Algebra
(Autumn Term)

Christopher Hughes (with some updates by Emilie Dufresne)

Autumn Term 2022/23

Contents

DISCLAIMER

These notes cover the first part of the first year module *Algebra*, as taught in the Autumn Term of 2017/18. They mostly contain what I intend to say during the lectures, though they may offer some extra examples or further details, and the lectures might include more details and a few extra anecdotes. Only that which was given in lectures will be examinable; any extra material found in these notes is for your entertainment and edification only.

Emilie Dufresne, September 2022

Chapter 1

Complex Numbers

Before we define complex numbers, let us look at other, more familiar, types of number and how they relate to each other.

- Natural number, $\mathbb{N}$. These are the "counting numbers" (positive whole numbers).
- Integer, $\mathbb{Z}$. These are all whole numbers: positive, negative and the number zero.

  Remark. The notation $\mathbb{Z}$ comes from the German for number, "Zahlen".

  Remark. The number zero was historically not acknowledged as true number by humankind until very late in the development of mathematics. People struggled with the notation that "nothing" could be something.

- Rational number, $\mathbb{Q}$. These are ratios of an integer divided by a natural number.

  Remark. The notation $\mathbb{Q}$ comes from the word "quotient"

  Remark. As defined, there are many ways of expressing the same rational number. For example $-1/2 = -2/4$.

- Real number, $\mathbb{R}$. These can be quite hard to define rigorously (and that's not the subject of this module!). For our purposes we'll simply say they fill in all the gaps in the number line between the rational numbers. By this we mean if we have bounded increasing real sequence, then it will have a limit and that limit will be a real number. (More details in next term's *Real Analysis* module).
- Complex numbers, $\mathbb{C}$. These will be the subject of this chapter. They take the form $z = x + \mathrm{i}y$ where $x, y \in \mathbb{R}$ and i satisfies $\mathrm{i}^2 = -1$.

1.1 The algebra of complex numbers

Given any real number, x, we know that x^2 is positive. Therefore there are no real solutions to $x^2 = -1$. If we wanted to solve that equation, we would need to create a new number.

Definition. The number i has the property $\mathrm{i}^2 = -1$.

Remark. Appreciate the subtlety here: We don't say "i is the square root of -1" because we don't know what "square root of a negative number" means, and we want to avoid the distraction of the question "which square root?". Rather we have defined the number i in terms of its behaviour when multiplied by itself.

Definition. We say $z = x + \mathrm{i}y$ with $x, y \in \mathbb{R}$ is a complex number. The real part of z is x and is denoted $\mathrm{Re}(z)$, and the imaginary part is y and is denoted $\mathrm{Im}(z)$.

Definition. The set of all complex numbers is denoted $\mathbb{C}$.

We define the addition of two complex numbers by simply adding their real and imaginary parts.

Example. Let $z_1 = 1 + 2\mathrm{i}$ and let $z_2 = 3 + 4\mathrm{i}$. Then $z_1 + z_2 = (1 + 3) + (2 + 4)\mathrm{i} = 4 + 6\mathrm{i}$.

We define subtraction of two complex numbers similarly.

Example. Let $z_1 = x_1 + \mathrm{i}y_1$ and let $z_2 = x_2 + \mathrm{i}y_2$. Then $z_1 - z_2 = (x_1 - x_2) + (y_1 - y_2)\mathrm{i}$.

To multiply two complex numbers, use the rules for multiplying out brackets and remember that $\mathrm{i}^2 = -1$. For $z_1 = x_1 + \mathrm{i}y_1$ and $z_2 = x_2 + \mathrm{i}y_2$ we have

$$\begin{aligned} z_1 z_2 &= (x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) \\ &= x_1 x_2 + \mathrm{i}y_1 x_2 + \mathrm{i}x_1 y_2 + \mathrm{i}^2 y_1 y_2 \\ &= (x_1 x_2 - y_1 y_2) + \mathrm{i}(x_2 y_1 + x_1 y_2) \end{aligned}$$

Example. Let $z_1 = 1 + 2\mathrm{i}$ and let $z_2 = 3 + 4\mathrm{i}$. Then $z_1 z_2 = (1 \times 3 - 2 \times 4) + (1 \times 4 + 2 \times 3)\mathrm{i} = -5 + 10\mathrm{i}$.

Example. Let $z_1 = 3 + 4\mathrm{i}$ and let $z_2 = 3 - 4\mathrm{i}$. Then $z_1 z_2 = (3 \times 3 - 4 \times (-4)) + (4 \times 3 - 3 \times 4)\mathrm{i} = 25$.

The fact that the last answer is a positive real number is no fluke.

Definition. The complex conjugate of $z = x + \mathrm{i}y$ is $\bar{z} = x - \mathrm{i}y$.

Definition. The modulus (or absolute value) of z is $|z| = \sqrt{x^2 + y^2}$, where the positive square root is taken.

Lemma. *The absolute value of a complex number z is $|z| = \sqrt{z\bar{z}}$.*

Proof. Let $z = x + \mathrm{i}y$. We have

$$z\bar{z} = (x + \mathrm{i}y)(x - \mathrm{i}y) = x^2 - y(-y) + (yx - xy)\mathrm{i} = x^2 + y^2 = |z|^2$$

Taking the positive square root of both sides completes the proof. □

Remark. Notice that in the second example above, $z_2 = \bar{z}_1$ and so we showed $|3 + 4\mathrm{i}|^2 = 25$.

Finally, using this we can work out to divide two complex numbers. Since $w/z = w \times 1/z$ we actually only need to work out how to calculate $1/z$ given a complex number z.

Lemma. *Given a non-zero complex number $z = x + \mathrm{i}y$, then*

$$\frac{\bar{z}}{|z|^2} = \frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}\mathrm{i} \in \mathbb{C}$$

is the multiplicative inverse of z.

Proof. Indeed

$$z \times \frac{\bar{z}}{|z|^2} = \frac{z \times \bar{z}}{|z|^2} = 1,$$

and

$$\frac{\bar{z}}{|z|^2} \times z = \frac{\bar{z} \times z}{|z|^2} = \frac{z \times \bar{z}}{|z|^2} = 1.$$

□

Example. Let $z_1 = 1 + 2\mathrm{i}$ and let $z_2 = 3 - 4\mathrm{i}$. Then

$$\begin{aligned}
\frac{z_1}{z_2} &= \frac{1 + 2\mathrm{i}}{3 - 4\mathrm{i}} \\
&= (1 + 2\mathrm{i}) \times \frac{3 + 4\mathrm{i}}{|3 + 4\mathrm{i}|^2} \\
&= \frac{(1 + 2\mathrm{i})(3 + 4\mathrm{i})}{3^2 + 4^2} \\
&= \frac{-5 + 10\mathrm{i}}{25} && \text{(making use of a previous example to multiply the numerator)} \\
&= -\frac{1}{5} + \frac{2}{5}\mathrm{i}
\end{aligned}$$

where in the last line we divided a real number by a real number, which we know we can do.

Lemma (Results about the complex conjugate)**.** *Let $z, w \in \mathbb{C}$. We have*

- $z + \bar{z} = 2\,\mathrm{Re}(z)$ *and* $z - \bar{z} = 2\mathrm{i}\,\mathrm{Im}(z)$
- $\overline{z + w} = \bar{z} + \bar{w}$
- $\overline{zw} = \bar{z}\bar{w}$*, and*
- $\bar{\bar{z}} = z$*.*

Proof. All these results can be proved by writing the complex numbers in component form and applying the definitions given above:

- If $z = x + \mathrm{i}y$ then $\bar{z} = x - \mathrm{i}y$ so $z + \bar{z} = (x + \mathrm{i}y) + (x - \mathrm{i}y) = 2x = 2\,\mathrm{Re}(z)$.
  Similarly $z - \bar{z} = (x + \mathrm{i}y) - (x - \mathrm{i}y) = 2\mathrm{i}y = 2\mathrm{i}\,\mathrm{Im}(z)$.
- Let $w = a + \mathrm{i}b$, so we have

$$\begin{aligned}
\overline{z + w} &= \overline{(x + \mathrm{i}y) + (a + \mathrm{i}b)} && \text{writing } z \text{ and } w \text{ in their component form} \\
&= \overline{(x + a) + \mathrm{i}(y + b)} && \text{by definition of adding two complex numbers} \\
&= (x + a) - \mathrm{i}(y + b) && \text{by definition of complex conjugation} \\
&= (x - \mathrm{i}y) + (a - \mathrm{i}b) && \text{reordering the summands} \\
&= \bar{z} + \bar{w}
\end{aligned}$$

- Similarly we have

$$\overline{zw} = \overline{(x + \mathrm{i}y)(a + \mathrm{i}b)} = \overline{xa - yb + \mathrm{i}(xb + ya)} = xa - yb - \mathrm{i}(xb + ya) = (x - \mathrm{i}y)(a - \mathrm{i}b) = \bar{z}\bar{w}$$

- Finally, $\bar{\bar{z}} = \overline{x - \mathrm{i}y} = x + \mathrm{i}y = z$.

□

Lemma (Results about the modulus)**.** *For $z, w \in \mathbb{C}$ we have*

- $|\bar{z}| = |z|$,
- $|zw| = |z||w|$,
- $|z/w| = |z|/|w|$ *so long as* $w \neq 0$.

Proof. For $z = x + \mathrm{i}y$, we have

- $|\bar{z}|^2 = |x - \mathrm{i}y|^2 = x^2 + (-y)^2 = x^2 + y^2 = |x + \mathrm{i}y|^2 = |z|^2$ and taking the positive square root completes for proof for the first part.
- This can be shown as $|zw|^2 = zw\overline{zw} = z\bar{z}w\bar{w} = |z|^2|w|^2$.
- Finally, for non-zero w by definition we have $1/w = \bar{w}/|w|^2$ so by the first part,

$$\left|\frac{1}{w}\right| = \left|\frac{\bar{w}}{|w|^2}\right| = \frac{|\bar{w}|}{|w|^2} = \frac{1}{|w|}.$$

Therefore the result follows from the second part. □

Where next?

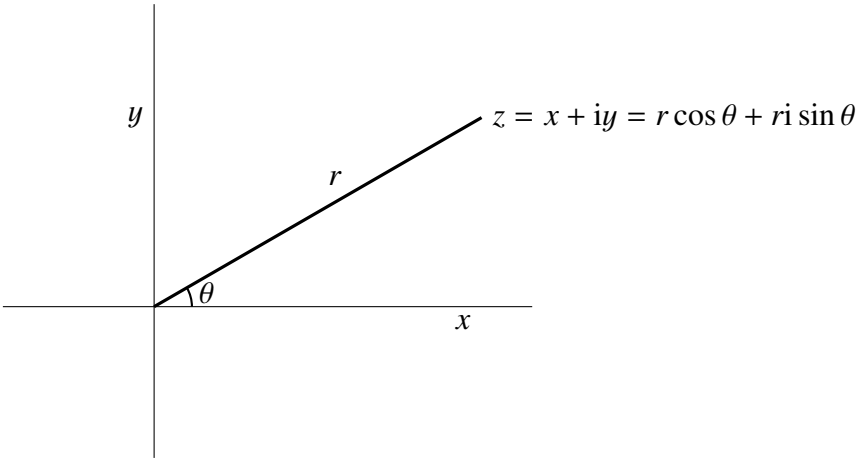
Complex numbers, as we have just defined them, have the following (very reasonable) properties:

- You can add two complex numbers, and the result is another complex number.
- You can multiply two complex numbers, and the result is another complex number.
- For every complex number z, there is another complex number $-z$ such that $z + (-z) = 0$. (Subtraction).
- Implicit in this rule is the fact there exists a complex number 0, and it has the property $z + 0 = z$ for all $z \in \mathbb{C}$.
- For every non-zero $z \in \mathbb{C}$, there is another complex number $1/z$ such that $z \times (1/z) = 1$. (Division).
- Implicit in this rule is there exists a complex number 1 such that $1 \times z = z$ for all $z \in \mathbb{C}$.

In the second-year *Pure Mathematics* stream, you will see that these rules define a *field*. (That is, real numbers and complex numbers are examples of fields). This will be discussed and developed in more detail in that course. In the third year, you can see such results be used powerfully in *Algebraic Number Theory*.

1.2 Geometric representation: Argand plane and polar form

Since a complex number has two real components (the real and imaginary parts) it cannot be drawn on a "number line", but rather it can be represented on a "number plane" known as the *Argand plane*, named after Jean-Robert Argand (1768–1822), a French bookstore owner and amateur mathematician, who first suggested a geometric interpretation of complex numbers.



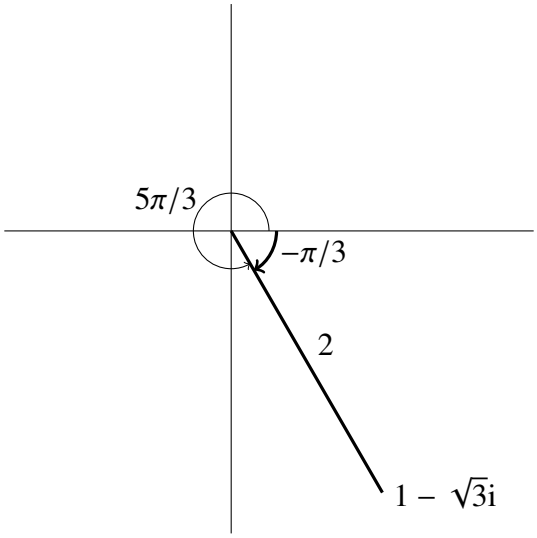
The Argand plane

Simple trigonometry gives $x = r\cos\theta$ and $y = r\sin\theta$, and $z = r(\cos\theta + \mathrm{i}\sin\theta)$. This is known as the polar representation of the complex number.

Note that $r = \sqrt{x^2 + y^2} = |z|$, the modulus.

The argument of z, θ, is the angle between the positive real axis and the line connecting 0 to z, measured in radians in the counterclockwise direction. Notice that you can add any multiple of 2π to the argument and still have a valid argument. By convention the principal value of the argument, $\arg(z)$, is defined to be the unique value of the argument such that $-\pi < \arg(z) \leq \pi$.

Example. The number $1 - \sqrt{3}\mathrm{i}$ has radius 2 and argument $-\pi/3$.



The principal value of the argument lies between $-\pi < \arg(z) \leq \pi$, so $\arg(1 - \sqrt{3}\mathrm{i}) = -\pi/3$.

It will get annoying to have to write $\cos\theta + \mathrm{i}\sin\theta$ so we will introduce a convenient notation:

Definition. We write

$$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

Remark. We will see in the second year (in *Functions of a Complex Variable* that $e^{i\theta}$ really does behave like the exponential function, but for the moment simply take is as a shorthand notation.

Example. In the previous example we showed that $1 - \sqrt{3}i = 2e^{5i\pi/3} = 2e^{-i\pi/3}$.

Example. The number i has modulus 1 and argument $\pi/2$. Therefore $i = e^{i\pi/2}$.

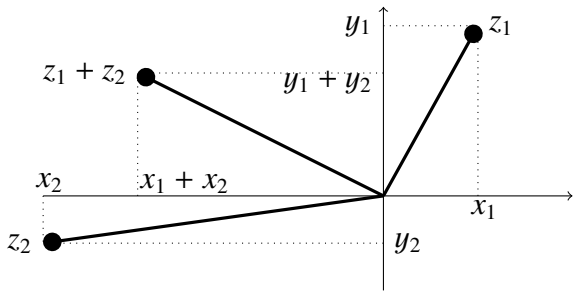
Example. The so-called "most important equation in mathematics" is $e^{i\pi} = -1$ as it contains five of the most significant symbols in mathematics.

Lemma. *If a complex number has argument θ, then its complex conjugate has argument $-\theta$. That is, if $z = re^{i\theta}$ then $\overline{z} = re^{-i\theta}$.*

Proof. Write $z = x + iy$. Recall that $z = re^{i\theta}$ means that $x = r\cos\theta$ and $y = r\sin\theta$. Also recall that $\overline{z} = x - iy$. Replacing θ with $-\theta$ achieves $x \mapsto x$ and $y \mapsto -y$ as required. □

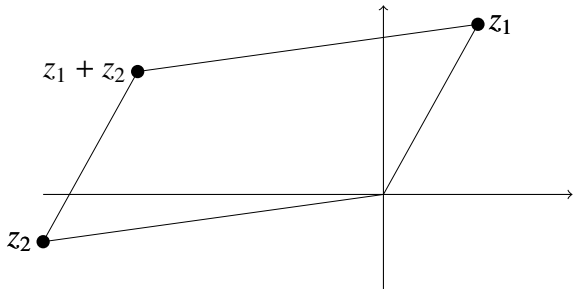
The algebra of complex numbers has a geometric interpretation in the complex plane:

Algebraically, addition of two complex numbers simply means adding the real and imaginary components separately.



Algebraically, adding two complex numbers simply means adding their real and imaginary components.

Geometrically, in the Argand plane the sum of two complex numbers can be found by starting at the origin, going along to z_1, and from there going along z_2. (Or alternatively, this turns out to be the same point you get from first going along z_2 and then along z_1 — see the two dotted paths).



This is the same diagram as above, but showing how joining a line representing the length and direction of z_1 with a line representing the length and direction of z_2 ends at the position of $z_1 + z_2$.

Remark. This process is identical to the addition (or subtraction) operations for two-dimensional vectors. This implies, for example, that $|z_1 - z_2|$ is the distance between z_1 and z_2 on the Argand diagram.

To obtain a geometric understand of multiplication and division, we need de Moivre's Theorem.

1.3 De Moivre's theorem

Theorem (De Moivre)**.** *Let w and z be complex numbers with polar form $w = \rho\,\mathrm{e}^{\mathrm{i}\varphi}$ and $z = r\,\mathrm{e}^{\mathrm{i}\theta}$. Then the product has polar form*

$$wz = \rho r e^{\mathrm{i}(\varphi+\theta)}$$

Remark. This tells us the modulus of zw is ρr. However, if we wanted to talk about the principal value of the argument, $\arg(zw)$, then it might not be $\arg(z) + \arg(w)$; We would need to check it lay between $-\pi$ and π, adding or subtracting multiples of 2π if necessary.

Proof. We write $w = \rho\,\mathrm{e}^{\mathrm{i}\varphi} = \rho(\cos(\varphi) + \mathrm{i}\sin(\varphi))$ and $z = r\,\mathrm{e}^{\mathrm{i}\theta} = r(\cos\theta + \mathrm{i}\sin\theta)$. Then

$$\begin{aligned} wz &= \rho e^{\mathrm{i}\varphi} r e^{\mathrm{i}\theta} \\ &= \rho r\left((\cos(\varphi)\cos(\theta) - \sin(\varphi)\sin(\theta)) + \mathrm{i}\,(\sin(\varphi)\cos(\theta) + \cos(\varphi)\sin(\theta))\right) \\ &= \rho r\,(\cos(\varphi+\theta) + \mathrm{i}\sin(\varphi+\theta)) \\ &= \rho r e^{\mathrm{i}(\varphi+\theta)} \end{aligned}$$

where we used the double-angle formulae in the penultimate line. □

This geometric interpretation of multiplication is illustrated in the following diagram:

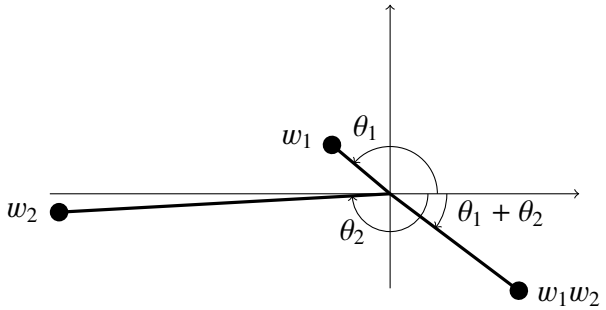


Diagram showing the product of two complex numbers in polar form. Note that θ_1 is positive, θ_2 and $\theta_1 + \theta_2$ are negative.

Remark. For division, note that if $z = re^{\mathrm{i}\theta}$ then

$$\frac{1}{z} = \frac{1}{re^{\mathrm{i}\theta}} = \frac{1}{r}e^{-\mathrm{i}\theta}$$

by de Moivre.

1.4 The complex exponential

Note that de Moivre's Theorem implies that

$$e^{\mathrm{i}\varphi}e^{\mathrm{i}\theta} = e^{\mathrm{i}(\varphi+\theta)}$$

demonstrating the similarity with this notation and the (real) exponential function. This can be taken further:

Let $z = x + \mathrm{i}y$ with $x, y \in \mathbb{R}$, so $x = \mathrm{Re}\,z$, $y = \mathrm{Im}\,z$.

Definition. The complex exponential function is defined to be

$$e^z = e^{x+\mathrm{i}y} = e^x e^{\mathrm{i}y} = e^x(\cos y + \mathrm{i}\sin y)$$

This (using known properties of the real exponential and trigonometric functions, such as the 2π-periodicity of trig functions and the fact that $e^x = 1$ if and only of $x = 0$) we immediately can see that the following properties hold:

1. $\left|e^{i\theta}\right| = 1$ if $\theta \in \mathbb{R}$.
2. $|e^z| = e^{\mathrm{Re}(z)}$.
3. $e^z = 1$ if and only if $x = 0$ and $y = 2\pi k$ for $k \in \mathbb{Z}$. That is $e^z = 1 \iff z = 2\pi i k$, $k \in \mathbb{Z}$.
4. $e^{z+2\pi i k} = e^z$ for $k \in \mathbb{Z}$ (that is, e^z is periodic in the imaginary direction with period 2π.

Remark. Next year you will see the complex exponential defined in a different manner (as a power series). This enables other useful facts, such as differentiability, to be shown. However, it comes with the cost that you must prove these facts above all over again, starting from the new definition. (They are, of course, equivalent).

We can write the sine and cosine functions in terms of the complex exponential function.

Lemma. *We have*

$$\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$$
$$\sin(x) = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right)$$

Proof. Recall that $e^{ix} = \cos(x) + i\sin(x)$, so if $x \in \mathbb{R}$ we have $\cos(x) = \mathrm{Re}(e^{ix})$ and $\sin(x) = \mathrm{Im}(e^{ix})$ Also recall that $\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z})$ and $\mathrm{Im}(z) = \frac{1}{2i}(z + \overline{z})$. The required result follows from the fact that $\overline{e^{i\theta}} = e^{-i\theta}$. □

Remark. Since we have just extended the complex exponential function to complex z, this gives us a way to extend sine and cosine to have complex inputs. Now it is possible to solve the equation $\sin(z) = 2$, for example! More details on this in the second year course *Functions of a complex variable*.

1.5 Roots of unity and non-integer powers

Definition. An n^{th} root of unity is a complex number ζ which satisfies $\zeta^n = 1$.

The n^{th} roots of unity and can be found using de Moivre's theorem.

Example. Find the cube roots of unity (or the 3^{rd} roots of unity).

Solution. We know from de Moivre's Theorem that if $\zeta = re^{i\theta}$ then $\zeta^3 = r^3 e^{3i\theta}$. We need to find values of r and θ such that this equals 1 (which has magnitude 1 and argument 0). The only positive real solution to $r^3 = 1$ is $r = 1$. However, the argument is more complicated, because you can wrap around the circle 2π, or 4π, or -2π etc, and end up back at the same place. That is for any integer n, $e^{2\pi i n} = e^0 = 1$. Therefore if $\theta = 2\pi n/3$ then $\zeta = e^{i\theta}$ satisfies $\zeta^3 = 1$. This suggests there are infinitely many cube roots of unity. However, consider $\theta = 0$ and $\theta = 6\pi/3 = 2\pi$. These represent the same complex number (namely $\zeta = 1$) and so they cannot count as "different" solutions. This is an example of modular arithmetic, where we are only interested in solutions modulo 2π. The next chapter will contain much more information on modular arithmetic.

Therefore there are three cube roots of unity,

$$\zeta_1^{(3)} = e^0 = 1$$
$$\zeta_2^{(3)} = e^{2\pi \mathrm{i}/3} = \cos(2\pi/3) + \mathrm{i}\sin(2\pi/3) = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}$$
$$\zeta_3^{(3)} = e^{-2\pi \mathrm{i}/3} = \cos(2\pi/3) + \mathrm{i}\sin(-2\pi/3) = -\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}$$

This argument generalises to show that there are exactly n n^{th} roots of unity. They are

$$\begin{aligned}\zeta_1^{(n)} &= e^0 = 1\\ \zeta_2^{(n)} &= e^{2\pi \mathrm{i}/n}\\ \zeta_3^{(n)} &= e^{4\pi \mathrm{i}/n}\\ &\vdots\\ \zeta_n^{(n)} &= e^{2(n-1)\pi \mathrm{i}/n}\end{aligned}$$

Remark. The value for $\zeta_3^{(3)}$ is not inconsistent with our previous example, since $e^{4\pi \mathrm{i}/3} = e^{4\pi \mathrm{i}/3 - 2\pi} = e^{-2\pi \mathrm{i}/3}$ as before.

Remark. Note that all the roots of unity lie on the unit circle, in the sense that $|\zeta_j^{(n)}| = 1$.

Lemma. *For any $n \geq 2$, the n^{th} roots of unity sum to zero, in the sense that*

$$\sum_{j=1}^{n} \zeta_j^{(n)} = 0$$

and they multiply together to give either -1 (if n is even) or $+1$ (if n is odd), in the sense that

$$\prod_{j=1}^{n} \zeta_j^{(n)} = (-1)^{n+1}$$

Example. We can check this for the cube roots of unity, found above:

$$1 + \left(-\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}\right) + \left(-\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}\right) = 0.$$

and

$$1 \times \left(-\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}\right) \times \left(-\frac{1}{2} - \mathrm{i}\frac{\sqrt{3}}{2}\right) = \frac{1}{4} + \frac{3}{4} = 1$$

Proof. The roots of unity are, by definition, the solutions to the polynomial equation $z^n - 1 = 0$. Therefore (see the later chapter on polynomials) we can write

$$z^n - 1 = (z - \zeta_1^{(n)})(z - \zeta_2^{(n)}) \ldots (z - \zeta_n^{(n)})$$

and expanding out the brackets and finding the coefficient of z^{n-1} and z^0 we see that

- The coefficient of the z^{n-1} term on the RHS is the sum of the roots of unity, and since for $n > 1$ there is no z^{n-1} term on the LHS the coefficient must be zero.
- The coefficient of the constant term on the RHS is the product $\prod_{j=1}^{n}(-\zeta_j)$ and the constant term in the defining equation is -1.

□

1.5.1 Square- and cube-roots

The same methodology allows one to calculate square- and cube-roots of complex z (and higher powers).

Example. Find the square-roots of i.

Solution. We have that $\mathrm{i} = e^{\mathrm{i}\pi/2}$ and we can also write this (by subtracting 2π from the argument) as $e^{-3\mathrm{i}\pi/2}$.

Therefore the square-roots of i are $e^{\mathrm{i}\pi/4}$ and $e^{-3\mathrm{i}\pi/4}$.

Remark. We can easily write the roots in cartesian form: We have the first one is

$$e^{\mathrm{i}\pi/4} = \frac{\sqrt{2}}{2}(1+\mathrm{i})$$

and the second one is

$$e^{-3\mathrm{i}\pi/4} = -\frac{\sqrt{2}}{2}(1+\mathrm{i})$$

Notice the familiar "two square roots, one the negative of the other" still holds true.

There is an alternative way to find these square roots. Say we wish to find the square root of $a + \mathrm{i}b$. Call the answer $x + \mathrm{i}y$. By the definition of square-root this means that

$$a + \mathrm{i}b = (x + \mathrm{i}y)^2 = x^2 - y^2 + 2\mathrm{i}xy$$

and identifying real and imaginary parts yields

$$\begin{aligned} a &= x^2 - y^2 \\ b &= 2xy \end{aligned}$$

The second equation says that $y = b/2x$ (so long as $x \neq 0$), and substituting that into the first yields

$$a = x^2 - \tfrac{b^2}{4x^2} \qquad \text{which rearranges to yield} \qquad x^4 - ax^2 - \tfrac{1}{4}b^2 = 0$$

This is a quadratic equation in x^2, so letting $t = x^2$ we have two solutions

$$t = \frac{a \pm \sqrt{a^2 + b^2}}{2}$$

One of these solutions will be positive, the other negative. Let t_+ be the positive one, then the real part of the square root of $a + \mathrm{i}b$ is equal to $\pm\sqrt{t_+}$ and the imaginary part can be read off from $y = b/2x$.

Example. Find the square roots of i using the above method.

Solution. In this case $a = 0$ and $b = 1$, so

$$t_+ = \frac{0 + \sqrt{0^2 + 1^2}}{2} = \frac{1}{2}$$

Hence we have either

$$x = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2} \qquad \text{and} \qquad y = \frac{1}{2\frac{\sqrt{2}}{2}} = \frac{\sqrt{2}}{2}$$

or

$$x = -\sqrt{\frac{1}{2}} = -\frac{\sqrt{2}}{2} \qquad \text{and} \qquad y = \frac{1}{-2\frac{\sqrt{2}}{2}} = -\frac{\sqrt{2}}{2}$$

as before.

Example. Find the three cube-roots of $1 + \mathrm{i}$.

Solution. We have $1 + \mathrm{i} = \sqrt{2}e^{\mathrm{i}\pi/4}$ so one cube-root is $2^{1/6}e^{\mathrm{i}\pi/12}$.

Adding 2π to the argument we have that $1 + \mathrm{i} = \sqrt{2}e^{9\mathrm{i}\pi/4}$ so another cube-root is $2^{1/6}e^{9\mathrm{i}\pi/12}$.

Finally subtracting 2π from the argument yields $1 + \mathrm{i} = \sqrt{2}e^{-7\mathrm{i}\pi/4}$ so the third cube-root is $2^{1/6}e^{-7\mathrm{i}\pi/12}$.

The point is, in all three cases, if you take the answer, raise it to the third power, you will have a polar form of the complex number $1 + \mathrm{i}$.

Remark. It's not immediately obvious, but these cube-roots can be evaluated in closed cartesian form too. We have

$$\begin{aligned} 2^{1/6}e^{\mathrm{i}\pi/12} &= \frac{1+\sqrt{3}}{2^{4/3}} + \mathrm{i}\frac{\sqrt{3}-1}{2^{4/3}} \\ 2^{1/6}e^{9\mathrm{i}\pi/12} &= -2^{-1/3} + 2^{-1/3}\mathrm{i} \\ 2^{1/6}e^{-7\mathrm{i}\pi/12} &= \frac{1-\sqrt{3}}{2^{4/3}} - \mathrm{i}\frac{\sqrt{3}+1}{2^{4/3}} \end{aligned}$$

(Once you've been told the answer, you can verify it by cubing the answer and getting $1 + \mathrm{i}$).

These results tell us, for instance, that $\cos(\frac{\pi}{12}) = \frac{1+\sqrt{3}}{2^{3/2}}$.

This can be generalised to the n^{th} roots. To find the n numbers that are the n^{th} roots of z, simply write $z = re^{\mathrm{i}\theta}$, then the n^{th} roots are

$$r^{1/n}e^{\mathrm{i}\theta/n}, r^{1/n}e^{\mathrm{i}(\theta+2\pi)/n}, r^{1/n}e^{\mathrm{i}(\theta+4\pi)/n}, \ldots, r^{1/n}e^{\mathrm{i}(\theta+2\pi(n-1))/n}$$

Notice how these can be written as $r^{1/n}e^{\mathrm{i}\theta/n}\zeta_j^{(n)}$ where $\zeta_j^{(n)}$ is one of the n^{th} roots of unity.

Finally, notice how we can now find rational powers of any complex z. For $z \in \mathbb{C}$, to calculate $z^{p/q}$ where $p \in \mathbb{Z}$ and $q \in \mathbb{N}$, first find the q^{th} roots of z, and for each answer raise it to the p^{th} power.

1.6 Curves and planar regions described using complex numbers

For $r, \theta \in \mathbb{R}$ the values of z which satisfy the equation $\mathrm{Re}(e^{\mathrm{i}\theta}z) = r$ yield a straight line in the Argand plane. To see this, expand out $z = x + \mathrm{i}y$ and $e^{\mathrm{i}\theta} = \cos(\theta) + \mathrm{i}\sin(\theta)$, and take the real parts:

$$\mathrm{Re}(e^{\mathrm{i}\theta}z) = \cos(\theta)x - \sin(\theta)y = r$$

which can be rearranged to

$$y = \cot(\theta)x - r\csc(\theta)$$

which is the equation of a straight line with gradient $\cot(\theta)$ and intercept $-r\csc(\theta)$.

Similarly, for $s \in \mathbb{R}$, $\mathrm{Im}(e^{\mathrm{i}\theta}z) = s$ also yields a straight line. The two lines cross at a point when $e^{\mathrm{i}\theta}z = r + \mathrm{i}s$, that is, when $z = (r + \mathrm{i}s)e^{-\mathrm{i}\theta}$. The two lines cross at right-angles (we'll delay the proof of this fact until we have covered the dot product in vectors).

For $\omega \in \mathbb{C}$ and $r \in \mathbb{R}$ the equation $|z - \omega| = r$ is an equation for the circle of radius r centred at ω.

Similarly, the region of the Argand plane where $|z - \omega| \leq r$ is the disc of radius r centered at ω.

Finally $\mathrm{Re}(z) > 1$ is the region of the Argand plane to the right of the vertical line with real part 1.

Further reading, and where's next?

In the second year *Pure Mathematics* stream, you will see how to create other systems of number that have the same properties as real and complex numbers (known as fields).

Algebraic Number Theory in the third year makes use of roots of unity when understanding properties (such as primality) of generalisations of integers.

Functions of a Complex Variable in the second year and *Applied Complex Analysis* in the third year show that imposing the structure of differentiability and being complex leads to many spectacular results (such as analytic continuation: If you know the values of a function in one region, then you know its values in a completely different region). These lead to many applications, such as integral transforms seen in the *Applied Mathematics* second year stream, which is the mathematics behind CAT scans.

It's worth remarking that electrical engineering uses j not i.

One can look for higher dimensional numbers. There are no numbers that have three real dimensions, but the quaternions are a four-real dimensional family. They were discovered by Hamilton in 1843.

Chapter 2

Integers and congruences

This chapter we will consider equations defined over the integers. We need two sets of integers:

- The set of all integers (positive, negative and zero) is $\mathbb{Z}$
- The set of natural numbers (positive integers) is $\mathbb{N}$

2.1 The greatest common divisor

Lemma 2.1. *For $a \in \mathbb{Z}$ and $b \in \mathbb{N}$, there exists unique $q, r \in \mathbb{Z}$ with $0 \leq r < b$ such that $a = qb + r$*

Example. If $a = 19$ and $b = 5$ then $a = 3b + 4$.

If $a = 20$ and $b = 5$ then $a = 4b$ (that is $r = 0$).

If $a = -19$ and $b = 5$ then $a = (-4)b + 1$ (note that $0 \leq r < b$).

Proof. Let

$$S = \{n \geq 0, n \in \mathbb{Z}, n = a - mb \text{ for some } m \in \mathbb{Z}\}$$

Note that S is not the empty set. (That is, it contains at least one element). This is because if $a \geq 0$ then $n = a - 0b = a \geq 0$ is an element of S. If $a < 0$, then since $b \geq 1$ we have $n = a - ab = (-a)(b-1) \geq 0$ and so is an element of S.

We now make use of the seemingly trivial fact that a non-empty set of integers which is bounded from below has a least element. Call that element r. Therefore $r = a - qb$ for some $q \in \mathbb{Z}$.

Note that $r \geq 0$ since it is an element of S.

Note that $r < b$ since $a - (q + 1)b < a - qb = r$ which means that $a - (q + 1)b \notin S$ since r is the least member. Therefore $a - (q + 1)b < 0$ which implies $r = a - qb < b$.

Now we show such q and r are unique.

Assume there is another solution. That is, assume we have $a = qb + r = q'b + r'$ with $0 \leq r, r' < b$. If $q = q'$ then $r = r'$ and the two solutions are the same. If $q \neq q'$ then one must be larger than the other, so assume (without loss of generality) that $q' > q$. Since they are integers, we must have $q' \geq q + 1$. Then we have

$$r - r' = q'b - qb = (q' - q)b \geq b$$

But this means $r \geq b + r' \geq b$ which contradicts the assumption $r < b$. □

Definition. Let $a, b \in \mathbb{Z}$. We say that *b divides a*, or $b \mid a$ if $a = qb$ for some $q \in \mathbb{Z}$. Alternatively we can say that *a is a multiple of b*, or that *b is a factor of a*.

Remark. In the notation of the previous lemma, this says $r = 0$. But note that divisibility doesn't have the constraint that $b \in \mathbb{N}$. Factors can be positive and negative.

Having said that, you will occasionally find mathematicians refer to the divisors of a number and mean only positive divisors.

Example. We have

- $3 \mid 12$ but $3 \nmid 13$.
- Negative numbers are fine: We have $-2 \mid 4$ and $2 \mid (-4)$.
- The integer 0 does not divide any integer. (The special case "zero dividing zero" is declared to be undefined).
- Every integer divides 0.

Definition. We say that c is a *common divisor* of a and b if $c \mid a$ and $c \mid b$.

Definition. We say that c is the *greatest common divisor* of a and b if c is the largest number such that $c \mid a$ and $c \mid b$ (that is, there is no c' such that $c' > c$ and $c' \mid a$ and $c' \mid b$).

This is written as $c = (a, b)$ or $c = \gcd(a, b)$.

Example. The divisors of 12 are $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$.

The divisors of 30 are $\pm 1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 10, \pm 15, \pm 30$

The common divisors of 12 and 30 are $\pm 1, \pm 2, \pm 3, \pm 6$.

The greatest common divisor of 12 and 30 is 6. (Note it is +6; the greatest common divisor is positive).

Definition. If $\gcd(a, b) = 1$ then we say that a and b are *coprime*

Lemma 2.2. *If $c \mid a$ and $c \mid b$ then $c \mid (sa + tb)$ for any $s, t \in \mathbb{Z}$.*

Proof. If $c \mid a$ and $c \mid b$ then $a = cn$ and $b = cm$ for some $n, m \in \mathbb{Z}$.

Therefore $sa + tb = scn + tcm = c(sn + tm)$. Since $sn + tm \in \mathbb{Z}$, we have $c \mid (sa + tb)$. □

This implies

Lemma 2.3. *If $a = qb + r$ then* $\gcd(a, b) = \gcd(b, r)$

Proof. If $c \mid a$ and $c \mid b$ then $c \mid (a - qb)$ (by the previous lemma) which is another way of saying $c \mid r$. Therefore a common divisor of both a and b is also a common divisor of b and r.

Conversely if $c \mid b$ and $c \mid r$ then $c \mid (qb + r)$, that is $c \mid a$. Therefore any common divisor of b and r is also a common divisor of b and a.

Since the common divisors are the same, in particular the greatest common divisor is the same. □

Lemma 2.4. *The greatest common divisor of a and b can be written as a linear multiple of a and b. That is there exist s and t such that* $\gcd(a, b) = sa + tb$.

Proof. Let $S = \{n \in \mathbb{N} : n = ma + \ell b \text{ for some } m, \ell \in \mathbb{Z}\}$.

The set S is not empty. To see this note that if $a > 0$ then $a = 1 \times a + 0 \times b \in S$. If $a < 0$ then $(-1) \times a + 0 \times b > 0$ and so is in S. If $a = 0$ then switch the role of a and b.

Since the set has integer elements, is non-empty, and is bounded from below, it has a least member. Let $d = sa + tb$ be that least member. We claim that $d = \gcd(a, b)$.

Write $a = qd + r$ with $0 \leq r < d$ (which we can do due to Lemma 2.1 above).

Note that $r = a - qd = a - q(sa + tb) = (1 - qs)a - qtb$. If $r > 0$ then $r \in S$ and $r < d$ contradicting the fact that d is the least member of S.

Therefore the only possibility is that $r = 0$. But that means $a = qd$ so d is a divisor of a.

Running the same argument but for $b = q'd + r'$ shows that $d \mid b$ too. Therefore d is a common divisor of both a and b.

Now assume d' is another common divisor of a and b. Then by Lemma 2.2 we have $d' \mid (sa + tb)$ which means $d' \mid d$ which means $d' \leq d$. Therefore d is the greatest common divisor. □

2.2 Euclid's algorithm

The previous results are existence proofs. But how can we find the greatest common divisor of two numbers? And how can we find the numbers s and t such that $\gcd(a, b) = sa + tb$?

The answer comes from *Euclid's algorithm*, which is a constructive method for finding $\gcd(a, b)$ and also for finding $s, t \in \mathbb{Z}$ such that $\gcd(a, b) = sa + tb$. Indeed, one could use Euclid's algorithm as an alternative proof for Lemma 2.4. This algorithm was first described by Euclid in Book VII of his Elements in about 300 BC.

Given $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ by Lemma 2.1 we can write

$$a = q_1 b + r_1$$

with $0 \leq r_1 < b$.

If $r_1 = 0$, stop and set $\gcd(a, b) = b$.

Otherwise, note that by Lemma 2.3 we have $\gcd(b, r_1) = \gcd(a, b)$.

Now set

$$b = q_2 r_1 + r_2 \qquad \text{with } 0 \leq r_2 < r_1$$

Again, if $r_2 = 0$ stop, and note that $r_1 = \gcd(b, r_1) = \gcd(a, b)$.

Otherwise repeat

$$\begin{aligned} r_1 &= q_3 r_2 + r_3 \\ r_2 &= q_4 r_3 + r_4 \\ &\vdots \\ r_{k-2} &= q_k r_{k-1} + r_k \\ r_{k-1} &= q_{k+1} r_k + 0 \end{aligned}$$

Since $b > r_1 > r_2 > r_3 > \cdots \geq 0$ eventually this algorithm must terminate with $r_{k+1} = 0$.

By repeated use of Lemma 2.3 we have $\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{k-1}, r_k) = r_k$.

That is, the last non-zero remainder is the gcd.

This is the first part of the algorithm, finding the gcd of two numbers.

Remark. We have assumed $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ here. If both a and b are negative, then one can use the algorithm to find the gcd of $-a$ and $-b$ instead. If a is positive and b is negative, switch the roles of a and b in the above algorithm.

The second part of the algorithm says that one can write the gcd as a linear combination of a and b. To see this, simply implement the algorithm backwards.

Note that

$$r_k = r_{k-2} - q_k r_{k-1}$$

But from the previous line

$$r_{k-1} = r_{k-3} - q_{k-1} r_{k-2}$$

so substitute that to get

$$\begin{aligned} r_k &= r_{k-2} - q_k(r_{k-3} - q_{k-1} r_{k-2}) \\ &= (1 + q_k q_{k-1}) r_{k-2} - q_k r_{k-3} \end{aligned}$$

and then use the line above that to write r_{k-2} as a linear combination of r_{k-2} and r_{k-4}. Repeat this until r_k is written in terms of a and b.

Example. Find the gcd$(1152, 924)$ and write it as a linear combination of 1152 and 924.

$$\begin{aligned} 1152 &= 1 \times 924 + 228 \\ 924 &= 4 \times 228 + 12 \\ 228 &= 19 \times 12 + 0 \end{aligned}$$

therefore gcd$(1152, 924) = 12$ and we have

$$\begin{aligned} 12 &= 924 - 4 \times 228 \\ &= 924 - 4 \times (1152 - 1 \times 924) \\ &= (1 + 4) \times 924 - 4 \times 1152 \\ &= 5 \times 924 - 4 \times 1152 \end{aligned}$$

Remark. The coefficients in the linear combination are not unique.

Note that for any $n \in \mathbb{Z}$

$$\begin{aligned} 12 &= 5 \times 924 - 4 \times 1152 \\ &= (5 + 1152n) \times 924 - (4 + 924n)1152 \end{aligned}$$

2.3 Modular arithmetic

Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{N}$.

Definition. We say that *a is congruent to b modulo m*, written $a \equiv b \pmod{m}$ if $m \mid (a - b)$.

Example. We have

- $28 \equiv 13 \pmod{15}$ since $28 - 13 = 15$ is a multiple of 15.
- $9 \equiv -1 \pmod{5}$ since $9 - (-1) = 10$ is a multiple of 5.

- $1 \equiv 3 \equiv 5 \equiv 7 \equiv \ldots \pmod 2$ since odd-odd is even, and hence divisible by 2.

Lemma 2.5. *The following are equivalent:*

1. $a \equiv b \pmod m$
2. $m \mid (a - b)$
3. $a = qm + b$ *for some* $q \in \mathbb{Z}$
4. *a and b have the same remainder on division by m*

Proof. By definition, we have that (1) $\iff$ (2).

To see that (2) $\iff$ (3) note that

$$m \mid (a - b) \iff a - b = qm \iff a = qm + b$$

To see that (3) $\implies$ (4) note that if the remainder on dividing b by m is r then $b = q'm + r$ with $0 \leq r < m$. This implies that

$$a = qm + b = qm + q'm + r = (q + q')m + r$$

That is, the remainder on dividing a by m is also r.

To see that (4) $\implies$ (2) note that by (4) we have just shown we may write $a = q''m + r$ and $b = q'm + r$. Therefore

$$a - b = q''m + r - (q'm + r) = (q'' - q')m$$

That is, $m \mid (a - b)$ as required. □

Lemma 2.6. *Given $a \in \mathbb{Z}$ and $m \in \mathbb{N}$ there exists a unique r satisfying $0 \leq r < m$ such that $a \equiv r \pmod m$.*

Proof. This follows immediately from Lemma 2.1, which says that for any $a \in \mathbb{Z}$ and $m \in \mathbb{N}$ there exists a unique r with $0 \leq r < m$ such that $a = qm + r$. That is, any $a \in \mathbb{Z}$ is congruent to exactly one of $0, 1, 2, \ldots, (m - 1)$. □

Lemma 2.7. *Suppose $a \equiv r_1 \pmod m$ and $b \equiv r_2 \pmod m$ then*

- $a + b \equiv r_1 + r_2 \pmod m$
- $ab \equiv r_1 r_2 \pmod m$.

Proof. The assumptions imply that there exists q_1 and q_2 such that $a = q_1 m + r_1$ and $b = q_2 m + r_2$. (Note that we are not assuming $0 \leq r_1, r_2 < m$ here). Therefore

$$a + b = q_1 m + r_1 + q_2 m + r_2 = (q_1 + q_2)m + (r_1 + r_2)$$

and

$$ab = (q_1 m + r_1)(q_2 m + r_2) = (q_1 q_2 m + q_1 r_2 + q_2 r_1)m + r_1 r_2$$

which implies the results. □

Corollary. *If $a \equiv b \pmod m$ then $a^n \equiv b^n \pmod m$.*

This allows us to calculate various congruences more easily than one might initially suspect.

Lemma 2.8. *Every square number is congruent to* 0 *or* 1 *modulo 4.*

Proof. Every integer n is congruent to exactly one of $0, 1, 2, 3 \pmod 4$. Therefore n^2 is congruent to $0^2, 1^2, 2^2, 3^2 \equiv 0, 1, 0, 1 \pmod 4$. □

Corollary (1)**.** *If* $n \equiv 3 \pmod 4$ *then n is not the sum of two squares.*

Proof. By Lemma 2.8, the sum of two squares can only be $\equiv 0, 1, 2 \pmod 4$. □

Remark. This is the first hint of the power of congruence relations. It enables mathematicians to rule things out (that might be hard to prove) on purely congruence conditions (often known as local conditions). The third year module *Number Theory* will develop the mathematics behind representing numbers as sums of squares, cubes, etc.

Corollary (2)**.** *The number* $\sqrt{2}$ *is not rational.*

Proof. If $\sqrt{2}$ were rational, then it could be written as $\sqrt{2} = \frac{p}{q}$ with $\gcd(p, q) = 1$. That is, $p^2 = 2q^2$. Reducing this modulo 4 and using Lemma 2.8 we see that $p^2 \equiv 0, 1 \pmod 4$ and $2q^2 \equiv 0, 2 \pmod 4$. The only possible solution is $4 \mid p^2$ and $4 \mid 2q^2$, which means both p and q must be even which contradicts the requirement that $\gcd(p, q) = 1$. □

Remark. This proof is a pure mimic of the traditional one, but simply written in modular language.

Example. Find $2^{14} \pmod{15}$

Solution. We have

$$2^{14} = 2^{12} \times 2^2 = (2^4)^3 \times 2^2 = (16)^3 \times 4 \equiv 1^3 \times 4 \equiv 4 \pmod{15}$$

Example. Show that $8^{4n+2} + 4^{2n}$ is a multiple of 5 for all $n \in \mathbb{N}$.

Solution. We have that $4 \equiv -1 \pmod 5$ so $4^{2n} \equiv 1 \pmod 5$ for all $n \in \mathbb{N}$.

Similarly we have $8 \equiv 3 \pmod 5$ so $8^{4n+2} = (8^2)^{2n+1} \equiv 9^{2n+1} \equiv (-1)^{2n+1} \equiv -1 \pmod 5$.

Therefore

$$8^{4n+2} + 4^{2n} \equiv (-1) + 1 \equiv 0 \pmod 5$$

and this is the same as $5 \mid \left(8^{4n+2} + 4^{2n}\right)$

2.4 Congruence equations

A congruence equation is a problem of the sort: Given $a, b \in \mathbb{Z}$ and $m \in \mathbb{N}$ find $x \in \mathbb{Z}$ such that $ax \equiv b \pmod m$.

Example. Find solutions to $4x \equiv 3 \pmod{10}$.

Solution. Note that $4x$ is always even but 3 is odd. Therefore $4x - 3$ is odd, so can never be divisible by 10. Therefore there are no solutions.

Example. Find solutions to $6x \equiv 9 \pmod{15}$

Solution. Running through all possible values x can take mod 15, we see that $x = 4$ and $x = 9$ are solutions (as is any other x that is congruent to either 4 or 9).

Lemma 2.9. *The congruence equation* $ax \equiv b \pmod m$ *has solutions if and only if* $\gcd(a, m) \mid b$.

Proof. Throughout we let $d = \gcd(a, m)$.

$\Longrightarrow$ Let $x = c$ be a solution of $ax \equiv b \pmod{m}$, which means $ac = qm + b$. But rearranging we have $b = ac - qm$. However, since d is a gcd, we have $d \mid a$ and $d \mid m$ so also $d \mid ac - qm$ as it is a linear combination. That is, $d \mid b$.

$\Longleftarrow$ Note that if $d \mid b$ then $b = b'd$ for some $b' \in \mathbb{Z}$. By Lemma 2.4 we can write the gcd as $d = sa + tm$ for some $s, t \in \mathbb{Z}$. This means that

$$b = b'(sa + tm) = b'sa + b'tm \equiv (b's)a \pmod{m}$$

which tells us $x = sb'$ is a solution. □

Remark. This proof, whereas written as an existence proof, actually shows how to find a solution: Use Euclid's algorithm to write the gcd of a and m as a linear combination of a and m, and use that combination to find a solution.

Example. There are no solutions to $6x \equiv 4 \pmod{9}$ as $\gcd(6, 9) = 3$ and $3 \nmid 4$.

Example. Find an x so that $924x \equiv 36 \pmod{1152}$

Solution. In a previous example we showed that $\gcd(924, 1152) = 12$ and that

$$12 = 5 \times 924 - 4 \times 1152$$

Note that

$$36 = 3 \times 12 = 3 \times (5 \times 924 - 4 \times 1152) \equiv 15 \times 924 \pmod{1152}$$

That is, we've just shown that $x = 15$ is a solution.

(And one can indeed check that $1152 \mid (15 \times 924 - 36)$) □

There are other solutions to this congruence equation, though. For example one can check that $x = 111$ is also a solution to the congruence $924x \equiv 36 \pmod{1152}$ since $1152 \mid (111 \times 924 - 36)$. Since $111 \not\equiv 15 \pmod{1152}$ they are genuinely "different" solutions.

The following results give us the whole set of solutions

Lemma 2.10. *If $ax \equiv b \pmod{m}$ has a solution $x = c$, then any $x \equiv c \pmod{m/d}$ is also a solution, where $d = \gcd(a, m)$.*

Example. Notice that in our example above, $\frac{m}{d} = \frac{1152}{12} = 96$ and $111 \equiv 15 \pmod{96}$.

Proof of Lemma. Let $d = \gcd(a, m)$ and let $m = m'd$ and $a = a'd$. Since there is a solution, Lemma 2.9 tells us that $d \mid b$, so also let $b = b'd$.

As c is a solution, we have $m \mid (ac - b)$. In other words,

$$m'd \mid (a'dc - b'd).$$

Dividing both sides by d this becomes

$$m' \mid (a'c - b')$$

which is another way of saying that $a'c \equiv b' \pmod{m'}$.

If $x \equiv c \pmod{m'}$ then $x = rm' + c$ for some $r \in \mathbb{Z}$ and so

$$a'x = arm' + a'c \equiv a'c \pmod{m'} \equiv b' \pmod{m'}$$

That is, any $x \equiv c \pmod{m'}$ solves the equation $a'x \equiv b' \pmod{m'}$, and hence also solves the equation $ax \equiv b \pmod{m}$. □

Note that if $d = \gcd(a, m) > 1$ then setting $a' = a/d$ and $m' = m/d$ implies $\gcd(a', m') = 1$. The next lemma shows that is essentially only one solution to $a'x \equiv b' \pmod{m'}$.

Lemma 2.11. *If* $\gcd(a', m') = 1$*, then there is exactly one solution modulo* m' *to* $a'x \equiv b' \pmod{m'}$.

Proof. By Lemma 2.9 we know that the equation has solutions. Let x and y be two such solutions, so $a'x \equiv a'y \pmod{m'}$. That is, $m' \mid a'(x - y)$.

Since $\gcd(m', a') = 1$ it must be that $m' \mid (x - y)$, that is they are the same solution modulo m'. □

Remark. When you look back at these notes in your second year, you will realise that all that's going on is the solution is b' times the multiplicative inverse of a' modulo m'. That is, $x \equiv (a')^{-1}b' \pmod{m'}$.

Summary: To solve $ax \equiv b \pmod{m}$,

1. Find $d = \gcd(a, m)$
2. If $d \nmid b$ then there are no solutions.
3. If $d \mid b$ then write $b = b'd$ and write $d = sa + tm$.
4. $x = sb'$ is a solution
5. Let $m' = m/d$. The set of numbers congruent to $sb' \pmod{m'}$ is the set of all solutions.

Example. Find the general solution to $70x \equiv 42 \pmod{119}$.

Solution. First find the gcd of 70 and 119.

$$\begin{aligned} 119 &= 70 + 49 \\ 70 &= 49 + 21 \\ 49 &= 2 \times 21 + 7 \\ 21 &= 3 \times 7 \end{aligned}$$

so we conclude 7 is the gcd.

Since $7 \mid 42$ we know there is a solution. Let $b' = 42/7 = 6$.

Working back up the chain of equalities, we have

$$\begin{aligned} 7 &= 49 - 2 \times 21 \\ &= 49 - 2 \times (70 - 49) \\ &= 3 \times 49 - 2 \times 70 \\ &= 3 \times (119 - 70) - 2 \times 70 \\ &= 3 \times 119 - 5 \times 70 \end{aligned}$$

Therefore we have $s = -5$.

Therefore $sb' = -5 \times 6 = -30$ is a solution.

Note that $m' = 119/7 = 17$.

The general solution is the set of all $x \equiv -30 \pmod{17} \equiv 4 \pmod{17}$. □

2.5 Chinese Remainder Theorem

From the previous section we can solve congruence equations to get solutions in the form

$$x \equiv c \pmod{m}$$

The Chinese Remainder Theorem tells us how to combine two or more of these sorts of equations to get a common solution.

Theorem 2.12 (Chinese Remainder Theorem)**.** *If m_1 and m_2 are such that* $\gcd(m_1, m_2) = 1$*, then there is a unique $x \pmod{m_1 m_2}$ that simultaneously satisfies $x \equiv c_1 \pmod{m_1}$ and $x \equiv c_2 \pmod{m_2}$.*

Proof. By Lemma 2.4 since $\gcd(m_1, m_2) = 1$ we can write $1 = s_1 m_1 + s_2 m_2$ with $s_1, s_2 \in \mathbb{Z}$.

We claim that any integer

$$x \in \{c_1 m_2 s_2 + c_2 m_1 s_1 + q m_1 m_2 \ : \ q \in \mathbb{Z}\}$$

satisfies both these equations.

Note that modulo m_1 we have

$$\begin{aligned} x &\equiv c_1 m_2 s_2 \pmod{m_1} \\ &\equiv c_1(1 - s_1 m_1) \pmod{m_1} \\ &\equiv c_1 \end{aligned}$$

and modulo m_2 we have

$$\begin{aligned} x &\equiv c_2 m_1 s_1 \pmod{m_2} \\ &\equiv c_2(1 - s_2 m_2) \pmod{m_2} \\ &\equiv c_2 \end{aligned}$$

To see there are no other solutions modulo $m_1 m_2$, note that if y is a another solution of the two original equations, then $x \equiv y \pmod{m_1}$ so $x - y$ is a multiple of m_1. Similarly $x - y$ is also a multiple of m_2. Since m_1 and m_2 are coprime, $x - y$ must be a multiple of $m_1 m_2$. That means, $x \equiv y \pmod{m_1 m_2}$, so there is only one essential solution. □

Example. Find the solution to $x \equiv 4 \pmod{5}$ and $x \equiv 6 \pmod{7}$.

Solution. Euclid's algorithm gives

$$\begin{aligned} 7 &= 5 + 2 \\ 5 &= 2 \times 2 + 1 \end{aligned}$$

so

$$1 = 5 - 2 \times 2 = 5 - 2 \times (7 - 5) = 3 \times 5 - 2 \times 7$$

Therefore the solution is $4 \times (-2) \times 7 + 6 \times 3 \times 5 = 34$ and this is the unique solution modulo $5 \times 7 = 35$. □

Further reading, and where's next?

Modular arithmetic is developed in the second year *Pure Mathematics* stream, in particular in the number theory part, but also in the group theory part. In the third year, you can see those results used in the module *Number Theory*. Basically, if $\gcd(a, m) = 1$ then by Euclid's algorithm you can find an integer a^{-1} such that $aa^{-1} \equiv 1 \pmod{m}$. The integer a^{-1} is interpreted as a multiplicative inverse, and this is a crucial ingredient to building new number systems from modular arithmetic.

Example. Multiplicative inverses are helpful for interpreting what we did in, say, the Chinese Remainder Theorem. In effect, we discovered that the common solution to $x \equiv c_1 \pmod{m_1}$ and $x \equiv c_2 \pmod{m_2}$ is

$$x \equiv [\text{mult. inv. of } m_1 \pmod{m_2}] \cdot m_1 \cdot c_2 + [\text{mult. inv. of } m_2 \pmod{m_1}] \cdot m_2 \cdot c_2 \pmod{m_1 m_2}$$

Euclid's algorithm is also used extensively in the third year module *Cryptography*. Indeed, any modern public key cryptographic system will almost certainly make use of Euclid's algorithm.

Chapter 3

Real and complex polynomials

A polynomial is simply

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where the a_j are all real numbers, or all complex numbers, or all integers, etc. The variable x does not have to be the same type of number; for example you might wish to talk about complex roots of a real polynomial, or the real roots of a polynomial with integer coefficients (such solutions are called algebraic numbers).

If $a_n \neq 0$ then we say that the polynomial is of degree n.

If $a_n = 1$ then we say that it is a monic polynomial.

A polynomial of degree 0 is a non-zero constant. If that constant is zero, that is, the polynomial is identically equal to 0, then it doesn't really have a degree, although some people say it has degree "negative infinity".

Notice that if you multiply a degree n polynomial with a degree m polynomial, you have a degree $m + n$ polynomial. If you add a degree n polynomial to a degree m polynomial where $m < n$ then you still have a degree n polynomial. However if you add a degree n polynomial to a degree n polynomial you may have a polynomial of degree less than n.

3.1 Binomial Theorem

Recall (from school) the binomial theorem, that

$$\begin{aligned}(x+y)^n &= \underbrace{(x+y)(x+y)\ldots(x+y)}_{n \text{ times}} \\ &= x^n + nx^{n-1}y + \frac{n(n-1)}{2}x^{n-2}y^2 + \cdots + nxy^{n-1} + y^n \\ &= \sum_{j=0}^{n} \binom{n}{j} x^{n-j}y^j\end{aligned}$$

Here $\binom{n}{j}$ is the binomial coefficient, which equals the number of ways of choosing j unordered things from a set of n objects (without repetition). Combinatorially, therefore,

$$\binom{n}{j} = \frac{n(n-1)(n-2)\ldots(n-j+1)}{j!} = \frac{n!}{j!(n-j)!}.$$

The binomial coefficients can be found by Pascal's triangle

$$\binom{0}{0}$$

$$\binom{1}{0} \quad \binom{1}{1}$$

$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$

$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$

which equals

$$1$$

$$1 \quad 1$$

$$1 \quad 2 \quad 1$$

$$1 \quad 3 \quad 3 \quad 1$$

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

The binomial theorem leads to various identities for sums of binomial coefficients being easily proven

- By setting $x = y = 1$ we see that
$$\sum_{j=0}^{n} \binom{n}{j} = 2^n$$
- By setting $x = 1$ and $y = -1$ we see that for $n \geq 1$,
$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} = 0$$
- By differentiating both sides wrt x and then setting $x = y = 1$ we have
$$\sum_{j=0}^{n} j\binom{n}{j} = n2^{n-1}$$
- By switching the roles of x and y we see that $\binom{n}{j} = \binom{n}{n-j}$, although this symmetry is also obvious from its definition in terms of factorials, $\binom{n}{j} = \frac{n!}{j!(n-j)!} = \binom{n}{n-j}$.

3.2 The remainder theorem

Theorem 3.1 (The remainder theorem for polynomials)**.** *Let p be a (real or complex) polynomial of degree n, and let q be a (real or complex) polynomial of degree m with $m \leq n$. There exists a polynomial s of degree $n - m$ and a polynomial r of degree $< m$ such that*

$$p(x) = q(x)s(x) + r(x)$$

Definition. If $r(x) = 0$ for all x (i.e. r is identically equal to zero) then we say that q is a factor of p.

Proof. Consider the set

$$S = \{p(x) - q(x)s(x) \ : \ s(x) \text{ is a (real / complex) polynomial}\}$$

If the set contains the 0 element (that is, the polynomial which is identically equal to zero), then we are done.

Therefore we assume q is not a factor of p. This set can be partially ordered by degree. Since the degree is a non-negative integer, and the set is clearly non-empty, the set must have an element of least degree. Call that element $r(x)$ (we haven't shown it's unique yet).

Let the degree of p be n and the degree of q be m, where we are assuming $m \leq n$. Let the degree of r be ℓ.

First we show that $\ell < m$. Assume it's not (that is, assume $\ell \geq m$). Say $r(x) = a_\ell x^\ell + a_{\ell-1}x^{\ell-1} + \cdots + a_0$ and say $q(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$. Replacing $s(x)$ with $s_1(x) = s(x) + \frac{a_\ell}{b_m}x^{\ell-m}$ (and note we are assuming $\ell - m \geq 0$ here, so $s_1(x)$ is also a polynomial, and also note $b_m \neq 0$), we see that

$$\begin{aligned}
r_1(x) &= p(x) - s_1(x)q(x) \\
&= p(x) - q(x)s(x) - \frac{a_\ell}{b_m}x^{\ell-m}q(x) \\
&= \underbrace{a_\ell x^\ell + a_{\ell-1}x^{\ell-1} + \cdots + a_0}_{r(x)=p(x)-q(x)s(x)} - \frac{a_\ell}{b_m}x^{\ell-m}\Big(\underbrace{b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0}_{q(x)}\Big) \\
&= a_{\ell-1}x^{\ell-1} - \frac{a_\ell b_{m-1}}{b_m}x^{\ell-1} + \ldots
\end{aligned}$$

which is a polynomial of degree at most $\ell - 1$. That is, $r_1 \in S$ is a polynomial of smaller degree than r, contradicting the assumption r has smallest degree. Therefore it cannot be that $\ell \geq m$, and so $\ell < m$.

Now we show r and s are unique. Let r_1 and s_1 be two other polynomials with $p - s_1 q = r_1$ and degree $r_1 < m$. Consider their difference, that is

$$r - r_1 = (p - sq) - (p - s_1 q) = (s - s_1)q$$

Since q has degree m and $r - r_1$ has degree strictly less than m, the only possibility is that $s - s_1$ is identically equal to 0. That means $s = s_1$, and so $r = r_1$ too. □

Definition. If $p = qs + r$ we call the polynomial s the quotient of p/q and the polynomial r the remainder of p/q.

One way of finding $s(x)$ and $r(x)$ is via long division. This is best explained by way of examples.

Say we wished to find the $s(x)$ and $r(x)$ for $p(x) = x^3 + 2x^2 + 3x + 4$ and $q(x) = 5x^2 + 6x + 7$. The first step in long division is

$$\begin{array}{rl}
 & \frac{1}{5}x \\
5x^2 + 6x + 7 \ \Big) & \begin{array}{llll} x^3 & +2x^2 & +3x & +4 \end{array} \\
 & \begin{array}{llll} x^3 & +\frac{6}{5}x^2 & +\frac{7}{5}x & \end{array} \\
\hline
 & \begin{array}{llll} & \frac{4}{5}x^2 & +\frac{8}{5}x & +4 \end{array}
\end{array}$$

Here the bottom line is the difference of the two lines above it, and the middle line is $q(z)$ multiplied by the top term, and that term is chosen so that the x^3 term disappears in the difference.

The bottom line is not of smaller degree than $q(z)$, so we can continue this process and add a further step to the long division

$$\begin{array}{rl|llll} & & \frac{1}{5}x & +\frac{4}{25} & & \\ \hline 5x^2+6x+7 &) & x^3 & +2x^2 & +3x & +4 \\ & & x^3 & +\frac{6}{5}x^2 & +\frac{7}{5}x & \\ \hline & & & \frac{4}{5}x^2 & +\frac{8}{5}x & +4 \\ & & & \frac{4}{5}x^2 & +\frac{24}{25}x & +\frac{28}{25} \\ \hline & & & & \frac{16}{25}x & +\frac{72}{25} \end{array}$$

and the last line is now a linear equation, so of smaller degree than the quadratic $q(x)$. This process tells us that

$$x^3 + 2x^2 + 3x + 4 = \left(5x^2 + 6x + 7\right)\left(\frac{1}{5}x + \frac{4}{25}\right) + \left(\frac{16}{25}x + \frac{72}{25}\right)$$

that is $s(x) = \frac{1}{5}x + \frac{4}{25}$ and $r(x) = \frac{16}{25}x + \frac{72}{25}$.

Example. Let $p(x) = 2x^3 + 5x^2 + 4x + 1$ and let $q(x) = 2x + 1$. Find a polynomial $s(x)$ of degree 2 and a constant r such that $p(x) = s(x)q(x) + r$.

Solution. Long division yields

$$\begin{array}{rl|llll} & & x^2 & +2x & +1 & \\ \hline 2x+1 &) & 2x^3 & +5x^2 & +4x & +1 \\ & & 2x^3 & +x^2 & & \\ \hline & & & 4x^2 & +4x & +1 \\ & & & 4x^2 & +2x & \\ \hline & & & & 2x & +1 \\ & & & & 2x & +1 \\ \hline & & & & & 0 \end{array}$$

which says that $2x^3+5x^2+4x+1 = (2x+1)(x^2+2x+1)$. That is we have (partially) factored $p(x)$. □

Example. Let $p(x) = x^2 + 1$ and let $q(x) = x + 2\mathrm{i}$. Find a polynomial $s(x)$ of degree 1 and a constant r such that $p(x) = s(x)q(x) + r$.

Solution. Long division works for complex numbers too.

$$\begin{array}{rl|lll} & & x & -2\mathrm{i} & \\ \hline x+2\mathrm{i} &) & x^2 & & +1 \\ & & x^2 & +2\mathrm{i}x & \\ \hline & & & -2\mathrm{i}x & +1 \\ & & & -2\mathrm{i}x & +4 \\ \hline & & & & -3 \end{array}$$

and so we have $x^2 + 1 = (x + 2\mathrm{i})(x - 2\mathrm{i}) - 3$

3.3 Euclid's algorithm for polynomials

Definition. If p and q are polynomials such that there exists a polynomial s with $p = qs$ then we say that q is a *factor* of p.

Remark. Note that if a polynomial q is a factor of a polynomial p, then so is the polynomial cq for any real (or complex) constant c. (This is different to integer factors).

Definition. If p_1 and p_2 are polynomials, and both have a factor q, then we say that q is a *common factor* of p_1 and p_2.

Definition. The monic polynomial of highest degree that is a common factor of p_1 and p_2 is called the *highest common factor* of p_1 and p_2.

The highest common factor can be found by Euclid's algorithm for polynomials, which is the same sequence of steps as for Euclid's algorithm for integers, and relies on a similar lemma.

Lemma 3.2. *If polynomials p_1 and p_2 have a common factor q, and if $p_1 = sp_2 + r$ with s and r polynomials, then r also has q as a factor. In particular, the highest common factor of p_1 and p_2 is the highest common factor of p_2 and r.*

Remark. This Lemma should be compared to Lemma 2.3.

Proof. Since q is a factor of p_1 we can write $p_1 = s_1 q$ with s_1 a polynomial, and similarly we can write $p_2 = s_2 q$ with s_2 a polynomial. Therefore,

$$r = p_1 - sp_2 = s_1 q - ss_2 q = (s_1 - ss_2)q$$

and since $s_1 - ss_2$ is also a polynomial [because adding two polynomials yields a polynomial, and multiplying two polynomials also yields a polynomial], we have that q is a factor of r too. □

As for the integers (Lemma 2.4), Euclid's algorithm provides a constructive proof of the following:

Theorem 3.3. *Given two polynomials, p_1 and p_2, with highest common factor q, there exists polynomials s_1 and s_2 such that*

$$q = s_1 p_1 + s_2 p_2$$

Example. Find the greatest common factor of $p_1(x) = x^4 - x^3 - 13x^2 + x + 12$ and $p_2(x) = x^3 - 2x^2 - 5x + 6$.

Solution. Using long division we have

$$\begin{array}{rl|llll} & & x & +1 & & & \\ \hline x^3 - 2x^2 - 5x + 6 &) & x^4 & -x^3 & -13x^2 & +x & +12 \\ & & x^4 & -2x^3 & -5x^2 & +6x & \\ \hline & & & x^3 & -8x^2 & -5x & +12 \\ & & & x^3 & -2x^2 & -5x & +6 \\ \hline & & & & -6x^2 & & +6 \end{array}$$

which tells us that

$$p_1 = (x + 1)p_2 + (-6x^2 + 6)$$

Repeating this, we have

$$\begin{array}{rl|llll} & & -\frac{1}{6}x & +\frac{1}{3} & & \\ \hline -6x^2 + 6 &) & x^3 & -2x^2 & -5x & +6 \\ & & x^3 & & -x & \\ \hline & & & -2x^2 & -4x & +6 \\ & & & -2x^2 & & +2 \\ \hline & & & & -4x & +4 \end{array}$$

which says that

$$p_2(x) = (-\tfrac{1}{6}x + \tfrac{1}{3})(-6x^2 + 6) + (-4x + 4)$$

Finally, we have

$$\begin{array}{rr|l} & \tfrac{3}{2}x \quad +\tfrac{3}{2} & \\ \hline -4x+4 \;\big)\; & -6x^2 \qquad\quad +6 & \\ & -6x^2 \;+6x & \\ \hline & \quad -6x \;\; +6 & \\ & \quad -6x \;\; +6 & \\ \hline & \qquad\quad 0 & \end{array}$$

which says that

$$-6x^2 + 6 = (\tfrac{3}{2}x + \tfrac{3}{2})(-4x + 4)$$

Note that Lemma 3.2 this tells us $-4x+4$ is a factor of p_1 and p_2 and there is no other factor of higher degree. To turn it into a monic polynomial (and thus the highest common factor) we divide by -4, and get

$$\text{highest common factor} = x - 1$$

Example. Find two polynomials s_1 and s_2 such that

$$x - 1 = s_1(x)p_1(x) + s_2(x)p_2(x)$$

where

$$p_1(x) = x^4 - x^3 - 13x^2 + x + 12$$

and

$$p_2(x) = x^3 - 2x^2 - 5x + 6$$

Solution. We work our way back up the chain of Euclid's algorithm. From

$$p_2(x) = (-\tfrac{1}{6}x + \tfrac{1}{3})(-6x^2 + 6) + (-4x + 4)$$

a little rearranging, and multiplying by $\frac{1}{4}$, gives

$$x - 1 = -\tfrac{1}{4}p_2(x) + (-\tfrac{1}{24}x + \tfrac{1}{12})(-6x^2 + 6).$$

Using

$$p_1(x) = (x + 1)p_2(x) + (-6x^2 + 6)$$

to get rid of the $(-6x^2 + 6)$ term, we obtain

$$\begin{aligned} x - 1 &= -\tfrac{1}{4}p_2(x) + \left(-\tfrac{1}{24}x + \tfrac{1}{12}\right)\left\{p_1(x) - (x + 1)p_2(x)\right\} \\ &= \left(-\tfrac{1}{24}x + \tfrac{1}{12}\right)p_1(x) + \left(\tfrac{1}{24}x^2 - \tfrac{1}{24}x - \tfrac{1}{3}\right)p_2(x) \end{aligned}$$

That is, $s_1(x) = -\frac{1}{24}x + \frac{1}{12}$ and $s_2(x) = \frac{1}{24}x^2 - \frac{1}{24}x - \frac{1}{3}$.

3.4 Roots of a Polynomial

Definition. The *roots* of a polynomial p are the values of x such that $p(x) = 0$.

Remark. It makes a great deal of difference whether x is allowed to be complex, or is restricted to being real, as we shall see.

First, some asides:

Remark. A real root of a monic integer polynomial is called an algebraic integer. For example, $\sqrt{2}$ is a root of the polynomial $x^2 - 2$ and so is an algebraic integer.

Remark. If we drop the requirement the polynomial is monic, then we have an algebraic number. That is, an algebraic integer is a real root of an integer polynomial. For example, $-\sqrt{2/3}$ is a root of the polynomial $3x^2 - 2$ and so is an algebraic number.

Remark. If there is *no* polynomial with integer coefficients with the property that $x = \alpha$ is one of its root, then we call α a transcendental number. There are "many more" transcendental numbers than there are algebraic numbers, but proving any one particular number transcendental is exceedingly hard. (For example, Lindemann's 1882 proof that π is transcendental was a breakthrough, that showed the impossibility of several geometric constructs, such as "squaring the circle". His proof relies on that fact that $e^{i\pi} = -1$).

Lemma 3.4. *A real or complex polynomial of degree n can have at most n roots.*

Proof. The Remainder Theorem (Theorem 3.1) allows us to see that if α is a root, we can write $p(x) = (x-\alpha)s(x) + r(x)$ where the degree of $s(x)$ is $n-1$ and the degree of $r(x)$ is strictly less than 1, that is r is a constant (or zero). If it were a non-zero constant that the LHS would be zero when $x = \alpha$ but the RHS would not be. Therefore $r(x)$ is identically equal to zero.

Thus we have shown that if $p(x)$ is a degree n polynomial in x and $p(\alpha) = 0$ then $p(x) = (x-\alpha)s(x)$ where $s(x)$ is a degree $n-1$ polynomial. If $\beta \neq \alpha$ is another root of $p(x)$, then β is a root of $s(x)$, since $p(\beta) = 0$ but $\beta - \alpha \neq 0$ so it must be that $s(\beta) = 0$. This root can then be factored out in a similar manner. It might also be the case that $s(\alpha) = 0$, in which case $(x-\alpha)$ can be factored out of $s(x)$ as before, and α is then known as a repeated root of $p(x)$.

This process can be repeated at most n times, since the degree of the quotient polynomial cannot be negative. Therefore there are at most n roots of a degree n polynomial. □

3.5 Fundamental Theorem of Algebra

Theorem 3.5 (The Fundamental Theorem of Algebra)**.** *A degree n complex polynomial has exactly n (not necessarily distinct) complex roots.*

Remark. The proof of this theorem has a remarkably long history, which we won't have time to go into here, but is well worth looking up.

Rough sketch of a proof. We will show that a complex non-constant polynomial has at least one root in $\mathbb{C}$, and by the remarks from the previous section (where that root is factored out leaving a degree $n-1$ polynomial) this shows that we can inductively find n complex roots.

Assume $p(z)$ is such that $p(0) \neq 0$ (for it were, we have found a root). Consider the image of $p(re^{i\theta})$ as θ moves from 0 to 2π. If r is very small, this image will encircle the point $p(0)$. If r is very large this image will look "close to" $r^n e^{in\theta}$ which encircles the origin n times. By continuity, there must be an r where the image passes through the origin, which means (for that r) there is a θ such that $p(re^{i\theta}) = 0$. The complex number $re^{i\theta}$ is a desired root. □

A degree n real polynomial might not have real roots (for example $x^2 + 1$), though the FTA assures us it will have complex roots. We can however, deduce a little bit more:

Lemma 3.6. *A real polynomial can be factored into linear terms and quadratic terms*

Proof. Without loss of generality assume $p(x)$ is a monic polynomial. By the FTA, we can write

$$p(x) = \prod_{j=1}^{n} \big(x - z_j\big)$$

where the z_j are the (complex) roots. If $z_j \in \mathbb{R}$ then we have a real linear factor, $(x - z_j)$.

Note that since $p(x)$ is real (which means its coefficients are real), if we evaluate $p(x)$ at a root and take the complex conjugate we get

$$0 = \overline{p(z_j)} = \overline{\sum_{k=0}^{n} a_k z_j^k} = \sum_{k=0}^{n} a_k \overline{z_j}^k.$$

That is, if z_j is a root of p, then so is $\overline{z_j}$. If $z_j \notin \mathbb{R}$ then these two roots are distinct. That is, the complex roots for real polynomials come in complex conjugate pairs. Notice that

$$(x - z_j)(x - \overline{z_j}) = x^2 - 2\,\mathrm{Re}(z_j)x + |z_j|^2$$

and this is a real quadratic factor.

Summary: We used FTA to find the roots. If the root is real, then you get a linear factor. If the root is not real, then it has a complex conjugate pair, and those two combine to give a real quadratic factor. □

3.6 Lagrange interpolation formula

As hinted at in the previous section, given the roots of a polynomial you can reconstruct the polynomial up to a multiplicative factor in front. That is, given all the (complex) roots of $p(x)$, there exists an $\alpha \in \mathbb{C}$ such that

$$p(x) = \alpha \prod_{j=1}^{n} \big(x - z_j\big)$$

The α can be determined by looking at the coefficient of the x^n term. It can also be deduced from knowing the value of p at *any* single point that is not a root.

Example. Find the quadratic polynomial $p(x)$ with roots $x = 1$ and $x = \frac{3}{2}$, and satisfies $p(2) = 1$.

Solution. We know that $p(x) = \alpha(x - 1)(x - 3/2)$. Plugging in to both sides $x = 2$ we have $1 = \alpha \times 1 \times (1/2)$. Therefore $\alpha = 2$ and

$$p(x) = 2(x - 1)(x - 3/2) = (x - 1)(2x - 3) = 2x^2 - 5x + 3$$

Not formally part of this course, the Lagrange Interpolation Formula shows how to reconstruct a degree n polynomial from just knowing any of its $n + 1$ values (not just the roots).

Theorem 3.7 (Lagrange Interpolation Formula)**.** *Let $p(x)$ be a degree n polynomial. Given any $x_1, \ldots, x_{n+1}$, let $y_j = p(x_j)$. Then*

$$p(x) = \sum_{j=1}^{n+1} p_j(x)$$

where

$$p_j(x) = y_j \prod_{\substack{k=1 \\ k \neq j}}^{n+1} \frac{x - x_k}{x_j - x_k}$$

Proof. Each $p_j(x)$ is a polynomial of degree exactly n, so the sum of them is a polynomial of degree no more than n.

Substituting $x = x_\ell$ into $p_j(x)$ yields zero (as one of the terms in the product equals 0) unless $j = \ell$ when it yields y_ℓ (as all terms in the product are equal to 1). □

Example. Find the quadratic polynomial which has values $p(1) = -1$, $p(2) = 1$ and $p(3) = 7$.

Solution. We have

$$p_1(x) = -\frac{(x-2)(x-3)}{(1-2)(1-3)} = -\frac{x^2}{2} + \frac{5x}{2} - 3$$
$$p_2(x) = \frac{(x-1)(x-3)}{(2-1)(2-3)} = -x^2 + 4x - 3$$
$$p_3(x) = 7\frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{7x^2}{2} - \frac{21x}{2} + 7$$

and so we have

$$p(x) = p_1(x) + p_2(x) + p_3(x) = 2x^2 - 4x + 1$$

Further reading, and where's next?

In the second year *Pure Mathematic* stream, you will use polynomials to create new finite fields, and these fields are used in the third year to study more advanced number theory (including things such as algebraic numbers).

Finite fields and the Lagrange formula are used in the error-correcting codes found on CDs. This enables them to play even when scratched (the sound is encoded as points on a polynomial in a finite field, and the polynomial is oversampled. If some of the data is corrupted this is noticed by the formula and can be corrected).

The stability of the Lagrange Interpolation Formula is discussed in the third year course on *Numerical Analysis*.

In the third year *Cryptography* course, Euclid's algorithm is employed to calculate in the explicit finite field used in the Advanced Encryption Standard, which is the block cipher used in almost all secure internet sites.

Chapter 4

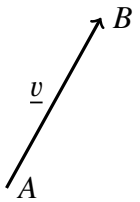
Vectors in 2 and 3 dimensions

A *scalar* is a quantity specified by a single value, such as length or temperature.

A *vector* is a quantity with both magnitude (its "size" or "magnitude") and direction (eg whether it's "up" or "down").

4.1 Definition and algebra

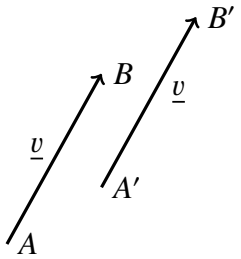
Geometrically, a vector can be thought of as a line segment going from one point to another. That is, given A and B, two points in space, the vector runs in the direction from A to B and its size (also known as its magnitude) is the distance from A to B.



$\underline{v} = \overrightarrow{AB}$

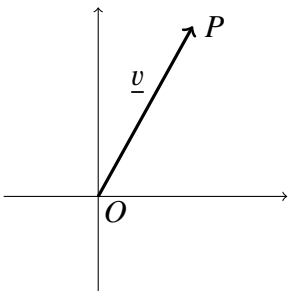
$|\underline{v}|$ is the distance from A to B

Two vectors are the same if they have the same size and the same direction, even if they are drawn in different parts of space.



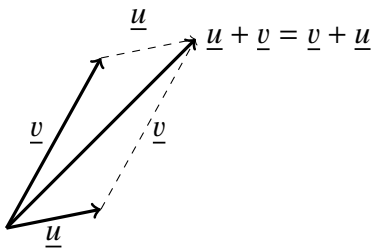
$\underline{v} = \overrightarrow{AB} = \overrightarrow{A'B'}$

In particular, any vector can be thought of starting at the origin O and going to a certain point P. Flipping this on its head, any point P in space has an associated position vector which runs from O to P.



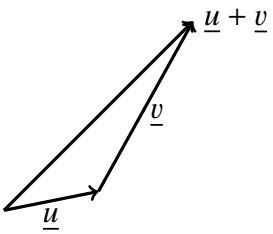
$\underline{v} = \overrightarrow{OP}$

A 2d vector is one that lives on a plane; a 3d vector is one that lives in 3-dimensional space. Two vectors of the same dimension can be added.



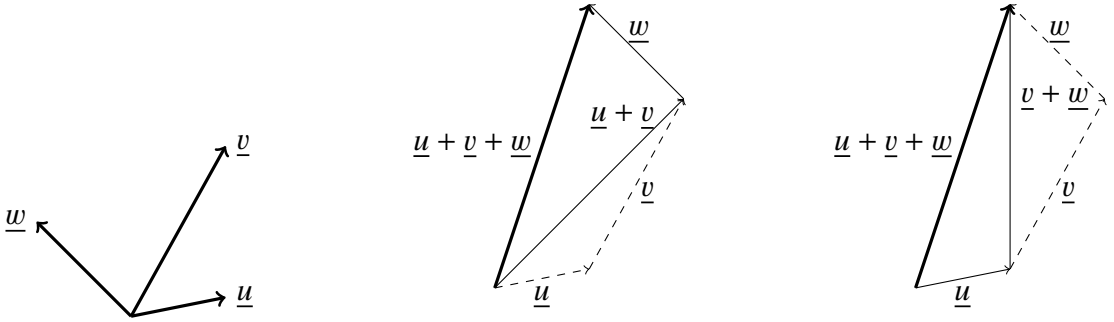
Geometrically, it is clear that you get the same effect as travelling along $\underline{v}$ and then along $\underline{u}$.

Think of $\underline{u} + \underline{v}$ as the "net effect" of going from travelling along $\underline{u}$ and then along $\underline{v}$.



We have the triangle inequality: $|\underline{u} + \underline{v}| \leq |\underline{u}| + |\underline{v}|$.

It doesn't matter the order in which we add three vectors together.

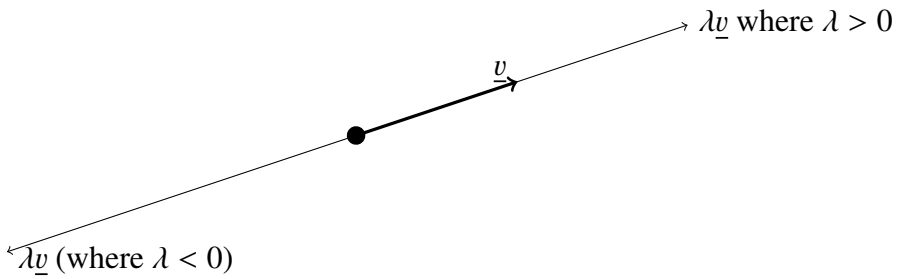


In conclusion we have seen the following

Lemma 4.1 (Properties of vector addition)**.** *Vector addition is:*

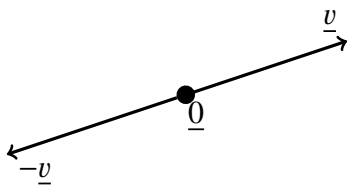
- *Commutative, since* $\underline{u} + \underline{v} = \underline{v} + \underline{u}$
- *Associative, since* $(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w})$

Vectors can be multiplied by scalars (real numbers) to get another vector.

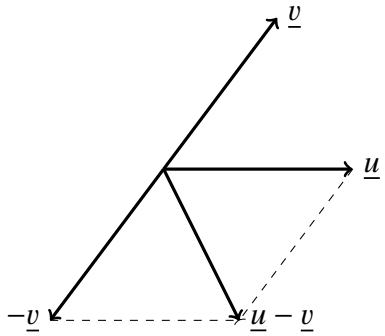


The new vector is in the same direction as the old vector if $\lambda > 0$ and is in the opposite direction to that of the old vector if $\lambda < 0$. The new vector has magnitude $|\lambda|$ times that of the old vector (so if $|\lambda| > 1$ it is longer; if $|\lambda| < 1$ it is shorter).

In particular we can subtract a vector by adding the vector multiplied by -1. Adding $\underline{v}$ to $-\underline{v}$ results in a vector with length zero, starting and finishing at the same point. That is, it is the zero vector $\underline{0}$.



We define $\underline{u} - \underline{v}$ to be $\underline{u} + (-\underline{v})$



Finally, this allows us to return to our initial definition. For a vector $\underline{v} = \overrightarrow{AB}$, if the point A has position vector $\underline{a}$ $(= \overrightarrow{OA})$ and the point B has position vector $\underline{b}$ $(= \overrightarrow{OB})$, then $\underline{v} = \underline{b} - \underline{a}$.

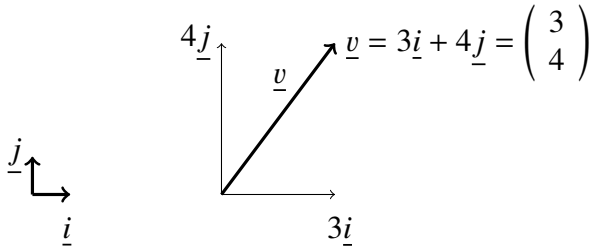
Here are some more algebraic laws which are geometrically obvious from how we have defined vectors.

Lemma 4.2 (Properties of scalar multiplication)**.** *Let $\lambda, \mu \in \mathbb{R}$ and $\underline{u}$ and $\underline{v}$ be both either 2- or 3-dimensional vectors. Then*

- $\lambda(\underline{u} + \underline{v}) = \lambda\underline{u} + \lambda\underline{v}$. *(That is, scalar multiplication is distributive over vector addition).*
- $(\lambda\mu)\underline{v} = \lambda(\mu\underline{v})$. *(That is, scalar multiplication is associative).*
- $(\lambda + \mu)\underline{v} = \lambda\underline{v} + \mu\underline{v}$. *(That is, scalar addition is distributive).*
- $1\underline{v} = \underline{v}$. *(that is, the number* 1 *is the multiplicative identity.*
- $0\underline{v} = \underline{0}$. *(Here the 0 on the LHS is the number zero, and the $\underline{0}$ on the RHS is the zero vector).*
- $\underline{v} + \underline{0} = \underline{v}$. *(That is, the zero vector is the additive identity).*

4.1.1 Coordinates

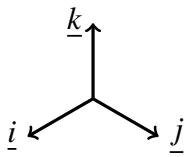
Fix special 2-d unit vectors which are perpendicular to each other, called $\underline{i}$ and $\underline{j}$, as your coordinate axes. A vector can then be written in terms of these coordinates.



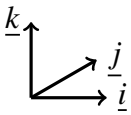
The modulus of a vector is its magnitude or its length. By Pythagoras, in this example we see that

$$|\underline{v}| = \sqrt{3^2 + 4^2} = 5$$

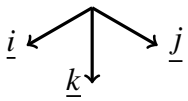
In 3-d, there are three special vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$ which all have unit length, are mutually orthogonal and arranged by the "right-hand rule"



Correct



Correct



Wrong

You can get to any point P in 3-d space by going distance a in the $\underline{i}$ direction, distance b in the $\underline{j}$ direction, and distance c in $\underline{k}$ direction. That is, a position vector $\underline{p}$ can be written as $\underline{p} = a\underline{i} + b\underline{j} + c\underline{k}$ for some $a, b, c \in \mathbb{R}$, or alternatively as

$$\underline{p} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

We call a, b, c the coordinates of the point P.

Pythagoras (twice) tells us that the magnitude or length of $\underline{p}$ is

$$|\underline{p}| = \sqrt{a^2 + b^2 + c^2}$$

Vector algebra is easy with coordinates. If $\underline{u} = a_1\underline{i} + b_1\underline{j} + c_1\underline{k}$ and $\underline{v} = a_2\underline{i} + b_2\underline{j} + c_2\underline{k}$, then

$$\underline{u} + \underline{v} = (a_1 + a_2)\underline{i} + (b_1 + b_2)\underline{j} + (c_1 + c_2)\underline{k} = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \\ c_1 + c_2 \end{pmatrix}$$

and

$$\lambda\underline{u} = \lambda a_1\underline{i} + \lambda b_1\underline{j} + \lambda c_1\underline{k} = \begin{pmatrix} \lambda a_1 \\ \lambda b_1 \\ \lambda c_1 \end{pmatrix}$$

The zero vector is

$$\underline{0} = 0\underline{i} + 0\underline{j} + 0\underline{k} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Remark. Sometimes the coordinates are written as row vectors to save space, eg $\underline{p} = (a, b, c)$.

Example. The point P has coordinates $(1, -2, 3)$ and the point Q has coordinates $(-4, -5, 6)$. The position vectors are

$$\underline{p} = \underline{i} - 2\underline{j} + 3\underline{k}$$

and

$$\underline{q} = -4\underline{i} - 5\underline{j} + 6\underline{k}$$

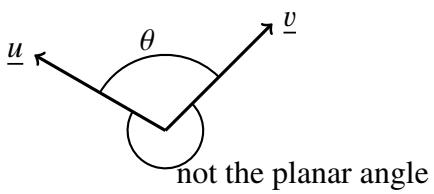
The displacement vector $\overrightarrow{PQ}$ is

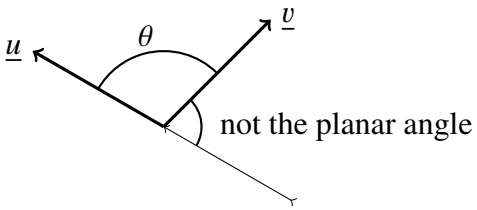
$$\underline{q} - \underline{p} = (-4 - 1)\underline{i} + (-5 + 2)\underline{j} + (6 - 3)\underline{k} = -5\underline{i} - 3\underline{j} + 3\underline{k}$$

and the distance between P and Q is the length of this vector, which is $\sqrt{(-5)^2 + (-3)^2 + 3^2} = \sqrt{43}$.

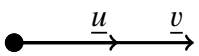
4.2 Dot product

Let $\underline{u}$ and $\underline{v}$ be vectors, and let θ be their planar angle (that is, the angle between them in the plane spanned by $\underline{u}$ and $\underline{v}$). Choose the planar angle θ such that $0 \leq \theta \leq \pi$, and it goes between the two vectors "moving outwards"

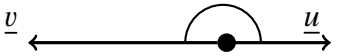




Remark. If the two vectors are going in exactly the same direction, then the angle between them is 0. If the two vectors are going in exactly opposite angles, then the angle between them is π.



Angle between $\underline{u}$ and $\underline{v}$ is 0



Angle between $\underline{u}$ and $\underline{v}$ is π

Definition. The dot product between $\underline{u}$ and $\underline{v}$ is

$$\underline{u} \cdot \underline{v} = |\underline{u}||\underline{v}| \cos\theta$$

Remark. Note that $\underline{u} \cdot \underline{v}$ is a scalar (a number). It is not a vector. Therefore $\underline{u} \cdot \underline{v} \cdot \underline{w}$ does not make sense.

Example (Very important example)**.** We have

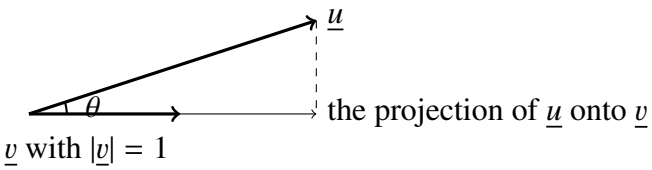
$$\underline{i} \cdot \underline{i} = \underline{j} \cdot \underline{j} = \underline{k} \cdot \underline{k} = 1$$

(This is because they have unit length). Further more

$$\underline{i} \cdot \underline{j} = \underline{i} \cdot \underline{k} = \underline{j} \cdot \underline{k} = 0$$

(This is because they are orthogonal, so the angle between them is $\pi/2$ and $\cos(\pi/2) = 0$).

One interpretation of the dot product is that if $\underline{v}$ has length 1, then $\underline{u} \cdot \underline{v}$ is "how much of $\underline{u}$ is in the direction of $\underline{v}$". This vector is known as the component of $\underline{u}$ in the direction of $\underline{v}$, or as the projection of $\underline{u}$ onto $\underline{v}$. Informally, you can think of it as the "shadow" cast by $\underline{u}$ on $\underline{v}$.



Simple trigonometry shows that the "amount" of $\underline{u}$ in the direction of $\underline{v}$ is $|\underline{u}|\cos(\theta)$ where θ is the angle between the two lines. Since $|\underline{v}| = 1$ we see that this is the same as $\underline{u} \cdot \underline{v}$. If $\underline{v}$ is a nonzero vector but doesn't have unit length, then

$$\frac{\underline{u} \cdot \underline{v}}{|\underline{v}|}$$

gives the projection of $\underline{u}$ onto $\underline{v}$, as $\frac{1}{|\underline{v}|}\underline{v}$ has unit length and is in the same direction as $\underline{v}$.

Remark. Sometimes unit vectors are written as

$$\hat{\underline{v}} = \frac{1}{|\underline{v}|}\,\underline{v}$$

Thus for any non-zero $\underline{v}$, the projection of $\underline{u}$ onto $\underline{v}$ equals $\underline{u} \cdot \hat{\underline{v}}$.

Lemma 4.3 (Properties of the dot product)**.** *We have*

- *The dot product is commutative, that is* $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$

  Proof. Follows from the definition, since the planar angle between $\underline{u}$ and $\underline{v}$ is the same as the planar angle between $\underline{v}$ and $\underline{u}$, and thus $\underline{u} \cdot \underline{v} = |\underline{u}||\underline{v}|\cos\theta = |\underline{v}||\underline{u}|\cos\theta = \underline{v} \cdot \underline{u}$. □

- *Two non-zero vectors* $\underline{u}$ *and* $\underline{v}$ *are orthogonal if and only if* $\underline{u} \cdot \underline{v} = 0$.

  Proof. There is exactly one $\theta \in [0, \pi]$ with $\cos(\theta) = 0$, namely $\theta = \pi/2$. If two lines are at right angles (have angle $\pi/2$ between them) then we say they are orthogonal. □

- *We have that* $\underline{u} \cdot \underline{u} = |\underline{u}|^2$

  Proof. This follows from $\cos(0) = 1$. □

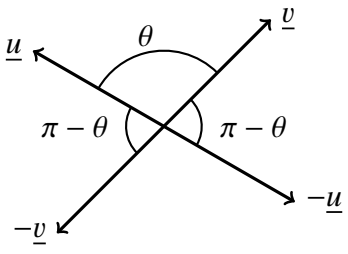
- *We have*

  $$(\lambda\underline{u}) \cdot \underline{v} = \lambda(\underline{u} \cdot \underline{v}) = \underline{u} \cdot (\lambda\underline{v})$$

  Proof. We have

  $$\underline{u} \cdot (\lambda\underline{v}) = \begin{Bmatrix} |\underline{u}||\underline{v}||\lambda|\cos\theta & \text{if } \lambda \geq 0 \\ |\underline{u}||\underline{v}||\lambda|\cos(\pi - \theta) & \text{if } \lambda < 0 \end{Bmatrix} = (\lambda\underline{u}) \cdot \underline{v}$$

  since



□

- *As a corollary of the above, for any non-zero* $\underline{v}$

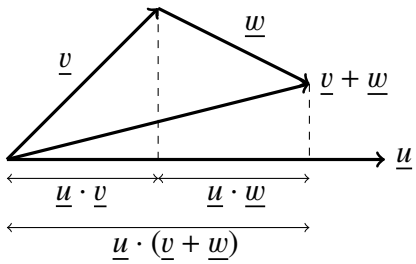
$$\frac{1}{|\underline{v}|}\underline{v} = \frac{\underline{v}}{\sqrt{\underline{v}\cdot\underline{v}}}$$

has length 1.

- *The dot product is distributive over vector addition,*

$$\underline{u}\cdot(\underline{v}+\underline{w}) = \underline{u}\cdot\underline{v} + \underline{u}\cdot\underline{w}$$

Proof. The following picture demonstrates the reason:



(Note that $\underline{u}\cdot\underline{w}$ can be negative, so some of these lengths might be subtracted). □

Here's a very important consequence to the distributive law.

Lemma 4.4 (Coordinate version of the dot product)**.** *Let* $\underline{u} = a_1\underline{i} + b_1\underline{j} + c_1\underline{k}$ *and* $\underline{v} = a_2\underline{i} + b_2\underline{j} + c_2\underline{k}$. *Then*

$$\underline{u}\cdot\underline{v} = a_1a_2 + b_1b_2 + c_1c_2$$

Proof. We have

$$\begin{aligned}\underline{u}\cdot\underline{v} &= \left(a_1\underline{i} + b_1\underline{j} + c_1\underline{k}\right)\cdot\left(a_2\underline{i} + b_2\underline{j} + c_2\underline{k}\right)\\ &= a_1\underline{i}\cdot\left(a_2\underline{i} + b_2\underline{j} + c_2\underline{k}\right) + b_1\underline{j}\cdot\left(a_2\underline{i} + b_2\underline{j} + c_2\underline{k}\right) + c_1\underline{k}\cdot\left(a_2\underline{i} + b_2\underline{j} + c_2\underline{k}\right)\end{aligned}$$

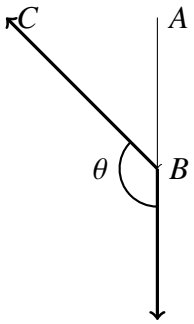
which follows from using the distributive law three times in the first term. Expanding each term out using the distributive law (this time in the second term),

$$\begin{aligned}\underline{u}\cdot\underline{v} = {} & a_1a_2\underline{i}\cdot\underline{i} + a_1b_2\underline{i}\cdot\underline{j} + a_1c_2\underline{i}\cdot\underline{k}\\ & + b_1a_2\underline{j}\cdot\underline{i} + b_1b_2\underline{j}\cdot\underline{j} + b_1c_2\underline{j}\cdot\underline{k}\\ & + c_1a_2\underline{k}\cdot\underline{i} + c_1b_2\underline{k}\cdot\underline{j} + c_1c_2\underline{k}\cdot\underline{k}\\ = {} & a_1a_2 + b_1b_2 + c_1c_2\end{aligned}$$

where the last line follows from $\underline{i}\cdot\underline{i} = 1$ and $\underline{i}\cdot\underline{j} = 0$ and similar. □

Example. Let A, B, C be points with coordinates $(2,4,2)$, $(2,4,1)$ and $(1,4,2)$ respectively. Find the angle between the lines $\overrightarrow{AB}$ and $\overrightarrow{BC}$.

Solution. The picture is

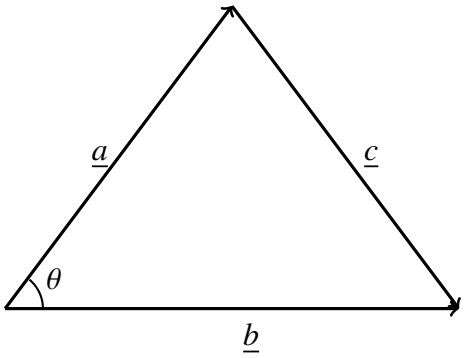


We have $\overrightarrow{AB} = \underline{b} - \underline{a} = -\underline{k}$ and $\overrightarrow{BC} = \underline{c} - \underline{b} = -\underline{i} + \underline{k}$. Therefore $\overrightarrow{AB} \cdot \overrightarrow{BC} = -1$ and $|\overrightarrow{AB}| = 1$ and $|\overrightarrow{BC}| = \sqrt{2}$. Therefore

$$\overrightarrow{AB} \cdot \overrightarrow{BC} = |\overrightarrow{AB}||\overrightarrow{BC}| \cos\theta = \sqrt{2}\cos(\theta) = -1$$

and so $\cos\theta = -1/\sqrt{2}$ implying $\theta = 3\pi/4$.

Example. The dot product enables a quick proof of the cosine rule.



As $\underline{a} + \underline{c} = \underline{b}$ we have

$$\begin{aligned}\underline{c} \cdot \underline{c} &= (\underline{b} - \underline{a}) \cdot (\underline{b} - \underline{a}) \\ &= (\underline{b} - \underline{a}) \cdot \underline{b} - (\underline{b} - \underline{a}) \cdot \underline{a} \\ &= \underline{b} \cdot \underline{b} + \underline{a} \cdot \underline{a} - 2\underline{a} \cdot \underline{b}\end{aligned}$$

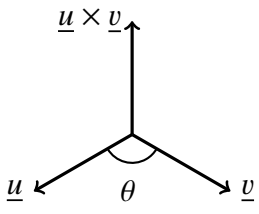
That is,

$$|\underline{c}|^2 = |\underline{a}|^2 + |\underline{b}|^2 - 2|\underline{a}||\underline{b}| \cos\theta$$

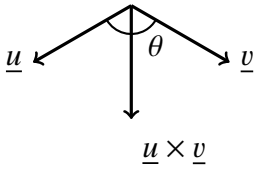
4.3 Cross product

There is another "multiplication" of two vectors, called the cross product. However, it is special to three dimensions.

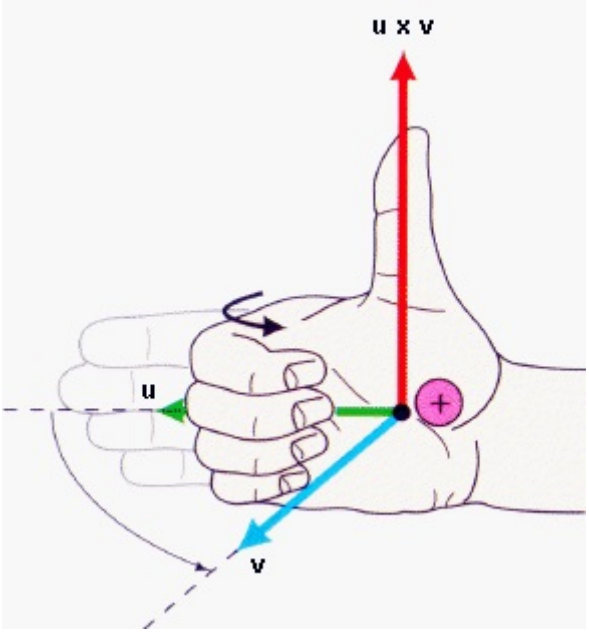
Definition. The cross product between two three dimensional vectors $\underline{u}$ and $\underline{v}$ is a vector with length $|\underline{u}||\underline{v}| \sin\theta$ where θ is the planar angle between $\underline{u}$ and $\underline{v}$, and in the direction orthogonal to both $\underline{u}$ and $\underline{v}$ according to the right hand rule.



Correct



Wrong

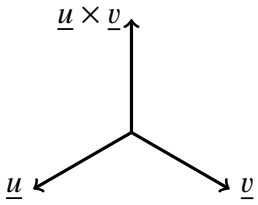


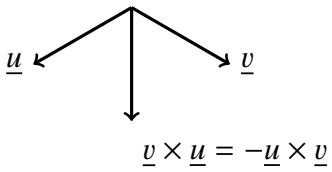
Taken from `sycode.com` with the explanation "flatten the right hand, extend it in the direction of $\underline{u}$, and then curling the fingers in the direction that the angle $\underline{v}$ makes with $\underline{u}$." The thumb then points in the direction of $\underline{u} \times \underline{v}$.

Some consequences of this geometric definition

1. Note that the right hand rule implies

$$\underline{u} \times \underline{v} = -\underline{v} \times \underline{u}$$





2. If $\underline{v} = \lambda \underline{u}$ then $\theta = 0$ if $\lambda > 0$ and $\theta = \pi$ is $\lambda < 0$. In both cases, $\sin(\theta) = 0$ so $\underline{v} \times \underline{u} = \underline{0}$. The way to think of this is that if $\underline{u}$ and $\underline{v}$ are parallel (that is, they point in either the same direction or in the opposite direction) then their cross product is the zero vector.

3. This implies $\underline{i} \times \underline{i} = \underline{j} \times \underline{j} = \underline{k} \times \underline{k} = \underline{0}$.

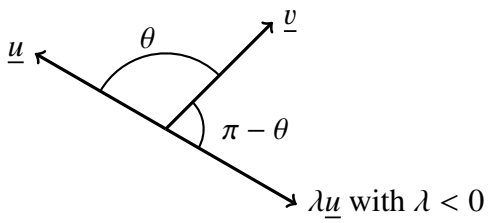
4. Recall that the three vectors $\underline{i}, \underline{j}, \underline{k}$ are chosen with the right hand rule, are of length 1 and are at mutual right angles. Since $\sin(\pi/2) = 1$ this implies $\underline{i} \times \underline{j} = \underline{k}$. Similarly $\underline{j} \times \underline{k} = \underline{i}$ and $\underline{k} \times \underline{i} = \underline{j}$.

5. We have for any real λ

$$(\lambda \underline{u}) \times \underline{v} = \lambda(\underline{u} \times \underline{v}) = \underline{u} \times (\lambda \underline{v})$$

Proof. If $\lambda > 0$ then $\lambda \underline{u}$ points in the same direction as $\underline{u}$ so the right hand rule points in the same direction, and thus $(\lambda \underline{u}) \times \underline{v}$ points in the same direction as $\lambda(\underline{u} \times \underline{v})$. Furthermore, note that the planar angle between $\lambda \underline{u}$ and $\underline{v}$ is the same as between $\underline{u}$ and $\underline{v}$. Since the length of $(\lambda \underline{u})$ equals $\lambda |\underline{u}|$, we see that $|(\lambda \underline{u}) \times \underline{v}| = \lambda |\underline{u}||\underline{v}| \sin(\theta) = \lambda |\underline{u} \times \underline{v}|$. Since the vectors point in the same direction and have the same length, they are equal. A similar argument works for the second equality in the case when $\lambda > 0$.

If $\lambda < 0$ then $\lambda\underline{u}$ points in the opposite direction to $\underline{u}$. Therefore the righthand rule means $(\lambda\underline{u})\times\underline{v}$ points in the opposite direction to $\underline{u}\times\underline{v}$, and therefore in the same direction as $\lambda(\underline{u}\times\underline{v})$. It also means that if the planar angle between $\underline{u}$ and $\underline{v}$ is θ, then the planar angle between $\lambda\underline{u}$ and $\underline{v}$ is $\pi-\theta$.



Since $\sin(\pi-\theta)=\sin(\theta)$ we have the length of $(\lambda\underline{u})\times(v)$ equals $|\lambda||\underline{u}||\underline{v}|\sin\theta = -\lambda|\underline{u}||\underline{v}|\sin\theta$. This is also the length of $\lambda(\underline{u}\times\underline{v})$ since $\lambda<0$. So the vectors have the same length and point in the same direction, which shows that $(\lambda\underline{u})\times\underline{v}=\lambda(\underline{u}\times\underline{v})$. A similar argument works for the second equality. □

6. We have the distributive law
$$\underline{u}\times(\underline{v}+\underline{w}) = \underline{u}\times\underline{v}+\underline{u}\times\underline{w}$$

Proof. Omitted. □

Similar to the dot product, the distributive law has a very important consequence in terms of evaluating the cross product when the vectors are written in their coordinates.

Lemma 4.5 (Coordinate version of the cross product)**.** *Let* $\underline{u}=a_1\underline{i}+b_1\underline{j}+c_1\underline{k}$ *and* $\underline{v}=a_2\underline{i}+b_2\underline{j}+c_2\underline{k}$. *Then*
$$\underline{u}\times\underline{v} = (b_1c_2-c_1b_2)\underline{i}+(c_1a_2-a_1c_2)\underline{j}+(a_1b_2-b_1a_2)\underline{k}$$

Proof. We have
$$\begin{aligned}\underline{u}\times\underline{v} &= \left(a_1\underline{i}+b_1\underline{j}+c_1\underline{k}\right)\times\left(a_2\underline{i}+b_2\underline{j}+c_2\underline{k}\right)\\ &= a_1\underline{i}\times\left(a_2\underline{i}+b_2\underline{j}+c_2\underline{k}\right)+b_1\underline{j}\times\left(a_2\underline{i}+b_2\underline{j}+c_2\underline{k}\right)+c_1\underline{k}\times\left(a_2\underline{i}+b_2\underline{j}+c_2\underline{k}\right)\end{aligned}$$
which follows from using the distributive law three times in the first term. Expanding each term out using the distributive law (this time in the second term),
$$\begin{aligned}\underline{u}\times\underline{v} &= a_1a_2\underline{i}\times\underline{i}+a_1b_2\underline{i}\times\underline{j}+a_1c_2\underline{i}\times\underline{k}\\ &\quad+b_1a_2\underline{j}\times\underline{i}+b_1b_2\underline{j}\times\underline{j}+b_1c_2\underline{j}\times\underline{k}\\ &\quad+c_1a_2\underline{k}\times\underline{i}+c_1b_2\underline{k}\times\underline{j}+c_1c_2\underline{k}\times\underline{k}\\ &= a_1b_2\underline{k}-a_1c_2\underline{j}-b_1a_2\underline{k}+b_1c_2\underline{i}+c_1a_2\underline{j}-c_1b_2\underline{i}\\ &= (b_1c_2-c_1b_2)\underline{i}+(c_1a_2-a_1c_2)\underline{j}+(a_1b_2-b_1a_2)\underline{k}\end{aligned}$$
where the last line follows from $\underline{i}\times\underline{i}=0$, $\underline{i}\times\underline{j}=\underline{k}$, $\underline{j}\times\underline{i}=-\underline{k}$ and similar. □

Example. Find $\underline{u}\times\underline{v}$ for $\underline{u}=\underline{i}+2\underline{j}+3\underline{k}$ and $\underline{v}=4\underline{i}+5\underline{j}+6\underline{k}$.

Solution. We have
$$\begin{aligned}\underline{u}\times\underline{v} &= (2\times6-3\times5)\underline{i}+(3\times4-1\times6)\underline{j}+(1\times5-2\times4)\underline{k}\\ &= -3\underline{i}+6\underline{j}-3\underline{k}\end{aligned}$$

Example. Find a unit vector perpendicular to $\underline{i} + 2\underline{j} + 3\underline{k}$ and $-\pi\underline{i} + \pi\underline{j} + 2\pi\underline{k}$.

Solution. We have their cross product equals

$$(2 * 2\pi - 3 * \pi)\underline{i} + (3 * (-\pi) - 1 * 2\pi)\underline{j} + (1 * \pi - 2 * (-\pi))\underline{k} = \pi\underline{i} - 5\pi\underline{j} + 3\pi\underline{k}$$

One can easily check that this vector is orthogonal to both the vectors by finding its dot product with them.

The question asks for a unit vector, so we divide by its length, which is

$$\sqrt{\pi^2 + 25\pi^2 + 9\pi^2}$$

Thus an orthogonal unit vector is

$$\frac{1}{\sqrt{35}}(\underline{i} - 5\underline{j} + 3\underline{k})$$

Note that an equally valid answer is

$$\frac{1}{\sqrt{35}}(-\underline{i} + 5\underline{j} - 3\underline{k})$$

Remark. Looking ahead to the next chapter on matrices (or for those who have seen determinants of 3×3 matrices before), note that the cross product has a nice interpretation: If $\underline{u} = a_1\underline{i} + b_1\underline{j} + c_1\underline{k}$ and $\underline{v} = a_2\underline{i} + b_2\underline{j} + c_2\underline{k}$, then

$$\begin{aligned}\underline{u} \times \underline{v} &= \det\begin{pmatrix} \underline{i} & \underline{j} & \underline{k} \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix} \\ &= \underline{i}\det\begin{pmatrix} b_1 & c_1 \\ b_2 & c_2 \end{pmatrix} - \underline{j}\det\begin{pmatrix} a_1 & c_1 \\ a_2 & c_2 \end{pmatrix} + \underline{k}\det\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\end{aligned}$$

4.4 Vector equations of lines

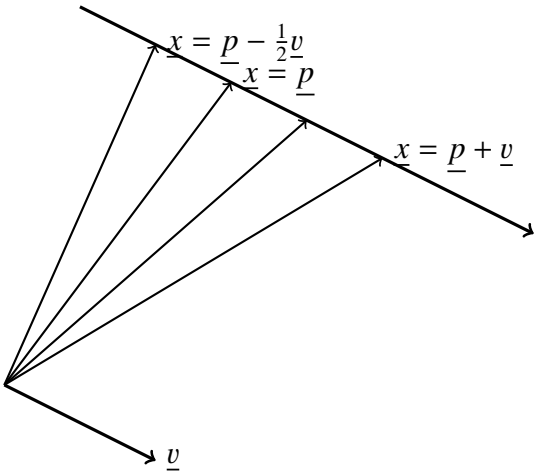
The vector equation of a line

Suppose we have a line in space through the point P (with position vector $\underline{p}$) and parallel to the vector $\underline{v}$. A vector equation for the line expresses the position vector $\underline{x}$ of an arbitrary point X on the line in terms of the vectors $\underline{p}$ and $\underline{v}$.

$$\underline{x} = \underline{p} + t\underline{v} \quad \text{for} \quad t \in \mathbb{R}$$

"To get to X, first travel to P then go along t units in the $\underline{v}$ direction."

Note that $\underline{p}$ and $\underline{v}$ are fixed, and $\underline{x}$ varies as t varies.



Example. Find the vector equation of the line going through the point $\underline{p} = 3\underline{i} + 4\underline{j}$ in the direction $\underline{v} = 2\underline{i} - \underline{j} + 3\underline{k}$.

Solution. The vector equation of the line is

$$\begin{aligned}\underline{x} &= \underline{p} + t\underline{v} = 3\underline{i} + 4\underline{j} + 2t\underline{i} - t\underline{j} + 3t\underline{k} \\ &= (3 + 2t)\underline{i} + (4 - t)\underline{j} + 3t\underline{k}\end{aligned}$$

The parametric equation of a line

Note that every point $\underline{x}$ on the line can be written as $\underline{x} = x\underline{i} + y\underline{j} + z\underline{k}$ for some $x, y, z \in \mathbb{R}$. Therefore one can write an equation of a line as

$$\begin{cases} x = p_1 + tv_1 \\ y = p_2 + tv_2 \\ z = p_3 + tv_3 \end{cases} \qquad \text{for } t \in \mathbb{R}$$

This is known as the parametric coordinate equation of the line. Here p_1, p_2, p_3 are the coordinates of a point on the line, and v_1, v_2, v_3 are the coordinates of the direction of the line.

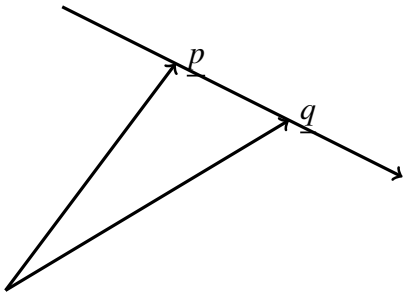
Example. The parametric coordinate equation of the line in the previous example is

$$\begin{aligned} x &= 3 + 2t \\ y &= 4 - t \\ z &= 3t \end{aligned}$$

Describing a line going through two points

An alternative way to describe a line is to give two points, P and Q, the line runs through. Recall that the direction vector $\overrightarrow{PQ} = \underline{q} - \underline{p}$, and thus the line is the line that passes through P in the direction $\overrightarrow{PQ}$. That is,

$$\underline{x} = \underline{p} + t(\underline{q} - \underline{p}) = t\underline{q} + (1 - t)\underline{p} \quad \text{for } t \in \mathbb{R}$$



Example. Find the equation of the line going through the points $\underline{p} = 7\underline{i} + 2\underline{j} + 6\underline{k}$ and $\underline{q} = \underline{i} + 5\underline{j} - 3\underline{k}$.

Solution. The vector equation of the line is

$$\underline{x} = t(\underline{i} + 5\underline{j} - 3\underline{k}) + (1 - t)(7\underline{i} + 2\underline{j} + 6\underline{k})$$

and the parametric coordinate equation of the line is

$$\begin{aligned} x &= t + 7(1 - t) = 7 - 6t \\ y &= 5t + 2(1 - t) = 2 + 3t \\ z &= -3t + 6(1 - t) = 6 - 9t \end{aligned}$$

Showing a particular point lies on a line

Example. Show that the point $3\underline{i} + 4\underline{j}$ lies on the line in the previous example.

Solution. To show the point $3\underline{i} + 4\underline{j}$ lies on the line we need to find a particular value of t such that

$$\begin{aligned} 3 &= 7 - 6t \\ 4 &= 2 + 3t \\ 0 &= 6 - 9t \end{aligned}$$

Note that (solving the simultaneous equations) $t = \frac{2}{3}$ works, so the point lies on the line.

Remark. Recall the first example, where we found the equation of a line going through $3\underline{i} + 4\underline{j}$ in the direction of $\underline{v} = 2\underline{i} - \underline{j} + 3\underline{k}$. Note that the direction vector of this line is $-6\underline{i} + 3\underline{j} - 9\underline{k}$ and that this equals $-3\underline{v}$. Therefore these two lines are exactly the same line!

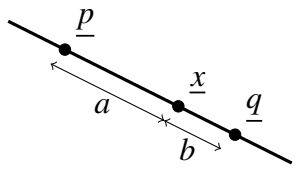
In other words, to show two vector equations $\underline{x} = \underline{p_1} + t\underline{v_1}$ for $t \in \mathbb{R}$ and $\underline{x} = \underline{p_2} + s\underline{v_2}$ for $s \in \mathbb{R}$ are the same line, show that p_1 lies on the second line (that is, there exists an s' such that $\underline{p_1} = \underline{p_2} + s'\underline{v_2}$ and that $\underline{v_1} = \lambda\underline{v_2}$ for some $\lambda \in \mathbb{R}$.

Dividing a line segment into two pieces of varying proportion

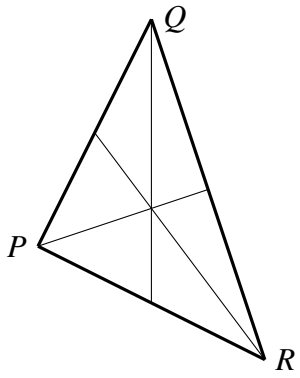
Dividing the line segment from P and Q in the ratio $a : b$ means finding the point between P and Q (call it X) such that $\frac{|PX|}{|XQ|} = \frac{a}{b}$. Note that $\underline{p} + t(\underline{q} - \underline{p}) = t\underline{q} + (1 - t)\underline{p}$ for $0 \leq t \leq 1$ yields the points on the line segment between P and Q. If we put $t = \frac{a}{a+b}$ into this equation we get the point

$$\underline{x} = \underline{p} + \frac{a}{a+b}(\underline{q} - \underline{p}) = \frac{b}{a+b}\underline{p} + \frac{a}{a+b}\underline{q}$$

Note that $|\underline{x} - \underline{p}| = \frac{a}{a+b}|\underline{q} - \underline{p}|$ and $|\underline{x} - \underline{q}| = \frac{b}{a+b}|\underline{p} - \underline{q}|$. That is, this point divides the line segment in the ratio $a : b$.



Example. Suppose P, Q, R are three points. The three lines connecting a point to the midpoint of the other two points all intersect at a common point, called the centroid of the triangle.



Solution. The midpoint of QR is the point that divides the line segment in the ratio $1 : 1$, which is

$$\frac{1}{2}\underline{q} + \frac{1}{2}\underline{r}$$

Thus the line from P to the midpoint of QR is

$$\underline{x_1} = t_1\underline{p} + (1 - t_1)(\tfrac{1}{2}\underline{q} + \tfrac{1}{2}\underline{r}) \quad \text{for } 0 \le t_1 \le 1$$

Similarly the other three lines are

$$\underline{x_2} = t_2\underline{q} + (1 - t_2)(\tfrac{1}{2}\underline{p} + \tfrac{1}{2}\underline{r}) \quad \text{for } 0 \le t_2 \le 1$$

$$\underline{x_3} = t_3\underline{r} + (1 - t_3)(\tfrac{1}{2}\underline{p} + \tfrac{1}{2}\underline{q}) \quad \text{for } 0 \le t_3 \le 1$$

Note that if you put $t_1 = t_2 = t_3 = \frac{1}{3}$ you get

$$\underline{x_1} = \underline{x_2} = \underline{x_3} = \frac{1}{3}\left(\underline{p} + \underline{q} + \underline{r}\right)$$

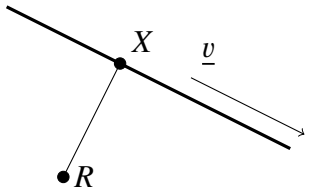
which shows there is a point lying on all three lines — the centroid.

The distance from a point to a line

The distance from a point R to a line is the minimum of $|\underline{x} - \underline{r}|$ where $\underline{x}$ lies on the the line. Geometrically it is clear this point will be when the vector from $\underline{r}$ to $\underline{x}$ is orthogonal to the direction of the line. That is, we are searching for an $\underline{x}$ that lies on the line such that

$$(\underline{x} - \underline{r}) \cdot \underline{v} = 0$$

where $\underline{v}$ is a vector parallel to the line.



Recall that a general point on the line can be written as $\underline{p} + t\underline{v}$ for some $t \in \mathbb{R}$ where $\underline{p}$ is the position vector of a point on the line. Thus we wish to find the specific $t \in \mathbb{R}$ such that

$$(\underline{p} + t\underline{v} - \underline{r}) \cdot \underline{v} = 0$$

Expanding out the dot product, this is

$$(\underline{p} - \underline{r}) \cdot \underline{v} + t\underline{v} \cdot \underline{v} = 0$$

so

$$t = -\frac{(\underline{p} - \underline{r}) \cdot \underline{v}}{|\underline{v}|^2}$$

Plugging this value of t in shows the point on the line closes to R has position vector

$$\underline{p} - \frac{(\underline{p} - \underline{r}) \cdot \underline{v}}{|\underline{v}|^2}\underline{v}$$

The square of the minimal distance is therefore $|\underline{x} - \underline{r}|^2 = (\underline{x} - \underline{r}) \cdot (\underline{x} - \underline{r})$, or

$$\left(\underline{p} - \underline{r} - \frac{(\underline{p} - \underline{r}) \cdot \underline{v}}{|\underline{v}|^2}\underline{v}\right) \cdot \left(\underline{p} - \underline{r} - \frac{(\underline{p} - \underline{r}) \cdot \underline{v}}{|\underline{v}|^2}\underline{v}\right) = |\underline{p} - \underline{r}|^2 - 2\frac{((\underline{p} - \underline{r}) \cdot \underline{v})^2}{|\underline{v}|^2} + \left(\frac{(\underline{p} - \underline{r}) \cdot \underline{v}}{|v|^2}\right)^2 |\underline{v}|^2$$
$$= \frac{|\underline{p} - \underline{r}|^2|\underline{v}|^2 - ((\underline{p} - \underline{r}) \cdot \underline{v})^2}{|\underline{v}|^2}$$

Note that if θ is the planar angle between the vector $\underline{p} - \underline{r}$ and the vector $\underline{v}$ then the above equals

$$\frac{|\underline{p} - \underline{r}|^2|\underline{v}|^2(1 - \cos^2\theta)}{|\underline{v}|^2} = \frac{|\underline{p} - \underline{r}|^2|\underline{v}|^2 \sin^2\theta}{|\underline{v}|^2}$$
$$= \frac{|(\underline{p} - \underline{r}) \times \underline{v}|^2}{|\underline{v}|^2}$$

and taking the square root shows that the minimal distance of R to a line going through the point P in the direction $\underline{v}$ equals

$$\frac{|(\underline{p} - \underline{r}) \times \underline{v}|}{|\underline{v}|}$$

4.5 Vector equations of planes

The vector equation for a plane is similar to that for a line, but it has two parameters rather than 1. Given a point P with position vector $\underline{p}$ and two vectors $\underline{u}$ and $\underline{v}$ not lying in the same line (that is, not co-linear) then there is a plane that passes through P parallel to both $\underline{u}$ and $\underline{v}$.

The position vector of an arbitrary point X in the plane is

$$\underline{x} = \underline{p} + s\underline{u} + t\underline{v} \quad \text{for } s, t \in \mathbb{R}$$

This is known as the plane spanned by $\underline{u}$ and $\underline{v}$ going through $\underline{p}$.

Example. The plane going through $\underline{i} + \underline{j} + \underline{k}$ parallel to both $\underline{i} + 2\underline{j} + \underline{k}$ and $3\underline{i} - \underline{j} + 2\underline{k}$ is

$$\begin{aligned}\underline{x} &= (\underline{i} + \underline{j} + \underline{k}) + s(\underline{i} + 2\underline{j} + \underline{k}) + t(3\underline{i} - \underline{j} + 2\underline{k}) \\ &= (1 + s + 3t)\underline{i} + (1 + 2s - t)\underline{j} + (1 + s + 2t)\underline{k}\end{aligned}$$

for $s, t \in \mathbb{R}$.

Parametric equation for a plane

The parametric equations for this plane are

$$\begin{cases} x = 1 + s + 3t \\ y = 1 + 2s - t \\ z = 1 + s + 2t \end{cases}$$

Defining a plane containing three points

An alterative way to define a plane is to give three non-co-linear points lying on the plane, $\underline{p}$, $\underline{q}$, $\underline{r}$. Note that both $\underline{p} - \underline{r}$ and $\underline{q} - \underline{r}$ are vectors parallel to the plane, and by definition the point with position vector $\underline{r}$ lies in the plane, so the vector equation of the plane can be written as

$$\begin{aligned}\underline{x} &= \underline{r} + s(\underline{p} - \underline{r}) + t(\underline{q} - \underline{r}) \\ &= s\underline{p} + t\underline{q} + (1 - s - t)\underline{r}\end{aligned}$$

This can also be written in a more symmetric form as

$$\underline{x} = a\underline{p} + b\underline{q} + c\underline{r}$$

where $a, b, c \in \mathbb{R}$ such that $a + b + c = 1$.

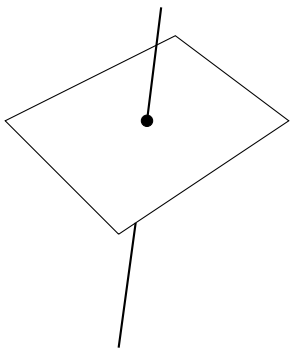
Example. Find the plane going through the points $\underline{p} = \underline{i} + 2\underline{j} + 3\underline{k}$, $\underline{q} = 3\underline{i} + 2\underline{j} + \underline{k}$ and $\underline{r} = \underline{i} + \underline{j} + \underline{k}$.

Solution. Note that $\underline{p} - \underline{r} = \underline{j} + 2\underline{k}$ and $\underline{q} - \underline{r} = 2\underline{i} + \underline{j}$ we have that the plane is

$$\underline{x} = (1 + 2t)\underline{i} + (1 + t + s)\underline{j} + (1 + 2s)\underline{k} \quad \text{for} \quad s, t \in \mathbb{R}$$

Finding the intersection of a line with a plane

Unless a line is parallel to the plane, it will intersect the plane at a point. To find that point solve the three relevant simultaneous equations (so find a point on the line (one parameter) that is also a point on the plane (two parameters).



Example. Find the point of intersection between the line through $3\underline{i}+6\underline{j}+2\underline{k}$ and parallel to $-\underline{i}+2\underline{j}-\underline{k}$, and the plane through $\underline{i}+\underline{j}+\underline{k}$ and parallel to $\underline{i}+2\underline{j}+\underline{k}$ and $3\underline{i}-\underline{j}+2\underline{k}$

Solution. We write the line and plane in parametric form: The line is

$$\begin{cases} x = 3 - t \\ y = 6 + 2t \quad \text{for } t \in \mathbb{R} \\ z = 2 - t \end{cases}$$

and the plane is

$$\begin{cases} x = 1 + s + 3r \\ y = 1 + 2s - r \quad \text{for } s, r \in \mathbb{R} \\ z = 1 + s + 2r \end{cases}$$

The common point is when

$$3 - t = 1 + s + 3r \tag{1}$$
$$6 + 2t = 1 + 2s - r \tag{2}$$
$$2 - t = 1 + s + 2r \tag{3}$$

Note that

$$\begin{cases} 2\times(1)+(2) \implies & 12 = 3 + 4s + 5r \\ 2\times(3)+(2) \implies & 10 = 3 + 4s + 3r \end{cases} \implies 2 = 2r \implies r = 1 \text{ and } s = 1$$

and hence $t = -2$. Thus the point in common (the intersection) is at $x = 5$, $y = 2$ and $z = 4$.

The normal vector to a plane

Another way to define a plane is by noticing that (in 3d) there is exactly one line which is perpendicular to the plane. A vector parallel to this line is called a *normal vector*. If it has unit length, it is called a *unit normal*. Thus if $\hat{\underline{n}}$ is a unit normal to a plane, there is exactly one other unit normal to that plane, namely $-\hat{\underline{n}}$.

If $\underline{p}$ is the position vector of a known point in the plane, and $\underline{x}$ is a position vector of an arbitrary point in the plane, then $\underline{x} - \underline{p}$ is parallel to the plane, and thus orthogonal to the normal. Therefore the equation of a plane can be given as

$$(\underline{x} - \underline{p}) \cdot \underline{n} = 0$$

or

$$\underline{x} \cdot \underline{n} = \underline{p} \cdot \underline{n}$$

Note that $\underline{p} \cdot \underline{n}$ is a number, not a vector.

The Cartesian equation of a plane

Writing $\underline{x} = x\underline{i} + y\underline{j} + z\underline{k}$ and $\underline{n} = a\underline{i} + b\underline{j} + c\underline{k}$ we have that

$$ax + by + cz = \underline{p} \cdot \underline{n}$$

This is known as the Cartesian equation of a plane.

Example. The plane passing through the point $2\underline{i} + \underline{j} + 3\underline{k}$ with normal vector $\underline{n} = \underline{i} - \underline{j} + 2\underline{k}$ has Cartesian equation

$$x - y + 2z = 2 * 1 + 1 * (-1) + 3 * 2 = 7$$

Example. The plane with Cartesian equation $2x + 2y - z = 9$ has a normal vector $\underline{n} = 2\underline{i} + 2\underline{j} - \underline{k}$. Note that $|\underline{n}| = \sqrt{4 + 4 + 1} = 3$ so its unit normal is either $\hat{\underline{n}} = \frac{2}{3}\underline{i} + \frac{2}{3}\underline{j} - \frac{1}{3}\underline{k}$ or $\hat{\underline{n}} = -\frac{2}{3}\underline{i} - \frac{2}{3}\underline{j} + \frac{1}{3}\underline{k}$.

Recall that the vector equation for a plane has two non-co-linear vectors, $\underline{u}$ and $\underline{v}$, parallel to the plane. Taking their cross product yields a vector orthogonal to them both, and thus orthogonal to the plane. Therefore the Cartesian equation of the plane that is spanned by $\underline{u}$ and $\underline{v}$ and goes through the point $\underline{p}$ is

$$\underline{x} \cdot (\underline{u} \times \underline{v}) = \underline{p} \cdot (\underline{u} \times \underline{v})$$

Example. Find the Cartesian equation for the plane through the point P with position vector $\underline{i} - \underline{k}$ and parallel to $\underline{u} = -\underline{i} + 2\underline{j} - \underline{k}$ and $\underline{v} = 2\underline{i} - \underline{j} + 3\underline{k}$.

Solution. We have

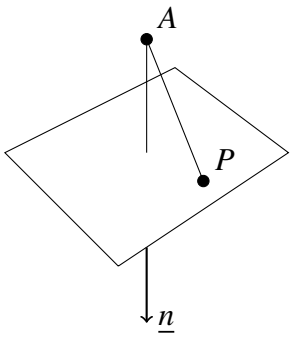
$$\begin{aligned}\underline{u} \times \underline{v} &= (2 * 3 - (-1) * (-1))\underline{i} + ((-1) * 2 - (-1) * 3)\underline{j} + ((-1) * (-1) - 2 * 2)\underline{k} \\ &= 5\underline{i} + \underline{j} - 3\underline{k}\end{aligned}$$

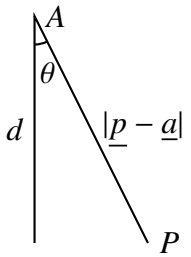
so the Cartesian equation is

$$5x + y - 3z = 1 * 5 + 0 * 1 + (-1) * (-3) = 8$$

Distance from a point to a plane

The distance from a point A to the plane with equation $\underline{x} \cdot \underline{n} = \underline{p} \cdot \underline{n}$ can be found by noticing that if θ is the planar angle between $\overrightarrow{AP}$ and the normal vector, then the distance from A to the plane is $\left||\underline{p} - \underline{a}| \cos\theta\right|$.





That is,

$$\text{distance from } A \text{ to plane } = \frac{1}{|\underline{n}|}\left|(\underline{p} - \underline{a}) \cdot \underline{n}\right| = \frac{\left|\underline{p} \cdot \underline{n} - \underline{a} \cdot \underline{n}\right|}{|\underline{n}|}$$

(the absolute value is needed because $(\underline{p} - \underline{a}) \cdot \underline{n}$ might be negative).

Example. Find the distance of $A = (2, -1, 3)$ to the plane $x - y + z = 4$.

Solution. The plane has normal vector $\underline{n} = \underline{i} - \underline{j} + \underline{k}$, with $|\underline{n}| = \sqrt{3}$. Furthermore, $\underline{p} \cdot \underline{n} = 4$ so we have the minimal distance is

$$\frac{|4 - 6|}{\sqrt{3}} = \frac{2}{\sqrt{3}}$$

Further reading, and where next?

The mathematics of using vectors is obviously important in the second year course *Vector Calculus*, and results from this course will be used in the third year *fluid dynamics* modules, and also within *electromagnetism*.

Using vectors to identify points in space, and their dynamics is fundamental to the development of *Classical mechanics* seen both next term and also in the second year in the *Applied Mathematics* stream.

Although this course has been concrete in how it has developed the notion of vectors, they have a rich mathematical underpinnings, that will be explored in more depth as abstract vector spaces in the second year course *Linear Algebra* (which will also develop the mathematics underpinning the material on matrices found in the next chapter).