MA135
Vectors and Matrices

Samir Siksek

Samir Siksek, Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

E-mail address: samir.siksek@gmail.com

Contents

CHAPTER 1

Prologue

1. What's This?

These are my lecture notes for MA135 Vectors and Matrices. Thanks to Mario Micallef and Miles Reid for substantial advice on how this subject should be taught, and to Owen Daniel, Jason Davison, Leslie Kanthan, Martin Gould and Lee McDonnell for corrections. Please send comments, misprints and corrections to samir.siksek@gmail.com.

2. A Jolly Good Read!

Linearity pervades Mathematics and Science—even non-linear processes are best studied through linear approximations. For pedagogical convenience, the mathematical study of linearity is divided into two settings:

(1) A concrete setting called 'Vectors and Matrices'.
(2) An abstract setting called 'Linear Algebra'.

The 'Vectors and Matrices' setting is the right place to compute, look at concrete examples and reinforce our intuition. Once we know how to compute and have built-up our intuition we can turn to 'Linear Algebra' with confidence, wherein, with the aid of a little abstraction, we gain deeper insights into linear phenomena.

The pleasures of Linear Algebra (the abstract setting) will have to wait till Term 2; this course is concerned with vectors and matrices (the concrete setting).

It is possible that you have seen at school much or most of what is in this course—then again, do not worry if this is not the case. The point is that we will look at vectors, matrices and complex numbers from a more critical point of view than that of school. Of course we want to calculate, but we want to know why the calculations give the correct answer to the problem at hand. In other words we want to prove things. Sometimes it will not be possible to prove what we want using the concrete tools at our disposal. Consequently, we will take a few things on trust and relegate their proofs to 'Linear Algebra' or some other course. However, as high-minded and morally-correct persons, it should always be clear to us what we have proved and what we are taking on trust, for neither do we want to delude ourselves nor would we allow ourselves to be deluded by someone else (even if this 'deluder' is the lecturer).

3. On Homework (a little preaching)

> The Pythagoreans divided their followers into two groups. One group, the $\mu\alpha\theta\eta\mu\alpha\tau\iota\kappa\eta$ learned the subject completely and understood all the details.. From them comes our word "mathematician," as you can see for yourself if you know the Greek alphabet (mu, alpha, theta, eta, . . .). The second group, the $\alpha\kappa o\upsilon\sigma\mu\alpha\tau\iota o\kappa\iota$,

> or "acusmatics," kept silent and merely memorized the master's words without understanding. The point I am making here is that if you want to be a mathematician, you have to participate, and that means doing the exercises.
>
> Jeffrey Stopple [1]

4. A Preamble on the $\mathbb{R}$eal World

The real numbers, we know and love. We often think of them as points on the real number line. Examples or real numbers are 0, -1, $1/2$, π, $\sqrt{2}\ldots$. The set of real numbers is given the symbol $\mathbb{R}$. Below we list some of their properties. There is no doubt that you are thoroughly acquainted with all these properties, and that you use these properties in your manipulations (both by conscious design and by reflex). The point of listing some of these properties is that we want to have them at hand so that we can ask, when we present complex numbers, vectors and matrices, whether these new objects have the same or similar properties. We also want to *learn the names* of these properties, which are presented in brackets below.

For all real numbers a, b, c

(i) $a + b = b + a$ (addition is commutative)
(ii) $(a + b) + c = a + (b + c)$ (addition is associative)
(iii) $a + 0 = a$ (0 is the additive identity element)
(iv) for any number a there is another number $-a$ (the additive inverse of a) such that $a + (-a) = 0$.
(v) $ab = ba$ (multiplication is commutative)
(vi) $(ab)c = a(bc)$ (multiplication is associative)
(vii) $a(b + c) = ab + ac$ (multiplication distributes over addition)
(viii) $a \cdot 1 = a$ (1 is the multiplicative identity element)
(ix) if $a \neq 0$, there is a number denoted by a^{-1} (the multiplicative inverse of a) such that $a \cdot a^{-1} = 1$.

We have not exhausted the properties of real numbers. For example, we can add

(x) If $a \geq b$ then $a + c \geq b + c$.
(xi) If $a \geq b$ and $c > 0$ then $ac \geq bc$. If $a \geq b$ and $c < 0$ then $ac \leq bc$.

One particularly important property that we will not write down, but which you will come to admire in the analysis courses is 'The Completeness Axiom'.

Exercise 4.1. You know that if $a, b \in \mathbb{R}$ and $ab = 0$ then either $a = 0$ or $b = 0$. Explain how this follows from property (ix) above.

5. Fields

The syllabus for this course mentions 'fields', and says that their definition is non-examinable. We shall be a little vague about fields, and give a heuristic definition, rather than a precise one. A **field** is a 'number system' where you can add, subtract, multiply and divide, and where these operations satisfy properties (i)–(ix) above. For example, the real number system is a field. The integers do not form a field because we cannot divide and stay inside the integers (e.g. 2 and 3 are integers but $2/3$ is not an integer). However, the rational numbers is a field, because we can add, subtract, multiply and divide and properties (i)–(ix) are satisfied.

[1] *A Primer of Analytic Number Theory*, Cambridge University Press, 2003.

The next chapter is devoted to the field of complex numbers $\mathbb{C}$.

Remark Look again at your answer to Exercise 4.1. You will find that all you have used is some of the properties (i)–(ix). Since the elements of any field satisfy (i)–(ix), what you have really shown in answering the exercise is the following: if a, b are elements (or numbers if you like) belonging to any field, and if $ab = 0$ then $a = 0$ or $b = 0$.

CHAPTER 2

The Delightful $\mathbb{C}$omplex Number System

1. What on Earth are $\mathbb{C}$omplex Numbers?

A *complex number* is represented by $a + bi$ where a and b are real numbers and i is a symbol that satisfies $i^2 = -1$. We add and multiply complex numbers as you would expect and every time we see an i^2 we replace it by -1.

Example 1.1. Let α and β be the complex numbers $\alpha = 2 + 3i$ and $\beta = -7 + 4i$. Addition is straightforward:

$$\alpha + \beta = -5 + 7i, \qquad \alpha - \beta = 9 - i.$$

Multiplication involves the usual expansion of brackets and then replacing i^2 by -1:

$$\begin{aligned} \alpha\beta &= (2 + 3i)(-7 + 4i) \\ &= -14 - 13i + 12i^2 && \text{usual expansion of brackets} \\ &= -14 - 13i + 12(-1) && \text{replace } i^2 \text{ by } -1 \\ &= -26 - 13i \end{aligned}$$

The set of complex numbers is denoted by $\mathbb{C}$. In set notation we can write

$$\mathbb{C} = \{a + bi \ : \quad a, b \in \mathbb{R}\}.$$

Definition. Let α be a complex number and write $\alpha = a + bi$ where a and b are real numbers. We call a the real part of α and b the imaginary part of α. We write $\mathrm{Re}(\alpha) = a$ and $\mathrm{Im}(\alpha) = b$.

Example 1.2. $\mathrm{Re}(2 - 4i) = 2$ and $\mathrm{Im}(2 - 4i) = -4$.

2. The Complex Plane

The complex number $a + bi$ is represented by the point (a, b) in the coordinate plane. The x-axis is called the real axis and the y-axis is called the imaginary axis. When used to represent complex numbers in this way, the coordinate plane is called 'The Argand diagram', or 'the complex plane'. See Figure 1.

Addition can be described geometrically (i.e. on the complex plane) by completing the parallelogram. If z and w are complex numbers, then the points representing 0 (the origin), z, w and $z + w$ form a parallelogram; see Figure 2.

3. Some Formal Definitions

Equality of complex numbers is straightforward.

Definition. Two complex numbers are equal if and only if their real parts are equal and their imaginary parts are equal. Another way of saying this is: if $\alpha = a + b\,i$

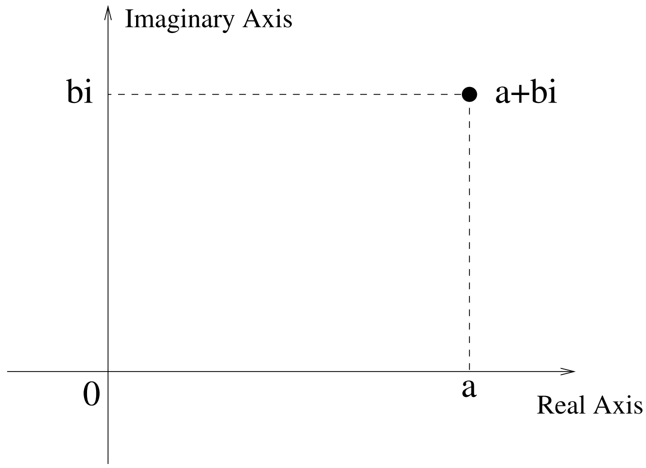

FIGURE 1. The Complex Plane (or The Argand Diagram).

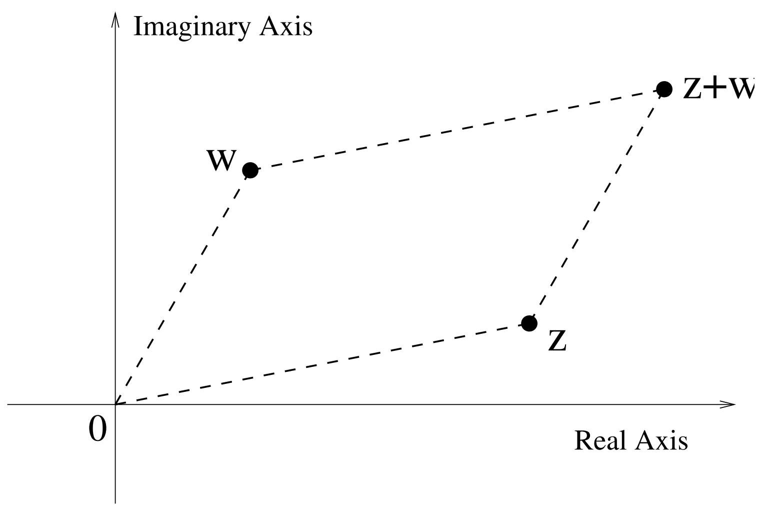

FIGURE 2. If z and w are complex numbers, then the points on the complex plane representing 0 (the origin), z, w and $z+w$ form a parallelogram.

and $\beta = c + d\,i$ are complex numbers (with a, b, c, d real) then $\alpha = \beta$ if and only if $a = b$ and $c = d$.

We saw examples of addition and multiplication above, but let us write the definition of addition and multiplication more formally.

Definition. Suppose $\alpha = a + bi$ and $\beta = c + di$ are complex numbers, where a, b, c, d are real numbers. We define the sum $\alpha + \beta$ by

$$\alpha + \beta = (a + c) + (b + d)i$$

and the product $\alpha\beta$ by

$$\alpha\beta = (ac - bd) + (ad + bc)i.$$

Notice that real numbers are also complex numbers. Indeed if a is a real number, we can think of it as the complex number $a + 0.i$. A complex number of the form bi (i.e. $0 + bi$), with b real, is called an imaginary number. Thus real numbers and imaginary numbers are special types of complex numbers.

Example 3.1. Suppose r is a positive real number. We denote by $\sqrt{r}$ the positive square-root of r. Notice that

$$(\pm\sqrt{r}i)^2 = -r.$$

We see that positive real numbers have two real square-roots, whereas negative real numbers have two imaginary square-roots.

Exercise 3.2. Which number is both real and imaginary?

4. ℂ is a Field

We can make a long list of properties of complex numbers as we did for real numbers. But it is quicker to say the following.

Theorem 4.1. *ℂ is a field.*

Recall that a field is 'number system' (here we are being vague) whose elements satisfy conditions (i)–(ix) of Section 4. Saying that ℂ is a field is an economical way of saying that ℂ satisfies properties (i)–(ix).

Proof of Theorem 4.1. We will actually prove only one of the required properties and **leave the others as an exercise**.

The property we prove is the commutativity of multiplication: if α and β are complex numbers then $\alpha\beta = \beta\alpha$. So suppose that α and β are complex numbers. Write $\alpha = a + bi$ and $\beta = c + di$ where a, b, c, d are real numbers. Then by the definition of multiplication

$$\alpha\beta = (ac - bd) + (ad + bc)i, \qquad \beta\alpha = (ca - db) + (da + cb)i.$$

But $ac = ca$, $bd = db$, $ad = da$, $bc = cb$. How do we know this; isn't this the same as what we want to prove? No, not really. We know this because a, b, c, d are real numbers and we are using the commutativity of multiplication for real numbers which we already know.

It follows that $\alpha\beta = \beta\alpha$ which we wanted to prove. The proof of the remaining properties is an exercise. The reader is encouraged to prove these (or at least some of them) using the proof of the commutativity of multiplication as a model. The only one of these that might prove slightly troublesome is the "existence of multiplicative inverse", because we haven't defined division yet. □

Notice that not all the properties of the real numbers listed in Section 4 carry over to the complexes. The properties involving inequalities do not. *That is because inequalities between complex numbers have no meaning.* This is a point that needs special care. Never write $\alpha > \beta$ if either α or β are complex.

5. An Example is NOT a Proof

In proving that $\alpha\beta = \beta\alpha$ for complex numbers, some may have been tempted to argue along the following lines. Let, say, $\alpha = 1 + 2i$ and $\beta = 3 + 4i$. Then we check that

$$\alpha\beta = -5 + 10i = \beta\alpha.$$

"So the multiplication of complex numbers is commutative". Is this an acceptable proof? *No, it is not.* This is merely an example. It shows that multiplication is commutative *for this particular pair* α, β. This is example might lead us to *suspect* that multiplication of complex numbers is commutative, but *it does not give us the right to conclude that it is commutative.* We want to know that commutativity of multiplication holds for every conceivable pair of complex numbers α, β. That is why, in the proof above, we took $\alpha = a + bi$ and $\beta = c + di$ where a, b, c, d were unspecified real numbers; every pair of complex numbers may be written so, and thus our calculation $\alpha\beta = \beta\alpha$ (in the proof) is valid for all pairs of complex numbers.

The avoidance of blurred impressions will do much good, and so I will go on at the risk of labouring the point. Suppose that a Martian landed on Earth and, by chance, had the misfortune of running into a couple of extremely dim people. Would you not feel offended if the Martian, on the basis of this limited interaction with mankind, was to conclude that all humans—including yourself—are zombies?

To sum up, an example is just that, it is circumstantial evidence that does not mount to a proof.

6. The Quaternionic Number System(do not read this)

This section is not examinable. It is here for the benefit of those few who believe that the above discussion of commutativity is overly pedantic. "Why should multiplication not be commutative? After all, it is just multiplication. You are wasting time on contrived pedanticisms".

Well, matrix multiplication is not commutative—as we will see in the near future. If A and B are matrices, then it is fairly likely that the products AB and BA will not be the same.

In the meantime, we exhibit the quaternionic number system where multiplication is not commutative. Quaternions were fashionable in the late 19th century and had substantial physical applications. Eventually it was discovered that vectors do a better job of just about anything you could do with quaternions, and they fell out of fashion.

Remember that the complex numbers are of the form $a + bi$ where a, b are real and i is a symbol satisfying $i^2 = -1$. Well, quaternions are of the form $a+bi+cj+dk$ where a, b, c, d are real and i, j, k are symbols satisfying

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$

You can already see that quaternionic multiplication is not commutative, since $ij \neq ji$. You might also calculate $(1 + i)(1 + j)$ and $(1 + j)(1 + i)$.

I had asked you not to read this section on quaternions, but obviously the temptation overwhelmed you. Anyway, it is best to erase what you have seen from your memory banks and concentrate on complex numbers. However, a lasting impression that it is better to check things than take them for granted will do you more good than harm.

7. Exponentiation

How do we define exponentiation? In other words, what does α^n mean? Well if n is a positive integer then this is easy to define.

Definition. If α is a complex number and n a positive integer then we define

$$\alpha^n = \underbrace{\alpha \cdot \alpha \cdots \alpha}_{n \text{ times}}.$$

The definition is correct, but involves a subtle point that needs to be checked. Before talking about this subtle point, let us see an example.

Example 7.1. Let $\alpha = 1 + i$. Then (check this):

$$\alpha^2 = 2i, \qquad \alpha^4 = (2i)^2 = -4.$$

Notice that -4 is not the square of a real number but it is the square and fourth power of certain complex numbers. We see from the calculation above that $\alpha = 1+i$ is a root of the polynomial $X^4 + 4$. This polynomial does not have any real roots but has 4 complex roots which are $\pm 1 \pm i$ (check). [1]

We mentioned that there is a subtle point involved in the definition of α^n. There is no problem with the definition for $n = 1, 2$. The problem starts when $n = 3$ and onward. For example we defined

$$\alpha^3 = \alpha \cdot \alpha \cdot \alpha.$$

Notice here that there are two multiplications involved. The problem or ambiguity is that of which multiplication we are doing first. In other words do we mean

$$\alpha^3 = \alpha \cdot (\alpha \cdot \alpha)$$

or do we mean

$$\alpha^3 = (\alpha \cdot \alpha) \cdot \alpha.$$

You would be right in expecting that this does not matter; both will give the same result and so there is no ambiguity in the definition. But how do we know this? Recall that the complex numbers form a field. One of the defining properties of fields is that multiplication is associative. So we know that if α, β, γ are complex numbers then

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma.$$

So, letting $\beta = \gamma = \alpha$ we see that

$$\alpha \cdot (\alpha \cdot \alpha) = (\alpha \cdot \alpha) \cdot \alpha.$$

Hence there is no ambiguity in the definition of α^3.

We saw that we can remove the ambiguity involved in the definition of α^3 using the associativity of multiplication of complex numbers. It can be shown that this associativity makes the definition of α^n unambiguous for all positive n. This is laborious, so we will not do it, but the reader can take it on trust.

Exercise 7.2. For which positive integral values of n is i^n real?

---

[1] An important theorem which we will see later is the Fundamental Theorem of Algebra which says that a polynomial of degree n has n complex roots (counting multiplicities). This is clearly not true if we work just with real numbers.

8. Conjugates

Definition. Let $\alpha = a + bi$ be a complex number, where a, b are real numbers. We define the conjugate of α (denoted by $\overline{\alpha}$) to be $\overline{\alpha} = a - bi$.

We will see later that the conjugate helps us in defining the division of complex numbers.

Theorem 8.1. *Suppose α, β are complex numbers. Then*

(i) *The equality $\alpha = \overline{\alpha}$ holds if and only if α is a real number.*
(ii) *Conjugation distributes over addition and multiplication: in other words*

$$\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$$

and

$$\overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}.$$

(iii) *If $\alpha = a + bi$ with a, b real numbers then*

$$\alpha \cdot \overline{\alpha} = a^2 + b^2.$$

In particular $\alpha \cdot \overline{\alpha}$ is a non-negative real number.

Proof. The proof is left as an exercise. □

Exercise 8.1. What is the geometric meaning of conjugation [2]? I.e. if z is a complex number, describe the geometric operation on the complex plane that takes z to its conjugate $\overline{z}$.

9. Reciprocals and Division

We would like to define reciprocals of complex numbers. In other words, if α is a non-zero complex number, what do we mean by $1/\alpha$? There are certain reasonable things that we should expect from this definition. Of course we want to define reciprocal in such a way that $\alpha \cdot 1/\alpha = 1$. The key to discovering the correct definition is part (iii) of Theorem 8.1. This can be rewritten as follows: if a, b are real then

$$(a + bi)(a - bi) = a^2 + b^2.$$

We instantly see that the following definition is reasonable.

Definition. Let α be a non-zero complex number and write $\alpha = a + bi$ where a, b are real. Define the reciprocal of α to be

$$\frac{1}{\alpha} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2} i.$$

[2] You should get used to thinking geometrically, and to drawing pictures. The true meaning of most mathematical concepts is geometrical. If you spend all your time manipulating symbols (i.e. doing algebra) without understanding the relation to the geometric meaning, then you will have very little in terms of mathematical insight.

The reader will quickly check that [3]

$$\alpha \cdot \frac{1}{\alpha} = 1$$

holds for non-zero α. It is now obvious how to define division: if α and β are complex and β is non-zero then we define $\alpha/\beta = \alpha \cdot 1/\beta$.

The reader will recall that we defined α^n only for positive integer values of n. Now if n is negative we can define $\alpha^n = 1/\alpha^{-n}$. Note the following standard way to divide complex numbers:

Example 9.1.

$$\begin{aligned}\frac{3+i}{2-4i} &= \frac{(3+i)(2+4i)}{(2-4i)(2+4i)}\\ &= \frac{2+14i}{2^2+4^2}\\ &= \frac{1}{10}+\frac{7}{10}i.\end{aligned}$$

To divide note that we first multiply the numerator and denominator by the conjugate of the denominator. *This makes the denominator real* and division becomes easy.

Exercise 9.2. Solve the equation $(5-i)X+7=2i$.

Exercise 9.3. Write

$$\frac{1}{\cos\theta + i\,\sin\theta}$$

in the form $a+i\,b$. (You might already know the answer, but do this question using the definition of reciprocal).

10. The Absolute Value

Let a be a real number. We recall the definition of the modulus (also called the absolute value) $|a|$ as follows:

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0. \end{cases}$$

From now on we say absolute value instead of modulus. We would like to extend the notion of absolute value to complex numbers. The above definition will not do because the inequalities $a \geq 0$ and $a < 0$ do not have a meaning when a is a complex number. There is, however, another–more geometric–definition of the absolute value of a real number: if a is a real number then $|a|$ is the distance on the real line between the numbers a and 0. This definition can be extended to complex numbers. In geometric terms we define, for a complex number α, its absolute value $|\alpha|$ to be the distance between α and 0 (the origin) in the complex plane. This definition is not suitable for calculations, however it is easy to see how to turn it

---

[3]Please don't say "Cancel"! We have given a definition of $1/\alpha$. We want to make sure that this definition does what we expect, so we just have to multiply out

$$(a+bi)\left(\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i\right)$$

and see that we get 1.

into an algebraic definition; if $\alpha = a + bi$ with a, b real then the distance of α from the origin is $\sqrt{a^2 + b^2}$. We finally arrive at our definition.

Definition. Let $\alpha = a + bi$ be a complex number with a, b real. We define the absolute value of α to be

$$|\alpha| = \sqrt{a^2 + b^2}.$$

Notice that whenever we speak of the square-root of a positive real, we mean the positive square-root.

Theorem 10.1. *Let α, β be complex numbers.*

(i) $\alpha\overline{\alpha} = |\alpha|^2$.
(ii) $|\alpha| = 0$ *if and only if* $\alpha = 0$.
(iii) $|\alpha\beta| = |\alpha||\beta|$.
(iv) $|\alpha + \beta| \leq |\alpha| + |\beta|$ *(this is the triangle inequality).*
(v) $|\alpha - \beta| \geq \big||\alpha| - |\beta|\big|$.

The proof of the above theorem is left as an exercise.

11. Sums of Squares

Let a, b, c, d be integers. Can we always find integers u, v such that

$$(a^2 + b^2)(c^2 + d^2) = u^2 + v^2?$$

In other words, if we multiply the sum of two integer squares with the sum of two integer squares, do we get an expression of the same form? The answer, surprisingly, is yes. Before you read on, you might try to prove this yourself.

The easy way to do this is to let

$$\alpha = a + bi, \qquad \beta = c + di.$$

Let $\gamma = \alpha\beta$. Then γ is a complex number and we can write it as $\gamma = u + vi$ with u and v real. Now it is easy to convince yourself that u and v must be integers because a, b, c, d are integers (multiply out $\alpha\beta$ as see what you get). Also

$$\begin{aligned}(a^2 + b^2)(c^2 + d^2) &= |\alpha|^2|\beta|^2 \\ &= |\alpha\beta|^2 \\ &= |\gamma|^2 \\ &= u^2 + v^2,\end{aligned}$$

which shows what we want.

Exercise 11.1. Write 185 as the sum of two integer squares. (Hint: factorize first!) Actually write 185 as the sum of two integers squares in as many ways as you can.

The same is true of sums of four squares: if you multiply a sum of four squares with a sum of four squares you get a sum of four squares. How can you prove this? The easiest way is to use quaternions, which you are not meant to know about because you did not read Section 6.

It turns out that not every positive integer is the sum of two integer squares (find an example). If p is an odd prime, then it can be written as the sum of two squares [4] if and only if p is of the form $1 + 4m$ for some integer m. This means for

[4] This should be proved in the number theory course.

example that 37 can be written as the sum of two squares because $37 = 1 + 4 \times 9$. But 11 cannot be written as the sum of two squares because $11 = 1 + 4 \times 5/2$.

One of the most beautiful theorems in mathematics states that every positive integer is the sum of four squares. This was proved by Joseph Louis Lagrange in 1770, though the theorem appears—without proof—in the Arithmetica of Diophantus (probably written around 250AD). Some of the proofs of this theorem use quaternions.

Another fascinating question is, in how many ways can we write a positive integer n as the sum of four squares. This was answered in 1834 by Carl Jacobi. He showed that this number is eight times the sum of the divisors of n if n is odd, and 24 times the sum of the odd divisors of n if n is even [5].

12. The Argument

Recall that there are two coordinate systems which one may employ to specify points in the plane. The first is the Cartesian system and the second the polar system. In the Cartesian system we represent a point by a pair (a, b): here a and b are distances we have to move parallel to the x- and y-axes to reach our point, having started at the origin. In the polar system we represent points by a pair (r, θ): here r is the distance of the point from the origin. Moreover, if we denote the point by P then θ is the angle [6] measured from the positive x-axis to the ray $\overrightarrow{OP}$ in an anti-clockwise direction. This the polar system.

Converting between Cartesian and polar coordinates is easy. Let (a, b) and (r, θ) represent the same point. We deduce from Figure 3 that

$$a = r\,\cos\theta, \qquad b = r\,\sin\theta.$$

Previously we used the Cartesian system to represent the complex number $a+bi$ on the complex plane. But we can also use the polar system. Now r is the distance from the origin, so $r = |a + bi|$ is just the absolute value. The θ has a special name called the argument.

Definition. Let $\alpha = a + bi$ be a non-zero complex number, and suppose that α is represented in the complex plane by the point P. Let θ be the angle the ray $\overrightarrow{OP}$ makes with the positive real axis (or the positive x-axis). We call θ the argument of α. Note that we can take $0 \le \theta < 2\pi$.

We collect the above facts in a useful Lemma.

Lemma 12.1. *If $\alpha = a + bi$ is a non-zero complex number, r is its absolue value, and θ is its argument, then*

$$a = r\,\cos\theta, \qquad b = r\,\sin\theta,$$

and

$$\alpha = r(\cos\theta + i\,\sin\theta). \tag{1}$$

Moreover,

$$r = |\alpha| = \sqrt{a^2 + b^2}, \qquad b = a\tan\theta.$$

The expression on the right-hand side of (1) is called the (r, θ)-form of α.

---

[5] Jacobi's theorem has remarkable proof using modular forms. Hopefully there will soon be a module on modular forms offered at Warwick.

[6] Normally, when we talk of angles, we are using the radian measure.

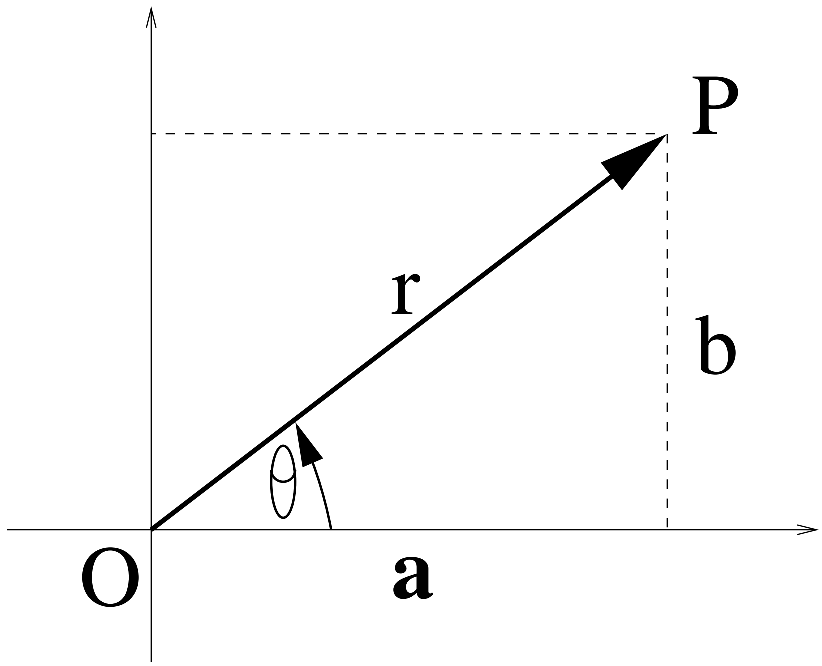

FIGURE 3. A point P can be specified by Cartesian (a, b) coordinates, or by polar (r, θ) coordinates.

Example 12.1. Write the following numbers in (r, θ)-form:

$$3i, \qquad -2, \qquad -5i, \qquad -1+i, \qquad \sqrt{3}+i.$$

Answer: For the first four of the complex numbers, a quick sketch will give us the argument and it is easy to get the (r, θ)-form. For example,

$$|-1+i| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}.$$

From the sketch, the argument of $-1+i$ is $\pi/4 + \pi/2 = 3\pi/4$. Thus

$$-1+i = \sqrt{2}\left(\cos\frac{3\pi}{4} + i\,\sin\frac{3\pi}{4}\right).$$

Similarly

$$3i = 3\left(\cos\frac{\pi}{2} + i\,\sin\frac{\pi}{2}\right), \qquad -2 = 2\left(\cos\pi + i\,\sin\pi\right),$$

$$-5i = 5\left(\cos\frac{3\pi}{2} + i\,\sin\frac{3\pi}{2}\right).$$

Now let $\alpha = \sqrt{3} + i$. We see that

$$r = |\alpha| = \sqrt{\left(\sqrt{3}\right)^2 + 1^2} = \sqrt{4} = 2.$$

A sketch will not immediately give us the value of θ, but it is useful to make one anyway. Note that $\sin\theta = 1/2$ and $\cos\theta = \sqrt{3}/2$. Thus $\theta = \pi/6$. Hence the

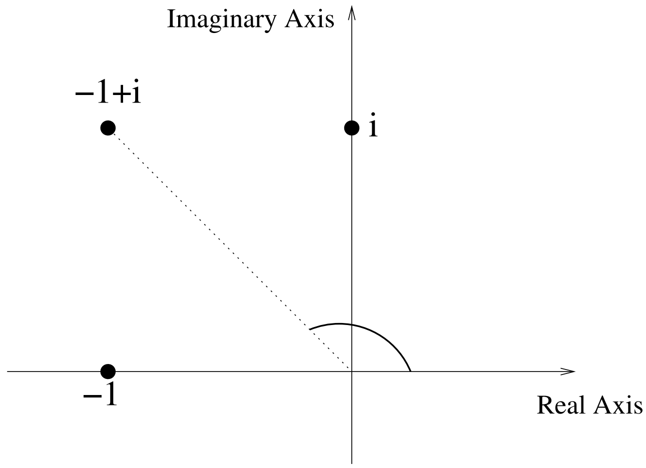

FIGURE 4. It is clear that the argument of $-1 + i$ is $3\pi/4$.

(r, θ)-form of α is

$$\alpha = 2\left(\cos\frac{\pi}{6} + i\,\sin\frac{\pi}{6}\right).$$

13. Multiplying and Dividing the (r, θ)-Form

Lemma 13.1. *Suppose* θ_1*,* θ_2*,* θ *are real. Then*

$$(\cos\theta_1 + i\,\sin\theta_1)(\cos\theta_2 + i\,\sin\theta_2) = \cos(\theta_1+\theta_2) + i\,\sin(\theta_1+\theta_2), \tag{2}$$

$$\frac{\cos\theta_1 + i\,\sin\theta_1}{\cos\theta_2 + i\,\sin\theta_2} = \cos(\theta_1-\theta_2) + i\,\sin(\theta_1-\theta_2), \tag{3}$$

and

$$\frac{1}{\cos\theta + i\,\sin\theta} = \cos\theta - i\,\sin\theta. \tag{4}$$

PROOF. We shall use the following pair of familiar identities which you will remember from school.

$$\sin(\theta_1+\theta_2) = \sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1,$$
$$\cos(\theta_1+\theta_2) = \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2.$$

Notice that

$$(\cos\theta_1 + i\,\sin\theta_1)(\cos\theta_2 + i\,\sin\theta_2) = (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i\,(\sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1) = \cos(\theta_1+\theta_2) + i\,\sin(\theta_1+\theta_2).$$

This proves (2). The proof of (3) is left as an exercise. You have already proved (4) as an exercise, but do it again using (3). □

Theorem 13.2. *(De Moivre's Theorem) Suppose* θ *is real and* n *is an integer. Then*

$$(\cos\theta + i\,\sin\theta)^n = \cos n\theta + i\,\sin n\theta. \tag{5}$$

PROOF. We shall first prove De Moivre's Theorem for non-negative n, using induction. It is clearly true for $n = 0$. Now suppose that (5) holds for a certain non-negative n. Then

$$\begin{aligned}(\cos\theta + i\,\sin\theta)^{n+1} &= (\cos\theta + i\,\sin\theta)^n(\cos\theta + i\,\sin\theta)\\ &= (\cos n\theta + i\,\sin n\theta)(\cos\theta + i\,\sin\theta)\\ &= \cos\{(n+1)\theta\} + i\,\sin\{(n+1)\theta\} \qquad \text{using (2).}\end{aligned}$$

This shows that (5) is true with n replaced by $n+1$. By induction, the identity (5) holds for all non-negative integers n.

It remains to prove (5) for negative n. Thus suppose that n is negative and let $m = -n$. Since m is certainly positive, we know that

$$(\cos\theta + i\,\sin\theta)^m = \cos m\theta + i\,\sin m\theta. \tag{6}$$

Thus

$$\begin{aligned}(\cos\theta + i\,\sin\theta)^n &= \frac{1}{(\cos\theta + i\,\sin\theta)^m} \qquad \text{using } n = -m\\ &= \frac{1}{\cos m\theta + i\,\sin m\theta} \qquad \text{using (6)}\\ &= \cos m\theta - i\,\sin m\theta \qquad \text{using (4)}\\ &= \cos(-n\theta) - i\,\sin(-n\theta) \qquad \text{again from } m = -n\\ &= \cos n\theta + i\,\sin n\theta,\end{aligned}$$

where for the last step we used the well-known identities:

$$\sin(-\theta) = -\sin(\theta), \qquad \cos(-\theta) = \cos(\theta).$$

This completes the proof. □

Example 13.1. Let n be an integer. We will show that

$$\left(\sqrt{3} + i\right)^n + \left(\sqrt{3} - i\right)^n = 2^{n+1}\cos\frac{1}{6}n\pi.$$

Let $\alpha = \sqrt{3} + i$. Since we will be exponentiating, it is convenient to use the (r, θ)-form for α, which we have already worked out in Example 12.1:

$$\alpha = 2\left(\cos\frac{\pi}{6} + i\,\sin\frac{\pi}{6}\right).$$

By De Moivre's Theorem

$$\alpha^n = 2^n\left(\cos\frac{n\pi}{6} + i\,\sin\frac{n\pi}{6}\right)$$

Hence

$$\begin{aligned}\left(\sqrt{3} + i\right)^n + \left(\sqrt{3} - i\right)^n &= \alpha^n + \overline{\alpha^n}\\ &= 2^n\left(\cos\frac{n\pi}{6} + i\,\sin\frac{n\pi}{6}\right) + 2^n\left(\cos\frac{n\pi}{6} - i\,\sin\frac{n\pi}{6}\right)\\ &= 2^n\left(2\cos\frac{n\pi}{6}\right)\\ &= 2^{n+1}\cos\frac{n\pi}{6}.\end{aligned}$$

Example 13.2. Simplify

$$\frac{(\cos\theta - i\,\sin\theta)^5}{\cos 7\theta + i\,\sin 7\theta}.$$

Answer: From (4)

$$(\cos\theta - i\,\sin\theta)^5 = (\cos\theta + i\,\sin\theta)^{-5}.$$

By De Moivre,

$$\cos 7\theta + i\,\sin 7\theta = (\cos\theta + i\,\sin\theta)^7.$$

Thus

$$\begin{aligned}\frac{(\cos\theta - i\,\sin\theta)^5}{\cos 7\theta + i\,\sin 7\theta} &= \frac{(\cos\theta + i\,\sin\theta)^{-5}}{(\cos\theta + i\,\sin\theta)^7} \\ &= (\cos\theta + i\,\sin\theta)^{-12} \\ &= \cos(-12\,\theta) + i\,\sin(-12\,\theta) \\ &= \cos(12\,\theta) - i\,\sin(12\,\theta).\end{aligned}$$

Example 13.3. De Moivre's Theorem is useful for reconstructing many formulae involving trigonometric functions. For example, letting $n = 2$ in De Moivre's Theorem we see that

$$\begin{aligned}\cos 2\theta + i\,\sin 2\theta &= (\cos\theta + i\,\sin\theta)^2 \\ &= \cos^2\theta - \sin^2\theta + i \cdot 2\sin\theta\cos\theta.\end{aligned}$$

Comparing the real and imaginary parts, we get the well-known identities

$$\cos 2\theta = \cos^2\theta - \sin^2\theta, \qquad \sin 2\theta = 2\,\sin\theta\cos\theta.$$

It is possible that you might forget one of these identities. There is however no excuse for being unable to reconstruct them using De Moivre.

It is useful to know that the identity for $\cos 2\theta$ is often given in an alternative form:

$$\cos 2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta.$$

It is straightforward to deduce this from the previous identity for $\cos 2\theta$ using $\cos^2\theta + \sin^2\theta = 1$.

Exercise 13.4. Let α, β be non-zero complex numbers. Suppose that the points P, Q represent α and β on the complex plane. Show that OP is perpendicular to OQ if and only if α/β is imaginary.

14. $e^{i\theta}$

Definition. Let θ be a real number. Define $e^{i\,\theta}$ by

$$e^{i\theta} = \cos\theta + i\,\sin\theta.$$

Let $\alpha = \phi + i\theta$, where ϕ and θ are real numbers. Define

$$e^{\alpha} = e^{\phi} \cdot e^{i\theta} = e^{\phi}\,(\cos\theta + i\,\sin\theta).$$

Here e^{ϕ} has the usual meaning.

Remark. You have probably seen the definition of $e^{i\theta}$ before, but it is surprising and in need of some justification. The following is a **heuristic** argument that is intended to convince you that such a definition is reasonable. Let

$$z = \cos\theta + i\,\sin\theta$$

and think of this z as a function of θ. Differentiating with respect to θ we get

$$\begin{aligned} z' &= -\sin\theta + i\,\cos\theta \\ &= i\,(\cos\theta + i\,\sin\theta) = i\,z. \end{aligned}$$

Thus $z'/z = i$; integrating both sides with respect to θ gives

$$\int \frac{z'}{z}\,\mathrm{d}\theta = \int i\,\mathrm{d}\theta,$$

and so [7]

$$\log z = i\theta + C$$

where C is the integration constant. To find the integration constant C, substitute $\theta = 0$ in both sides, noting that $z(0) = \cos 0 + i\sin 0 = 1$. Hence $C = 0$ which gives $\log z = i\theta$ and so $e^{i\theta} = z = \cos\theta + i\sin\theta$.

We should again reiterate that the above argument is heuristic; we haven't even defined what integration in the complex number setting means. It only suggests that our definition is plausible.

Exercise 14.1. Let $z = \pi/6 + i\log 2$. Write $e^{i\,z}$ in the form $a + b\,i$. (Careful! This is a trick question.)

Exercise 14.2. Let $\alpha = \phi + i\,\theta$ where ϕ and θ are real.

(i) Simplify $|e^{\alpha}|$ and $|e^{i\,\alpha}|$.
(ii) Show that the conjugate of e^{α} is $e^{\overline{\alpha}}$.

15. Expressing $\cos^n\theta$ and $\sin^n\theta$ in terms of Multiple Angles

Let $z = \cos\theta + i\sin\theta$. By De Moivre's Theorem

$$z^n = \cos n\theta + i\sin n\theta, \qquad \frac{1}{z^n} = \cos n\theta - i\sin n\theta.$$

Hence

$$z^n + \frac{1}{z^n} = 2\,\cos n\theta, \qquad z^n - \frac{1}{z^n} = 2i\,\sin n\theta \tag{7}$$

This can help us derive expressions for $\cos^n\theta$ and $\sin^n\theta$ which involve multiple angles but no powers. Make sure you understand the following example.

Example 15.1. Express $\cos^5\theta$ in terms of multiple angles. Hence calculate

$$\int \cos^5\theta\mathrm{d}\theta.$$

[7] In university Mathematics, $\log x$ is almost always $\log_e x$ or what you called $\ln x$ at school.

Answer: Write $z = \cos\theta + i\sin\theta$. Thus $z + 1/z = 2\cos\theta$. Hence

$$\begin{aligned} 2^5\cos^5\theta &= \left(z + \frac{1}{z}\right)^5 \\ &= z^5 + 5z^3 + 10z + \frac{10}{z} + \frac{5}{z^3} + \frac{1}{z^5} \\ &= \left(z^5 + \frac{1}{z^5}\right) + 5\left(z^3 + \frac{1}{z^3}\right) + 10\left(z + \frac{1}{z}\right) \\ &= 2\cos 5\theta + 10\cos 3\theta + 20\cos\theta. \end{aligned}$$

Hence

$$\cos^5\theta = \frac{1}{16}\cos 5\theta + \frac{5}{16}\cos 3\theta + \frac{5}{8}\cos\theta.$$

Thus

$$\int \cos^5\theta \mathrm{d}\theta = \frac{1}{80}\sin 5\theta + \frac{5}{48}\sin 3\theta + \frac{5}{8}\sin\theta + K.$$

Exercise 15.2. Express $\sin^4\theta$ in terms of multiple angles and hence evaluate

$$\int_0^{\pi/2} \sin^4\theta \mathrm{d}\theta.$$

16. The Exponential Form of Complex Numbers

Let α be a non-zero complex number. Lemma 12.1 tells us that we may write

$$\alpha = r(\cos\theta + i\,\sin\theta)$$

where r and θ are respectively the absolute value and the argument of α. We also recall that $e^{i\theta} = \cos\theta + i\,\sin\theta$. Thus we arrive at a very convenient representation of complex numbers.

Lemma 16.1. *Let α be a non-zero complex number. Then*

$$\alpha = re^{i\,\theta} \tag{8}$$

where $r = |\alpha|$ and θ is the argument of α.

We call $re^{i\,\theta}$ the *exponential form* of the complex number α. The exponential form of complex numbers is very useful for multiplication, division and exponentiation of complex numbers.

Lemma 16.2. *Suppose r_1, r_2, r, θ_1, θ_2, θ_3 are real. Then*

$$r_1e^{i\theta_1} \cdot r_2e^{i\theta_2} = r_1r_2e^{i\,(\theta_1+\theta_2)}, \tag{9}$$

$$\frac{r_1e^{i\theta_1}}{r_2e^{i\theta_2}} = \frac{r_1}{r_2}e^{i\,(\theta_1-\theta_2)}, \tag{10}$$

and

$$\overline{(re^{i\,\theta})} = re^{-i\,\theta}. \tag{11}$$

Moreover, for n an integer,

$$\left(re^{i\,\theta}\right)^n = r^n e^{i\,n\theta}. \tag{12}$$

PROOF. You should be able to deduce this theorem from Lemma 13.1 and Theorem 13.2. □

Example 16.1. In this example, we calculate

$$\int e^{\theta} \sin\theta \mathrm{d}\theta. \tag{13}$$

The calculation is again not rigorous, but we can check the answer at the end as we shall see. Notice that $\sin\theta$ is the imaginary part of $e^{i\,\theta}$. Thus

$$e^{\theta}\sin\theta = \operatorname{Im}\left(e^{\theta}e^{i\,\theta}\right),$$

and so

$$\int e^{\theta}\sin\theta \mathrm{d}\theta = \operatorname{Im}\left(\int e^{\theta}e^{i\,\theta}\mathrm{d}\theta\right).$$

Hence the desired integral in (13) is the imaginary part of the integral

$$\begin{aligned}
\int e^{\theta}e^{i\,\theta}\mathrm{d}\theta &= \int e^{(i+1)\,\theta}\mathrm{d}\theta \\
&= \frac{1}{i+1}e^{(i+1)\,\theta} + C \\
&= \frac{(1-i)}{2}e^{\theta}\left(\cos\theta + i\sin\theta\right) + C \\
&= \frac{1}{2}e^{\theta}\left((\cos\theta + \sin\theta) + i(\sin\theta - \cos\theta)\right) + C.
\end{aligned}$$

Write the integration constant $C = L + iK$ where L and K are real. Thus we believe on the basis of the above (non-rigorous) argument that

$$\int e^{\theta}\sin\theta\mathrm{d}\theta = \frac{1}{2}e^{\theta}(\sin\theta - \cos\theta) + K. \tag{14}$$

But now, having guessed what the integral should be, we can check it by differentiating:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{1}{2}e^{\theta}(\sin\theta - \cos\theta) + K\right) = e^{\theta}\sin\theta.$$

We have therefore proved that the correctness of (14).

Example 16.2. Use what you know about $e^{i\,\theta}$ to simplify

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{2^n}.$$

Answer: Note that the required sum is the real part of

$$\begin{aligned}
\sum_{n=0}^{\infty}\frac{\cos(n\theta) + i\sin(n\theta)}{2^n} &= \sum_{n=0}^{\infty}\left(\frac{e^{i\theta}}{2}\right)^n \\
&= \frac{1}{1-\frac{e^{i\theta}}{2}} \\
&= \frac{2}{2-\cos\theta - i\sin\theta} \\
&= \frac{2(2-\cos\theta + i\sin\theta)}{(2-\cos\theta)^2 + \sin^2\theta} \\
&= \frac{2(2-\cos\theta + i\sin\theta)}{5-4\cos\theta}.
\end{aligned}$$

Hence

$$\sum_{n=0}^{\infty} \frac{\cos(n\theta)}{2^n} = \frac{4 - 2\cos\theta}{5 - 4\cos\theta}.$$

Example 16.3. Use what you know about $e^{i\theta}$ to simplify

$$1 + \cos\left(\frac{\pi}{10}\right) + \cos\left(\frac{2\pi}{10}\right) + \cdots + \cos\left(\frac{9\pi}{10}\right). \tag{15}$$

Answer: We know that $\cos n\theta$ is the real part of $e^{in\theta}$. Thus the sum (15) is the real part of

$$\begin{aligned}
1 + e^{\pi i/10} + e^{2\pi i/10} + \cdots + e^{9\pi i/10} &= \frac{e^{10\pi i/10} - 1}{e^{\pi i/10} - 1} \\
&= \frac{-2}{\cos(\pi/10) - 1 + i\sin(\pi/10)} \\
&= -2\frac{\cos(\pi/10) - 1 - i\sin(\pi/10)}{(\cos(\pi/10) - 1)^2 + \sin^2(\pi/10)} \\
&= -2\frac{\cos(\pi/10) - 1 - i\sin(\pi/10)}{\cos^2(\pi/10) + \sin^2(\pi/10) - 2\cos(\pi/10) + 1} \\
&= -2\frac{\cos(\pi/10) - 1 - i\sin(\pi/10)}{2 - 2\cos(\pi/10)} \\
&= \frac{\cos(\pi/10) - 1 - i\sin(\pi/10)}{\cos(\pi/10) - 1}.
\end{aligned}$$

Taking the real part we get

$$1 + \cos\left(\frac{\pi}{10}\right) + \cos\left(\frac{2\pi}{10}\right) + \cdots + \cos\left(\frac{9\pi}{10}\right) = \frac{\cos(\pi/10) - 1}{\cos(\pi/10) - 1} = 1.$$

Exercise 16.4. Let α be a complex number. Describe geometrically what happens to α (in the complex plane) when it is multiplied by $e^{i\phi}$ (where ϕ is real).
Hint: write α in exponential form.

17. Quadratic Equations—Read on your own

Here is some 'baby-stuff' for you to read on your own (and **yes it is examinable**, though very easy). The methods that you learned at school for solving quadratic equations still work at university—even when dealing with equations with complex coefficients.

Theorem 17.1. *(Quadratic formula) Suppose a, b, c are complex numbers with $a \neq 0$. Then the solutions to the quadratic equation*

$$ax^2 + bx + c = 0$$

are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

PROOF. Our first step is to divide by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Next we complete the square, which means that we write the expression on the left as a perfect square, plus a constant:

$$\left(x+\frac{b}{2a}\right)^2-\frac{b^2}{4a^2}+\frac{c}{a}=0.$$

Reorganizing, we get

$$\left(x+\frac{b}{2a}\right)^2=\frac{b^2-4ac}{4a^2}.$$

Square-rooting both sides we get

$$x+\frac{b}{2a}=\pm\frac{\sqrt{b^2-4ac}}{2a}.$$

Taking $b/2a$ to right-hand side gives us the quadratic formula. □

Make sure you have understood the argument in the proof of the quadratic formula before you answer this question.

Exercise 17.1. Solve the equation $(x-i+1)^2=-4$. (**Please, please, don't say, "expand the brackets, rearrange and use the quadratic formula"!!!**).

18. The Fundamental Theorem of Algebra

Theorem 18.1. *Suppose f is a polynomial of degree n with coefficients in $\mathbb{C}$. Then f has n roots in $\mathbb{C}$ (counting multiplicities).*

You do not need to know a proof of the Fundamental Theorem of Algebra for this course. You will find a proof in the Lecture Notes for the Foundations course. However, if in your third year you take the Complex Analysis course you will see a one line proof of this theorem.

The Fundamental Theorem of Algebra is very useful to know, but does not tell you how to find the roots. For this you have to turn to other methods. Suppose however that somehow (e.g. by searching) you have managed to find n roots of f; then the Fundamental Theorem tells you that you have found them all.

Example 18.1. Find the roots of the polynomial

$$f(X)=X^4-4\,X^3+4\,X^2+4\,X-5.$$

Answer: We do a little search for roots and find that $f(1)=f(-1)=0$. Hence $(X-1)(X+1)=X^2-1$ is a factor of f. Dividing f by this factor we discover that

$$f(X)=(X-1)(X+1)(X^2-4\,X+5).$$

Let $g(X)=X^2-4\,X+5$. Using the 'quadratic formula' we find that the roots of g are

$$\frac{4\pm\sqrt{16-20}}{2}=\frac{4\pm 2\,i}{2}=2\pm i.$$

Hence the roots of $f(X)$ are

$$-1,\quad 1,\quad 2+i,\quad 2-i.$$

Exercise 18.2. Let f be a polynomial with real coefficients.

(i) Show that α is a root of f if and only if $\overline{\alpha}$ is also a root.
(ii) Show that the number of non-real roots of f must be even.
(iii) If f has an odd degree, show that f must have a real root.

Remark. Part (iii) has profound implications for geometry. For example, it follows from (iii) that every object rotating in three dimensions must have some fixed axis about which it rotates.

18.1. Formulae for solutions of polynomial equations? You might be wondering if there are formulae for solving polynomial equations of degree ≥ 3. We have to be a little careful by what we mean by 'solving', since polynomial equations can always be solved by numerical methods, which give an approximate solution to whatever degree of accuracy we like. But what you probably have in mind is a formula like the quadratic formula above, in which you would substitute the coefficients of the equation and after a few operations of addition, subtraction, multiplication and extractions of n-th roots will give us the solutions. This is what historically was called 'solubility by radicals'; a **radical** is just an n-th root. It turns out that there are such formulae for cubic and quartic equations (i.e. degrees 3 and 4), called Cardano's formulae, and have been known since the 16th Century. A major question for about two centuries thereafter was whether the quintic (i.e. degree 5) can be solved by radicals, until Niels Henrik Abel (1802–1829) and Évariste Galois (1811–1832) showed that this is impossible [8].

19. n-th Roots

Just as the exponential form makes it easy to multiply and divide complex numbers, so it also makes it easy to find the n-th roots of complex numbers.

The trigonometric function $\sin\theta$ is periodic with period 2π. Thus if $\theta_1 - \theta_2 = 2\pi k$ where k is an integer, then $\sin\theta_1 = \sin\theta_2$. However the converse does not have to be true. By this we mean, if $\sin\theta_1 = \sin\theta_2$ then it is not necessarily true that $\theta_1 - \theta_2 = 2\pi k$ for some integer k. For example $\sin\pi/4 = \sin 3\pi/4$.

However, the function $e^{i\theta}$ has a very attractive property.

Lemma 19.1. *The function $e^{i\theta}$ is periodic with period 2π. Moreover, $e^{i\theta_1} = e^{i\theta_2}$ if and only if $\theta_1 - \theta_2 = 2\pi k$ for some integer k.*

The lemma follows from the properties of sin and cos.

Lemma 19.2. *Suppose α and β are non-zero complex numbers with exponential forms*

$$\alpha = re^{i\theta}, \qquad \beta = se^{i\phi}.$$

Suppose that n is a positive integer. Then $\alpha^n = \beta$ if and only if

$$r = s^{1/n}, \qquad \theta = \frac{\phi + 2\pi k}{n} \tag{16}$$

for some integer k.

Proof. Suppose that $\alpha^n = \beta$. Note that

$$r^n = |\alpha^n| = |\beta| = s.$$

But r and s are positive, so $r = s^{1/n}$. Canceling $r^n = s$ from $\alpha^n = \beta$, we get

$$e^{in\theta} = e^{i\phi}.$$

[8] This is proved in the remarkable Galois Theory (third year) module. Galois Theory answers—almost effortlessly—many classical problems that had vexed mathematicians for thousands of years; for example, is it possible to trisect an angle using a ruler and compass?

From Lemma 19.1 we see that

$$n\theta = \phi + 2\pi k,$$

for some integer k. Dividing by n gives (16).

Conversely, suppose that (16) holds for some integer k. Then

$$\alpha^n = r^n e^{i\,n\theta} = s\,e^{i\,\phi} \cdot e^{2\pi i\,k} = s\,e^{i\,\phi} = \beta,$$

as required. □

Apparently, the Lemma gives us infinitely many n-th roots of a complex number β: one for each value of k. This is not so, for n-th roots of β are the roots of the polynomial $X^n - \beta$ and so (by the Fundamental Theorem of Algebra) there are n of them. In fact there is repetition. The following theorem gives the n-th roots without repetition.

Theorem 19.3. *Let β be a non-zero complex number and let its exponential form be*

$$\beta = se^{i\,\phi}.$$

The n-th roots of β are

$$s^{1/n} \exp\left(\frac{(\phi + 2\pi\,k)i}{n}\right), \qquad k = 0, 1, 2, \ldots, n-1.$$

Proof. Let

$$\alpha_k = s^{1/n} \exp\left(\frac{(\phi + 2\pi\,k)i}{n}\right).$$

It is clear from the above lemma that α_k is an n-th root of β, and that any n-th root of β is of this form. We want to show that

$$\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \tag{17}$$

are all the nth roots of β and there are none missing. Thus we want to show that if m is any integer then α_m is equal to one of the alphas in the list (17). But notice that

$$\alpha_{m+n} = \alpha_m \exp(2\pi\,i) = \alpha_m.$$

So we find

$$\cdots = \alpha_{m-2n} = \alpha_{m-n} = \alpha_m = \alpha_{m+n} = \alpha_{m+2n} = \ldots$$

Now if m is any integer, we can add or subtract a multiple of n to get an integer between 0 and $n-1$. So α_m does really belong to the list (17). □

Example 19.1. Find the cubic roots of -2.

Answer: We note first that

$$-2 = 2\exp(\pi i).$$

Thus from the Theorem, the cube roots of -2 are

$$2^{1/3} \exp\left(\frac{(\pi + 2\pi k)i}{3}\right), \qquad k = 0, 1, 2.$$

These are

$$2^{1/3}\exp\left(\frac{\pi i}{3}\right) = 2^{1/3}\left(\cos\pi/3 + i\,\sin\pi/3\right) = 2^{1/3}\left(\frac{1}{2} + i\,\frac{\sqrt{3}}{2}\right)$$

$$2^{1/3}\exp\left(\pi\,i\right) = -2^{1/3}$$

$$2^{1/3}\exp\left(\frac{5\pi i}{3}\right) = 2^{1/3}\left(\cos 5\pi/3 + i\,\sin 5\pi/3\right) = 2^{1/3}\left(\frac{1}{2} - i\,\frac{\sqrt{3}}{2}\right).$$

20. The n-th Roots of Unity

It is worthwhile looking a little more closely at the n-th roots of 1. We can write 1 in exponential form as $1 = 1\exp(0 \cdot i)$. Theorem 19.3 tells us that the n-th roots of 1 are

$$\exp\left(\frac{2\pi\,k\,i}{n}\right) = \cos\frac{2\pi\,k}{n} + i\,\sin\frac{2\pi\,k}{n}, \qquad k = 0, 1, 2, \ldots, n-1.$$

If we write

$$\zeta = \exp\left(\frac{2\pi\,i}{n}\right)$$

then we see that the n-th roots of unity are

$$1, \zeta, \zeta^2, \ldots, \zeta^{n-1}.$$

It is easy to sketch the n-th roots of unity on the complex plane. They all have absolute value 1, so they lie on the circle with radius 1 and centre at the origin. The first one to draw is 1; you know where that one is. The next one is ζ. This is the one you get if start at 1 go around the circle in an anticlockwise direction through an angle of $2\pi/n$. To get ζ^2, start at ζ and go around the circle in an anticlockwise direction through an angle of $2\pi/n$, and so on. The points $1, \zeta, \ldots, \zeta^{n-1}$ are equally spaced around the circle with an angle $2\pi/n$ between each one and the next. See Figure 5 for the cube and fourth roots of unity.

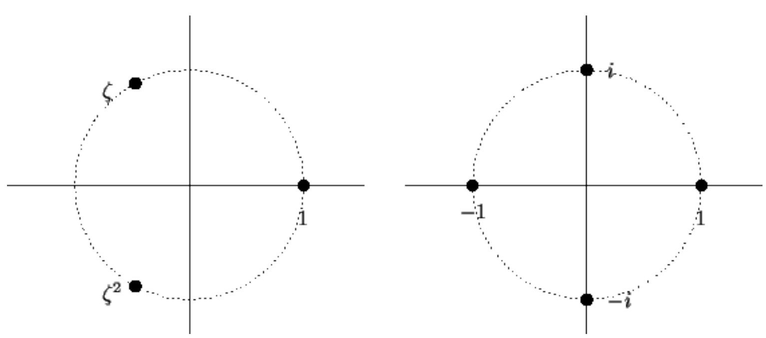

FIGURE 5. On the left, the three cube roots of unity: here $\zeta = e^{2\pi i/3}$. On the right, the fourth roots of unity. Note that $e^{2\pi i/4} = e^{\pi i/2} = i$, so the fourth roots of unity are 1, i, $i^2 = -1$, and $i^3 = -i$.

Example 20.1. What is the sum of the n-th roots of unity?

Answer:

$$1+\zeta+\zeta^2+\cdots+\zeta^{n-1} = \frac{\zeta^n - 1}{\zeta - 1} = \frac{1-1}{\zeta-1} = 0.$$

Example 20.2. Write down the cube roots of unity.

Answer: We can (and will) use the above recipe to write the cube roots of unity. But there is another (easier) way: the cube roots of unity are the roots of the polynomial $X^3 - 1$. Note

$$X^3 - 1 = (X-1)(X^2 + X + 1).$$

Thus (using the quadratic formula), the cube roots of unity are

$$1, \qquad \frac{-1+i\sqrt{3}}{2}, \qquad \frac{-1-i\sqrt{3}}{2}.$$

Having found them, what is the point of using the above recipe to find the cube roots of unity? Well, knowing the solution beforehand will allow us to check that the recipe that we wrote down is correct.

Using the above recipe we find that the cube roots of unity are

$$\exp\left(\frac{2\pi\, k\, i}{3}\right) = \cos\frac{2\pi\, k}{3} + i\,\sin\frac{2\pi\, k}{3}, \qquad k = 0, 1, 2, \ldots, n-1.$$

These are

$$\begin{aligned} \cos 0 + i\,\sin 0 &= 1, \\ \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} &= \frac{-1+i\sqrt{3}}{2}, \\ \cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} &= \frac{-1-i\sqrt{3}}{2}. \end{aligned}$$

Whilst it is always true that the n-th roots of unity are the roots of $X^n - 1$, for large values of n it is not convenient to use this fact to write down the n-th roots.

A Very Important Summary. Repeat after me: [9] There are two square-roots of unity [10] and they add up to 0; there are three cube-roots of unity and they add up to 0; there are four fourth-roots of unity and they add up to 0; there are five fifth-roots of unity and they add up to 0; ...

Exercise 20.3. Sketch the fifth and sixth roots of unity.

21. Why ℂomplicate Things?

I will not inflict on you a long and eloquent speech on the merits of the complex number system. But there is perhaps a persistent doubt at the back of your mind as to their usefulness. You might be saying to yourself, "we live in the real world where physical quantities are represented by real numbers". In truth, complex numbers are not esoteric objects that became fashionable through the efforts of

[9] This bit might seem patronizing—sorry! But a favourite trick question of several maths lecturers is "how many ... and what is their sum?". If you like to vex them, make sure you give the wrong answer, but be prepared for the consequences ...

[10] Just in case you still don't know, unity is another name for 1.

nerdy mathematicians with heads in the clouds. They are used by practical (hard-nosed) people to solve real world problems. Indeed, it often turns out that a solution to a real world problem is through complex numbers. If unconvinced, talk to a friend in electrical engineering. But before you do that, try this question.

Exercise 21.1. You are going to solve the cubic equation $x^3 + x^2 - 2x - 1 = 0$. You know from the Fundamental Theorem of Algebra that you are looking for 3 solutions.

(i) You might want to try first whatever tricks you learned from school—if any.

(ii) Now, assuming that does not work, try the following substitution: $x = y + 1/y$ the clear denominators and simplify (the secret behind hitting on the right substitution lies in algebraic number theory). You will get

$$y^6 + y^5 + y^4 + y^3 + y^2 + y + 1 = 0.$$

(iii) Can you see why y must be a 7th root of unity that is not 1?

(iv) By writing down the 7th roots of unity show that x is one of

$$2\cos\left(\frac{2\pi}{7}\right), \quad 2\cos\left(\frac{4\pi}{7}\right), \quad \dots, \quad 2\cos\left(\frac{12\pi}{7}\right).$$

(v) Altogether we have found 6 solutions to our cubic equation in x whilst we were expecting 3. There is in fact some repetition in the above list that is a side effect of our non-linear substitution $x = y + 1/y$. Eliminate the repetition to find the 3 roots.

Moral: The moral of the above is that to write down the 3 real roots of a cubic equation we had to go through complex numbers. This is not the first time where we deduced a 'real' fact on the basis of 'complex' arguments. Recall that in Example 16.1 we determined the integral of a real function using the complexes.

CHAPTER 3

Vectors (at Last)

1. Vectors in $\mathbb{R}^n$

We define $\mathbb{R}^2$ to be the set

$$\mathbb{R}^2 = \{(x_1, x_2) \ : \quad x_1,\, x_2 \in \mathbb{R}\}.$$

That is, $\mathbb{R}^2$ is the set of ordered pairs of real numbers. We call $\mathbb{R}^2$ *Euclidean 2-space*. There are two ways of thinking about an element $(a, b) \in \mathbb{R}^2$.

- The first is to think of (a, b) as the point $P(a, b)$ in the coordinate plane with x-coordinate a and y-coordinate b.
- The second is to think of (a, b) as the vector $\overrightarrow{OP}$ (directed line segment) which starts at the origin O and ends at the point $P(a, b)$.

When referring to an element of $\mathbb{R}^2$, it will probably be clear which interpretation we have in mind. When we talk about points we use uppercase letters to represent them, as in "the point P". When we talk about vectors we use bold lowercase letters, as in "the vector $\mathbf{v}$".

Likewise, elements of $\mathbb{R}^3$ can be thought of as points or vectors in 3-dimensional space. Of course $\mathbb{R}^3$ is just:

$$\mathbb{R}^3 = \{(x_1, x_2, x_3) \ : \quad x_1,\, x_2,\, x_3 \in \mathbb{R}\}.$$

Euclidean n-space is just the set $\mathbb{R}^n$ defined by

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \ : \quad x_1, x_2, \dots, x_n \in \mathbb{R}\}.$$

You should not be scared of Euclidean n-space for $n \geq 4$. Working in 4- or 5-dimensional space is not really any more demanding than 2- or 3-dimensional space. The intuition that we need for thinking about $\mathbb{R}^n$ will come from $\mathbb{R}^2$ and $\mathbb{R}^3$. Practically all concepts, operations, theorems etc. on $\mathbb{R}^n$ are fairly trivial extensions of those on $\mathbb{R}^2$ and $\mathbb{R}^3$.

We call the vector $(x_1, x_2, \dots, x_n)$ the **position vector** of the point $P(x_1, x_2, \dots, x_n)$ in Euclidean n-space.

Example 1.1. $P(1, -1, 3, 4)$ and $Q(-1, 0, -1, 1)$ are points in $\mathbb{R}^4$ (or Euclidean 4-space). But also $\mathbf{u} = (1, -1, 3, 4)$ and $\mathbf{v} = (-1, 0, -1, 1)$ are vectors in $\mathbb{R}^4$. In fact $\mathbf{u}$ and $\mathbf{v}$ are respectively the position vectors of P and Q. We will see later that the vectors $\mathbf{u}$ and $\mathbf{v}$ are perpendicular. This we will show using the dot product which you have learned at school for $\mathbb{R}^2$ and $\mathbb{R}^3$, but works in the same way for $\mathbb{R}^4$.

2. First Definitions

Two vectors $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$ are said to be equal if their corresponding components are equal. For example, in $\mathbb{R}^2$ this means the following: suppose $\mathbf{u} = (u_1, u_2)$ and

$\mathbf{v} = (v_1, v_2)$ then $\mathbf{u} = \mathbf{v}$ if and only if $u_1 = v_1$ and $u_2 = v_2$. In $\mathbb{R}^n$, we say that $\mathbf{u} = (u_1, \dots, u_n)$ and $\mathbf{v} = (v_1, \dots, v_n)$ are equal if $u_i = v_i$ for $i = 1, 2, \dots, n$.

The *zero vector* (written $\mathbf{0}$) is the vector with all components equal to 0. Thus the zero vector in $\mathbb{R}^n$ is

$$\mathbf{0} = (\underbrace{0, 0, \dots, 0}_{n \text{ times}}).$$

The *length* (or *norm*) of a vector $\mathbf{u} = (u_1, u_2)$ in $\mathbb{R}^2$ is defined by

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2}.$$

For $\mathbf{u} = (u_1, u_2, u_3)$ in $\mathbb{R}^3$ the length is defined by

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$

In general, the length of $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$ is

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

If $\|\mathbf{u}\| = 1$, the vector $\mathbf{u}$ is said to be a *unit vector*.

Example 2.1. Find the length of the vector $\mathbf{u} = (-2, 1, -1, 3, 0)$.

Answer:

$$\|\mathbf{u}\| = \sqrt{(-2)^2 + 1^2 + (-1)^2 + 3^2 + 0^2} = \sqrt{15}.$$

Example 2.2. Find the values of a making the vector $\mathbf{v} = (1/5, 2/5, 2/5, a)$ a unit vector.

Answer: Note

$$\|\mathbf{v}\| = \sqrt{(1/5)^2 + (2/5)^2 + (2/5)^2 + a^2} = \sqrt{9/25 + a^2}.$$

Then $\mathbf{v}$ is a unit vector if and only if $\|\mathbf{v}\| = 1$, or in other words:

$$9/25 + a^2 = 1.$$

This gives $a = \pm 4/5$.

3. Addition and Scalar Multiplication

When we talk about vectors we call real numbers *scalars*.

Definition. Let $\mathbf{u} = (u_1, u_2, \dots, u_n)$ and $\mathbf{v} = (v_1, v_2, \dots, v_n)$ be two vectors in $\mathbb{R}^n$. We define the **sum**

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n).$$

Let λ be a scalar (i.e. a real number). We define the **scalar multiple** of $\mathbf{u}$ by λ

$$\lambda\, \mathbf{u} = (\lambda\, u_1, \lambda\, u_2, \dots, \lambda\, u_n).$$

Example 3.1. Let $\mathbf{u} = (3, -1, -2, -3)$ and $\mathbf{v} = (-3, -2, 5, 1)$. Then $\mathbf{u} + \mathbf{v} = (0, -3, 3, -2)$ and $2\,\mathbf{u} = (6, -2, -4, -6)$.

Geometrically, the sum $\mathbf{u} + \mathbf{v}$ can be described by the familiar ‘completing the parallelogram’ process. See Figure 3.

If the scalar λ is positive then $\lambda\mathbf{u}$ has the same direction as $\mathbf{u}$, and if λ is negative $\lambda\mathbf{u}$ has the opposite direction to $\mathbf{u}$.

Lemma 3.1. *Let* $\mathbf{u}$ *be a vector and* λ *a scalar. Then* $\|\lambda\, \mathbf{u}\| = |\lambda|\, \|\mathbf{u}\|$.

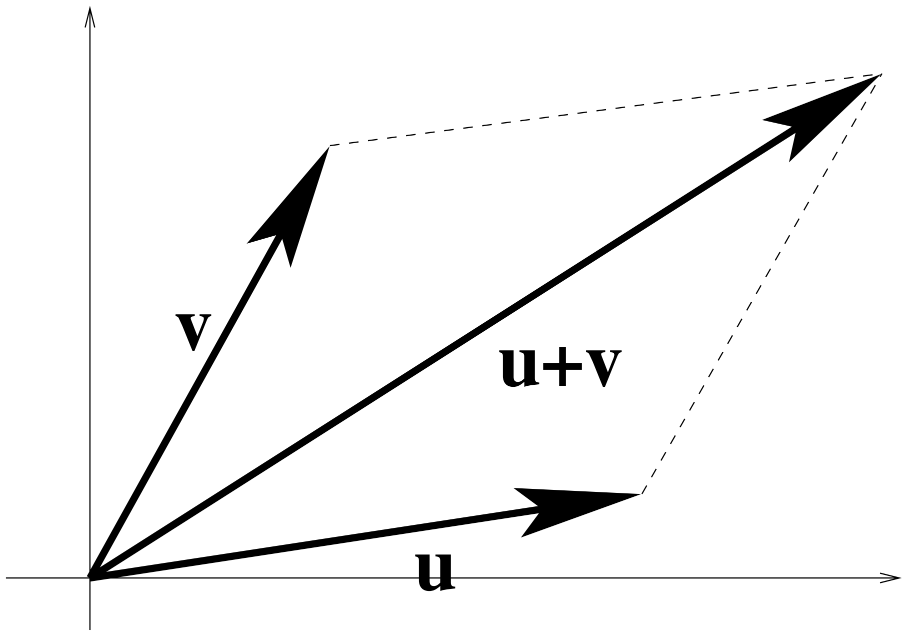

FIGURE 1. Geometrically, the sum $\mathbf{u}+\mathbf{v}$ can be described by the familiar 'completing the parallelogram' process.

PROOF. Write $\mathbf{u} = (u_1, u_2, \dots, u_n)$. Then $\lambda\,\mathbf{u} = (\lambda\,u_1, \lambda\,u_2, \dots, \lambda\,u_n)$, so

$$\begin{aligned}\|\lambda\mathbf{u}\| &= \sqrt{(\lambda\,u_1)^2 + (\lambda\,u_2)^2 + \cdots + (\lambda\,u_n)^2} \\ &= \sqrt{\lambda^2}\sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = |\lambda|\,\|\mathbf{u}\|.\end{aligned}$$

□

Example 3.2. For a non-zero vector $\mathbf{v}$, find a unit vector with the same direction.

Answer: We know that if the scalar λ is positive then $\lambda\mathbf{v}$ has the same direction as $\mathbf{v}$. We must choose $\lambda > 0$ so that $\lambda\mathbf{v}$ is a unit vector. In other words, we want $\|\lambda\mathbf{v}\| = 1$. But $\|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\| = \lambda\,\|\mathbf{v}\|$. So we must choose $\lambda = 1/\|\mathbf{v}\|$ and we see that $\dfrac{1}{\|\mathbf{v}\|}\mathbf{v}$ is a unit vector with the same direction as $\mathbf{v}$.

For a vector $\mathbf{u} \in \mathbb{R}^n$ we define $-\mathbf{u} = (-1)\mathbf{u}$. Thus if $\mathbf{u} = (u_1, \dots, u_n)$ then $-\mathbf{u} = (-u_1, \dots, -u_n)$. From what we know about scalar multiplication $-\mathbf{u}$ has the same length as $\mathbf{u}$ but opposite direction. For vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$ we define their difference to be $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$.

Exercise 3.3. Let

$$\mathbf{u} = (-1, 2, 1, 0), \qquad \mathbf{v} = (0, 1, 3, -1), \qquad \mathbf{w} = (-2, 3, 0, 5).$$

Perform the following operations.

(i) $\mathbf{u}+\mathbf{v}$ (ii) $2\mathbf{u}-\mathbf{v}+\mathbf{w}$ (iii) $3(\mathbf{u}-\mathbf{v})+\mathbf{w}$

(iv) $\|\mathbf{v}+\mathbf{w}\|$ (v) $\dfrac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|}$.

Exercise 3.4. Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be vectors in $\mathbb{R}^7$ and λ be a scalar. Which of the following operations is **not** defined?

(i) $2\lambda+\mathbf{v}$ (ii) $\mathbf{w}+\lambda\mathbf{v}$

(iii) $\|\mathbf{v}\|-\lambda\mathbf{u}$ (iv) $\|\mathbf{u}\|\,\mathbf{v}-\lambda\mathbf{u}$

Exercise 3.5. Let $\mathbf{v}$ be a non-zero vector and $\mathbf{w}$ the unit vector of the opposite direction to $\mathbf{v}$. Write $\mathbf{w}$ in terms of $\mathbf{v}$.

4. Geometric Interpretation of Subtraction

Recall that geometrically a vector is a directed line segment $\overrightarrow{OP}$ with O being the origin. For any two points P, Q, we can also consider directed line segments of the from P to Q which we denote by $\overrightarrow{PQ}$. We will say that two directed line segments $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ are **equivalent** if they have the same length and direction. See Figure 2.

The following theorem shows that any directed line segment is equivalent to a vector.

Theorem 4.1. *Let P and Q be points in $\mathbb{R}^n$ and let $\mathbf{u}$ and $\mathbf{v}$ be respectively the vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$. Then the directed line segment $\overrightarrow{PQ}$ is equivalent to $\mathbf{v}-\mathbf{u}$. In other words $\overrightarrow{PQ}$ is parallel to the vector $\mathbf{v}-\mathbf{u}$ and the distance from P to Q is equal to $\|\mathbf{v}-\mathbf{u}\|$.*

Proof. Note that $\mathbf{u}+(\mathbf{v}-\mathbf{u})=\mathbf{v}$. Now think back to the geometric interpretation of vector addition in terms of the parallelogram. The vectors $\mathbf{u}$ and $\mathbf{v}-\mathbf{u}$ form two adjacent sides of a parallelogram with $\mathbf{v}$ (their sum) in the diagonal; see Figure 3. It is clear that $\overrightarrow{PQ}$ and $\mathbf{v}-\mathbf{u}$ have the same direction and length: in other words they are equivalent. □

Example 4.1. Find the distance between the point $P(2,3,1,-1)$ and the point $Q(1,2,-1,1)$.

Answer: Notice that $\overrightarrow{PQ}$ is equivalent to the vector $(1,2,-1,1)-(2,3,1,-1)=(-1,-1,-2,2)$. Hence the distance from P to Q is equal to

$$\|(-1,-1,-2,2)\|=\sqrt{(-1)^2+(-1)^2+(-2)^2+2^2}=\sqrt{10}.$$

Exercise 4.2. Let $\mathbf{r}_0=(-1,1)$. Find all vectors $\mathbf{r}=(x,y)$ satisfying

$$\|\mathbf{r}\|=\|\mathbf{r}-\mathbf{r}_0\|=\sqrt{5}.$$

Draw a picture.

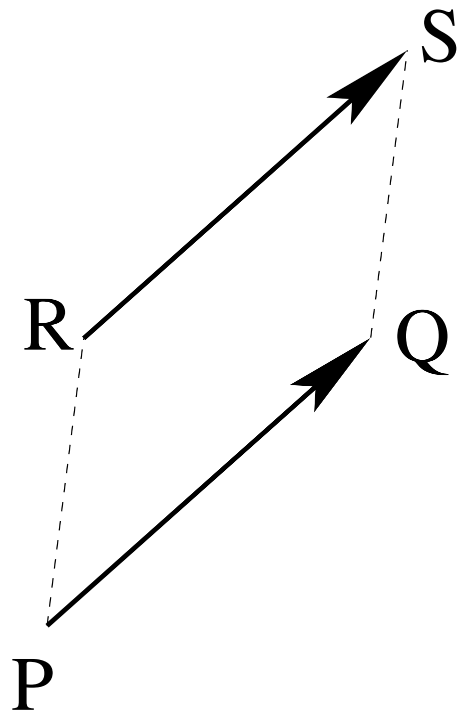

FIGURE 2. The directed line segments are equivalent because they have the same length and direction.

5. Properties of Vector Addition and Scalar Multiplication

Theorem 5.1. *Let* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$, *and let* λ, μ *be scalars. Then*

(a) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ *(vector addition is commutative)*
(b) $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$ *(vector addition is associative)*
(c) $\mathbf{u}+\mathbf{0}=\mathbf{u}$ *($\mathbf{0}$ is the additive identity)*
(d) $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$ *($-\mathbf{u}$ is the additive inverse of $\mathbf{u}$)*
(e) $(\lambda\,\mu)\mathbf{u}=\lambda(\mu\,\mathbf{u})$
(f) $(\lambda+\mu)\mathbf{u}=\lambda\,\mathbf{u}+\mu\,\mathbf{u}$ *(distributive law)*
(g) $\lambda(\mathbf{u}+\mathbf{v})=\lambda\,\mathbf{u}+\lambda\,\mathbf{v}$ *(another distributive law)*
(h) $1\,\mathbf{u}=\mathbf{u}$
(i) $(-1)\mathbf{u}=-\mathbf{u}$
(k) $0\,\mathbf{u}=\mathbf{0}$

PROOF. Every part of the theorem follows from the properties of real numbers. We will prove part (g) to illustrate this and leave the other parts as an exercise for the reader.

For the proof of (g) let

$$\mathbf{u}=(u_1,u_2,\ldots,u_n),\qquad \mathbf{v}=(v_1,v_2,\ldots,v_n).$$

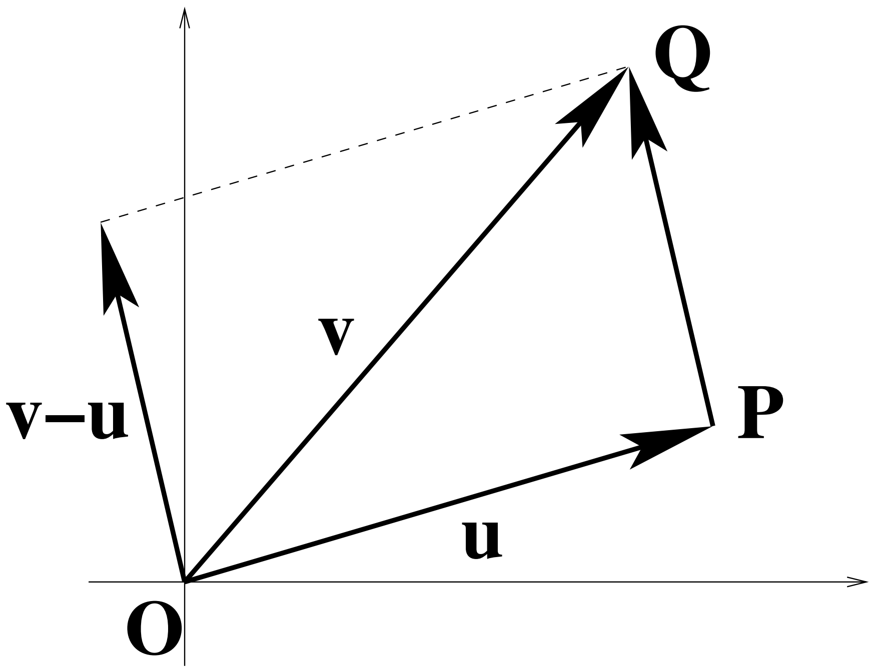

FIGURE 3. As $\mathbf{u} + (\mathbf{v} - \mathbf{u}) = \mathbf{v}$ then $\mathbf{v}$ is the diagonal of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v} - \mathbf{u}$.

Then

$$\begin{aligned}
\lambda(\mathbf{u}+\mathbf{v}) &= \lambda\,(u_1+v_1,\dots,u_n+v_n) && \text{by defn of vector addition}\\
&= (\lambda(u_1+v_1),\dots,\lambda(u_n+v_n)) && \text{by defn of scalar mult.}\\
&= (\lambda u_1+\lambda v_1,\dots,\lambda u_n+\lambda v_n)\\
&= (\lambda u_1,\dots,\lambda u_n)+(\lambda v_1,\dots,\lambda v_n) && \text{by defn of vector addition}\\
&= \lambda(u_1,\dots,u_n)+\lambda(v_1,\dots,v_n) && \text{by defn of scalar mult.}\\
&= \lambda\mathbf{u}+\lambda\mathbf{v}
\end{aligned}$$

Notice that in all the steps we are using the definitions of vector addition and scalar multiplication, except for the third step where we used the distributive property for **real numbers**:

$$\lambda(u_i + v_i) = \lambda u_i + \lambda v_i.$$

□

6. The i, j, k notation

In $\mathbb{R}^2$ we write

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

We can then express every vector $(a, b) \in \mathbb{R}^2$ in terms of $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j}.$$

Similarly, every vector in $\mathbb{R}^3$ can be written in terms of $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ where

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

For example $(3,-1,4) = 3\mathbf{i} - \mathbf{j} + 4\mathbf{k}$.

7. Row Vectors and Column Vectors

There are two ways of representing vectors in $\mathbb{R}^2$. One way is to write (a,b) for the vector $\overrightarrow{OP}$ which starts at the origin and ends at the point $P(a,b)$. The other way is to write

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

for the vector $\overrightarrow{OP}$. We call (a,b) a row vector, and

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

a column vector. At the moment the difference between row vectors and column vectors is **only notational**. Notice that row vectors are more economical on space. But there are good psychological reasons for using column vectors which will be explained in due course (page 75) when we define matrices and come to think of them as functions.

Likewise, we can think of vectors in $\mathbb{R}^n$ as row vectors and column vectors. We have the same notions of length, addition, subtraction, scalar multiplication and dot product for column vectors that we do for row vectors. However, we never think of adding a row vector to a column vector; the difference may be notational, but this addition would be in extremely bad taste.

Exercise 7.1. Let

$$\mathbf{u} = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix}.$$

Perform the following operations.

(i) $\mathbf{u} + \mathbf{v}$ (ii) $2\mathbf{u} - \mathbf{v} + \mathbf{w}$ (iii) $3(\mathbf{u} - \mathbf{v}) + \mathbf{w}$

(iv) $\|\mathbf{v} + \mathbf{w}\|$ (v) $\dfrac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|}$.

8. Dot Product

Recall our geometric interpretation of a vector. A vector is a directed line segment $\overrightarrow{OP}$ starting at the origin and ending at some point P. Any two non-zero vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ meet at the origin and so form an angle. The dot product will allow us to calculate this angle.

Definition. Let

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \qquad \mathbf{v} = (v_1, v_2, \ldots, v_n)$$

be two vectors in $\mathbb{R}^n$. We define the **dot product** of $\mathbf{u}$ and $\mathbf{v}$ (denoted by $\mathbf{u} \cdot \mathbf{v}$) as follows:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Example 8.1. Let $\mathbf{u} = (1, 0, 4, 0, 1)$ and $\mathbf{v} = (-1, 1, 0, 3, 4)$. Then

$$\mathbf{u} \cdot \mathbf{v} = 1 \times -1 + 0 \times 1 + 4 \times 0 + 0 \times 3 + 1 \times 4 = 3.$$

Exercise 8.2. Let

$$\mathbf{u} = (-1, 2, 1, 0), \qquad \mathbf{v} = (0, 1, 3, -1), \qquad \mathbf{w} = (-2, 3, 0, 5).$$

Perform the following operations.

$$\text{(i)}\ \mathbf{u} \cdot \mathbf{v} \qquad \text{(ii)}\ (2\mathbf{u} + \mathbf{w}) \cdot \mathbf{v}$$

Exercise 8.3. Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be vectors in $\mathbb{R}^n$. Explain why the expressions

$$(\mathbf{u} \cdot \mathbf{v}) \cdot \mathbf{w}, \qquad \|\mathbf{u} \cdot \mathbf{v}\|$$

are not defined.

9. Properties of the Dot Product

Theorem 9.1. *Let* $\mathbf{u}$*,* $\mathbf{v}$*,* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$*, and let* λ *be a scalar. Then*

(i) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ *(dot product is commutative)*
(ii) $(\lambda \mathbf{u}) \cdot \mathbf{v} = \lambda(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (\lambda \mathbf{v})$
(iii) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ *(distributive law)*
(iv) $\mathbf{u} \cdot \mathbf{0} = 0$
(v) $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$

Proof. Again the proof follows from the definition and the properties of the real numbers. For illustration, we prove part (i) and leave the rest as an easy exercise. So suppose that

$$\mathbf{u} = (u_1, u_2, \dots, u_n), \qquad \mathbf{v} = (v_1, v_2, \dots, v_n).$$

Then

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n && \text{defn of dot product} \\ &= v_1 u_1 + v_2 u_2 + \cdots + v_n u_n && \text{mult. of reals is commutative} \\ &= \mathbf{v} \cdot \mathbf{u} && \text{defn of dot product} \end{aligned}$$

□

Exercise 9.1. Suppose $\mathbf{v} \in \mathbb{R}^3$. Show that

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{i})\mathbf{i} + (\mathbf{v} \cdot \mathbf{j})\mathbf{j} + (\mathbf{v} \cdot \mathbf{k})\mathbf{k}.$$

10. Angle Between Vectors

Theorem 10.1. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be non-zero vectors in* $\mathbb{R}^n$*, and let* θ *be the angle between* $\mathbf{u}$ *and* $\mathbf{v}$*. Then*

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta. \tag{18}$$

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be the directed line segments $\overrightarrow{OP}$ and $\overrightarrow{OQ}$. By Theorem 4.1, the directed line segment $\overrightarrow{PQ}$ is equivalent to the vector $\mathbf{v} - \mathbf{u}$. See Figure 4.

Let a, b and c be the length of $\overrightarrow{OP}$, $\overrightarrow{OQ}$ and $\overrightarrow{PQ}$ respectively. Thus

$$a = \|\mathbf{u}\|, \qquad b = \|\mathbf{v}\|, \qquad c = \|\mathbf{v} - \mathbf{u}\|.$$

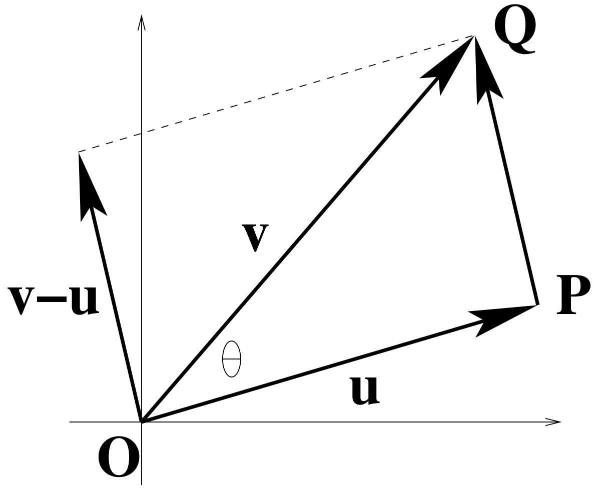

FIGURE 4

Applying the law of cosines to the triangle OPQ gives

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

In other words,

$$\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta. \tag{19}$$

But there is another way of evaluating $\|\mathbf{v} - \mathbf{u}\|^2$. From part (v) of Theorem 9.1, we see that

$$\|\mathbf{v} - \mathbf{u}\|^2 = (\mathbf{v} - \mathbf{u}) \cdot (\mathbf{v} - \mathbf{u}) = \mathbf{v} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{u}.$$

Now by part (a) of Theorem 9.1 we know that $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ and so we get

$$\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\mathbf{u} \cdot \mathbf{v}. \tag{20}$$

Comparing (19) and (20) we see that

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta,$$

which, after a little cancellation, gives (18). □

Suppose that $\mathbf{u}$, $\mathbf{v}$ are non-zero vectors and θ is the angle between them. Recall that $\cos\theta$ is positive for $0 \le \theta < \pi/2$ and negative for $\pi/2 < \theta \le \pi$. Thus we know the following:

- If $\mathbf{u} \cdot \mathbf{v} = 0$ then $\theta = \pi/2$ and the vectors are orthogonal.
- If $\mathbf{u} \cdot \mathbf{v} > 0$ then the angle θ is acute.
- If $\mathbf{u} \cdot \mathbf{v} < 0$ then the angle θ is obtuse.

Example 10.1. Let us return to the two vectors in $\mathbb{R}^4$ that we met in Example 1.1: $\mathbf{u} = (1, -1, 3, 4)$ and $\mathbf{v} = (-1, 0, -1, 1)$. We see that $\mathbf{u} \cdot \mathbf{v} = 0$ and so the two vectors are perpendicular.

Example 10.2. Let $\mathbf{u} = \mathbf{i}$ and $\mathbf{v} = \mathbf{j} - \mathbf{i}$ and let θ be the angle between them. Then

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{-1}{\sqrt{2}}.$$

Thus $\theta = 3\pi/4$. Make a sketch to satisfy yourself that this answer is reasonable.

Example 10.3. Let $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Show that

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\,\|\mathbf{u}\|^2 + 2\,\|\mathbf{v}\|^2.$$

Answer: Recall from the proof of Theorem 10.1 that

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\,\mathbf{u}\cdot\mathbf{v}.$$

In the same way

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})\\ &= \mathbf{u}\cdot\mathbf{u} + 2\,\mathbf{u}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v}\\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\,\mathbf{u}\cdot\mathbf{v}.\end{aligned}$$

Adding the two identities we obtain the desired result.

Exercise 10.4. Determine the cosine of the angle determined by the given pair of vectors, and state where the angle is acute, obtuse or a right-angle:

(i) $(1,0,2),\ (-2,1,1)$ (ii) $(1,-1),\ (1,-2)$
(iii) $(4,1,-1,2),\ (1,-4,2,-1)$

Exercise 10.5. Let $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and θ be the angle between them. Prove that

$$\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2 = 4\|\mathbf{u}\|\,\|\mathbf{v}\|\,\cos\theta.$$

Exercise 10.6. Find the area of the parallelogram that has **unit** vectors $\mathbf{u}$ and $\mathbf{v}$ as adjacent sides, if $\mathbf{u}\cdot\mathbf{v} = \sqrt{3}/2$.

Exercise 10.7. Suppose that $\mathbf{u}$, $\mathbf{v}$ are **unit** vectors in $\mathbb{R}^n$. Show that

$$\|\mathbf{u}+2\,\mathbf{v}\|^2 + \|\mathbf{u}-2\,\mathbf{v}\|^2 = 10.$$

11. Two Important Inequalities

In this section you will prove two important inequalities which you will need again and again in your mathematical career.

Exercise 11.1. (The Cauchy-Schwartz Inquality) Suppose $u_1,\dots,u_n$ and $v_1,\dots,v_n$ are real numbers. Show that

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \le \left(u_1^2 + u_2^2 + \cdots + u_n^2\right)^{1/2}\left(v_1^2 + v_2^2 + \cdots + v_n^2\right)^{1/2}.$$

Hint: Think about what the inquality is saying in terms of vectors.

Exercise 11.2. (The Triangle Inequality) Let $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Show that

$$\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Hint: Start with $\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})$ and after expanding the brackets use the Cauchy-Schwartz inequality.

12. Orthogonality

Definition. We say that non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** (or **perpendicular**) if the angle between them is $\pi/2$. By Theorem 10.1, non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\mathbf{u}\cdot\mathbf{v} = 0$.

We say that non-zero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **orthogonal** if every pair is orthogonal. Using more notation, we can say that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **orthogonal** if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$.

We say that vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **orthonormal** if they are orthogonal and each is a unit vector.

Notice that if the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are orthogonal, and we let $\mathbf{u}_i = \mathbf{v}_i/\|\mathbf{v}_i\|$, then the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are orthonormal.

Example 12.1. Obviously the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in $\mathbb{R}^3$ are orthonormal. It is easy to check that $\mathbf{i}+\mathbf{j}$, $\mathbf{i}-\mathbf{j}$, $\mathbf{k}$ are orthogonal. To get orthonormal vectors we can scale them, diving each vector by its length. We obtain orthonormal vectors

$$\frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j}), \quad \frac{1}{\sqrt{2}}(\mathbf{i}-\mathbf{j}), \quad \mathbf{k}.$$

Exercise 12.2. Find a vector that is orthogonal to both $(1, 0, 3)$, $(0, 1, -1)$.

Exercise 12.3. Let $\mathbf{u} = (a, b)$ be a non-zero vector in $\mathbb{R}^2$. Show that $\mathbf{v} = (b, -a)$ is orthogonal to $\mathbf{u}$. Hence find two unit vectors that are orthogonal to $\mathbf{u}$. Are there any others?

Exercise 12.4. Suppose that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are orthogonal non-zero vectors in Euclidean n-space, and that a vector $\mathbf{v}$ is expressed as

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_m\mathbf{v}_m.$$

Show that the scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$ are given by

$$\lambda_i = \frac{\mathbf{v}\cdot\mathbf{v}_i}{\|\mathbf{v}_i\|^2}, \qquad i = 1, 2, \ldots, m.$$

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are orthonormal, what would the answer be?

13. Vector Equation of the Line

You should know that any straight line in the plane can be written either as $y = mx + c$ if it isn't vertical, and as $x = b$ if it is vertical. We would like to write down the equation of a straight line in terms of vectors. Working with vectors we will see that there is none of the awkward subdivision of cases: vertical and non-vertical. Moreover, our equation is of the same form regardless of whether the line is in the plane $\mathbb{R}^2$, 3-space $\mathbb{R}^3$, or any Euclidean space $\mathbb{R}^n$.

Take any straight line L in $\mathbb{R}^n$ (if you are happier with $\mathbb{R}^2$ and $\mathbb{R}^3$ then think that n is 2 or 3). Pick any point Q on L and any non-zero vector $\mathbf{v}$ parallel to L. If P is a point on L then the vector $\overrightarrow{QP}$ is parallel to L and so parallel to $\mathbf{v}$. Hence

$$\overrightarrow{QP} = t\mathbf{v}$$

for some scalar t. See Figure 5.

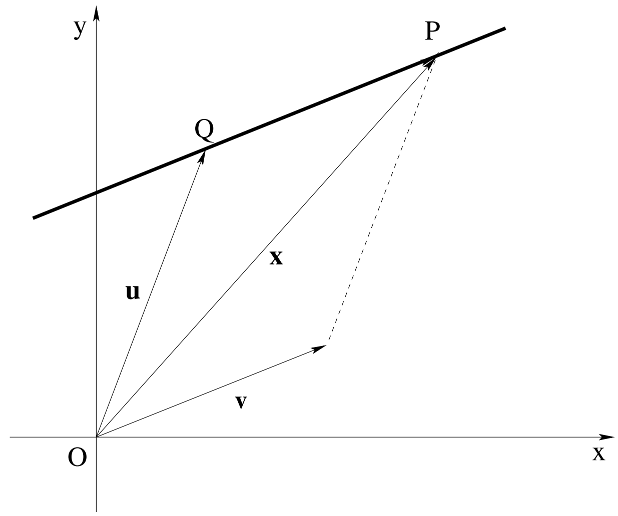

FIGURE 5

As usual, write O for the origin. Then

$$\begin{aligned}\overrightarrow{OP} &= \overrightarrow{OQ} + \overrightarrow{QP} \quad \text{using the definition of vector addition}\\ &= \overrightarrow{OQ} + t\mathbf{v}.\end{aligned}$$

Now let $\mathbf{u} = \overrightarrow{OQ}$ and $\mathbf{x} = \overrightarrow{OP}$. Then

$$\mathbf{x} = \mathbf{u} + t\mathbf{v}.$$

This is the vector equation for the line L. Note that we think of $\mathbf{u}$ and $\mathbf{v}$ as fixed vectors; the first is the position vector of a fixed point Q on L and the second is a vector parallel to L. We think of $\mathbf{x}$ as a variable vector. As t varies along the real line (recall that t is a scalar), $\mathbf{x}$ is the position vector of a point varying along L. So we think of $\mathbf{x}$ as a function of t and write

$$\mathbf{x}(t) = \mathbf{u} + t\mathbf{v}.$$

Example 13.1. Find the vector equation of the line L passing through the points $(0,1,1)$ and $(1,3,2)$ in $\mathbb{R}^3$.

Answer: We can take $Q = (0,1,1)$ as a point on the line, with position vector $\mathbf{u} = \overrightarrow{OQ} = (0,1,1)$. We also note that $\mathbf{v} = (1,3,2) - (0,1,1) = (1,2,1)$ is a vector parallel to the line. So the vector equation of L is

$$L \ : \qquad \mathbf{x} = (0,1,1) + t(1,2,1).$$

We can also write this without vectors by letting $\mathbf{x} = (x,y,z)$ and we see that

$$L \ : \begin{cases} x = t, \\ y = 1 + 2t, \\ z = 1 + t. \end{cases}$$

This is the **parametric form** of the line L. It gives us the coordinates x, y, z in terms of a parameter t.

Exercise 13.2. Let L_1, L_2 be the two lines in the plane $\mathbb{R}^2$ given by the equations

$$L_1 \;:\; \mathbf{x} = (0,1) + t(1,1), \qquad L_2 \;:\; \mathbf{x} = (1,0) + t(1,-1).$$

The lines are non-parallel; why? We know that any two non-parallel lines in the plane must intersect. We want to find the point of intersection of L_1 and L_2. **What is wrong with the following argument:**

To find the point of intersection, equate

$$(0,1) + t(1,1) = (1,0) + t(1,-1).$$

Thus $t(0,2) = (1,-1)$ which is impossible???

Exercise 13.3. Let L be the line in the plane given by the equation $y = mx + c$. What is the vector equation of L?

Exercise 13.4. Let $L_1 : \mathbf{x} = \mathbf{u}_1 + t\mathbf{v}_1$ and $L_2 : \mathbf{x} = \mathbf{u}_2 + t\mathbf{v}_2$ be straight lines in $\mathbb{R}^n$, where $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2 \neq \mathbf{0}$ and $\mathbf{u}_1 \neq \mathbf{u}_2$. Show that L_1 and L_2 are the same line if and only if the three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{u}_1 - \mathbf{u}_2$ are parallel.

13.1. A Very Important Warning. A straight line has infinitely many vector equations and infinitely many parametric equations. Just because two lines have different vector equations, doesn't give us the right to conclude that the lines are different. Do this exercise and you will see what I mean.

Exercise 13.5. Let L_1, L_2 be the pair of lines

$$L_1 \;:\; \mathbf{x} = (1,1,2) + t(1,0,1), \quad L_2 \;:\; \mathbf{x} = (0,1,1) + t(2,0,2).$$

Show that L_1, L_2 both pass through the pair of points $(1,1,2)$ and $(0,1,1)$. But you know that passing through any two distinct points is a unique line, so ...

Exercise 13.6. Let L be the line with vector equation $\mathbf{x} = (2,1,0) + t(1,3,1)$. Which of the following lines is parallel to L but passess through the point $P(1,1,-1)$:

(a) $\mathbf{x} = (2,1,0) + t(1,1,-1)$ (b) $\mathbf{x} = (4,2,0) + t(1,3,1)$
(c) $\mathbf{x} = (-1,-5,-3) + t(2,6,2)$ (d) $\mathbf{x} = (1,1,-1) + t(3,4,1)$

14. Vector Equation of the Plane

We shall only be concerned with planes in $\mathbb{R}^3$. A plane is normally denoted by the symbol Π (a capital π). You will need to be comfortable with four different ways of representing a plane in $\mathbb{R}^3$.

I. The Point-Normal Form. Let Π be a plane in $\mathbb{R}^3$. Let $\mathbf{n}$ be a vector normal to Π (by normal to Π we simply mean perpendicular to Π). Choose and fix a point Q on the plane Π and let $\mathbf{u} = \overrightarrow{OQ}$ be the position vector of Q. Suppose now that P is any point on Π and let $\mathbf{x} = \overrightarrow{OP}$ be its position vector. Note that the vector $\overrightarrow{QP} = \mathbf{x} - \mathbf{u}$ is parallel to the plane and so perpendicular to $\mathbf{n}$. Hence $\mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0$. This is the point-normal equation for the plane:

$$\Pi \;:\; \mathbf{n} \cdot (\mathbf{x} - \mathbf{u}) = 0. \tag{21}$$

Here $\mathbf{n}$ is any (non-zero) vector normal to the plane, and $\mathbf{u}$ is the position vector of any point on the plane.

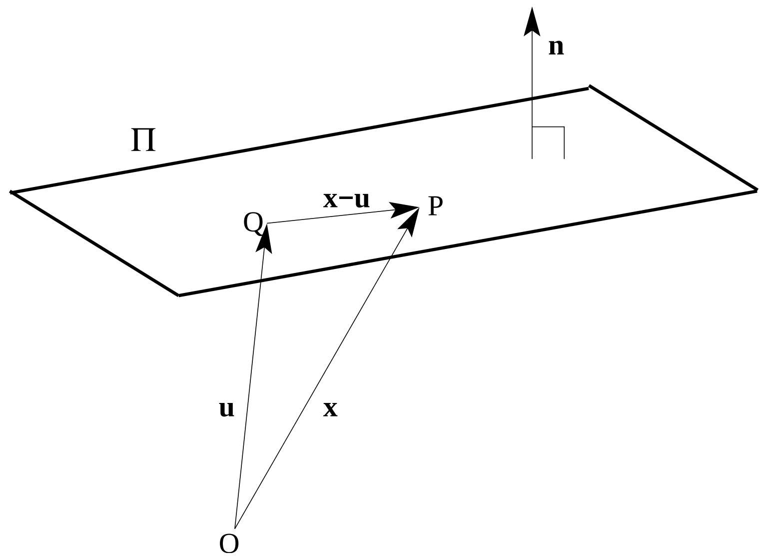

FIGURE 6

II. The Standard Form. You have met this form at school (and probably the other three). A plane is represented by the equation

$$\Pi \ : \ ax + by + cz = d \tag{22}$$

where a, b, c, d are (real) numbers, and not all of a, b, c are zero. Let us see how to get from the point-normal form to the standard form. Suppose Π is the plane given by (21) where $\mathbf{n} = (a, b, c)$ is a non-zero vector normal to the plane and $\mathbf{u} = (u, v, w)$ is the position vector of some point on the plane. Write $\mathbf{x} = (x, y, z)$. Then equation (21) can be rewritten as $\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{u}$, or

$$ax + by + cz = au + bv + cw.$$

Writing $d = au + bv + cw$ we get the equation for Π in its standard form (22).

III. The Vector Form.

Let $\mathbf{u}$ be the position vector of a point on Π. Let $\mathbf{v}$ and $\mathbf{w}$ be non-zero non-parallel vectors [1], both parallel to Π. Then the vector equation of the plane is

$$\Pi \ : \ \mathbf{x} = \mathbf{u} + s\mathbf{v} + t\mathbf{w}. \tag{23}$$

Note that the vector equation of the plane is given in terms of two parameters s, t, because the plane is a 2-dimensional object.

IV. The Parametric Form.

[1] The statement that $\mathbf{v}$ and $\mathbf{w}$ are non-zero and non-parallel means that neither $\mathbf{v}$ nor $\mathbf{w}$ is a scalar multiple of the other.

In the vector form (23) write $\mathbf{x} = (x, y, z)$, $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, $\mathbf{w} = (w_1, w_2, w_3)$. Then (23) can be rewritten as

$$\Pi \ : \ \begin{cases} x = u_1 + sv_1 + tw_1, \\ y = u_2 + sv_2 + tw_2, \\ z = u_3 + sv_3 + tw_3. \end{cases}$$

Example 14.1. Let Π be the plane given by standard equation

$$\Pi \ : \ x + y + z = 1.$$

Write equations for Π in point-normal and vector forms.
Answer: Let $\mathbf{n} = (1, 1, 1)$, $\mathbf{x} = (x, y, z)$. Note that the equation can be rewritten as $\mathbf{n} \cdot \mathbf{x} = 1$. Now the point $P(1, 0, 0)$ is on the plane and so its position vector $\mathbf{u} = (1, 0, 0)$ satisfies $\mathbf{n} \cdot \mathbf{u} = 1$ (which is clear anyway). Hence we can rewrite the equation as $(1, 1, 1) \cdot \mathbf{x} = (1, 1, 1) \cdot (1, 0, 0)$, or

$$(1, 1, 1) \cdot (\mathbf{x} - (1, 0, 0)) = 0,$$

which is the equation in point-normal form.

We now want the vector equation for Π. Note that the points $Q(0, 1, 0)$ and $R(0, 0, 1)$ are on the plane. Hence the vectors $\mathbf{v} = \overrightarrow{PQ} = (-1, 1, 0)$ and $\mathbf{w} = \overrightarrow{PR} = (-1, 0, 1)$ are parallel to the plane. Moreover, these vectors are not parallel to each other (they are not multiples of one another). Hence a vector equation for Π is

$$\mathbf{x} = (1, 0, 0) + s(-1, 1, 0) + t(-1, 0, 1).$$

Exercise 14.2. Let Π_1, Π_2 be the two planes in $\mathbb{R}^3$ given by

$$\Pi_1 \ : \ x + y + z = 1, \qquad \Pi_2 \ : \ x - 2y - z = 0.$$

Find the vector equation of the straight line given by the intersection of Π_1 and Π_2.

Exercise 14.3. Find the vector equation of the plane in $\mathbb{R}^3$ passing through the point $(0, 1, 1)$ and containing the line $L \ : \ \mathbf{x} = (1, 0, 0) + t(0, 0, 1)$.

Exercise 14.4. In the vector equation of the plane (23) we insisted that the vectors $\mathbf{v}$ and $\mathbf{w}$ are non-parallel. Show that if $\mathbf{v}$ and $\mathbf{w}$ are parallel then the equation gives a line and not a plane.

Exercise 14.5. Let

$$\Pi_1 : \ \mathbf{x} \cdot (1, 1, 1) = \gamma, \qquad \Pi_2 : \ \mathbf{x} = s(-1, \alpha, 0) + t(1, 1, \beta) + (-1, -1, -1).$$

For which triples of values α, β, γ will Π_1, Π_2 be the same plane?

Exercise 14.6. For which pairs of values α, β does the line

$$L \ : \quad \mathbf{x} = (-1, \beta, \alpha) + t(7, \beta, 4)$$

lie on the plane $x + y + z = 3$?

15. Subspaces of $\mathbb{R}^n$

Definition. A subset V of $\mathbb{R}^n$ is said to be a **subspace of** $\mathbb{R}^n$ if it satifies these three conditions:

- $\mathbf{0} \in V$;
- for all vectors $\mathbf{u}$, $\mathbf{v}$ in V, their sum $\mathbf{u} + \mathbf{v}$ must be in V;
- for all vectors $\mathbf{v}$ in V and scalars λ, the product $\lambda\mathbf{v}$ must be in V.

In other words, in a subspace we can add and multiply by scalars.

Example 15.1. Let $V = \{(a, a) : a \in \mathbb{R}\}$. In other words V is the subset of $\mathbb{R}^2$ where the x-coordinate equals the y-coordinate. Thus V is the the line $y = x$ in $\mathbb{R}^2$. It is geometrically obvious that V contains the origin; that if we add two vectors belonging to it the result also belongs to it; and that if scale any vector belonging to this diagonal, by any scalar we please, the result also belongs to V. But at this stage in your academic career, you are expected to write a proof in symbols. Let us do that:

First note that $\mathbf{0} = (0, 0) \in V$. Secondly, suppose $\mathbf{u} \in V$ and $\mathbf{v} \in V$. By definition of V, $\mathbf{u} = (a, a)$ and $\mathbf{v} = (b, b)$ for some $a, b \in \mathbb{R}$. Thus $\mathbf{u}+\mathbf{v} = (a+b, a+b)$ which again belongs to V. Finally, suppose that $\mathbf{v} \in V$ and $\lambda \in \mathbb{R}$. By definition of V, $\mathbf{v} = (a, a)$. So $\lambda\mathbf{v} = (\lambda a, \lambda a)$ which is in V. This shows that V is a subspace of $\mathbb{R}^2$.

Example 15.2. This time we take $W = \{(a, a) : a \in \mathbb{R},\ a \geq 0\}$. The set W is not all the line $y = x$ but a 'ray' as in Figure 7. Note that W does not satisfy the last

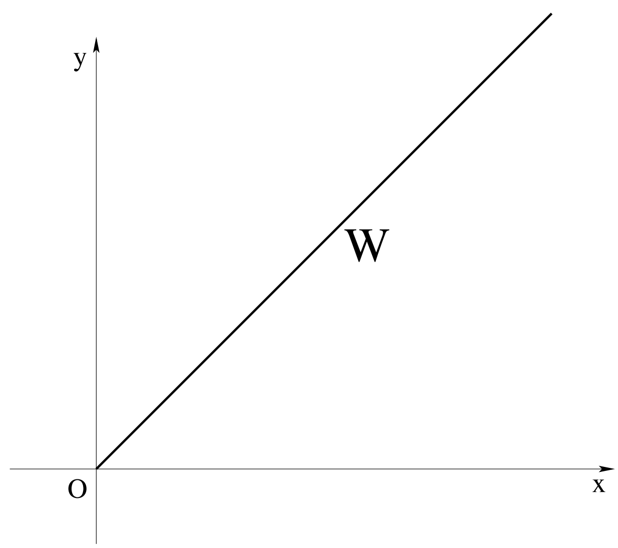

FIGURE 7. $W = \{(a, a) : a \in \mathbb{R},\ a \geq 0\}$ is ray.

condition for being a subspace. For example, take $\mathbf{v} = (1, 1)$ and $\lambda = -1$. Note $\mathbf{v}$ is a vector belonging to W and λ is a scalar, but $\lambda\mathbf{v} = (-1, -1)$ does not belong to W. Hence W is not subspace of $\mathbb{R}^2$.

Notice that to show that W is not a subspace, we gave a **counterexample**. This means that we gave an example to show that at least one of the requirements in the definition is not always satisfied.

What do subspaces of $\mathbb{R}^2$ look like? In fact there are exactly three possibilities:

(1) either the subspace is just the origin $\{\mathbf{0}\}$,
(2) or it is a straight line passing through the origin,
(3) or it is the whole of $\mathbb{R}^2$.

Now what do subspaces of $\mathbb{R}^3$ look like? There are four possibilities. What do you think they are?

CHAPTER 4

Matrices

1. What are Matrices?

Let m, n be positive integers. An $m \times n$ matrix (or a matrix of size $m \times n$) is a rectangular array consisting of $m\,n$ numbers arranged in m rows and n columns:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

Example 1.1. Let

$$A = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 7 & 14 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & -2 \\ -1 & 8 \\ 2 & 5 \end{pmatrix}, \qquad C = \begin{pmatrix} 3 & 1 & 5 \\ -6 & -8 & 12 \\ 2 & 5 & 0 \end{pmatrix}.$$

A, B, C are matrices. The matrix A has size 2×3 because it has 2 rows and 3 columns. Likewise B has size 3×2 and C has size 3×3.

Displaying a matrix A by writing

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}.$$

wastes a lot of space. It is convenient to abbreviate this matrix by the notation $A = (a_{ij})_{m \times n}$. This means that A is a matrix of size $m \times n$ (i.e. m rows and n columns) and that we shall refer to the element that lies at the intersection of the i-th row and j-th column by a_{ij}.

Example 1.2. Let $A = (a_{ij})_{2 \times 3}$. We can write A out in full as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

Notice that A has 2 rows and 3 columns. The element a_{12} belongs to the 1st row and the 2nd column.

The abbreviated notation $(a_{ij})_{m \times n}$ for matrices is particularly convenient when we have a recipe for the elements of the matrix. As we see in the following example.

Example 1.3. Let $C = (c_{ij})_{2 \times 2}$ where $c_{ij} = i - j$. Calculate C.

Answer: C has 2 rows and 2 columns. Thus

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$

We need the values of the c_{ij} which we get from the recipe $c_{ij} = i - j$. Hence

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} 1-1 & 1-2 \\ 2-1 & 2-2 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Exercise 1.4. Let $B = (b_{ij})_{3\times 4}$ where

$$b_{ij} = \begin{cases} i\,j & \text{if } i \le j \\ i + j & \text{otherwise.} \end{cases}$$

Calculate B.

2. Matrix Operations

Definition. Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{p\times q}$ be matrices. We say that A and B are **equal** (and write $A = B$) if the following conditions are satisfied:

- $m = p$ and $n = q$;
- $a_{ij} = b_{ij}$ for all pairs i, j.

In other words, for two matrices to be equal we require them to have the same size, and we require that corresponding elements are equal.

Example 2.1. Let

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0.1 \end{pmatrix}.$$

Then $A \neq B$ because A is 2×3 and B is 3×2. For the same reason $A \neq C$. Although B and C have the same size, they are not equal because the elements in the $(3, 2)$-position are not equal.

Definition. Given $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we define the **sum** $A + B$ to be the $m \times n$ matrix whose (i, j)-th element is $a_{ij} + b_{ij}$. We define the **difference** $A - B$ to be the $m \times n$ matrix whose (i, j)-th element is $a_{ij} - b_{ij}$.

Let λ be a scalar. We define λA to be the $m \times n$ matrix whose (i, j)-th element is λa_{ij}.

We let $-A$ be the $m \times n$ matrix whose (i, j)-th element is $-a_{ij}$. Thus $-A = (-1)A$.

Note that the sum $A + B$ is defined only when A and B have the same size. In this case $A + B$ is obtained by adding the corresponding elements.

Example 2.2. Let

$$A = \begin{pmatrix} 2 & -5 \\ -2 & 8 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix}, \qquad C = \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix}.$$

Then $A + B$ is undefined because A and B have different sizes. Similarly $A + C$ is undefined. However $B + C$ is defined and is easy to calculate:

$$B + C = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix} + \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ 1 & 6 \\ 8 & 3 \end{pmatrix}.$$

Likewise $A - B$ and $A - C$ are undefined, but $B - C$ is:

$$B - C = \begin{pmatrix} 4 & 3 \\ 1 & 0 \\ -1 & 2 \end{pmatrix} - \begin{pmatrix} -4 & 2 \\ 0 & 6 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 1 \\ 1 & -6 \\ -10 & 1 \end{pmatrix}.$$

Scalar multiplication is always defined. Thus, for example

$$-A = \begin{pmatrix} -2 & 5 \\ 2 & -8 \end{pmatrix}, \qquad 2\,B = \begin{pmatrix} 8 & 6 \\ 2 & 0 \\ -2 & 4 \end{pmatrix}, \qquad 1.5C = \begin{pmatrix} -6 & 3 \\ 0 & 9 \\ 13.5 & 1.5 \end{pmatrix}.$$

Definition. The zero matrix of size $m \times n$ is the unique $m \times n$ matrix whose entries are all 0. This is denoted by $0_{m\times n}$, or simply 0 if no confusion is feared.

Example 2.3.

$$0_{2\times 4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Exercise 2.4. Let

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

Which of the following operations is defined; if defined, give the result.

(i) $A + 2B$ (ii) $2C - A$

(iii) $B + 0_{2\times 2}$ (iv) $C + 0_{2\times 2}$.

Exercise 2.5. Let

$$A = \begin{pmatrix} x & y & 1 \\ z & -1 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & z & -1 \\ -y & 1 & 2 \end{pmatrix}.$$

Find the values of x, y and z if

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

Definition. Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$. We define the **product** AB to be the matrix $C = (c_{ij})_{m\times p}$ such that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}.$$

Note the following points:

(1) For the product AB to be defined we demand that the number of columns of A is equal to the number of rows of B.
(2) The ij-th element of AB is obtained by taking the dot product of the i-th row of A with the j-th column of B.

Example 2.6. Let

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

Both A and B are 2×2. From the definition we know that $A \times B$ will be a 2×2 matrix. We see that

$$AB = \begin{pmatrix} 1\times 5 + 2\times 0 & 1\times -3 + 2\times -2 \\ -1\times 5 + 3\times 0 & -1\times -3 + 3\times -2 \end{pmatrix} = \begin{pmatrix} 5 & -7 \\ -5 & -3 \end{pmatrix}.$$

Likewise

$$BA = \begin{pmatrix} 5 \times 1 + -3 \times -1 & 5 \times 2 - 3 \times 3 \\ 0 \times 1 - 2 \times -1 & 0 \times 2 + -2 \times 3 \end{pmatrix} = \begin{pmatrix} 8 & 1 \\ 2 & -6 \end{pmatrix}.$$

We make a very important observation: $AB \neq BA$ in this example. So **matrix multiplication is not commutative**.

Example 2.7. Let A be as in the previous example, and let

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

Then

$$AC = \begin{pmatrix} 8 & -7 & 3 \\ 7 & -13 & -3 \end{pmatrix}.$$

However, CA is not defined because the number of columns of C is not equal to the number of rows of A.

Definition. Let $A = (a_{ij})_{m \times n}$ be a matrix (of size $m \times n$). The transpose of A (written A^t) is the $n \times m$ matrix whose j-th column is the j-th row of A. Another way of saying this is: if $B = (b_{ij})_{n \times m} = A^t$ then $b_{ij} = a_{ji}$.

Example 2.8. The transpose of

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

is

$$C^t = \begin{pmatrix} 2 & 3 \\ 1 & -4 \\ 3 & 0 \end{pmatrix}.$$

Exercise 2.9. Commutativity—What can go wrong?

(1) Give a pair of matrices A, B, such that AB is defined but BA isn't.
(2) Give a pair of matrices A, B, such that both AB and BA are defined but they have different sizes.
(3) Give a pair of matrices A, B, such that AB and BA are defined and of the same size but are unequal.
(4) Give a pair of matrices A, B, such that $AB = BA$.

2.1. Where do matrices come from? No doubt you are wondering where matrices come from, and what is the reason for the weird definition of matrix multiplication. Matrices orginate from linear substitutions. Let a, b, c, d be fixed numbers, x, y some variables, and define x', y' by the linear substitutions

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy. \end{aligned} \tag{24}$$

The definition of matrix multiplication allows us to express this pair of equations as one matrix equation

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{25}$$

You should multiply out this matrix equation and see that it is the same as the pair of equations (24).

Now suppose moreover that we define new quantaties x'' and y'' by

$$x'' = \alpha x' + \beta y' \\ y'' = \gamma x' + \delta y', \tag{26}$$

where α, β, γ, δ are constants. Again we can rewrite this in matrix form as

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{27}$$

What is the relation between the latest quantaties x'', y'' and our first pair x, y? One way to get the answer is of course to substitute equations (24) into (26). This gives us

$$x'' = (\alpha a + \beta c)x + (\alpha b + \beta d)y \\ y'' = (\gamma a + \delta c)x + (\gamma b + \delta d)y. \tag{28}$$

This pair of equations can re-expressed in matrix form as

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{29}$$

Another way to get x'', y'' in terms of x, y is to substitute matrix equation (25) into matrix equation (27):

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{30}$$

If the definition of matrix multiplication is sensible, then we expect that matrix equations (29) and (30) to be consistent. In other words, we would want that

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix}.$$

Check that this is indeed the case.

3. Some Special Matrices

A *square matrix* is one which has the same number of rows as columns. So if A is a square matrix, it will have size $n \times n$ for some positive integer n; *we say A is a square matrix of size n.* If A is a square matrix and k is a positive integer we define k-th power of A to be

$$A^k = \underbrace{AA \cdots A}_{k}.$$

We note that this is only defined for square matrices (why?). The reader will also note that for this definition to be sensible we need the fact that matrix multiplication is associative. For this see Theorem 4.1.

A **diagonal matrix** is a square matrix where all the entries that are not on the diagonal are zero. For example

$$A = \begin{pmatrix} 1 & 4 \\ 0 & 7 \end{pmatrix}$$

is **not** a diagonal matrix since the element 4 in the $(1, 2)$-position is a non-zero element that is not on the diagonal. However

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 7 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

are diagonal matrices. We refer to the diagonal matrix with diagonal entries a_1, $a_2, \dots, a_n$ with the notation $\operatorname{diag}(a_1, a_2, \dots, a_n)$. Thus the diagonal matrices B and C can be written more economically as $B = \operatorname{diag}(1, 7)$ and $C = \operatorname{diag}(1, 2, -3)$.

The **identity matrix** of size n written I_n is

$$I_n = \operatorname{diag}(\underbrace{1, 1, 1, \dots, 1}_{n}).$$

We define the **Kronecker delta** by the formula:

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

Thus, for example, $\delta_{23} = 0$ but $\delta_{44} = 1$. Then the identity matrix of size n can also be written as $I_n = (\delta_{ij})_{n \times n}$.

Exercise 3.1. Show that $\operatorname{diag}(a_1, a_2)^k = \operatorname{diag}(a_1^k, a_2^k)$ for positive integers k. Extend to diagonal matrices of size n.

Exercise 3.2. Let $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Use induction to show $A^n = \begin{pmatrix} 3^n & 0 & 0 \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$ for every positive integer n.

4. Properties of Matrix Operations

Theorem 4.1. *Let A, B, C be matrices and λ, μ scalars. The following properties hold (provided that the dimensions of the matrices are such that the operations are defined):*

(i) $A + B = B + A$ *(matrix addition is commutative)*
(ii) $A + (B + C) = (A + B) + C$ *(matrix addition is associative)*
(iii) $A + 0_{m \times n} = A$
(iv) $A + (-A) = 0_{m \times n}$
(v) $A(BC) = (AB)C$ *(matrix multiplication is associative)*
(vi) $AI_m = A$, $I_n B = B$
(vii) $A(B + C) = AB + AC$, $(A + B)C = AC + BC$ *(distributive laws)*
(viii) $\lambda(A + B) = \lambda A + \lambda B$, $(\lambda + \mu)A = \lambda A + \mu A$ *(distributive laws)*
(ix) $(\lambda\mu)A = \lambda(\mu A)$
(x) $1A = A$, $0A = 0_{m \times n}$
(xi) $\lambda 0_{m \times n} = 0_{m \times n}$
(xii) $A0_{n \times p} = 0_{m \times p}$, $0_{p \times m} A = 0_{p \times n}$
(xiii) $\lambda(AB) = (\lambda A)B = A(\lambda B)$
(xiv) $(A + B)^t = A^t + B^t$
(xv) $(AB)^t = B^t A^t$.

5. Matrix Inverses

We talked about matrix addition, subtraction and multiplication. The reader may be wondering if there is an analogue of division for matrices or at least a concept of a reciprocal for matrices. If λ is real, then the reciprocal of λ is a number μ that satisfies

$$\lambda\mu = \mu\lambda = 1.$$

Of course the reciprocal does not exist for $\lambda = 0$ but does exist otherwise, and we denote it by λ^{-1}.

We would like to define a similar concept to reciprocal for matrices. Before doing that we have to ask what is the analogue of 1 for matrix multiplication.

Let A be a square matrix of size n. Recall, from Theorem 4.1, that

$$AI_n = I_n A = A$$

where I_n is the identity matrix. In other words, the identity matrix acts in the same way as 1 does for real number multiplication. You should now have guessed what the analogue of reciprocal for matrices is. In fact we call it the inverse and not the reciprocal.

Definition. Let A be square matrix of size n. We say that A is **invertible** (or **non-singular**) if there exists a matrix B (called the inverse) such that

$$AB = BA = I_n.$$

Clearly, if the inverse B exists it must be square of size n.

Example 5.1. For example, if

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

then we can take B to be

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

we see that $AB = BA = I_2$, so that B is the inverse of A.

The reader will of course expect that the zero matrix does not have an inverse. But there are also non-zero matrices that do not have inverses. These are called **singular** or **non-invertible**.

Example 5.2. Let

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

We show that A is non-invertible (or singular) by contradiction. So suppose it is invertible. That is, there is a 2×2 matrix B such that $AB = BA = I_2$. Let

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

Then

$$AB = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}.$$

Clearly $AB \neq I_2$ whatever the values of a and b are. Thus we have a contradiction and we see that A is non-invertible.

We have learned that some matrices are invertible and others are not. There is a hidden assumption in the definition of inverse given above. Recall that we called B the inverse of A if $AB = BA = I_n$. The hidden assumption is in the word 'the', which implies uniqueness. How do we know that there aren't two or more inverse to A? Of course inverses (i.e. reciprocals) of real numbers are unique, but we have seen that matrices don't have some of the properties of real numbers (e.g. matrix multiplication is non-commutative). Let us prove that the inverse (if it exists) is unique.

Theorem 5.1. *Let A be a square matrix of size n. If A is invertible then it has a unique inverse.*

Proof. Our proof is a standard uniqueness proof that you will see again and again during your undergraduate career. Almost all uniqueness proofs follow the same pattern: suppose that there are two of the thing that we want to prove unique; show that these two must be equal; therefore it is (whatever it is) unique.

For our proof we suppose that A has two inverses, and call them B and C. We want to show that $B = C$. By definition of inverse we know that

$$AB = BA = I_n, \qquad AC = CA = I_n.$$

Thus

$$\begin{aligned} B &= BI_n && \text{from Theorem 4.1} \\ &= B(AC) && \text{from the above } AC = I_n \\ &= (BA)C && \text{matrix multiplication is associative (Theorem 4.1)} \\ &= I_nC && \text{from the above } BA = I_n \\ &= C && \text{from Theorem 4.1.} \end{aligned}$$

Thus $B = C$. Since every two inverses must be equal, we see that the inverse must be unique. ☐

Now we have shown that the inverse, if it exists, must be unique, we introduce a suitable notation for it. If the square matrix A is invertible, we write A^{-1} for the (unique) inverse of A.

6. Some Peculiarities of Matrix Arithmetic

Manipulations involving matrices need particular care because the matrix multiplication need not be commutative. You must be on guard for subconscious assumptions that are based on your experience with real numbers and which do not hold for matrices. Here is a sample of pitfalls:

- For real or complex numbers α and β, we know that $(\alpha+\beta)(\alpha-\beta) = \alpha^2 - \beta^2$. However for square matrices A, B of the same dimension, we note that $(A + B)(A - B) = A^2 - AB + BA - B^2$; since $AB \neq BA$ in general, we cannot cancel to obtain $A^2 - B^2$.
- Again for square matrices of the same dimension $(A + B)^2 = A^2 + AB + BA + B^2$ and so in general $(A + B)^2 \neq A^2 + 2AB + B^2$.
- Again for for square matrices A, B of the same dimension $(AB)^2 = ABAB$ and so in general $(AB)^2 \neq A^2B^2$.

We must never talk of dividing matrices; if we write A/B do we mean $B^{-1}A$ or AB^{-1}? In the matrix world, the two do not have to be same. For reals and complexes there is no ambiguity in writing α/β since $\beta^{-1}\alpha$ and $\alpha\beta^{-1}$ are same. Study the proof of the following lemma.

Lemma 6.1. *Let A be a square matrix of size m, B and C be matrices of size $m \times n$. Suppose A is invertible. If $AB = AC$ then $B = C$.*

Proof. If we were dealing with real or complex numbers we would just say 'divide both sides by A', but we are dealing with matrices. One of the hypotheses of the lemma is that A is invertible. So premultiply [1] both sides of $AB = AC$ by A^{-1}:

$$A^{-1}AB = A^{-1}AC.$$

But $A^{-1}A = I_m$ (the identity matrix) and $I_mB = B$, $I_mC = C$ so we get $B = C$. □

Exercise 6.1. Give an example to show that the above lemma fails if A is not invertible.

Another of the properties of real and complex numbers that does not extend to matrices is the following: if $\alpha\beta = 0$ then $\alpha = 0$ or $\beta = 0$. As we know, this property of real and complex numbers is a consequence of the existence of multiplicative inverses of non-zero numbers. But we know that there are non-zero square matrices that are not invertible. So we should also expect this property to fail in the matrix setting, and indeed it does. For example, let

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

Then

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

but A and B are non-zero.

7. Determinants

Definition. Let A be a 2×2 matrix and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We define the **determinant** of A, written $\det(A)$ to be

$$\det(A) = ad - bc.$$

Another common notation for the determinant of the matrix A is the following

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

Theorem 7.1. *(Properties of Determinants) Let A, B be 2×2 matrices.*

(i) $\det(I_2) = 1$.
(ii) $\det(AB) = \det(A)\det(B)$.
(iii) *If A is invertible then* $\det(A) \neq 0$ *and* $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
(iv) *Let n be a positive integer. Then* $\det(A^n) = \det(A)^n$. *If A is invertible then this is also true for negative n.*
(v) $\det(A^t) = \det(A)$.

[1] Premultiply means multiply on the left and postmultiply means multiply on the right. For real and complex numbers this vocabulary is not used since multiplication is commutative, and so it does not matter on which side we multiply.

Proof. The proof is mostly left as an exercise for the reader. Parts (i), (ii) and (v) follow from the definition and effortless calculations. For (iii) note that

$$\det(A^{-1}A) = \det(I_2) = 1.$$

Now applying (ii) we have $1 = \det(A^{-1})\det(A)$. We see that $\det(A) \neq 0$ and $\det(A^{-1}) = 1/\det(A)$.

For positive n, the formula $\det(A^n) = \det(A)^n$ can be proved by a straightforward induction. To prove it for negative n we argue as follows. Suppose n is negative and write $m = -n$, which is clearly positive. Since you have already proved (iv) for positive exponents, we know that $\det(A^m) = \det(A)^m$. Now suppose that A is invertible. From (iii) we know that $\det(A^{-1}) = 1/\det(A)$ and

$$\det(A^{-m}) = \det((A^m)^{-1}) = 1/\det(A^m).$$

Thus

$$\det(A^n) = \det(A^{-m}) = 1/\det(A^m) = 1/\det(A)^m = \det(A)^{-m} = \det(A)^n.$$

This proves (iv) for negative exponents. (Go through the argument carefully and make sure you understand each step). □

Exercise 7.1. (The Geometric Meaning of Determinant) You might be wondering (in fact should be wondering) about the geometric meaning of the determinant. This exercise answers your question. Let Let A be a 2×2 matrix and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Let $\mathbf{u} = \begin{pmatrix} a \\ c \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} b \\ d \end{pmatrix}$; in otherwords, $\mathbf{u}$ and $\mathbf{v}$ are the columns of A. Show that $|\det(A)|$ is the area of the parallelogram with adjacent sides $\mathbf{u}$ and $\mathbf{v}$ (See Figure 1). This tells you the meaning of $|\det(A)|$, but what about the sign of $\det(A)$? What does it mean geometrically? Write down and sketch a few examples and see if you can make a guess. Can you prove your guess?

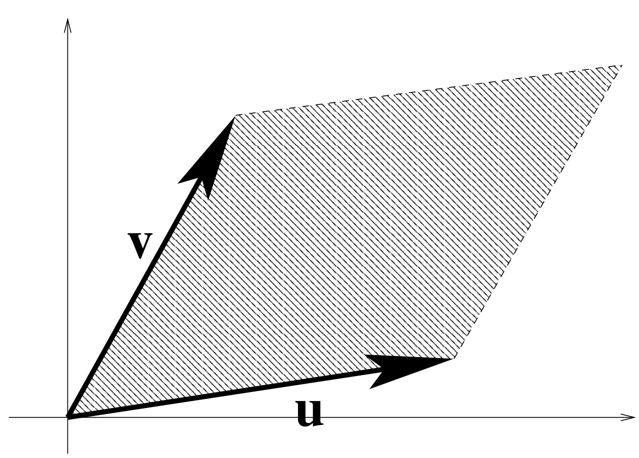

Figure 1. If $\mathbf{u}$ and $\mathbf{v}$ are the columns of A then the shaded area is $|\det(A)|$.

Exercise 7.2. Suppose $\mathbf{u} = \begin{pmatrix} a \\ c \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} b \\ d \end{pmatrix}$ are non-zero vectors, and let A be the matrix with columns $\mathbf{u}$ and $\mathbf{v}$; i.e. $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show (algebraically) that $\det(A) = 0$ if and only if $\mathbf{u}$, $\mathbf{v}$ are parallel. Explain this geometrically.

Exercise 7.3. Let $A = \begin{pmatrix} \alpha & 1 \\ 1 & 1 \end{pmatrix}$. For which values of α is $\det(A^5 - A^4) = -16$? (**Warning: This will be very messy unless you use the properties of determinants**)

8. Inverses

One reason why determinants are important lies the following theorem.

Theorem 8.1. *Let A be a 2×2 matrix, and write*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then A is invertible if and only if $\det(A) \neq 0$, in which case

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

Proof. Let

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

Then we check that

$$AB = \begin{pmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{pmatrix} = \begin{pmatrix} \det(A) & 0 \\ 0 & \det(A) \end{pmatrix} = \det(A) I_2.$$

Similarly (check), $BA = \det(A) I_2$. Suppose $\det(A)$ is non-zero. Then we see that

$$A \left(\frac{1}{\det(A)} B \right) = \left(\frac{1}{\det(A)} B \right) A = I_2.$$

Hence A is invertible and its inverse is

$$A^{-1} = \frac{1}{\det(A)} B = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

as required.

We have shown that if $\det(A) \neq 0$ then A is invertible. By Theorem 7.1, part (ii), we know that if A is invertible then $\det(A) \neq 0$. Thus A is invertible if and only if $\det(A) \neq 0$. This completes the proof. $\square$

Remark. The notion of determinant generalizes to arbitrary square matrices. It is also true that if A is a square matrix, then A is invertible if and only if $\det(A) \neq 0$. We leave this generalization to the Linear Algebra course, but it is helpful to be aware of this.

9. Linear Systems of Equations

We intend to apply what we have learned about matrices to solving systems of two linear equations equations in two variables. We begin this section with an example.

Example 9.1. Solve the system of simultaneous equations

$$2x + 3y = 8, \qquad x - 5y = -9. \tag{31}$$

Answer: Let

$$A = \begin{pmatrix} 2 & 3 \\ 1 & -5 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 8 \\ -9 \end{pmatrix}.$$

Think of the column vectors $\mathbf{x}$ and $\mathbf{b}$ as 2×1 matrices. Thus we can form the product $A\mathbf{x}$:

$$A\mathbf{x} = \begin{pmatrix} 2x + 3y \\ x - 5y \end{pmatrix}.$$

We note that $A\mathbf{x} = \mathbf{b}$ if and only if the unknowns x, y satisfy the pair of equations (31). Thus we can replace our two simultaneous equations with a single matrix equation $A\mathbf{x} = \mathbf{b}$ in unknown vector $\mathbf{x}$. Now $\det(A) = -10 - 3 = -13 \neq 0$. Hence A is invertible. Note $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x} = A^{-1}\mathbf{b}$. This means that the simultaneous system has a unique solution. We can get the solution by calculating

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{-13}\begin{pmatrix} -5 & -3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 8 \\ -9 \end{pmatrix} = \frac{1}{-13}\begin{pmatrix} -13 \\ -26 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

In other words, $x = 1$, $y = 2$ is the unique solution to (31).

You are probably thinking that the method used in the example above isn't the fastest way to solve a pair of equations in two variables. This is correct; you would have probably got the answer quicker by substitution from one equation into the other.

- To solve a system of m linear equations in n variables, the most efficient and reliable method is echelon reduction (also called Gaussian elimination). This method systematizes the ad hoc substitution method that you learned at school. Echelon reduction will be done in Linear Algebra.
- But echelon reduction, and the ad hoc method are a poor way to **think** about the solutions of systems of linear equations. Every system of m linear equations in n unknowns can be written in the form $A\mathbf{x} = \mathbf{b}$ where A is an $m \times n$ matrix, $\mathbf{b}$ is a column vector in $\mathbb{R}^m$ and $\mathbf{x}$ is an unknown column vector in $\mathbb{R}^n$. This is the best way to think about the linear system of equations. By 'think about', we mean 'study theoretically', 'prove theorems about', etc. Have a look at Theorems 9.2 and 9.3 below. These theorems would be difficult to formulate, let alone prove, if we weren't using matrix notation.

Consider a general system of m linear equations in n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{32}$$

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix},$$

and

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Then the system (32) can be rewritten as a single matrix equation $A\mathbf{x} = \mathbf{b}$. We still think of the single equation $A\mathbf{x} = \mathbf{b}$ as a system of linear equations.

The following terminology is standard, and the reader should know it.

Definition. We call the system of linear equation $A\mathbf{x} = \mathbf{b}$ **homogeneous** if $\mathbf{b} = \mathbf{0}$; otherwise we call it **inhomogeneous**. The system $A\mathbf{x} = \mathbf{b}$ is called **consistent** if it has a solution; otherwise, it is called **inconsistent**.

9.1. Homogeneous Systems. A homogeneous system $A\mathbf{x} = \mathbf{0}$ always has the solution $\mathbf{x} = \mathbf{0}$, which is called the **trivial solution**. Any non-zero solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is called **non-trivial**.

Lemma 9.1. *Suppose A is an $m \times n$ matrix.*

(a) $\mathbf{0}$ *is a solution to the homogeneous system* $A\mathbf{x} = \mathbf{0}$.
(b) *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are solutions to* $A\mathbf{x} = \mathbf{0}$ *then* $(\mathbf{u} + \mathbf{v})$ *is also a solution to* $A\mathbf{x} = \mathbf{0}$.
(c) *If* $\mathbf{u}$ *is a solution to* $A\mathbf{x} = \mathbf{0}$ *and* λ *is a scalar, then* $\lambda\mathbf{u}$ *is also a solution to* $A\mathbf{x} = \mathbf{0}$.

PROOF. The proof is a very easy exercise [2] □

Theorem 9.2. *Suppose that A is an $n \times n$ (i.e. square) matrix.*

(i) *If A is invertible then the homogeneous system* $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution.*
(ii) *If A is non-invertible, then the homogeneous system* $A\mathbf{x} = \mathbf{0}$ *has non-trivial solutions.*

PROOF. Suppose A is invertible. Premultiplying $A\mathbf{x} = \mathbf{0}$ by A^{-1} gives $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$. This shows that the only solution is the trivial one and so proves (i).

---

[2] The lemma shows that the set of solutions of $A\mathbf{x} = \mathbf{0}$ is a subset S of $\mathbb{R}^n$ satisfying three properties:

- $\mathbf{0} \in S$;
- if $\mathbf{u}, \mathbf{v} \in S$ then $\mathbf{u} + \mathbf{v} \in S$;
- if $\lambda \in \mathbb{R}$ and $\mathbf{u} \in S$ then $\lambda\mathbf{u} \in S$.

You will recall that any subset of $\mathbb{R}^n$ satisfying these three properties is called a *subspace of* $\mathbb{R}^n$ (we covered subspaces in section 15 of Chapter 3). Thus S is a subspace of $\mathbb{R}^n$; this particular subspace is called the *kernel of the matrix A*, and is more usually denoted by $\mathrm{Ker}(A)$. You will learn a lot more about subspaces and kernels in Linear Algebra.

We prove part (ii) only for the 2×2 case. Suppose now that A is non-invertible, and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Thus $\det(A) = ad - bc = 0$. We want to show that the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution. If $A = 0_{2\times 2}$ then the system has non-trivial solutions; e.g. $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. So we may suppose that A is non-zero. Let

$$\mathbf{v}_1 = \begin{pmatrix} -b \\ a \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} d \\ -c \end{pmatrix}.$$

Note that

$$A\mathbf{v}_1 = \begin{pmatrix} -ab + ba \\ -cb + da \end{pmatrix} = \begin{pmatrix} 0 \\ \det(A) \end{pmatrix} = \mathbf{0}.$$

Likewise $A\mathbf{v}_2 = \mathbf{0}$. Hence both $\mathbf{v}_1$ and $\mathbf{v}_2$ are solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$. If both were zero then $a = b = c = d = 0$ and so $A = 0_{2\times 2}$; this is impossible as we are assuming that $A \neq 0_{2\times 2}$. Thus at least one of $\mathbf{v}_1$ and $\mathbf{v}_2$ is a non-trivial solution to $A\mathbf{x} = \mathbf{0}$. This completes the proof. □

9.2. Inhomogeneous Systems.

Theorem 9.3. *Suppose that A is an $n \times n$ (i.e. square) matrix.*

(i) *If A is invertible then the system $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.*
(ii) *If A is non-invertible, then the system $A\mathbf{x} = \mathbf{b}$ either has no solutions, or it has infinitely many solutions.*

Proof. Suppose A is invertible. Premultiplying $A\mathbf{x} = \mathbf{b}$ by A^{-1} gives $\mathbf{x} = A^{-1}\mathbf{b}$. This proves (i).

Let us prove part (ii). Here we are supposing that A is non-invertible. We want to show that $A\mathbf{x} = \mathbf{b}$ either has no solutions or has infinitely many solutions.

Case I: The system $A\mathbf{x} = \mathbf{b}$ has no solutions. Then this is already one of the two possibilities that we want to prove and we're finished.

Case II: The system $A\mathbf{x} = \mathbf{b}$ has solutions. Let $\mathbf{v}$ be one of them. Thus $A\mathbf{v} = \mathbf{b}$.

We now apply part (ii) of Theorem 9.2. As A is not invertible, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has some non-trivial solution; call it $\mathbf{w}$. Thus $\mathbf{w} \neq \mathbf{0}$ (it is non-trivial) and $A\mathbf{w} = \mathbf{0}$.

We know that

$$A\mathbf{v} = \mathbf{b}, \qquad A\mathbf{w} = \mathbf{0}.$$

Now if t is a scalar then

$$A(\mathbf{v} + t\mathbf{w}) = A\mathbf{v} + tA\mathbf{w} = \mathbf{b} + t\mathbf{0} = \mathbf{b}.$$

In other words, for every t, the vector $\mathbf{v} + t\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$. But $\mathbf{w} \neq \mathbf{0}$. Thus we see that there are infinitely many solutions to $A\mathbf{x} = \mathbf{b}$. This proves (ii). □

Example 9.2. Consider the linear system

$$2x + 3y = 1,$$
$$4x + 6y = 2.$$

This can be rewritten as $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Notice that $\det(A) = 0$ and hence A is not invertible. You can check (by substitution) that $x = -1+3t$, $y = 1-2t$ gives a solution to the linear system for all values of t. Thus we obtain infinitely many solutions. It is actually possible to show that any solution must be of the form $x = -1+3t$, $y = 1-2t$ for some scalar t; we will not do this, but it will be easy for you once you have studied echelon reduction in Linear Algebra.

Now consider the linear system

$$2x + 3y = 1,$$
$$4x + 6y = 3.$$

This can be rewritten as $A\mathbf{x} = \mathbf{b}'$ where A is exactly the same (non-invertible) matrix above, and

$$\mathbf{b}' = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Notice that this system does not have any solutions. To see this multiply the first equation by 2 and subtract it from the second equation; you will get $0 = 1$ which is impossible.

9.3. A challenge. Perhaps you still don't see the point of matrices. If so, the following challenge will help you respect them. Theorem 9.3 is a theorem about systems of linear equations, but expressed in the language of matrices. Can you reformulate and prove this theorem using the old language of systems of linear equations.

10. Eigenvalues and Eigenvectors

Definition. Let A be an $n \times n$ matrix. Suppose that $\mathbf{v}$ is a non-zero vector and λ is a scalar such that $A\mathbf{v} = \lambda\mathbf{v}$. In this case we say that λ is an **eigenvalue** of A and $\mathbf{v}$ is an **eigenvector** that corresponds to the **eigenvalue** λ.

Note that an eigenvector of A is a vector whose direction is preserved (or reversed) on multiplying by A.

The assumption that $\mathbf{v} \neq \mathbf{0}$ in the definition is very important. If we allowed $\mathbf{0}$ as an eigenvector we have to allow all scalars as eigenvalues, and we don't want that.

Geometrically, an eigenvector is just a vector which when multiplied by the matrix keeps its direction (or just reverses it).

Example 10.1. Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then

$$A\mathbf{v} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\mathbf{v}.$$

Since $\mathbf{v}$ is non-zero, we see that 2 is an eigenvalue and $\mathbf{v}$ is an eigenvector that corresponds to the eigenvalue 2.

Definition. Suppose A is an $n \times n$ matrix. We define the *characteristic polynomial* of A to be the polynomial

$$\chi_A(x) = \det(xI_n - A).$$

Remark. We have only defined determinants of 2×2 matrices, but we did mention that the notion of determinant generalizes to arbitrary square matrices. The only property of $n \times n$ determinants that we will need is the following: an $n \times n$ matrix B is invertible if and only if $\det(B) \neq 0$.

The following theorem will tell us how to compute eigenvalues and eigenvectors of square matrices.

Theorem 10.1. *Suppose A is an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial $\chi_A(x)$. If λ is an eigenvalue of A, then $\mathbf{v}$ is an eigenvector corresponding to λ if and only if $\mathbf{v}$ is a non-trivial solution to the homogeneous system $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$.*

PROOF. Note that

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A &\iff A\mathbf{v} = \lambda\mathbf{v} \text{ for some } \mathbf{v} \neq \mathbf{0} \\
&\iff \text{there is a non-zero vector } \mathbf{v} \text{ such that } A\mathbf{v} = \lambda I_n \mathbf{v} \\
&\iff \text{there is a non-zero vector } \mathbf{v} \text{ such that } \lambda I_n \mathbf{v} - A\mathbf{v} = \mathbf{0} \\
&\iff \text{there is a non-zero vector } \mathbf{v} \text{ such that } (\lambda I_n - A)\mathbf{v} = \mathbf{0} \\
&\iff \text{the homogeneous system } (\lambda I_n - A)\mathbf{x} = \mathbf{0} \text{ has} \\
&\qquad \text{some non-trivial solution} \\
&\iff \text{the matrix } (\lambda I_n - A) \text{ is not invertible (Theorem 9.2)} \\
&\iff \det(\lambda I_n - A) = 0 \\
&\iff \lambda \text{ is a root of the polynomial } \chi_A(x) = \det(xI_n - A).
\end{aligned}$$

□

Example 10.2. Let

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

Find the eigenvalues and eigenvectors of $\mathbf{v}$.

Answer: The characteristic polynomial of A is

$$\begin{aligned}
\chi_A(x) = \det(xI_2 - A) &= \begin{vmatrix} x-1 & -2 \\ -2 & x-4 \end{vmatrix} \\
&= (x-1)(x-4) - 4 \\
&= x^2 - 5x = x(x-5).
\end{aligned}$$

So the eigenvalues of A are $\lambda_1 = 0$ and $\lambda_2 = 5$. To find an eigenvector corresponding to λ_1 we need to find a non-trivial solution to the homogeneous system $(\lambda_1 I_2 - A)\mathbf{x} = \mathbf{0}$. This can be rewritten as

$$\begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

In other words we want any non-trivial solution to

$$-x - 2y = 0, \qquad -2x - 4y = 0. \tag{33}$$

We see that $x = -2$, $y = 1$ is a non-trivial solution and so we can take

$$\mathbf{v}_1 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

as an eigenvector corresponding to $\lambda_1 = 0$. Notice that there are in fact infinitely many non-trivial solutions to the homogeneous system (33): namely we can take $x = -2t$ and $y = 4t$ for any non-zero value of t. **In this course**, we will only need to find one eigenvector corresponding to a particular eigenvalue. Thus, **in this course**, if a question says "find the eigenvalues and eigenvectors of (some matrix) A", you are expected to find all the eigenvalues and, for each eigenvalue, find one corresponding eigenvector. There are situations you will see in Linear Algebra where you will need to find more than one eigenvector for certain eigenvalues as we explain below, though this is not required for the Vectors and Matrices tests.

Similarly we find that

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

is an eigenvector that corresponds to λ_2.

Exercise 10.3. Let $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$. Use the definition of eigenvalue to show that $\alpha_1, \ldots, \alpha_n$ are eigenvalues of A (this means that you must find vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$such that $A\mathbf{v}_i = \alpha_i \mathbf{v}_i$).

Exercise 10.4. Let A, B be $n \times n$ matrices. Suppose that $\mathbf{v}$ is an eigenvector to both A and B. Show that $\mathbf{v}$ is an eigenvector to AB and to $A + B$. (**Hint**: use the definition of eigenvector).

Exercise 10.5. Suppose A is a square matrix and λ is an eigenvalue of A.

(i) Show that λ^n is an eigenvalue of A^n for all positive integers n.
(i) Suppose A is invertible. Show λ is non-zero and that λ^{-1} is an eigenvalue of A^{-1}.

11. Eigenspaces

This section on eigenspaces is not examinable, but reading it now will help you understand what is to come in Linear Algebra. Let A be an $n \times n$ matrix, and λ and eigenvalue of A. In other words λ is a root of the characteristic polynomial $\chi_A(x)$. Of course, when we count roots of a polynomial, we always care about multiplicity. The *algebraic multiplicity* of λ is the exact power of $(x - \lambda)$ that divides $\chi_A(x)$. A simple eigenvalue is one that has algebraic multiplicity 1. In Example 10.2 above both eigenvalues had algebraic multiplicity 1. Whereas in Example 11.1 below we have one eigenvalue of algebraic multiplicity 2.

Lemma 11.1. *Let λ be a simple eigenvalue of A. If $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of A corresponding to λ then $\mathbf{u} = \alpha\mathbf{v}$ for some non-zero scalar α.*

In other words, we only care to find one eigenvector for a simple eigenvalue because all the others are multiples of it and so knowing one means we know them all.

Let λ be an eigenvalue of A. We define the *eigenspace* of A corresponding to λ to be the set

$$V_\lambda = \{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \lambda\mathbf{v}\}.$$

In other words, the eigenspace of λ is the set of eigenvectors of λ together with the vector $\mathbf{0}$. It is easy for you to check that it is a subspace of $\mathbb{R}^n$ according to the definition given in Section 15 of Chapter 3.

We can rewrite the Lemma above as follows.

Lemma 11.2. *Let λ be a simple eigenvalue of A. Then its eigenspace V_λ is a straight line through the origin.*

What happens when λ is not simple. In this case the eigenspace can be bigger that just a line. Define the *geometric multiplicity* of λ to be the dimension of the eigenspace V_λ. What does dimension here mean? For now, just think of it intuitively: the origin on its own is 0-dimensional, a straight line through the origin is 1-dimensional, and plane through the origin is 2-dimensional, and so on. The whole story is

$$\text{geometric multiplicity} \leq \text{algebraic multiplicty}$$

and

$$\text{very often, though not always,} \quad \text{geometric multiplicity} = \text{algebraic multiplicty}.$$

Example 11.1. Let $A = I_2$. We find $\chi_A(x) = (x-1)^2$. Hence A has one eigenvalue $\lambda = 1$ of algebraic multiplicity 2. Now consider the corresponding eigenspace V_1. By definition it is

$$V_1 = \{\mathbf{v} \in \mathbb{R}^2 : A\mathbf{v} = 1\mathbf{v}\}.$$

But $A = I_2$ and so $A\mathbf{v} = I_2\mathbf{v} = \mathbf{v} = 1\mathbf{v}$ for all vectors $\mathbf{v} \in \mathbb{R}^2$. Hence $V_1 = \mathbb{R}^2$. In other words, the eigenspace is the whole of the plane $\mathbb{R}^2$ and so is 2-dimensional. We see that the geometric multiplicity of the eigenvalue $\lambda = 1$ is also 2.

Exercise 11.2. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Show that

(1) The only eigenvalue is $\lambda = 1$ and it has algebraic multiplicity 2.
(2) The corresponding eigenspace is the line $y = 0$ (in other words it is the x-axis). What is the geometric multiplicity?

12. Similarity and Diagonalization

Definition. Suppose A, B are $n \times n$ matrices. We say that A and B are **similar** if there is an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. We say that A is **diagonalizable** if it is similar to a diagonal matrix. In other words, A is **diagonalizable** if there is an invertible $n \times n$ matrix P such that $P^{-1}AP$ is a diagonal matrix. In this case we say that the matrix P **diagonalizes** the matrix A.

Theorem 12.1. *Suppose A is a 2×2 matrix with distinct [3] eigenvalues λ_1 and λ_2. Let $\mathbf{v}_1$, $\mathbf{v}_2$ be eigenvectors of A corresponding respectively to the eigenvalues λ_1 and λ_2. Let $P = (\mathbf{v}_1, \mathbf{v}_2)$ (this is the 2×2 matrix with columns $\mathbf{v}_1$ and $\mathbf{v}_2$). Then P is invertible and*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

[3] λ_1, λ_2 are **distinct** means that $\lambda_1 \neq \lambda_2$.

PROOF. Note that

$$AP = A(v_1, v_2) = (Av_1, Av_2) = (\lambda_1 v_1, \lambda_2 v_2) = (v_1, v_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

In other words

$$AP = P \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

We will prove in a minute that P is invertible. Then we can just premultiply both sides by P^{-1} to get $P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, which is exactly what we wanted to prove. Notice we haven't used the assumption that the eigenvalues are distinct, so we must use it in showing that P is invertible.

Let's show that P is invertible, which we do by contradition. So suppose it isn't invertible. By exercise 7.2, the columns $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel. Recall that they are eigenvectors and so, by definition, neither is zero. Hence $\mathbf{v}_1 = \alpha \mathbf{v}_2$ for some scalar $\alpha \neq 0$. Now $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ since $\mathbf{v}_1$ is the eigenvector corresponding to λ_1. However, pre-multiplying both sides of $\mathbf{v}_1 = \alpha \mathbf{v}_2$ with A we get

$$A\mathbf{v}_1 = A(\alpha \mathbf{v}_2) = \alpha A \mathbf{v}_2 = \alpha \lambda_2 \mathbf{v}_2 = \lambda_2(\alpha \mathbf{v}_2) = \lambda_2 \mathbf{v}_1.$$

In other words $\lambda_1 \mathbf{v}_1 = \lambda_2 \mathbf{v}_1$ and as $\mathbf{v}_1$ is non-zero, $\lambda_1 = \lambda_2$. This contradicts the assumption that the eigenvalues are distinct. □

Example 12.1. Let

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

Find a matrix P that diagonalizes A. Hence compute A^n for positive n.

Answer: In Example 10.2 we computed the eigenvalues and eigenvectors of A. We found that the eigenvalues are $\lambda_1 = 0$, $\lambda_2 = 5$, and that corresponding to these were respectively the eigenvectors

$$\mathbf{v}_1 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

We now let

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

this is the matrix with $\mathbf{v}_1$ an $\mathbf{v}_2$ as its columns. According to Theorem 12.1, the matrix P diagonalizes A. In fact the theorem says that

$$P^{-1}AP = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}. \tag{34}$$

To compute A^n we need the following observation:

$$\begin{aligned} (P^{-1}AP)^n &= (P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP) \\ &= P^{-1}A(PP^{-1})A(PP^{-1})A\cdots(PP^{-1})AP \\ &= P^{-1}AI_nAI_nA\cdots I_nAP \\ &= P^{-1}A^nP. \end{aligned}$$

Raising both sides of (34) to the power n we deduce,

$$(P^{-1}AP)^n = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}^n$$

and so

$$P^{-1}A^nP = \begin{pmatrix} 0 & 0 \\ 0 & 5^n \end{pmatrix}.$$

Pre-multiplying by P and post-multiplying by $P^{-1} = \begin{pmatrix} -2/5 & 1/5 \\ 1/5 & 2/5 \end{pmatrix}$ we get that

$$A^n = \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 5^n \end{pmatrix} \begin{pmatrix} -2/5 & 1/5 \\ 1/5 & 2/5 \end{pmatrix} = \begin{pmatrix} 5^{n-1} & 2 \times 5^{n-1} \\ 2 \times 5^{n-1} & 4 \times 5^{n-1} \end{pmatrix}.$$

Exercise 12.2. Let $A = \begin{pmatrix} 0 & -1 \\ 2 & 3 \end{pmatrix}$.

(i) Calculate the eigenvalues of A and corresponding eigenvectors.
(ii) Give a matrix P the diagonalizes A.
(iii) Calculate A^n for positive n.

Exercise 12.3. Show that similarity is an equivalence relation on $n \times n$ matrices. In other words show that

- A is similar to A;
- if A is similar to B then B is similar to A;
- if A is similar to B and B is similar to C the A is similar to C.

Exercise 12.4. Let A and B be similar 2×2 matrices.

(i) Show that A, B have the same characteristic polynomial (**Hint:** write $B = P^{-1}AP$ and $I_2 = P^{-1}I_2P$ in the definition of $\chi_B(x)$ and show that this is equal to $\chi_A(x)$.)
(ii) Part (i) shows that similar matrices have the same eigenvalues. Can you show this directly from the definition of eigenvalue?

13. Matrices as Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let A be an $m \times n$ matrix. Think of vectors in $\mathbb{R}^n$, $\mathbb{R}^m$ as column vectors. If $\mathbf{u}$ is a (column) vector in $\mathbb{R}^n$ then we can think of it as an $n \times 1$ matrix, and form the product $A\mathbf{u}$. The matrix $A\mathbf{u}$ is an $m \times 1$ matrix and so we can think of it as a (column) vector in $\mathbb{R}^m$. Thus multiplication by A gives a function $\mathbb{R}^n \to \mathbb{R}^m$, which we denote by T_A. In other words, we let

$$T_A : \mathbb{R}^n \to \mathbb{R}^m,$$

be given by $T_A(\mathbf{u}) = A\mathbf{u}$.

Example 13.1. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then A defines a function $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_A(\mathbf{u}) = A\mathbf{u}$. Let us calculate T_A explicitly:

$$T_A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

We note that, geometrically speaking, T_A represents reflection in the line $y = x$.

Example 13.2. Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then $T_A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}$. Thus geometrically, T_A represents projection onto the x-axis.

Theorem 13.1. *(Properties of T_A) Suppose that A is an $m \times n$ matrix, $\mathbf{u}_1$, $\mathbf{u}_2 \in \mathbb{R}^n$ and λ is a scalar. Then*

$$T_A(\mathbf{u}_1 + \mathbf{u}_2) = T_A(\mathbf{u}_1) + T_A(\mathbf{u}_2) \tag{35}$$

and

$$T_A(\lambda\, \mathbf{u}_1) = \lambda\, T_A(\mathbf{u}_1). \tag{36}$$

Proof. Note that

$$\begin{aligned} T_A(\mathbf{u}_1 + \mathbf{u}_2) &= A(\mathbf{u}_1 + \mathbf{u}_2) && \text{by definition of } T_A \\ &= A\mathbf{u}_1 + A\mathbf{u}_2 && \text{distributive property for matrices} \\ &= T_A(\mathbf{u}_1) + T_A(\mathbf{u}_2) && \text{by definition of } T_A \text{ again.} \end{aligned}$$

This proves (35); the proof of (36) is left as an exercise. □

The above theorem motivates the following definition.

Definition. A linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2) \tag{37}$$

for all vectors $\mathbf{u}_1$, $\mathbf{u}_2 \in \mathbb{R}^n$, and

$$T(\lambda\, \mathbf{u}) = \lambda\, T(\mathbf{u}) \tag{38}$$

for all $\mathbf{u} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

Now Theorem 13.1 can be rephrased as the first part of the following theorem.

Theorem 13.2. *Suppose A is an $m \times n$ matrix. Then the function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Conversely, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation then there is an $m \times n$ matrix A such that $T(\mathbf{u}) = T_A(\mathbf{u}) = A\mathbf{u}$ for all vectors $\mathbf{u} \in \mathbb{R}^n$. (We call A the matrix associated with the linear transformation T.)*

Proof. The first part follows from Theorem 13.1 and the definition of linear transformations. Let us prove the second part. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the following vectors in $\mathbb{R}^n$:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then $T(\mathbf{e}_1), \dots, T(\mathbf{e}_n)$ will be vectors in $\mathbb{R}^m$ and we can write

$$T(\mathbf{e}_1) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad T(\mathbf{e}_2) = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, T(\mathbf{e}_n) = \begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Notice that any vector in $\mathbb{R}^n$ can be rewritten in terms of $\mathbf{e}_1, \ldots, \mathbf{e}_n$ as follows:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

Thus

$$\begin{aligned}
T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\
&= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + \cdots + T(x_n\mathbf{e}_n) \\
&= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) \\
&= x_1\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}.
\end{aligned}$$

In other words, if we write

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}$$

then $T\mathbf{u} = A\mathbf{u} = T_A(\mathbf{u})$ for all vector $\mathbf{u} \in \mathbb{R}^n$. This completes the proof the second part of the theorem. □

Example 13.3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

We see straightaway that T represents reflection in the x-axis. The function T is actually a linear transformation. Let us prove this. First we want to prove (37) for all vectors $\mathbf{u}_1$, $\mathbf{u}_2$ in $\mathbb{R}^2$. So suppose $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^2$ and write

$$\mathbf{u}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

Then

$$\begin{aligned} T(\mathbf{u}_1 + \mathbf{u}_2) &= T \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} && \text{by definition of vector addition} \\ &= \begin{pmatrix} x_1 + x_2 \\ -(y_1 + y_2) \end{pmatrix} && \text{by the definition of } T \\ &= \begin{pmatrix} x_1 \\ -y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ -y_2 \end{pmatrix} && \text{by defining of vector addition.} \end{aligned}$$

But

$$T(\mathbf{u}_1) = \begin{pmatrix} x_1 \\ -y_1 \end{pmatrix}, \qquad T(\mathbf{u}_2) = \begin{pmatrix} x_2 \\ -y_2 \end{pmatrix}.$$

So $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$. This proves (37) algebraically, but see figure 2 for a geometric reason why this should be true.

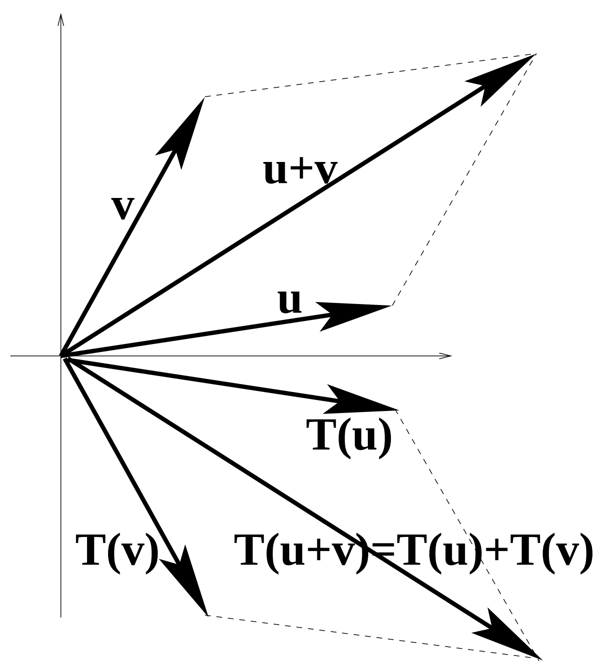

FIGURE 2. The end points of the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}+\mathbf{v}$, together with the origin form a parallelogram. We reflected the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$ in the x-axis. We obtain vectors $T(\mathbf{u})$, $T(\mathbf{v})$ and $T(\mathbf{u}+\mathbf{v})$. Note the end points of these three new vectors, together with the origin still form a parallelogram, simply because the reflection of a parallelogram is a parallolgram. This explains geometrically that $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.

Before declaring that T is a linear transformation we must show that it also satisfies (38) for all scalars λ and all vectors $\mathbf{u} \in \mathbb{R}^2$. So suppose that λ is a scalar and

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\begin{aligned} T(\lambda\mathbf{u}) &= T\begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix} && \text{by defn of mult of matrix by scalar} \\ &= \begin{pmatrix} \lambda x \\ -\lambda y \end{pmatrix} && \text{by definition of } T \\ &= \lambda \begin{pmatrix} x \\ -y \end{pmatrix} && \text{by defn of mult of matrix by scalar} \\ &= \lambda T\mathbf{u} && \text{by definition of } T. \end{aligned}$$

This proves (38) and completes that proof that T is a linear transformation. We have been told above that every linear transformation comes from a matrix. What is the matrix associated to this particular linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$? The matrix that we are looking for will be a 2×2 matrix. Let us call it A and write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We want $T = T_A$. In other words, we want

$$T\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} \tag{39}$$

to hold **for all** vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$. We can rewrite the identity (39) as

$$\begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Since this is to hold **for all** x, y, we see that $a = 1$, $b = 0$, $c = 0$, $d = -1$. In other words, the matrix A associated with the linear transformation T is

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

Example 13.4. Not every map $\mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation. For example, let $S: \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$S\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+1 \\ y \end{pmatrix}.$$

The map S represents a translation of 1 unit to the right. This S is not a linear transformation. To prove this we only need to give one counterexample to either condition (37) or to condition (38). For example, let us take

$$\mathbf{u} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \lambda = 2.$$

Then

$$S(\lambda\mathbf{u}) = S\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

whereas

$$\lambda S(\mathbf{u}) = 2\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}4\\2\end{pmatrix}.$$

So $S(\lambda\mathbf{u}) \neq \lambda S(\mathbf{u})$. This shows that S is not a linear transformation as condition (38) does not hold.

Exercise 13.5. Let $A = 2I_n$, $B = -I_n$ and $C = I_n$. What do the linear transformations T_A, T_B and T_C represent geometrically?

Exercise 13.6. Write down the 2×2 matrix corresponding to projection from the coordinate plan onto the y-axis.

Exercise 13.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}x\\y\\0\end{pmatrix}.$$

This is projection from 3-space onto the xy-plane. Show that T is a linear transformation. What is the matrix associated to T?

14. Why Column Vectors?

You will have noticed that of late we have taken a preference for column vectors as opposed to row vectors. Let us see how things are different if we stuck with row vectors. So for the moment think of elements of $\mathbb{R}^n$, $\mathbb{R}^m$ as row vectors. Let A be an $m \times n$ matrix.

If $\mathbf{u}$ is a (row) vector in $\mathbb{R}^n$ or $\mathbb{R}^m$ then $A\mathbf{u}$ is undefined. But we find that $\mathbf{u}A$ is defined if $\mathbf{u}$ is a (row) vector in $\mathbb{R}^m$ and gives a (row) vector in $\mathbb{R}^n$. Thus we get a function

$$S_A : \mathbb{R}^m \to \mathbb{R}^n$$

given by $S_A(\mathbf{u}) = \mathbf{u}A$. It is now a little harder to think of the matrix A as a function since we have written it on the right in the product $\mathbf{u}A$ (remember that when we thought of vectors as columns we wrote $A\mathbf{u}$).

Some mathematicians write functions on the right, so instead of writing $f(x)$ they will write xf. They will be happy to think of matrices as functions on row vectors because they can write the matrix on the right. Most mathematicians write functions on the left. They are happier to think of matrices as functions on column vectors because they can write the matrix on the left.

15. Rotations

We saw above some examples of transformations in the plane or in 3-space. In this section we take a closer look at rotations. Suppose that $P = \begin{pmatrix}x\\y\end{pmatrix}$ is a point in $\mathbb{R}^2$. Suppose that this point is rotated anticlockwise about the origin through an angle of θ. We want to write down the new point $P'\begin{pmatrix}x'\\y'\end{pmatrix}$ in terms of x, y and θ. Now suppose that the distance of P from the origin O is r and that the angle $\overrightarrow{OP}$ makes with the positive x-axis is ϕ. Thus

$$x = r\cos\phi, \qquad y = r\sin\phi.$$

Since we rotated P anticlockwise about the origin through an angle θ to obtain P', we see that the distance OP' is also r and the angle $\overrightarrow{OP'}$ makes with the positive x-axis is $\phi + \theta$. Thus

$$x' = r\cos(\phi + \theta), \qquad y' = r\sin(\phi + \theta).$$

We expand $\cos(\phi + \theta)$ to obtain

$$\begin{aligned} x' &= r\cos(\phi+\theta) \\ &= r\cos\phi\cos\theta - r\sin\phi\sin\theta \\ &= x\cos\theta - y\sin\theta. \end{aligned}$$

Similarly

$$y' = x\sin\theta + y\cos\theta.$$

We can rewrite the two relations

$$x' = x\cos\theta - y\sin\theta, \qquad y' = x\sin\theta + y\cos\theta,$$

in matrix notation as follows

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus anticlockwise rotation about the origin through an angle θ can be achieved by multiplying by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

This shows that rotation about the origin is really a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$.

We have shown algebraically that rotation about the origin is a linear transformation. But let us understand why this is so geometrically [4]. Pick an angle θ and let T be the map that rotates every point about the origin anticlockwise through this angle θ. This means that if we start with a vector $\mathbf{u}$ and we rotate anticlockwise about the origin by θ radians then we denote the new vector obtained by $T(\mathbf{u})$. Now suppose that we are given two vectors $\mathbf{u}$, $\mathbf{v}$ and let us apply T to three vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$. We know that the end points of the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} + \mathbf{v}$, together with the origin, form a parallelogram, as in left-half of Figure 3.

When we rotate the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}+\mathbf{v}$ we get $T(\mathbf{u})$, $T(\mathbf{v})$, $T(\mathbf{u}+\mathbf{v})$. Of course the whole parallelogram gets rotated, so the end points of $T(\mathbf{u})$, $T(\mathbf{v})$ and $T(\mathbf{u}+\mathbf{v})$ together with the origin form a new parallelogram, as in the right-half of Figure 3.

We see from the picture that the sum of the vectors $T(\mathbf{u})$ and $T(\mathbf{v})$ is $T(\mathbf{u}+\mathbf{v})$. Hence we have shown that

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}).$$

In the same way we can argue geometrically that $T(\lambda\mathbf{u}) = \lambda T(\mathbf{u})$. In other words, it doesn't matter if we stretch the vector first and then rotate it, or rotate the vector first and then stretch it. The end result will be the same. Since T satisfies the two defining properties of linear transformation, we see that T is a linear transformation.

[4] I told you long ago: "You should get used to thinking geometrically, and to drawing pictures. The true meaning of most mathematical concepts is geometric. If you spend all your time manipulating symbols (i.e. doing algebra) without understanding the relation to the geometric meaning, then you will have very little in terms of mathematical insight". No doubt you have taken my advice on board and so there is no need for me to repeat it.

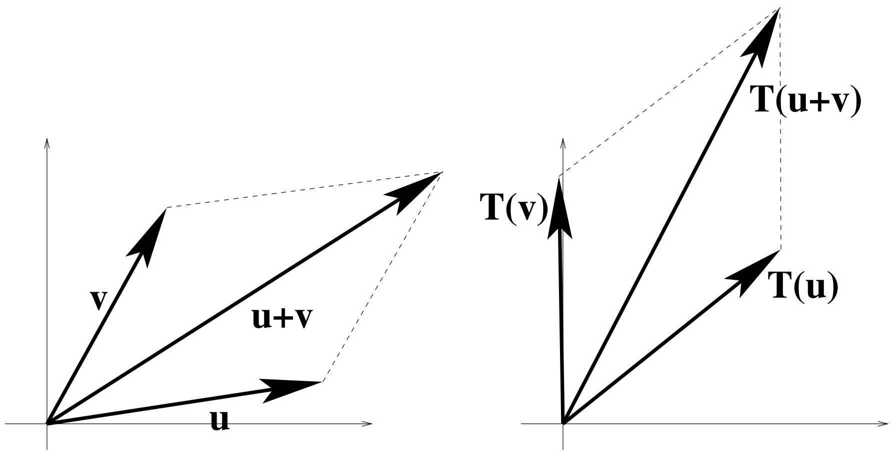

FIGURE 3. The end points of the vectors **u**, **v**, **u** + **v**, together with the origin form a parallelogram. We rotated the vectors **u**, **v** and **u** + **v** through an angle θ about the origin. We obtain vectors $T(\mathbf{u})$, $T(\mathbf{v})$ and $T(\mathbf{u}+\mathbf{v})$. Note the end points of these three new vectors, together with the origin still form a parallelogram.

Exercise 15.1. You know that R_θ represents anticlockwise rotation about the origin through angle θ. Describe in words the linear transformation associated to $-R_\theta$. (**Warning: don't be rash!**)

16. Orthogonal Matrices

Definition. A $n \times n$ matrix A is **orthogonal** if $A^t A = AA^t = I_n$.

In other words, an orthogonal matrix is invertible and its inverse is its own transpose.

Exercise 16.1. Let $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

(i) What is $\det(R_\theta)$?
(ii) Show that R_θ is orthogonal (recall that an $n \times n$ matrix A is **orthogonal** if $A^t A = AA^t = I_n$).
(iii) Show algebraically that $R_\phi R_\theta = R_{\phi+\theta}$.
(iv) Use the geometric interpretation of the matrix R_θ to explain (iii).
(v) Use the geometric interpretation of the matrix $R_{\pi/2}$ to explain why it cannot have real eigenvalues and eigenvectors.
(vi) Compute the eigenvalues and corresponding eigenvectors for $R_{\pi/2}$. Hence diagonalize $R_{\pi/2}$.

Exercise 16.2. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an orthogonal matrix. We will show in steps that either $A = \pm R_\theta$ or $A = SR_\theta$ for some θ, where $S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

(i) Show that $a^2 + c^2 = 1$, $b^2 + d^2 = 1$ and $ab + cd = 0$.

(ii) Deduce the existence of angles ϕ, ψ such that $a = \cos\phi$, $c = \sin\phi$, $b = \cos\psi$, $d = \sin\psi$.
(iii) Substitute into $ab + cd = 0$ and deduce that $\phi = \psi \pm \pi/2$.
(iv) Deduce that $A = R_\theta$ or $A = SR_\theta$ for some θ.
(v) You know that R_θ represents anti-clockwise rotation about the origin through an angle θ. Describe in words the linear transformation associated with the matrix SR_θ.

17. More on Rotations

Recall that to rotate a vector in $\mathbb{R}^2$ anticlockwise about the origin through an angle θ we multiply by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Let $\mathbf{v} \in \mathbb{R}^2$. What is the vector $R_\theta R_\phi \mathbf{v}$? It is the vector we get if we rotate $\mathbf{v}$ anticlockwise through an angle ϕ and then again through an angle θ. In other words $R_\theta R_\phi \mathbf{v}$ is what we obtain by rotating $\mathbf{v}$ through an angle $\theta + \phi$. Hence, we expect that

$$R_{\theta+\phi} = R_\theta R_\phi.$$

Verify this algebraically. Now this identity turns addition into multiplication, and so it should remind you of the identity $e^{\theta+\phi} = e^\theta e^\phi$. In fact, a more accurate analogy is identity

$$e^{i(\theta+\phi)} = e^{i\theta} e^{i\phi}.$$

The reason is because multiplying a complex number by $e^{i\theta}$ rotates it about the origin anticlockwise through the angle θ (prove this using the exponential form for complex numbers).

Now that you know that R_θ and $e^{i\theta}$ are analogues, you should have no trouble guessing what the matrix analogues of the n-th roots of unity are. If we let

$$\mathcal{Z} = R_{2\pi/n} = \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix},$$

then $I_2, \mathcal{Z}, \dots, \mathcal{Z}^{n-1}$ all satisfy the relationship $A^n = I_2$. What do you expect their sum to be? Can you prove it? **Why is it incorrect to write**

$$I_2 + \mathcal{Z} + \cdots + \mathcal{Z}^{n-1} = \frac{I_2 - \mathcal{Z}^n}{I_2 - \mathcal{Z}} = \frac{I_2 - I_2}{I_2 - \mathcal{Z}} = 0\ ?$$

18. The Image of a Linear Transformation

Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. The **image** of T, denoted by $\text{Im}(T)$, is defined in set-theoretic notation as follows:

$$\text{Im}(T) = \{T(\mathbf{v}) \ : \ \mathbf{v} \in \mathbb{R}^n\}.$$

Here is the geometric meaning: if you imagine the vector $\mathbf{v}$ moving in $\mathbb{R}^m$, then $\text{Im}(T)$ is the shape cut out by $T(\mathbf{v})$.

An equivalent way of defining $\text{Im}(T)$ is

$$\text{Im}(T) = \{\mathbf{w} \in \mathbb{R}^m \ : \ \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in \mathbb{R}^n\ \}$$

In other words, $\text{Im}(T)$ is the collection of those things in the codomain that are 'hit' by T.

Here are some examples which should clarify this idea.

Example 18.1. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in \mathbb{R}^2$. Then $\operatorname{Im}(T) = \{\mathbf{0}\}$; i.e. it is just the origin.

Example 18.2. Let $T: \mathbb{R} \to \mathbb{R}^2$ be given by $T(\alpha) = (\alpha, \alpha)$. You can check that T is a linear transformation. What is the shape cut out by $T(\alpha)$ as α ranges through the real numbers? We can rewrite

$$T(\alpha) = (\alpha, \alpha) = \alpha(1, 1).$$

Thus, as α ranges through the real numbers, $T(\alpha)$ ranges through the multiples of the vector $(1, 1)$. Hence we see that $\operatorname{Im}(T)$ is just the diagonal straight line $y = x$.

Example 18.3. Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(\alpha, \beta) = (\alpha, \beta, 0)$. It is easy to see that $\operatorname{Im}(T)$ is the plane in $\mathbb{R}^3$ given by $z = 0$.

Example 18.4. Here is a more serious example. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation. What can we say about its image? Note that any vector in $\mathbb{R}^2$ can be written as $(\alpha, \beta) = \alpha\mathbf{i} + \beta\mathbf{j}$. Since T is a linear transformation,

$$T(\alpha, \beta) = T(\alpha\mathbf{i} + \beta\mathbf{j}) = \alpha T(\mathbf{i}) + \beta T(\mathbf{j}).$$

Let $\mathbf{v} = T(\mathbf{i})$, $\mathbf{w} = T(\mathbf{j})$ and note that these are vectors in $\mathbb{R}^3$. Then

$$T(\alpha, \beta) = \alpha\mathbf{v} + \beta\mathbf{w}.$$

There are now three cases:

Case (i) $\mathbf{v} = \mathbf{w} = \mathbf{0}$. In this case $\operatorname{Im} T = \{\mathbf{0}\}$ is just the origin.

Case (ii) $\mathbf{v}$ is a multiple of $\mathbf{w}$ or $\mathbf{w}$ is a multiple of $\mathbf{v}$ (in other words, they are parallel). Let us say that $\mathbf{w} = \lambda\mathbf{v}$. Then

$$T(\alpha, \beta) = \alpha\mathbf{v} + \beta\mathbf{w} = \alpha\mathbf{v} + \lambda\beta\mathbf{v} = t\mathbf{v},$$

where $t = \alpha + \lambda\beta$ is a scalar. We see that $\operatorname{Im}(T)$ is the set of all scalar multiples of the vector $\mathbf{v}$. This is a straight line through the origin.

Case (iii) Neither $\mathbf{v}$ is a multiple of $\mathbf{w}$ nor $\mathbf{w}$ is a multiple of $\mathbf{v}$. Then the image is the plane passing through the origin which contains the vectors $\mathbf{v}$ and $\mathbf{w}$ (see Section 14 of Chapter 3).

Exercise 18.5. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(x, y) = (x+y, x+y, x+y)$. Show that the image of T is the line $x = y = z$.

Exercise 18.6. Give an explicit linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ whose image is the plane $x + y + z = 0$ (**Hint:** It would help to write the equation of the plane in vector form).

CHAPTER 5

Vector Product (or Cross Product) in $\mathbb{R}^3$

1. Not another chapter? Will this course ever finish?

In $\mathbb{R}^3$, apart from the dot product there is another product called the vector product. Unlike the dot product, which produces a scalar, the vector product gives a vector.

Definition. Let $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^3$ and write $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$. Define the vector product (or cross product) of $\mathbf{u}$ and $\mathbf{v}$ (written as $\mathbf{u} \times \mathbf{v}$) to be

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1).$$

This can be rewritten in determinant notation as

$$\mathbf{u} \times \mathbf{v} = \left(\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}, - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}, \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \right). \tag{40}$$

We emphasize that the vector product is only defined for vectors in $\mathbb{R}^3$.

The definition is a little awkward to use as it is a little hard to remember. Here is an easy way to remember it:

- Form the 2×3 matrix
$$\begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}$$
whose first row is $\mathbf{u}$ and second row is $\mathbf{v}$.
- To find the first coordinate of $\mathbf{u} \times \mathbf{v}$ delete the first column and take the determinant.
- To find the second coordinate of $\mathbf{u} \times \mathbf{v}$ delete the second column and take **MINUS** the determinant.
- To find the third coordinate of $\mathbf{u} \times \mathbf{v}$ delete the third column and take the determinant.

Example 1.1. Let $\mathbf{u} = (7, 2, -3)$ and $\mathbf{v} = (1, 0, 2)$. Find $\mathbf{u} \times \mathbf{v}$.

Answer: We first write
$$\begin{pmatrix} 7 & 2 & -3 \\ 1 & 0 & 2 \end{pmatrix}.$$
Following the recipe above, we see that
$$\mathbf{u} \times \mathbf{v} = \left(\begin{vmatrix} 2 & -3 \\ 0 & 2 \end{vmatrix}, - \begin{vmatrix} 7 & -3 \\ 1 & 2 \end{vmatrix}, \begin{vmatrix} 7 & 2 \\ 1 & 0 \end{vmatrix} \right) = (4, -17, -2).$$

Theorem 1.1. (Properties of the vector product) *Suppose that* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ *are vectors in* $\mathbb{R}^3$ *and* λ *is a scalar. Then*

(i) $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$ *(vector product is anti-commutative)*
(ii) $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$ *(distributive law)*

(iii) $(\mathbf{v}+\mathbf{w})\times\mathbf{u}=\mathbf{v}\times\mathbf{u}+\mathbf{w}\times\mathbf{u}$ *(distributive law)*
(iv) $\lambda(\mathbf{u}\times\mathbf{v})=(\lambda\mathbf{u})\times\mathbf{v}=\mathbf{u}\times(\lambda\mathbf{v})$.
(v) $\mathbf{u}\times\mathbf{u}=\mathbf{0}$
(vi) $\mathbf{u}\times\mathbf{0}=\mathbf{0}\times\mathbf{u}=\mathbf{0}$
(vii) $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=0$ *and* $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})=0$ *($\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$)*
(viii) $\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2$.

PROOF. All of the above can be proved directly using the definitions. We prove (i) and leave the rest as exercises. Note that when $\mathbf{u}$ and $\mathbf{v}$ are swapped in the vector product, we swap the rows of the determinants in (40). Now swapping the rows of a 2×2 determinant changes its sign. To see this, note that

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

This shows that $\mathbf{v}\times\mathbf{u}=-\mathbf{u}\times\mathbf{v}$. □

Exercise 1.2. Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be vectors in $\mathbb{R}^3$. Explain why $(\mathbf{u}\cdot\mathbf{v})\times\mathbf{w}$ is undefined and why $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ is defined.

Exercise 1.3. Let $\mathbf{w}\in\mathbb{R}^3$. Show that the map $S_{\mathbf{w}}:\ \mathbb{R}^3\to\mathbb{R}^3$ given by $S_{\mathbf{w}}(\mathbf{u})=\mathbf{w}\times\mathbf{u}$ is a linear transformation (**Hint:** use the properties of the vector product). What is the matrix associated with $S_{\mathbf{i}}$? Describe the image of $S_{\mathbf{i}}$.

2. Physical/Geometric Interpretation of the Vector Product

Theorem 2.1. *Let $\mathbf{u}$, $\mathbf{v}$ be non-zero vectors in $\mathbb{R}^3$, and let θ be the angle between them. Then $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Moreover,*

$$\|\mathbf{u}\times\mathbf{v}\|=\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta.$$

PROOF. The first part of the theorem is a repetition of part (vii) of Theorem 1.1, which you have already proved as an exercise.

For the second part of the theorem, we use the identity in part (viii) of Theorem 1.1: $\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2$. Now recall that $\mathbf{u}\cdot\mathbf{v}=\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$. Thus

$$\begin{aligned}\|\mathbf{u}\times\mathbf{v}\|^2&=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-\Big(\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta\Big)^2\\&=\|\mathbf{u}\|^2\|\mathbf{v}\|^2(1-\cos^2\theta)\\&=\|\mathbf{u}\|^2\|\mathbf{v}\|^2\sin^2\theta.\end{aligned}$$

Now $\|\mathbf{u}\|$, $\|\mathbf{v}\|$ are positive, since $\mathbf{u}$, $\mathbf{v}$ are non-zero vectors. Moreover, $0\le\theta\le\pi$. Thus $\sin\theta\ge 0$. Hence

$$\|\mathbf{u}\times\mathbf{v}\|=\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta.$$

Here we needed the positivity so we do not end up with two square-roots. □

2.1. The Right-Hand Rule. We know from the above that $\mathbf{u}\times\mathbf{v}$ is a vector that is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. There are in fact two opposite directions that are orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, and for some applications it is important to know which of these $\mathbf{u}\times\mathbf{v}$ takes. To determine this we use what is known as 'the right-hand rule'. Imagine that the thumb of your right hand is pointing in the direction of $\mathbf{u}$ and the index finger of your right-hand is pointing in the direction of $\mathbf{v}$. Then the middle finger will point in the direction of $\mathbf{u}\times\mathbf{v}$.

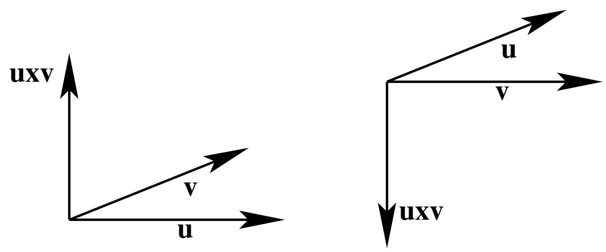

FIGURE 1. Right-Hand Rule: imagine that the thumb of your right-hand is pointing in the direction of $\mathbf{u}$ and the index finger of your right-hand is pointing in the direction of $\mathbf{v}$. Then the middle finger will point in the direction of $\mathbf{u}\times\mathbf{v}$.

2.2. Area of the Parallelogram. The reader will no doubt recall the formula for the area of the parallelogram with adjacent sides of length a, b and angle θ between them: it is $ab\sin\theta$. We instantaneously deduce the following corollary.

Corollary 2.2. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* $\mathbb{R}^3$*, then* $\|\mathbf{u}\times\mathbf{v}\|$ *is equal to the area of the parallelogram with sides* $\mathbf{u}$ *and* $\mathbf{v}$*.*

Exercise 2.1.
- (i) Compute $(\mathbf{i}\times\mathbf{i})\times\mathbf{j}$ and $\mathbf{i}\times(\mathbf{i}\times\mathbf{j})$. What do you notice? What is wrong with writing $\mathbf{u}\times\mathbf{v}\times\mathbf{w}$?
- (ii) Let $\mathbf{u}=(1,0,3)$, $\mathbf{v}=(0,1,-1)$. Compute $\mathbf{u}\times\mathbf{v}$. Hence find the two unit vectors that are orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.
- (iii) Find the area of the parallelogram that has **unit** vectors $\mathbf{u}$ and $\mathbf{v}$ as adjacent sides, if $\mathbf{u}\cdot\mathbf{v}=\sqrt{3}/2$.

Exercise 2.2. Find the two unit vectors parallel to the xy-plane that are perpendicular to the vector $(-2,3,5)$.

3. Triple Scalar Product

What does $\mathbf{u}\cdot\mathbf{v}\times\mathbf{w}$ mean? Does it mean $(\mathbf{u}\cdot\mathbf{v})\times\mathbf{w}$, or does it mean $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$? After a little thought we notice that $\mathbf{u}\cdot\mathbf{v}$ is a scalar, and we cannot takes its vector product with a vector. However, $\mathbf{v}\times\mathbf{w}$ is a vector. Thus $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ does have a meaning and gives a scalar. Thus when we write $\mathbf{u}\cdot\mathbf{v}\times\mathbf{w}$ we mean $\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$ and we call this the triple scalar product because its result is a scalar. The triple scalar product has an important geometric interpretation.

Theorem 3.1. *Let* $\mathbf{u}$*,* $\mathbf{v}$*,* $\mathbf{w}$ *be vectors in* $\mathbb{R}^3$*. The volume of the parallelepiped with sides* $\mathbf{u}$*,* $\mathbf{v}$*,* $\mathbf{w}$ *is* $|\mathbf{w}\cdot\mathbf{u}\times\mathbf{v}|$*.*

PROOF. You will remember from school that the formula:

$$\text{volume of parallelepiped} = \text{area of base} \times \text{height}.$$

We will use this formula to prove that the volume also equals $|\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}|$. Study Figure 2 carefully.

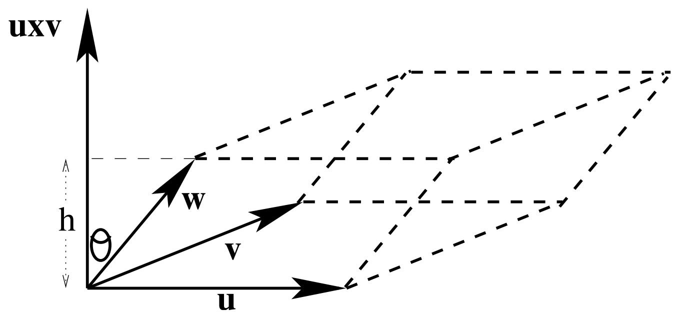

FIGURE 2

The area of the base is just the area of the parallelogram with sides $\mathbf{u}$ and $\mathbf{v}$. Hence

$$\text{area of base} = \|\mathbf{u} \times \mathbf{v}\|.$$

Recall that the direction of the vector $\mathbf{u} \times \mathbf{v}$ is perpendicular to the plane containing $\mathbf{u}$ and $\mathbf{v}$. In the figure, we denote the height by h and the angle the vector $\mathbf{w}$ makes with the $\mathbf{u} \times \mathbf{v}$ by θ. In other words, θ is the angle $\mathbf{w}$ makes with the perpendicular to the plane containing $\mathbf{u}$ and $\mathbf{v}$. Note that height$h = \|\mathbf{w}\| \cos\theta$. Hence

$$\begin{aligned} \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) &= \|\mathbf{w}\| \, \|\mathbf{u} \times \mathbf{v}\| \, \cos\theta \\ &= \|\mathbf{u} \times \mathbf{v}\| \, (\|\mathbf{w}\| \, \cos\theta) \\ &= \text{area of base} \times \text{height} = \text{volume of parallelepiped}. \end{aligned}$$

Is this the end of the proof? There is an unfinished point. Why do we have an absolute value in $|\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})|$ in the statement of the theorem? We didn't seem to use this in the proof. The truth is that the figure contains a subtle assumption: the subtle assumption is that $\mathbf{u} \times \mathbf{v}$ and $\mathbf{w}$ are both on the same side of the base. Note that the direction of $\mathbf{u} \times \mathbf{v}$ is controlled by the right-hand rule. What happens if you swap $\mathbf{u}$ and $\mathbf{v}$? Then $\mathbf{u} \times \mathbf{v}$ will be pointing downwards as in Figure 3.

Then

$$\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = \|\mathbf{w}\| \, \|\mathbf{u} \times \mathbf{v}\| \, \cos(\pi - \theta),$$

and we will have to make use of the fact that $\cos(\pi - \theta) = -\cos(\theta)$. Thus $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$ is the volume of the parallelepiped *up to sign*. □

No doubt, it has occured to you in the above proof, that it doesn't matter which two vectors we take as base—this should always give us the same answer. Hence you will have noted that

$$|\mathbf{w} \cdot \mathbf{u} \times \mathbf{v}| = |\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}| = |\mathbf{v} \cdot \mathbf{u} \times \mathbf{w}|.$$

If you are a little more careful, you will reach the following conclusion:

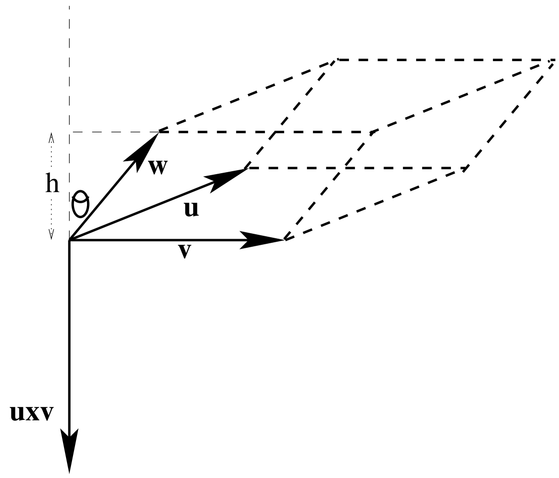

FIGURE 3

Theorem 3.2. *Cyclic permutations (see below) of the vectors* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ *do not affect the triple vector product* $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}$. *Non-cyclic permutations simply change its sign.*

What's a cyclic permutation of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$? Write the symbols $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ as in Figure 4 (i.e. arranged clockwise round a circle in that order). The cyclic permutations of $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are

(1) $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$;
(2) $\mathbf{v}$, $\mathbf{w}$, $\mathbf{u}$;
(3) $\mathbf{w}$, $\mathbf{u}$, $\mathbf{v}$.

In other words, these are the permutations where we pick one of them as a starting point and move clockwise round the circle. The non-cyclic permutations of $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are the other three permutations (recall that 3 objects have 6 permutations).

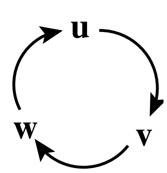

FIGURE 4

All that Theorem 3.2 is saying is

$$\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = \mathbf{v} \cdot \mathbf{w} \times \mathbf{u} = \mathbf{w} \cdot \mathbf{u} \times \mathbf{w},$$

and

$$\mathbf{u} \cdot \mathbf{w} \times \mathbf{v} = -\mathbf{u} \cdot \mathbf{v} \times \mathbf{w};$$

the latter is obvious from the properties of the vector product.

4. 3×3 Determinants

It is time to talk briefly about 3×3 determinants. But before we do this it helps to recall the geometric interpretation of determinant which we covered in Chapter 4. Let $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^2$. Let A be a 2×2 matrix with rows u and v. We know that $|\det(A)|$ is the area of the parallelogram with adjacent sides $\mathbf{u}$ and $\mathbf{v}$ (See Figure 5).

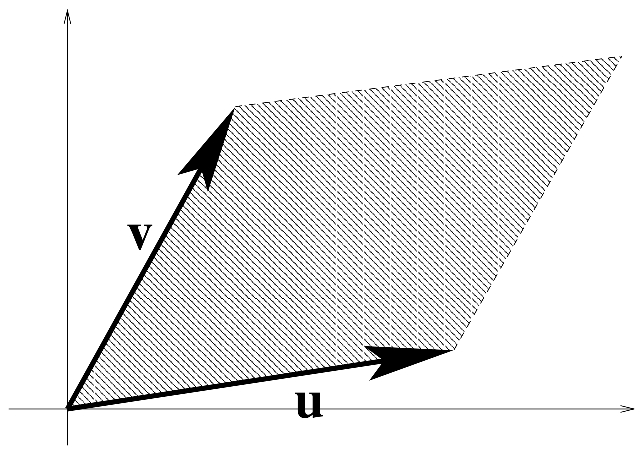

FIGURE 5. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^2$ then the shaded area is $|\det(A)|$ where A is the matrix with rows $\mathbf{u}$ and $\mathbf{v}$.

This suggests to us a sensible generalisation of 2×2 determinants to 3×3 determinants, for surely we suspect that (up to sign) the 3×3 determinant represents the volume of a parallelepiped.

Definition. Let A be a 3×3 matrix. Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be the vectors in $\mathbb{R}^3$ that form the rows of A (in that order). Define

$$\det(A) = \mathbf{u} \cdot \mathbf{v} \times \mathbf{w}.$$

Note the following important property of determinants:

Theorem 4.1. (i) *Let A be a 2×2 matrix with rows $\mathbf{u}$ and $\mathbf{v}$. $\det(A) = 0$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are parallel.*

(ii) *Let A be a 3×3 matrix with rows $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$. $\det(A) = 0$ if and only if $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are coplanar.*

PROOF. For (i), $\det(A) = 0$ if and only if the parallelogram in Figure 5 has area 0. This happens exactly when the parallelogram 'collapses'; that is when $\mathbf{u}$ and $\mathbf{v}$ are parallel.

For (ii), you should ponder over Figure 2 and convince yourself that this is true. The word 'coplanar' means that the three vectors are in the same plane. $\square$

CHAPTER 6

Complements (Secret Knowledge)

This chapter contains secret knowledge about linear algebra and other subjects that you're unlikely to learn about from any other source. Because I am a really sporting kind of person, I am letting you know some of the secrets of the trade. However, secrets are so much fun and exams are meant to be tedious and boring, so the stuff here is not examinable.

The subject of vectors and matrices (and linear algebra) works pretty much the same regardless of whether we are working over the rationals, real numbers or complex numbers. For example, take the system of two linear equations in four variables

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

Once you have learnt echelon reduction in Term 2 it will be very easy to solve this system. You can show that any solution is of the form

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \lambda \begin{pmatrix} 2 \\ 3 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 3 \\ 0 \\ 1 \end{pmatrix},$$

for some 'numbers' λ, μ. This means that the rational solutions are given by vectors of this form with λ, μ rational numbers. The real solutions are given by vectors of this form with λ, μ real numbers. The complex solutions ....

So you see, linear algebra works pretty much the same regardless of the field we are working over. We have to be a little careful if we are working with finite fields (which you will meet eventually), but not much more careful.

Why does linear algebra work pretty much the same regardless of the field? Think again about how you solve a system of linear equations. The operations you use are addition, subtraction, multiplication and division; these are operations that work in any field (remember the field properties you saw in Chapter 1). The situation is very different if, instead of solving a system of linear equations, we wanted to solve a system of polynomials equations. In this situation it matters very much what field we are working over. For example, take the polynomial equation $x^2 - 2 = 0$. This has two solutions in $\mathbb{R}$ but none in $\mathbb{Q}$ (you should know by now that $\sqrt{2}$ is irrational, i.e. it is not in $\mathbb{Q}$). Likewise the equation $x^2 + 1 = 0$ has two solutions in $\mathbb{C}$ but none in $\mathbb{R}$. Notice that to solve these last two equations we need to 'take square-roots', an operation that is different from addition, subtraction, multiplication and division.

Let us take another example—look at the system of two quadratic equations in four variables:

$$x^2 + y^2 = z^2, \qquad x^2 - y^2 = w^2. \tag{41}$$

This has infinitely many complex solutions and infinitely many real solutions; can you prove this? Fibonacci (1170 – 1250) showed that the only rational solutions are proportional to $(x, y, z, w) = (\pm 1, 0, \pm 1, \pm 1)$. But Fibonacci knew much less mathematics than you do, so you should be able to prove this!

This system of equations (41) can be looked at in many ways. The study of its complex solutions is part of algebraic geometry and Riemann surfaces; the study of its rational solutions is part of number theory. These are very different subjects, so for systems of polynomial equations the way we approach them differs very much with the field that we are working over.

Into the clouds—read at your own peril!

Recall the Fundamental Theorem of Algebra. That says that any polynomial (in one variable) of degree n has n roots in $\mathbb{C}$ (counting multiplicities). The field $\mathbb{C}$ is not the only field to have this property. We call a field 'algebraically closed' if it has this property. Thus $\mathbb{C}$ is an example of an algebraically closed field. We know that $\mathbb{Q}$ and $\mathbb{R}$ are not algebraically closed, because the polynomial $x^2 + 1$ doesn't have roots in $\mathbb{Q}$ and $\mathbb{R}$. It turns out that algebraically closed fields are the easiest setting over which one can study systems of polynomial equations in several variables. We call this subject 'algebraic geometry'. Just as linear algebra works pretty much the same regardless of the field, so algebraic geometry works pretty much the same regardless of which *algebraically closed* field we are using; this is the gist of a very deep theorem called *The Lefschetz Principle*.

Those brave souls who have ventured this deep are dying for another example of an algebraically closed field. Let us construct one. Inside $\mathbb{C}$ are two types of numbers: *algebraic* and *transcendental*. An algebraic number is the root of a polynomial with rational coefficients. For example, 5 is an algebraic number since it is the root of $X - 5$ which has rational coefficients. Likewise the 7-th roots of unity are algebraic because they are roots of $X^7 - 1$ which has rational coefficients. A complex number that is not algebraic is called *transcendental*. Now let $\overline{\mathbb{Q}}$ be the set of all algebraic numbers. It turns out that $\overline{\mathbb{Q}}$ is a field and that it is algebraically closed—these facts are not obvious, but they are proved in the third year Galois Theory course.

The shrewd reader is suspicious. I have defined a subset of $\mathbb{C}$ and said that it is an algebraically closed field: "maybe the subset $\overline{\mathbb{Q}}$ is equal to the whole of $\mathbb{C}$?" If that was the case then I haven't given an example of another algebraically closed field. "Can you show that $\overline{\mathbb{Q}}$ and $\mathbb{C}$ are different? Are there really complex numbers that are transcendental?". We can give two answers:

(i) Those who by the end of Term 1 have really, really, digested the Foundations course will be able to see that $\overline{\mathbb{Q}}$ is countable and $\mathbb{C}$ is uncountable. That means that there must be transcendental numbers, and that in fact, most complex numbers are trascendental.

(ii) The way (i) shows the existence of transcendental numbers is a little bit like witchcraft—we get a lot for almost no work. But still, it doesn't give a single example of a transcendental number. The first example was given by Lioville in 1850. He showed that the number $\sum_{n=0}^{\infty} 10^{-n!}$ is transcendental. Hermite (1873) and Lindemann (1882) showed the trancendence of e and π respectively. Whilst the proofs are a little involved, they are not beyond the abilities of an A-level student. If you feel adventurous have a look at the Galois theory books in the library, or google trancendental numbers.

APPENDIX A

MA135 Vectors and Matrices 2005–2006 Test

INSTRUCTIONS

CALCULATORS MAY NOT BE USED

1. Read each question carefully and decide which of the statements (a) – (d) best represents a correct answer. Black out the corresponding box (blue or black pen please) on the Answer Sheet, or alternatively, black out the 'no attempt' box.

2. Black-out **exactly one** *of the five boxes for each question.*

3. The scoring is: 3 for a correct answer, 1 for no attempt, and 0 *for a wrong answer. If T is your total score, your final mark out of 25 is:* $\text{Max}\{0,\, T-8\}$ *[so that people who guess all the answers get zero on average, whilst you still get full marks (25 out of 25) if you answer all the questions correctly.]*

Q1 Suppose X is a complex number satisfying $(2+i)X + (1+i)(1-2i) = -2+4i$. How many of the following statements are true?

- $\text{Re}(X) = -1$
- $X/\overline{X} = -(4+3i)/5$
- $X = \sqrt{10}e^{i\theta}$ where $\tan\theta = -3$

(a) 0 (b) 1 (c) 2 (d) 3

Q2 How many of the following statements are true?

- There are 6 cube-roots of unity.
- The 4-th roots of unity add up to 1.
- The 7-th roots of unity are $\cos(2\pi k/7) + i\sin(2\pi k/7)$ where $k = 0, \ldots, 6$.

(a) 0 (b) 1 (c) 2 (d) 3

Q3 Suppose $B = (b_{ij})_{3\times 2}$ where $b_{ij} = i + 2j$. Which of the following is equal to B?

(a) $\begin{pmatrix} 3 & 5 \\ 4 & 6 \\ 5 & 7 \end{pmatrix}$ (b) $\begin{pmatrix} 3 & 5 & 7 \\ 4 & 6 & 8 \end{pmatrix}$ (c) $\begin{pmatrix} 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix}$ (d) $\begin{pmatrix} 3 & 4 & 5 \\ 5 & 6 & 7 \end{pmatrix}$

Q4 Suppose A, B, C are 2×2 matrices satisfying $AB = C$, where $B = \begin{pmatrix} 4 & 3 \\ 6 & 5 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$. Then $A + A^t$ is

(a) $\begin{pmatrix} 8 & -9/2 \\ -9/2 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 2 & 18 \\ 18 & 54 \end{pmatrix}$ (c) $\begin{pmatrix} -7 & -9 \\ -9 & 15 \end{pmatrix}$ (d) none of the others.

Q5 Suppose A, B are 2×2 matrices. Suppose P is an invertible 2×2 matrix satisfying $B = P^{-1}AP$. How many of the following statements are true?

- $\det(xI_2 - A) = \det(xI_2 - B)$
- A and B have the same eigenvalues
- $A^n = PB^nP^{-1}$ for all positive integers n

(a) 0 (b) 1 (c) 2 (d) 3

Q6 Consider the following system of linear equations in two unknowns x, y

$$\alpha x - \zeta = -\beta y, \qquad \delta y + \gamma x = \xi.$$

Which of the following is a necessary and sufficient condition for this system to have a unique solution?

(a) $\alpha\delta \neq \beta\gamma$ (b) $\zeta \neq 0, \quad \xi \neq 0$ (c) $\alpha\beta \neq \gamma\delta$ (d) $\alpha\delta = \beta\gamma$.

Q7 Suppose $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are non-zero vectors in $\mathbb{R}^3$. How many of the following statements are true?

- $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$
- $|\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}|$ is the volume of the parallelepiped with adjacent sides $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$
- $(-1, 2, 3) \times (2, 1, 6) = (-9, -12, 5)$

(a) 0 (b) 1 (c) 2 (d) 3

Q8 How many of the following facts are used in the proof of the identity $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$?

- The property $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$.
- The property $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
- The cosine rule for triangles $c^2 = a^2 + b^2 - 2ab\cos\theta$.

(a) 0 (b) 1 (c) 2 (d) 3.

Q9 Suppose the vector $\mathbf{u} = (x, y)$ has length 5 units and satisfies $\mathbf{u} \cdot \mathbf{i} = 4$ and $\mathbf{u} \cdot \mathbf{j} < 0$. Then $\mathbf{u} \cdot (-4, 3)$ is equal to

(a) 0 (b) -25 (c) 25 (d) -7

Q10 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ y \end{pmatrix}$. Which of the following is the matrix associated with T?

(a) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

Q11 Which of the following matrices will rotate a column vector in $\mathbb{R}^2$ anticlockwise through an angle of $90°$?

(a) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

APPENDIX B

MA135 Vectors and Matrices 2006–2007 Test 1

INSTRUCTIONS

CALCULATORS MAY NOT BE USED

1. Read each question carefully and decide which of the statements (a) – (d) best represents a correct answer. Black out the corresponding box (blue or black pen please) on the Answer Sheet, or alternatively, black out the 'no attempt' box.

2. Black-out **exactly one** *of the five boxes for each question.*

3. The scoring is: 3 for a correct answer, 1 for no attempt, and 0 *for a wrong answer. If T is your total score, your final mark out of 25 is:* $\text{Max}\{0, T-8\}$ *[so that people who guess all the answers get zero on average, whilst you still get full marks (25 out of 25) if you answer all the questions correctly.]*

Q1 How many of the following statements are true?

- There are eight 4-th roots of unity.
- The cube roots of unity add up to -1.
- The 11-th roots of unity are $\exp(2\pi ik/11)$ where $k = 0, \dots, 10$.
- The 9-th roots of unity are $1, \zeta, \dots, \zeta^8$ where $\zeta = \cos(2\pi/9) + i\sin(2\pi/9)$.

(a) 1 (b) 2 (c) 3 (d) 4

Q2 Suppose X is a complex number satisfying $(1+i)X = (1-2i)(\overline{2-i})$, and let θ be its argument. Then $\tan\theta$ is

(a) $\sqrt{3}$ (b) 1 (c) -7 (d) $13/2$

Q3 Let α, β and γ be the roots of $x^3 + x^2 - 2$. Then $\alpha^2 + \beta^2 + \gamma^2$ is equal to

(a) $3+i$ (b) $11-2i$ (c) $\cos(\pi/7) + i\sin(\pi/7)$ (d) 1

Q4 The expression $\exp(2i\theta)\dfrac{(\cos\theta + i\sin\theta)^5}{\cos(4\theta) - i\sin(4\theta)}$ simplifies to

(a) $\exp(3i\theta)$ (b) $\exp(-14\theta)$ (c) $\exp(-3i\theta)$ (d) $\exp(11i\theta)$.

Q5 Let L be the line with vector equation $\mathbf{x} = (2,1,0) + t(1,3,1)$. Which of the following lines is parallel to L but passess through the point $P(1,1,-1)$:

(a) $\mathbf{x} = (2,1,0) + t(1,1,-1)$ (b) $\mathbf{x} = (4,2,0) + t(1,3,1)$
(c) $\mathbf{x} = (-1,-5,-3) + t(2,6,2)$ (d) $\mathbf{x} = (1,1,-1) + t(3,4,1)$

Q6 Let $\mathbf{r}_0 = (1,1)$. How many vectors $\mathbf{r} = (x,y)$ in $\mathbb{R}^2$ satisfy $\|\mathbf{r}\| = \|\mathbf{r} - \mathbf{r}_0\| = 1$?

(a) 0 (b) 1 (c) 2 (d) ∞

Q7 Suppose $\mathbf{u}$ is a non-zero vector in $\mathbb{R}^2$. Which of the following has length 3 but opposite direction to $\mathbf{u}$?

(a) $\dfrac{-3}{\|\mathbf{u}\|}\mathbf{u}$ (b) $\dfrac{-9}{\|\mathbf{u}\|^2}\mathbf{u}$ (c) $\dfrac{-\|\mathbf{u}\|}{3}\mathbf{u}$ (d) none of the others

Q8 How many of the following facts are used in the proof of the identity $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$?

- The property $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$.
- Commutativity of the dot product.
- The cosine rule for triangles.
- The formula for the area of the triangle $\frac{1}{2}ab\sin\theta$.

(a) 1 (b) 2 (c) 3 (d) 4.

Q9 Let $\mathbf{u}$ and $\mathbf{v}$ be (non-zero) orthogonal vectors in $\mathbb{R}^n$. What is the angle that the vector $\dfrac{1}{\|\mathbf{u}\|}\mathbf{u} + \dfrac{1}{\|\mathbf{v}\|}\mathbf{v}$ makes with $\mathbf{u}$?

(a) $\pi/2$ (b) 0 (c) $\pi/4$ (d) depends on $\mathbf{u}$, $\mathbf{v}$

Q10 Suppose that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are orthonormal in $\mathbb{R}^5$. Then $(5\mathbf{u}+\mathbf{v}+\mathbf{w}) \cdot (3\mathbf{u}-\mathbf{v}+2\mathbf{w})$ is

(a) 16 (b) 0 (c) 12 (d) undefined

Q11 How many complex roots (counting multiplicities) does the polynomial $f(x) = (x^4 - x - 1)(x^5 - x - 1)$ have?

(a) 4 (b) 5 (c) 9 (d) depends on x

MA135 Vectors and Matrices 2006–2007 Test 2

INSTRUCTIONS

CALCULATORS MAY NOT BE USED

1. Read each question carefully and decide which of the statements (a) – (d) best represents a correct answer. Black out the corresponding box (blue or black pen please) on the Answer Sheet, or alternatively, black out the 'no attempt' box.

2. Black-out **exactly one** *of the five boxes for each question.*

3. The scoring is: 3 for a correct answer, 1 for no attempt, and 0 *for a wrong answer. If T is your total score, your final mark out of 25 is:* Max$\{0, T-8\}$ *[so that people who guess all the answers get zero on average, whilst you still get full marks (25 out of 25) if you answer all the questions correctly.]*

Q1 Let $A = \begin{pmatrix} 4 & 2 \\ 1 & 1 \end{pmatrix}$. Then $\det(A^{11} - A^{10})$ is

(a) -17 (b) $-2^9 \times 17$ (c) -2^{11} (d) -24

Q2 Let

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad S = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Let $\mathbf{x}$ be a column vector in $\mathbb{R}^2$. To obtain $RS\mathbf{x}$ from $\mathbf{x}$ we

(a) reflect in the y-axis then rotate anti-clockwise about the origin through angle θ;
(b) rotate anti-clockwise about the origin through angle θ then reflect in the y-axis;
(c) reflect in the origin then rotate clockwise about the origin through angle θ;
(d) rotate clockwise about the origin through angle θ then reflect in the origin.

Q3 Let

$$\Pi_1 : \ \mathbf{x} \cdot (1,1,1) = \gamma, \qquad \Pi_2 : \ \mathbf{x} = s(-1,\alpha,0) + t(1,1,\beta) + (-1,-1,-1).$$

For which triple of values α, β, γ will Π_1, Π_2 be the same plane?

(a) $\alpha = 2$, $\beta = 1$, $\gamma = -3$;
(b) $\alpha = 1$, $\beta = -2$, $\gamma = -3$;
(c) $\alpha = 2$, $\beta = 1$, $\gamma = 5$;
(d) $\alpha = 1$, $\beta = -2$, $\gamma = 5$.

Q4 The eigenvalues of $A = \begin{pmatrix} 0 & -5 \\ -1 & 4 \end{pmatrix}$ are

(a) $3, -2$ (b) $-4, 3$ (c) $1+i, 1-i$ (d) $5, -1$.

Q5 Let $\mathbf{u} = (3,5,1)$, $\mathbf{v} = (1,2,-1)$, $\mathbf{w} = (1,1,1)$. Then the volume of the parallelpiped with adjacent sides $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is

(a) 10 (b) 2 (c) 5 (d) $\sqrt{7}$.

Q6 Let T_1, T_2, $T_3 : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$T_1(x,y,z) = (x+1, y+z), \qquad T_2(x,y,z) = (2x, y), \qquad T_3(x,y,z) = (x^2, y+z).$$

Which of T_1, T_2, T_3 is a linear transformation?

(a) T_1 (b) T_2 (c) T_3 (d) none of them.

Q7 The linear system in x, y

$$(1+\lambda)x - \mu y = \delta, \qquad (1-\lambda)x + \mu y = 2,$$

has a unique solution precisely when

(a) $\lambda = \pm\mu\delta$ (b) $\mu \neq \delta$ (c) $\lambda\mu = 2\delta$ (d) $\mu \neq 0$.

Q8 Let A, B be 2×2 matrices satisfying

$$A^{-1} + B = \begin{pmatrix} 4 & 2 \\ 7 & 2 \end{pmatrix}, \qquad A^{-1} - B = \begin{pmatrix} 4 & 0 \\ 5 & 2 \end{pmatrix}.$$

Then $2BA$ is

(a) undefined (b) $\begin{pmatrix} -6 & 4 \\ 2 & -1 \end{pmatrix}$ (c) $\begin{pmatrix} 0 & 3 \\ -7 & 5 \end{pmatrix}$ (d) $\begin{pmatrix} 2 & 1 \\ 3 & -5 \end{pmatrix}$.

Q9 Let $\omega = \exp(2\pi i/3)$. Then $i\omega$ is an n-th root of unity where n is

(a) 12 (b) 7 (c) 4 (d) 3.

Q10 Suppose that $\mathbf{u}$, $\mathbf{v}$ are unit vectors in $\mathbb{R}^4$. Then $\|\mathbf{u}+3\mathbf{v}\|^2 + \|3\mathbf{u}-\mathbf{v}\|^2$ is equal to

(a) $2\sqrt{7}$ (b) 20 (c) 12 (d) 16

Q11 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(x,y) = (x+y, x+y, x+y)$. The image of T is

(a) a point (b) a line (c) a plane (d) depends on x and y.

APPENDIX D

MA135 Vectors and Matrices 2007–2008 Test 1

INSTRUCTIONS

CALCULATORS MAY NOT BE USED

1. Read each question carefully and decide which of the statements (a) – (d) best represents a correct answer. Black out the corresponding box (blue or black pen please) on the Answer Sheet, or alternatively, black out the 'no attempt' box.

2. Black-out **exactly one** *of the five boxes for each question.*

3. The scoring is: 3 for a correct answer, 1 for no attempt, and 0 *for a wrong answer. If T is your total score, your final mark out of 25 is:* Max{*0*, $T-8$} *[so that people who guess all the answers get zero on average, whilst you still get full marks (25 out of 25) if you answer all the questions correctly.]*

Q1 How many of the following statements are true?

- There are five 5-th roots of unity.
- The 11-th roots of unity add up to 0.
- The 4-th roots of unity are $1, i, -1, -i$.
- The 7-th roots of unity are $1, \zeta, \dots, \zeta^6$ where $\zeta = \cos(2\pi/7) + i\sin(2\pi/7)$.

(a) 1 (b) 2 (c) 3 (d) 4

Q2 Suppose X is the complex number satisfying $(1+i)X = (2i-1)(\overline{2+i})$. The argument of X is

(a) $\pi/6$ (b) $\pi/4$ (c) $\pi/3$ (d) $\pi/2$

Q3 Let α, β and γ be the roots of x^3+x+2. Then $\alpha^{-1}+\beta^{-1}+\gamma^{-1}$ is equal to

(a) $-1/2$ (b) $-i/2$ (c) $\cos(\pi/3)+i\sin(\pi/3)$ (d) $1+i\sqrt{3}$

Q4 The expression $\dfrac{\cos\theta + i\sin\theta}{1+i\sqrt{3}}$ can be rewritten as

(a) $2^{-1/2}\exp(i(\theta-\pi/3))$ (b) $2^{-1/2}\exp(i(\theta-2\pi/3))$

(c) $2^{-1}\exp(i(\theta-2\pi/3))$ (d) $2^{-1}\exp(i(\theta-\pi/3))$.

Q5 Let L be the line with vector equation $\mathbf{x} = (2,1)+t(1,3)$. Which of the following lines is perpendicular to L and passess through the origin:

(a) $\mathbf{x} = t(-1,2)$ (b) $\mathbf{x} = (6,-2)+t(-3,1)$

(c) $\mathbf{x} = (0,0)+t(1,1)$ (d) none of the others.

Q6 Write $\mathbf{x} = (x, y)$ in $\mathbb{R}^2$. The two equations

$$\|\mathbf{x}\| = 1, \qquad \|\mathbf{x} - (0,2)\| = 1,$$

geometrically represent

(a) two overlaping circles; (b) two touching circles;

(c) two completely disjoint circles; (d) none of the others.

Q7 Let $\mathbf{v}$ is a non-zero vector in $\mathbb{R}^5$. Which of the following has length π but opposite direction to $\mathbf{v}$?

(a) $\dfrac{-\pi}{\|\mathbf{v}\|}\mathbf{v}$ (b) $\dfrac{-\|\mathbf{v}\|}{\pi}\mathbf{v}$ (c) $\dfrac{\pi^2}{\|\mathbf{v}\|}\mathbf{v}$ (d) none of the others

Q8 How many of the following facts are used in the proof of the identity $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$?

- The property $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$.
- Commutativity of the dot product.
- The cosine rule for triangles.
- De Moivre's Theorem.

(a) 1 (b) 2 (c) 3 (d) 4.

Q9 Let $\mathbf{x}$ and $\mathbf{y}$ be non-parallel non-zero vectors in $\mathbb{R}^n$ of **equal length**. What is the angle that the vector $\mathbf{x} + \mathbf{y}$ makes with the vector $\mathbf{x} - \mathbf{y}$?

(a) $\pi/2$ (b) 0 (c) $2\pi/3$ (d) depends on $\mathbf{x}$, $\mathbf{y}$

Q10 Let $\mathbf{u} = (1,0,0)$, $\mathbf{v} = (0,1/2,1/2)$, $\mathbf{w} = (0,1/2,-1/2)$, $\mathbf{x} = (1/2,0,1/2)$. Which of the following triples in $\mathbb{R}^3$ are orthonormal?

(a) $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ (b) $\mathbf{u}$, $\mathbf{v}$, $\mathbf{x}$ (c) $\mathbf{u}$, $\sqrt{2}\mathbf{v}$, $\sqrt{2}\mathbf{w}$ (d) $\mathbf{u}$, $\sqrt{2}\mathbf{v}$, $\sqrt{2}\mathbf{x}$

Q11 Which of the following planes is perpendicular to the line $\mathbf{x} = (0,1,1) + t(1,2,-2)$ and passes through the point $(1,1,1)$?

(a) $2x + 4y - 4z = 2$ (b) $-2x + 3y + 2z = 1$

(c) $2x + 4y - 4z = 0$ (d) $-2x + 3y + 2z = 3$.

APPENDIX E

MA135 Vectors and Matrices 2007–2008 Test 2

INSTRUCTIONS

CALCULATORS MAY NOT BE USED

1. Read each question carefully and decide which of the statements (a) – (d) best represents a correct answer. Black out the corresponding box (blue or black pen please) on the Answer Sheet, or alternatively, black out the 'no attempt' box.

2. Black-out **exactly one** *of the five boxes for each question.*

3. The scoring is: 3 for a correct answer, 1 for no attempt, and 0 *for a wrong answer. If T is your total score, your final mark out of 25 is:* $\mathrm{Max}\{0,\, T-8\}$ *[so that people who guess all the answers get zero on average, whilst you still get full marks (25 out of 25) if you answer all the questions correctly.]*

Q1 Let $A = \begin{pmatrix} 4 & \alpha \\ 1 & 1 \end{pmatrix}$. For which real numbers α is $\det(A^5) = -32$?

(a) -2 (b) 2 (c) 6 (d) -6

Q2 Let

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad S = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let $\mathbf{u}$ be a column vector in $\mathbb{R}^2$. To obtain $SR_\theta\mathbf{u}$ from $\mathbf{u}$ we

(a) rotate anti-clockwise about the origin through angle θ then reflect in the origin;

(b) reflect in the y-axis then rotate anti-clockwise about the origin through angle θ;

(c) reflect in the origin then rotate clockwise about the origin through angle θ;

(d) rotate clockwise about the origin through angle θ then reflect in the x-axis.

Q3 Let

$$\Pi_1 :\ \mathbf{x}\cdot(1,0,1) = 1, \qquad \Pi_2 :\ \mathbf{x} = s(1,0,a) + t(b,1,-1) + (2,1,c).$$

For which triple of values a, b, c will Π_1, Π_2 be the same plane?

(a) $a=-1,\ b=1,\ c=-3$ (b) $a=1,\ b=-1,\ c=-3$

(c) $a=1,\ b=-1,\ c=-1$ (d) $a=-1,\ b=1,\ c=-1.$

Q4 Let A be a 2×2 matrix satisfying $A\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}2\\2\end{pmatrix}$, $A\begin{pmatrix}2\\-1\end{pmatrix}=\begin{pmatrix}-6\\3\end{pmatrix}$. The characteristic polynomial of A is

(a) x^2+x-6 (b) x^2-5x+6 (c) x^2+5x+6 (d) x^2-x-6.

Q5 In $\mathbb{R}^3$ the volume of the parallelopiped with adjacent sides $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ is $|\mathbf{u}\cdot\mathbf{v}\times\mathbf{w}|$. How many of the following facts are used in the proof of this formula:

- volume of the parallelopiped = (area of base) $\times$ height;
- area of a parallelogram with adjacent sides $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{x}\times\mathbf{y}\|$;
- $\mathbf{x}\times\mathbf{y}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$;
- if θ is the angle between vectors $\mathbf{x}$ and $\mathbf{y}$ then $\mathbf{x}\cdot\mathbf{y}=\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$.

(a) 1 (b) 2 (c) 3 (d) 4.

Q6 Let T_1, T_2, $T_3 : \mathbb{R}^3\to\mathbb{R}^2$ be given by

$$T_1(x,y,z)=(0,y-z), \qquad T_2(x,y,z)=(2x+1,y), \qquad T_3(x,y,z)=(xy,0).$$

Which of T_1, T_2, T_3 is a linear transformation?

(a) T_1 (b) T_2 (c) T_3 (d) none of them.

Q7 The linear system in x, y

$$(1+\lambda)x+\mu y=\delta, \qquad (1-\lambda)x+\mu y=0,$$

is a homogeneous system with non-trivial solutions precisely when

(a) $\lambda=\mu\delta$ (b) $\lambda\mu=\delta=0$ (c) $\lambda\mu=0$, $\delta\neq 0$ (d) $\mu\neq 0$.

Q8 Let A, B be 2×2 matrices satisfying $AB=\begin{pmatrix}-3&7\\-4&1\end{pmatrix}$, $B^{-1}=\begin{pmatrix}1&-1\\0&1\end{pmatrix}$. Then $25A^{-1}$ is

(a) $\begin{pmatrix}-6&4\\2&-1\end{pmatrix}$ (b) $\begin{pmatrix}0&3\\-2&5\end{pmatrix}$ (c) $\begin{pmatrix}5&-10\\4&-3\end{pmatrix}$ (c) undefined.

Q9 Let $\zeta=\exp(2\pi i/5)$. Then $-\zeta$ is an n-th root of unity where n is

(a) 5 (b) 6 (c) 10 (d) 4.

Q10 Which of the following are necessary and sufficient conditions for the matrices $A=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)$ and $B=\left(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}\right)$ to commute?

(a) $A=\left(\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right)$ (b) $A=B$ (c) $b=c+d-a=0$ (d) $a+b=c+d$

Q11 Let $T : \mathbb{R}^3\to\mathbb{R}^3$ be given by $T(x,y,z)=(0,x+z,y)$. The image of T is the

(a) xy-plane (b) yz-plane (c) xz-plane (d) plane $x+z=y$.