbf{u} + (1-\lambda)\mathbf{v} : \lambda \in \mathbb{R} \,\}$ in $\mathbb{R}^n$ to either a straight line (if $t(\mathbf{u}) \neq t(\mathbf{v})$) or a point (if $t(\mathbf{u}) = t(\mathbf{v})$) in $\mathbb{R}^m$. In either case, the image of the straight line under t is $\{\, \lambda t(\mathbf{u}) + (1-\lambda)t(\mathbf{v}) : \lambda \in \mathbb{R} \,\}$.

Notice also, that Theorem 9.2 (i) combines both conditions (as per Definition 9.1) for t to be a linear transformation into a single condition (with more variables). It is also possible for a map $t : \mathbb{R}^n \to \mathbb{R}^n$ to satisfy both Parts (ii) and (iii) of Theorem 9.2 and still *not* be a linear transformation. One example is given by the map from $\mathbb{R}^2$ to itself given by

$$t\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a^3 \\ b^3 \end{pmatrix}.$$

9.4 Matrices and linear transformations

Let A be an $m \times n$ matrix. We define t_A to be the function

$$t_A : \mathbb{R}^n \to \mathbb{R}^m, \quad \mathbf{u} \mapsto t_A(\mathbf{u}) := A\mathbf{u}.$$

By properties of multiplication of matrices we have proved earlier, we know that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$, we have:

$$\begin{aligned} t_A(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = t_A(\mathbf{u}) + t_A(\mathbf{v}) \\ \text{and} \quad t_A(\alpha\mathbf{u}) = A(\alpha\mathbf{u}) = \alpha(A\mathbf{u}) = \alpha t_A(\mathbf{u}). \end{aligned}$$

So t_A is a linear transformation: we call t_A the linear transformation *represented by A*.

Example. Let A be the 3×3 matrix

$$\begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_3 \\ u_3 & v_3 & w_3 \end{pmatrix}.$$

Then $t_A : \mathbb{R}^3 \to \mathbb{R}^3$, defined by $t_A(\mathbf{u}) = A\mathbf{u}$ is a linear transformation. It has

$$t_A\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad t_A\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{and} \quad t_A\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

Is every linear transformation $t : \mathbb{R}^n \to \mathbb{R}^m$ represented by some matrix? The answer is yes. Rather than write out a formal proof for general m and n, let us think about the case of a linear transformation $t : \mathbb{R}^3 \to \mathbb{R}^3$. Suppose that

$$t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

But then, since t is a linear transformation, we know that for any $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$ we have

$$t\begin{pmatrix} a \\ b \\ c \end{pmatrix} = t\left(a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) = at\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + bt\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + ct\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$
$$= a\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + b\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} + c\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

and so the linear transformation t is represented by the matrix which has as its columns

$$t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

9.5 Composition of linear transformations and multiplication of matrices

Suppose that $s : \mathbb{R}^n \to \mathbb{R}^m$ and $t : \mathbb{R}^p \to \mathbb{R}^n$ are linear transformations. Define their *composition* $s \circ t : \mathbb{R}^p \to \mathbb{R}^m$ by:

$$(s \circ t)(\mathbf{u}) = s(t(\mathbf{u})) \text{ for all } \mathbf{u} \in \mathbb{R}^p.$$

Now suppose s is represented by the $m \times n$ matrix A and t is represented by the $n \times p$ matrix B. Then for all $\mathbf{u} \in \mathbb{R}^p$ we have

$$(s \circ t)(\mathbf{u}) = s(t(\mathbf{u})) = A(B\mathbf{u}) = (AB)\mathbf{u},$$

where the last equality follows from the associativity of matrix multiplication. We conclude the $s \circ t$ is a linear transformation, and that it is represented by the $m \times p$ matrix AB, which explains why we define matrix multiplication the way that we do!

Example. Let $s : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation which has

$$s\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad s\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \quad \text{and} \quad s\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$

and let $t : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation which has

$$t\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad t\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Then s is represented by $A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & -3 & 0 \end{pmatrix}$, and t is represented by $B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \\ 1 & 3 \end{pmatrix}$.

Moreover, $s \circ t : \mathbb{R}^2 \to \mathbb{R}^2$ is represented by $AB = \begin{pmatrix} 1 & 0 \\ 4 & -4 \end{pmatrix}$.

Thus if $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, then

$$s\left(t\begin{pmatrix} x \\ y \end{pmatrix}\right) = (s \circ t)\begin{pmatrix} x \\ y \end{pmatrix} = (AB)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 4 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 4x - 4y \end{pmatrix}.$$

9.6 Rotations and reflexions of the plane

Let r_θ denote a rotation of the plane, keeping the origin fixed, through an angle θ. If $\theta > 0$ this is an *anticlockwise* rotation. (In mathematics the convention is that anticlockwise is the positive direction when it comes to measuring angles.) It is relatively easy to prove that r_θ is a linear transformation of the plane, as follows:

(i) Consider the parallelogram defining the sum $\mathbf{u} + \mathbf{v}$ of the position vectors $\mathbf{u}$ and $\mathbf{v}$ of any two points in the plane $\mathbb{R}^2$. If we rotate this parallelogram through an angle θ we obtain a new parallelogram, congruent to the original one. This new parallelogram has vertices with position vectors $\mathbf{0}, r_\theta(\mathbf{u}), r_\theta(\mathbf{v}), r_\theta(\mathbf{u} + \mathbf{v})$, and the very fact that it is a parallelogram tells us that $r_\theta(\mathbf{u} + \mathbf{v}) = r_\theta(\mathbf{u}) + r_\theta(\mathbf{v})$.

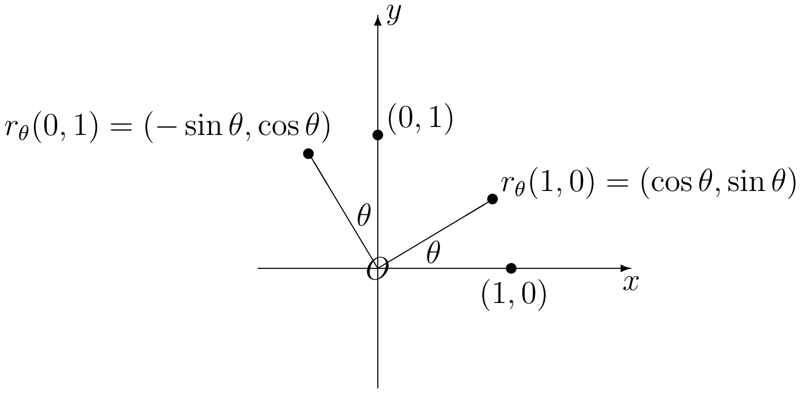
(ii) Given any vector $\mathbf{u}$ and scalar $\lambda \in \mathbb{R}$, we must show that $r_\theta(\lambda\mathbf{u}) = \lambda r_\theta(\mathbf{u})$. If $\lambda = 0$ or $\mathbf{u} = \mathbf{0}$ then $r_\theta(\lambda\mathbf{u}) = \mathbf{0} = \lambda r_\theta(\mathbf{u})$, as required. Now suppose that $\mathbf{u} \neq 0$ and $\lambda \neq 0$, and consider the line ℓ through the origin in the direction of $\mathbf{u}$. Let P be the point on this line with position vector $\mathbf{u}$ and Q be the point with position vector $\lambda\mathbf{u}$. Now r_θ sends O, P, Q to the points $O' = O, P', Q'$ (respectively) on $r_\theta(\ell)$, which is also a straight line. So we get that $|OQ'|/|OP'| = |OQ|/|OP| = |\lambda|$, since rotation preserves distances.

Rotations also preserve betweenness: that is if X, Y, Z are points on ℓ such that Y is between X and Z then $r_\theta(Y)$ is between $r_\theta(X)$ and $r_\theta(Z)$. Thus $O = O'$ is (strictly) between P' and Q' if and only if O is between P and Q, which is if and only if $\lambda < 0$. Hence $r_\theta(\lambda\mathbf{u}) = \lambda r_\theta(\mathbf{u})$ for all $\lambda \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^2$.

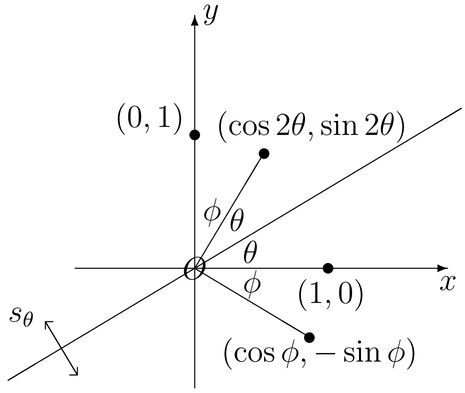
Since r_θ sends $(1, 0)$ to $(\cos\theta, \sin\theta)$ and sends $(0, 1)$ to $(-\sin\theta, \cos\theta)$ (see the picture overleaf, or draw your own picture), it is the linear transformation represented by the matrix:

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

It is easily checked that $(R_\theta)^{-1} = R_{-\theta}$ and that $R_{\phi+\theta}$ is the product of the matrices R_ϕ and R_θ. (For those of you who know about polar coördinates, the transformation r_θ maps the point with polar coördinates (r, ψ) to $(r, \psi + \theta)$.)



For this year [as I was away], we denote by s_θ the linear transformation which reflects the (x, y)-plane, with mirror the line through the origin at (anticlockwise) angle θ to the x-axis. Just as with a rotation r_θ, it is easy to prove geometrically that s_θ is a linear transformation (the details are left as an exercise for you).



From the illustration above it is apparent that $s_\theta(1,0) = (\cos 2\theta, \sin 2\theta)$ and that $s_\theta(0,1) = (\cos\phi, -\sin\phi)$, where ϕ is the angle shown. But $\phi + 2\theta = \frac{\pi}{2}$, so $\cos\phi = \sin 2\theta$ and $\sin\phi = \cos 2\theta$. Hence $s_\theta(0,1) = (\sin 2\theta, -\cos 2\theta)$, and we deduce that the reflexion s_θ is represented by the matrix

$$S_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$

(In terms of polar coördinates, s_θ maps (r, ψ) to $(r, 2\theta - \psi)$.)

Exercise. Check that $(S_\theta)^{-1} = S_\theta$ by showing that $(S_\theta)^2 = \mathrm{I}_2$. (This verifies that applying the same reflexion twice brings every point of the (x, y)-plane back to itself.)

Remark. My preferred notation is that the mirror of the reflexion s_θ should have angle $\theta/2$ with the x-axis. The matrix of s_θ would then be the same as $S_{\theta/2}$ above, but would now be denoted S_θ, and we would have

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

In terms of polar coördinates, this version of s_θ maps (r, ψ) to $(r, \theta - \psi)$.

Remark. Rotations and reflexions are examples of *orthogonal* linear transformations of the plane, that is to say linear transformations s that have the property that for every $\mathbf{u}$ and $\mathbf{v}$ the dot products $\mathbf{u}\cdot\mathbf{v}$ and $s(\mathbf{u})\cdot s(\mathbf{v})$ are equal. In geometric terms this means that s preserves lengths of vectors and angles between vectors: in other words s is a 'rigid motion' of the plane (with the origin a fixed point). It can be shown that a linear transformation s of $\mathbb{R}^n$ is orthogonal if and only the matrix S that represents it has the property that $SS^\mathsf{T} = \mathrm{I}_n = S^\mathsf{T}S$ (where S^T is the transpose of S). In the module MTH5112: Linear Algebra I you will study orthogonal matrices, but we remark here that every 2×2 matrix which is orthogonal is the matrix either of a rotation or of a reflexion (depending on whether its determinant is $+1$ or -1).

9.7 Other linear transformations of the plane

We consider some examples. Other examples may obtained by composing linear transformations already listed. (This new linear transformation need not be one of the types we have listed.) For example, the composition $s_\phi \circ s_\theta$ of two reflexions is a rotation r_ψ, as you can check by multiplying the corresponding matrices (as an exercise, compute the angle ψ in terms of θ and ϕ). The composition $r_\phi \circ t_{2\mathrm{I}_2}$ is not in general one of listed types of linear transformation of $\mathbb{R}^2$.

9.7.1 Linear 'stretches' of axes and dilations

The linear transformation t_A of the plane represented by the matrix

$$A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$$

sends the point $(x, 0)$ on the x-axis to the point $(ax, 0)$ on the x-axis, so it sends the x axis to itself, stretching it by a factor a (or contracting it if $|a| < 1$, and reflecting it if $a < 0$). We say that the x-axis is a *fixed line* of t_A, since every point on this line is mapped to a point on the same line. Similarly the y-axis is a fixed line for this t_A: it is stretched by a factor d. If $a \neq d$, the x-axis and the y-axis are the *only* fixed lines for this t_A. When $a = d$, *every* line through the origin is a fixed line and the effect of t_A is to apply the same stretching factor in all directions: such a t_A is called a *dilation* when $a = d > 0$. More generally, one can have stretches which fix some pair of lines other than the axes.

9.7.2 Shears (or transvections)

The linear transformation t_B represented by

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

is an example of a *shear*. In some areas of Mathematics (such abstract algebra) shears are known as *transvections* instead. Here each point (x, y) is mapped to the point $(x+y, y)$. So each line $y = c$ (constant) is a fixed line for the transformation. Each point on the line $y = c$ is translated to the right by a distance c (which of course is negative when c is negative). Notice that the x-axis is fixed *pointwise* (every point on this axis is mapped to itself), and the y-axis is not even a fixed line: it is mapped to the line with equation $x = y$ which passes through the origin and has angle $\frac{\pi}{4}$ with the x-axis. More generally, for any nonzero constants b and c

$$B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

are shears. The most general form of a shear in 2 dimensions is the matrix

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $a+d = 2$ and $ad-bc = 1$, except that (by convention) the case $T = \mathrm{I}_2$ is excluded. There is some line that T fixes pointwise; this line is unique (since $T \neq \mathrm{I}_2$), and need not be one of the axes.

Shears of $\mathbb{R}^3$ fix a plane Π pointwise and send a (particular) vector $\mathbf{x} \notin \Pi$ to $\mathbf{x} + \mathbf{u}$ where $\mathbf{0} \neq \mathbf{u} \in \Pi$. The general vector $\mathbf{v} \in \mathbb{R}^3$ has form $\mathbf{v} = \lambda\mathbf{x} + \mathbf{w}$ for some (unique) $\lambda \in \mathbb{R}$ and $\mathbf{w} \in \Pi$. Since shears are linear maps, the shear must send $\mathbf{v}$ to $\mathbf{v} + \lambda\mathbf{u}$, and I leave it as an exercise for you to verify this. (I write $\mathbf{w} \in \Pi$ to mean that the position vector of $\mathbf{w}$ is in Π.) Shears of $\mathbb{R}^n$ are like shears of $\mathbb{R}^3$, except that Π is now a hyperplane that is fixed pointwise.

9.7.3 Singular transformations

The $n \times n$ matrix A is called *non-singular* if $\det A \neq 0$ and *singular* if $\det A = 0$. So a non-singular matrix is invertible, and a singular matrix is not. We call the linear transformation $t_A : \mathbb{R}^n \to \mathbb{R}^n$ *non-singular* if A is a non-singular matrix, and we call it *singular* if A is a singular matrix. So far all our examples of linear transformations of the plane have been non-singular. What about the case that A is singular?

The simplest singular case is when A is the zero 2×2 matrix (or zero $n \times n$ matrix for any $n \geqslant 1$). There is not much to say about this case geometrically, just that the corresponding linear transformation sends every point of the plane to the origin. But what about the case when

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

with $ad - bc = 0$ but a, b, c, d not all zero? Suppose for example that the first column of A is nonzero. Then $ad - bc = 0$ implies that $b = (d/c)a$, $d = (d/c)c$ if $c \neq 0$ and $b = (b/a)a$, $d = (b/a)c$ if $a \neq 0$. So the second column of A is a scalar multiple of the first column. Writing $\mathbf{u}$ for first column of A (so that $\mathbf{u} \neq \mathbf{0} = \mathbf{0}_2$) and $\lambda\mathbf{u}$ for the second column of A we see that

$$A\mathbf{i} = A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathbf{u} \quad \text{and} \quad A\mathbf{j} = A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda\mathbf{u}$$

and so

$$A\begin{pmatrix} x \\ y \end{pmatrix} = A\left(x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = (x + \lambda y)\mathbf{u}$$

for every point (x, y) of the plane. Thus we find that

$$A(x\mathbf{v}) = A\begin{pmatrix} \lambda x \\ -x \end{pmatrix} = \mathbf{0} \text{ for all } x \in \mathbb{R}, \text{ where } \mathbf{v} = \begin{pmatrix} \lambda \\ -1 \end{pmatrix}.$$

So the whole of the plane is mapped onto the straight line through the origin which consists of the set of all scalar multiples of $\mathbf{u}$. If we take any point $\mathbf{c} = \alpha\mathbf{u}$ on this line then the equations $A\mathbf{x} = \mathbf{c}$ have infinitely many solutions (all the points on the straight line $\mathbf{r} = \alpha\mathbf{i} + \mu\mathbf{v}$ through $\alpha\mathbf{i}$ parallel to $\mathbf{v}$), and if we take any point $\mathbf{c}$ not on this line $[\mathbf{r} = \mu\mathbf{u}]$ then the equations $A\mathbf{x} = \mathbf{c}$ have no solution. (Note that in general $\mathbf{u}$ and $\mathbf{v}$ are neither perpendicular nor parallel, though of course they may be sometimes.)

A similar situation arises when $A : \mathbb{R}^3 \to \mathbb{R}^3$ is a singular matrix. Then it can be shown that A maps the whole of $\mathbb{R}^3$ onto a plane through the origin, a line through the origin, or a single point (the origin).

9.7.4 Translations and affine maps

In general, these are *not* linear maps, but are a slight generalisation thereof. An *affine (linear) transformation* or *affine map* of $\mathbb{R}^n$ is a map from $\mathbb{R}^n$ to itself of the form:

$$f(A, \mathbf{b}) : \mathbf{r} \mapsto A\mathbf{r} + \mathbf{b},$$

where A is an $n \times n$ real matrix and $\mathbf{b}, \mathbf{r} \in \mathbb{R}^n$ with $\mathbf{b}$ fixed. Despite the use of the word 'linear' in the terminology, the map $f(A, \mathbf{b})$ is linear only when $\mathbf{b} = \mathbf{0}[= \mathbf{0}_n]$. Composing two affine maps gives us an affine map, for we can calculate that $f(A, \mathbf{b}) \circ f(C, \mathbf{d}) = f(AC, A\mathbf{d} + \mathbf{b})$, which I leave it to you to check. The map $f(A, \mathbf{b})$ is invertible if and only if A is, and in that case $f(A, \mathbf{b}) = f(A^{-1}, -A^{-1}\mathbf{b})$.

A special case of affine maps occurs when $A = \mathrm{I}_n$. Such maps are *translations* of $\mathbb{R}^n$. The most general form of a translation is $T_{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T_{\mathbf{b}} : \mathbf{r} \mapsto \mathbf{r} + \mathbf{b}$, that is $T_{\mathbf{b}}$ has the effect of adding $\mathbf{b}$ to each vector of $\mathbb{R}^n$. So $T_{\mathbf{b}} = f(\mathrm{I}_n, \mathbf{b})$. It is linear if and only if $\mathbf{b} = \mathbf{0}$. Composing two translations gives a translation: we have $T_{\mathbf{b}} \circ T_{\mathbf{c}} = T_{\mathbf{c}} \circ T_{\mathbf{b}} = T_{\mathbf{b}+\mathbf{c}}$. Also, $(T_{\mathbf{b}})^{-1} = T_{-\mathbf{b}}$ for all $\mathbf{b} \in \mathbb{R}^n$.

In fact, translations and affine maps of $\mathbb{R}^n$ can be regarded as linear maps of $\mathbb{R}^{n+1}$. A standard way in which this can achieved is to let $f(A, \mathbf{b})$ correspond to the following $(n + 1) \times (n + 1)$ real matrix:

$$\begin{pmatrix} & A & & \mathbf{b} \\ 0 & \ldots & 0 & 1 \end{pmatrix}.$$

Chapter 10

Eigenvectors and Eigenvalues

10.1 Definitions

Let A be an $n \times n$ matrix. An *eigenvector* of A is a vector $\mathbf{v} \in \mathbb{R}^n$, with $\mathbf{v} \neq \mathbf{0}_n$, such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar $\lambda \in \mathbb{R}$ (which might be 0). The scalar λ is called the *eigenvalue* of A associated to the eigenvector $\mathbf{v}$.

Example. Let $A = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$. The vector $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is an eigenvector of A, with corresponding eigenvalue 5, since

$$\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \end{pmatrix} = 5\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Geometrically, it is clear that the eigenvectors of the linear transformation $t_A : \mathbf{x} \mapsto A\mathbf{x}$ are the position vectors of points on fixed lines through the origin (except for the origin itself), and the eigenvalues are the corresponding stretch factors, at least in the case of eigenvalues $\lambda \neq 0$.

10.2 Eigenvectors corresponding to the eigenvalue 0

From the definitions, saying that 0 is an eigenvalue of A means the same as saying that there exists a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}_n$. But this implies that A is not invertible (since if A is invertible then $A\mathbf{x} = \mathbf{0}_n$ has *unique* solution $\mathbf{x} = A^{-1}\mathbf{0}_n = \mathbf{0}_n$). In fact the converse is also true, namely that if A is not invertible then $A\mathbf{x} = \mathbf{0}_n$ has a nonzero solution $\mathbf{x}$ (we are not going to prove this here, but we saw it was true in the case $n = 2$, when we examined *singular* 2×2 matrices). Thus we have the

Useful Fact. An $n \times n$ matrix has 0 as one of its eigenvalues if and only if A is not invertible.

10.3 Finding all eigenvalues

Given a matrix, it turns out to be easiest to first calculate the eigenvalues, and then the eigenvectors.

Theorem 10.1. Let A be an $n \times n$ matrix. Then the eigenvalues of A are those real numbers λ which have the property that $\det(A - \lambda \mathrm{I}_n) = 0$.

Proof. Strictly speaking, we have only proved all properties of determinants (namely that A is invertible if and only if $\det A = 0$) we need for $n \leqslant 2$, and stated that the same holds for $n = 3$, and for general n if you read the optional section (§8.4) of Chapter 8. So strictly speaking we will only have proved this theorem for $n \leqslant 2$, though the same argument works for general n, as you will see in MTH5112: Linear Algebra I:

$$\begin{aligned} \lambda \text{ is an eigenvalue of } A &\iff \exists \mathbf{v} \neq \mathbf{0}_n \text{ such that } A\mathbf{v} - \lambda\mathbf{v} = (A - \lambda \mathrm{I}_n)\mathbf{v} = \mathbf{0}_n \\ &\iff (A - \lambda \mathrm{I}_n) \text{ is not invertible} \\ &\iff \det(A - \lambda \mathrm{I}_n) = 0. \end{aligned}$$

□

The expression $\det(A - x\mathrm{I}_n)$ is called the *characteristic poynomial* of A (it is a polynomial in x, of degree n), so another way to state the Theorem above is to say that *the eigenvalues of A are the zeros of the characteristic polynomial of A.* (Recall that the *zeros* of a function $f(x)$ are the values of x such that $f(x) = 0$.) Eigenvalues are usually counted by multiplicity, so that if the characteristic polynomial is, say, $-(x-4)^2(x+3)$ we would say that the eigenvalues are $-3, 4, 4$.

Example 1. Let $A = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$. Then the characteristic polynomial of A is

$$\begin{aligned} f(x) = \det(A - x\mathrm{I}_2) &= \det\left(\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} - \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}\right) = \det\begin{pmatrix} 1-x & 2 \\ 4 & 3-x \end{pmatrix} \\ &= (1-x)(3-x) - 8 = x^2 - 4x - 5 = (x+1)(x-5). \end{aligned}$$

The only zeros of $f(x)$ are -1 and 5, so by Theorem 10.1 the only eigenvalues of A are -1 and 5.

Example 2. Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{pmatrix}$. Then the characteristic polynomial of A is

$$\begin{aligned} f(x) = \det(A - x\mathrm{I}_3) &= \begin{vmatrix} 1-x & 2 & -1 \\ 0 & 1-x & 4 \\ 0 & 0 & 3-x \end{vmatrix} = (1-x)\begin{vmatrix} 1-x & 4 \\ 0 & 3-x \end{vmatrix} - 0 + 0 \\ &= (1-x)((1-x)(3-x) - 0) = -(x-1)^2(x-3). \end{aligned}$$

The only zeros of $f(x)$ are 1 and 3, so by Theorem 10.1 the only distinct eigenvalues of A are 1 and 3. In this example, the characteristic polynomial has a repeated root, and the eigenvalues of A are actually $1, 1, 3$. This does have some mathematical significance. Using language that will be defined in MTH5112: Linear Algebra I, this means that the eigenvectors having eigenvalue 1, together with the zero vector, could potentially form a 2-dimensional subspace of $\mathbb{R}^3$, though in this example they only form a 1-dimensional subspace.

Remarks.

1. Let $n = 2$ or 3, and let A be an $n \times n$ matrix, with characteristic polynomial $f(x)$. It is not difficult to see that $f(x)$ is a polynomial with coefficients in $\mathbb{R}$, and that the degree of $f(x)$ is n (the highest degree term in $f(x)$ comes from the product of the terms down the main diagonal of A, so the coefficient of x^n is ± 1 [$(-1)^n$ in fact]). Such an $f(x)$ has at most n real zeros, and so A has at most n distinct real eigenvalues.

2. A polynomial of odd degree with coefficients in $\mathbb{R}$ always has at least one real zero. (This follows from the Intermediate Value Theorem [see your MTH4100: Calculus I notes], since the graph of $y = f(x)$ is either above the x-axis as x tends to $+\infty$ and below the x-axis as x tends to $-\infty$, or vice versa.) Hence it follows that every 3×3 (real) matrix has at least one (real) eigenvalue, and so every linear transformation of $\mathbb{R}^3$ (except for zero transformation) fixes at least one line through the origin.

10.4 Finding eigenvectors

Let $n = 2$ or 3, and let A be an $n \times n$ matrix with an eigenvalue λ. How do we find one (or all) eigenvectors $\mathbf{v}$ of A with $A\mathbf{v} = \lambda\mathbf{v}$? The answer is that we solve a system of n equations in n unknowns. We have:

$$\begin{aligned} A\mathbf{v} = \lambda\mathbf{v} &\iff A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0}_n \\ &\iff A\mathbf{v} - \lambda \mathrm{I}_n\mathbf{v} = \mathbf{0}_n \\ &\iff (A - \lambda\mathrm{I}_n)\mathbf{v} = \mathbf{0}_n. \end{aligned}$$

If $n = 2$, to obtain the eigenvectors of A corresponding to the eigenvalue λ we solve $(A - \lambda\mathrm{I}_2)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ with $\begin{pmatrix} x \\ y \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

If $n = 3$, to obtain the eigenvectors of A corresponding to the eigenvalue λ we solve $(A - \lambda\mathrm{I}_3)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ for $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ with $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

Example 1. Let

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

In the previous subsection we found that the eigenvalues of A are -1 and 5. We now determine the eigenvectors corresponding to eigenvalue -1. We solve:

$$(A-(-1)\mathrm{I}_2)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

that is to say,

$$\left(\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} - \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

in other words,

$$\begin{pmatrix} 2 & 2 \\ 4 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This is the system of equations:

$$\left.\begin{aligned} 2x+2y&=0 \\ 4x+4y&=0 \end{aligned}\right\},$$

which, when reduced to echelon form is:

$$\left.\begin{aligned} 2x+2y&=0 \\ 0&=0 \end{aligned}\right\}.$$

Thus y can be any real number r, and then $2x+2r=0$, so $x=-r$. Thus the set of all eigenvectors of A corresponding to the eigenvalue -1 is

$$\left\{\begin{pmatrix} -r \\ r \end{pmatrix} : r \in \mathbb{R} \;\middle|\; r \neq 0\right\}.$$

One such eigenvector is $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. We can check our calculation, as shown below.

$$A\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (-1)\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

It is left as an exercise for you to compute the eigenvectors of A corresponding to the other eigenvalue, 5.

Example 2. Let

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

We have found that the characteristic polynomial of A is $f(x) = -(x-1)^2(x-3)$, and thus that the eigenvalues of A are 1 and 3 (the real zeros of $f(x)$), with 1 being repeated. We now find all the eigenvectors of A corresponding to the eigenvalue 3. We solve:

$$(A-3\mathrm{I}_3)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

that is to say,

$$\begin{pmatrix} 1-3 & 2 & -1 \\ 0 & 1-3 & 4 \\ 0 & 0 & 3-3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which, in other words, is

$$\begin{pmatrix} -2 & 2 & -1 \\ 0 & -2 & 4 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is the system of equations:

$$\left.\begin{aligned} -2x + 2y - z &= 0 \\ -2y + 4z &= 0 \\ 0 &= 0 \end{aligned}\right\}.$$

These equations are already in echelon form, so we can solve them by setting z to be r (representing any real number) and deducing by back substitution that $y = 2r$ (from the second equation) and then that $x = \frac{3}{2}r$ from the first equation. Thus the set of all eigenvectors of A corresponding to the eigenvalue 3 is

$$\left\{ \begin{pmatrix} \frac{3}{2}r \\ 2r \\ r \end{pmatrix} : r \in \mathbb{R} \;\middle|\; r \neq 0 \right\}.$$

One such eigenvector is $\mathbf{b} := \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}$. (Check this!)

Thus the line through the origin with vector equation $\mathbf{r} = \mu\mathbf{b}$ ($\mu \in \mathbb{R}$) is fixed by the linear transformation t_A represented by A, and a point on this line with position vector $\mathbf{v}$ is mapped by t_A to the point that has position vector $3\mathbf{v}$.

It is left as an exercise for you to compute the eigenvectors of A corresponding to the other eigenvalue, 1.

10.5 Eigenvectors and eigenvalues for linear transformations of the plane

We revisit rotations and reflexions, axes stretches, dilations and shears in $\mathbb{R}^2$, to see how eigenvectors and eigenvalues are involved.

Rotations. No real eigenvalues, as no fixed lines, except for the case of a rotation through $m\pi$ for some $m \in \mathbb{Z}$, when every nonzero vector is an eigenvector with eigenvalue $(-1)^m$, and the rotation matrix is $(-1)^m \mathrm{I}_2$. (In general, an anticlockwise rotation through θ has complex eigenvalues $\mathrm{e}^{\mathrm{i}\theta}$ and $\mathrm{e}^{-\mathrm{i}\theta}$.)

Reflexions. Here every (nonzero) vector in the direction $\mathbf{u}$ of the mirror is an eigenvector with eigenvalue $+1$, and every (nonzero) vector orthogonal to $\mathbf{u}$ is an eigenvector with eigenvalue -1.

Axes stretches. The matrix $\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$ (with a and d nonzero and $a \neq d$) has the position vector of every point on the x-axis (except the origin) as an eigenvector with eigenvalue a, and the position vector of every point on the y-axis (except the origin) as an eigenvector with eigenvalue d.

Dilations. The matrix $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ (with $a > 0$) has a as an eigenvalue and every nonzero vector as an eigenvector corresponding to a.

Shears. Consider for example $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. By an easy calculation the only eigenvalue is $+1$, and the eigenvectors corresponding to this eigenvalue are the vectors $\left\{ \begin{pmatrix} t \\ 0 \end{pmatrix} : t \in \mathbb{R} \;\middle|\; t \neq 0 \right\}$, that is, the position vectors of all points on the x-axis other than the origin. In general, all shears (in 2 dimensions) have characteristic polynomial $x^2 - 2x + 1 = (x-1)^2$, and thus eigenvalues $1, 1$. However, all eigenvectors of a shear lie along a single line (origin excluded).

10.6 Rotations and reflexions in $\mathbb{R}^3$

It was mentioned in an earlier remark (without proof) that every rigid motion of $\mathbb{R}^3$ which fixes the origin is represented by an *orthogonal* matrix, that is to say a matrix A with the property that $AA^\mathsf{T} = \mathrm{I}_3 = A^\mathsf{T}A$ (where A^T denotes the *transpose* of A). It follows that $(\det A)^2 = 1$ and hence that $\det A = \pm 1$.

Rigid motions of $\mathbb{R}^3$ represented by matrices A which have $\det A = +1$ are called *orientation-preserving* (they send right-handed triples of vectors to right-handed triples of vectors), and those which have $\det A = -1$ are called *orientation-reversing*.

As was mentioned in another earlier remark, every linear transformation of $\mathbb{R}^3$ has a fixed line (since the characteristic polynomial is a cubic and therefore has a real root). If the linear transformation is a rigid motion, the corresponding eigenvalue must be $+1$ or -1, since rigid motions preserve distances.

Let A be a 3×3 matrix representing a rigid motion of $\mathbb{R}^3$ fixing the origin.

Case 1. A has $+1$ as an eigenvalue.

Let $\mathbf{u} \neq \mathbf{0}_3$ be an eigenvector with eigenvalue $+1$. Now A maps the plane Π through the origin orthogonal to $\mathbf{u}$ to itself. If this map of Π to itself is a *rotation* then A represents a rotation of $\mathbb{R}^3$ around the axis which has direction $\mathbf{u}$. If the map of Π to itself is a reflexion, with mirror a line in Π having direction $\mathbf{v}$, then A represents a reflexion of $\mathbb{R}^3$ with mirror the plane containing $\mathbf{u}$ and $\mathbf{v}$.

Case 2. A has -1 as an eigenvalue.

Let $\mathbf{u} \neq \mathbf{0}_3$ be an eigenvector with eigenvalue -1. Once again, A maps the plane Π through the origin orthogonal to $\mathbf{u}$ to itself. If this map of Π to itself is a reflexion, then Π contains an eigenvector $\mathbf{v}$ of A with eigenvalue $+1$, corresponding to the direction of the mirror in Π, and an eigenvector $\mathbf{w}$ orthogonal to $\mathbf{v}$ with eigenvalue -1. Thus $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are an orthogonal set of vectors and A sends $\mathbf{u} \mapsto -\mathbf{u}$, $\mathbf{v} \mapsto \mathbf{v}$, $\mathbf{w} \mapsto -\mathbf{w}$, and so A is a *rotation* about the direction of $\mathbf{v}$ through an angle π. If, on the other hand, the map of Π to itself is a rotation, there is not much more to say except that A represents the composition of a reflexion in the plane Π followed by a rotation about the axis orthogonal to Π.

Remark. It follows from the analysis above that every orientation-preserving rigid motion of $\mathbb{R}^3$ which fixes the origin is a rotation about some axis, but that (unlike the case for $\mathbb{R}^2$) an orientation-reversing rigid motion of $\mathbb{R}^3$ which fixes the origin is not necessarily a reflexion.

We finish by exhibiting examples of 3×3 matrices representing rotations and reflexions.

Example 1. The 3×3 matrix which represents a rotation through angle θ around the z-axis is

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and similarly the 3×3 matrices which represent rotations through angle θ around the x- and y-axes are respectively.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}.$$

In these examples we look at the origin from a point somewhere along the positive portion of the relevant axis. These are rotations through an angle of θ anticlockwise from this point of view. (If we look the origin from the opposite direction then these rotations have an anticlockwise angle of $-\theta$.)

Example 2. The reflexion $s_{\mathbf{n}} = s_\Pi$ in the plane Π through the origin with normal vector $\mathbf{n} \neq \mathbf{0} = \mathbf{0}_3$ sends each vector $\mathbf{x}$ to:

$$s_\Pi(\mathbf{x}) = \mathbf{x} - 2\frac{(\mathbf{x}\cdot\mathbf{n})}{|\mathbf{n}|^2}\mathbf{n}.$$

To see this, draw a picture like the one we used in our calculation of the distance from the point X with position vector $\mathbf{x}$, to the plane Π (the point with position vector $\mathbf{x} - \frac{(\mathbf{x}\cdot\mathbf{n})}{|\mathbf{n}|^2}\mathbf{n}$ is the point of Π which is closest to X).

If the equation of the plane Π is $ax + by + cz = 0$, we can take $\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, and so:

$$s_\Pi(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} - 2\frac{ax + by + cz}{a^2 + b^2 + c^2}(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}).$$

Thus in particular

$$s_\Pi(\mathbf{i}) = \mathbf{i} - 2\frac{a}{a^2 + b^2 + c^2}(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) = \frac{1}{a^2 + b^2 + c^2}\left((b^2 + c^2 - a^2)\mathbf{i} - 2ab\mathbf{j} - 2ac\mathbf{k}\right),$$

and so the first column of the matrix S_Π representing s_Π is

$$\frac{1}{a^2 + b^2 + c^2}\begin{pmatrix} b^2 + c^2 - a^2 \\ -2ab \\ -2ca \end{pmatrix}.$$

Similarly, one can compute the second and third columns, to give the following matrix, which you are not expected to memorise for the examination for this module!

$$S_\Pi = \frac{1}{a^2 + b^2 + c^2}\begin{pmatrix} b^2 + c^2 - a^2 & -2ab & -2ac \\ -2ab & c^2 + a^2 - b^2 & -2bc \\ -2ac & -2bc & a^2 + b^2 - c^2 \end{pmatrix}.$$

Note that

$$S_\pi(\mathbf{n}) = S_\Pi\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -a \\ -b \\ -c \end{pmatrix} = -\mathbf{n},$$

which is exactly what one would expect since $\mathbf{n}$ is orthogonal to the plane Π. Now each vector $\mathbf{x}$ in $\mathbb{R}^3$ can be written uniquely as $\mathbf{x} = \mathbf{u}+\mathbf{v}$ where $\mathbf{u}$ and $\mathbf{n}$ are collinear and $\mathbf{v}$ is parallel to (and thus in) Π, so that $\mathbf{v}$ is orthogonal to $\mathbf{n}$. (We have $\mathbf{u} = \frac{(\mathbf{x}\cdot\mathbf{n})}{|\mathbf{n}|^2}\mathbf{n} = (\mathbf{x}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$ and $\mathbf{v} = \mathbf{x} - (\mathbf{x}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector in the direction of $\mathbf{n}$. You should prove the assertions of the previous sentence, and also try to prove that $\mathbf{u}$ and $\mathbf{v}$ are the *unique* vectors with this property.) Then $s_\Pi(\mathbf{u}) = -\mathbf{u}$ and $s_\Pi(\mathbf{v}) = \mathbf{v}$, so that $\mathbf{u}$ is an eigenvector of s_Π with corresponding eigenvalue -1 (unless $\mathbf{u} = \mathbf{0}$), and $\mathbf{v}$ is an eigenvector of s_Π with corresponding eigenvalue 1 (unless $\mathbf{v} = \mathbf{0}$). Moreover, $s_\Pi(\mathbf{x}) = s_\Pi(\mathbf{u}+\mathbf{v}) = -\mathbf{u}+\mathbf{v}$, so that (exercise for you) $\mathbf{x}$ is an eigenvector of s_Π when *exactly one* of the conditions $\mathbf{u} = 0$ and $\mathbf{v} = 0$ holds.

All of what we have done above generalises readily to reflexions in a hyperplane Π of $\mathbb{R}^n$ orthogonal to the vector $\mathbf{n} \neq \mathbf{0}_n$. I leave the reader to work out the necessary details.

Exercise. Check that the formula for S_Π gives the right answer when Π is the (x, y) plane (the plane defined by the equation $z = 0$). Also check that the transformation s_Π defined at the start of Example 2 is indeed a linear map.