MT181 NUMBER SYSTEMS

MARK WILDON

These notes are intended to give the logical structure of the course; proofs and further examples and remarks will be given in lectures. Further installments will be issued as they are ready. All handouts and problem sheets will be put on Moodle.

These notes are based in part on notes for similar courses run by Prof. R. Schack and Prof. P. J. Cameron. I would very much appreciate being told of any corrections or possible improvements.

You are warmly encouraged to ask questions in lectures, and to talk to me after lectures and in my office hours. I am also happy to answer questions about the lectures or problem sheets by email. My email address is `mark.wildon@rhul.ac.uk`.

Lectures in BLT1: Tuesday 1pm, Thursday 9am and Friday 2pm.

Office hours in McCrea 240: Monday 4pm, Wednesday 10am and Friday 4pm.

Solution class for 171/181: Friday 11am, from week 3 (18th October)

Workshops for 171/181: Mondays, from week 2 (8th October)

Date: First term 2013/14.

Number systems

This course is on the fundamental number systems used in mathematics: the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the integers modulo a prime $\mathbb{Z}_p$, and others. In parallel, we will develop the basic language of pure mathematics: sets, functions, relations, propositions, etc. There will be many interesting examples.

Outline.

(A) Sets, functions and complex numbers

§1 *Introduction*: sets of numbers and basic operations on sets.

§2 *Functions*: injective, surjective, bijective, compositions and inverse functions.

§3 *Complex numbers*: cartesian, polar and exponential forms. Solving equations in $\mathbb{C}$.

(B) Natural numbers and induction

§4 *Induction:* examples. Σ notation.

§5 *Prime numbers:* division, unique factorization, infinitely many primes. Binary and other bases.

(C) Logic and sets

§6 *Logic:* more on proofs, propositional logic, truth tables.

§7 *Sets:* further results including principle of inclusion and exclusion.

(D) Integers and rings

§8 *Integers:* Euclid's algorithm, congruences.

§9 *Relations:* equivalence relations, $\mathbb{Z}_n$.

§10 *Rings:* axioms and basic properties of rings, polynomial rings.

Recommended Reading.

[1] *How to think like a mathematician*. Kevin Houston, Cambridge University Press, 2009.

[2] *A concise introduction to pure mathematics*. Martin Liebeck, Chapman and Hall, 2000.

[3] *Discrete Mathematics*. Norman L. Biggs, Oxford University Press, 2002.

As part of problem sheets you will be asked to do some reading from *How to think like a mathematician*. The library has copies of this book on short-term loan. You can also buy it from the College Bookshop: they will match the price on amazon.co.uk.

Problem sheets. There will be 8 marked problem sheets; the first will be due in on Tuesday 8th October. To encourage you to work hard during the term, **each problem sheet is worth 1.25% of your overall grade**. Note that this mark is awarded for *any reasonable attempt* at the sheet. (There is a link on Moodle to the document explaining this policy in more detail.) Answers to problem sheets will be put on Moodle after the submission deadline.

Moodle. All handouts, problem sheets and answers will be posted on Moodle. You should find a link under 'My courses', but if not, go to `moodle.rhul.ac.uk/course/view.php?id=407`.

Exercises in these notes. Exercises set in these notes are mostly simple tests that you are following the material. Some will be used for quizzes in lectures. Doing the others will help you to review your notes.

Optional questions and extras. The 'Bonus question' at the end of each problem sheet and any 'extras' in these notes are included for interest only, and to show you some mathematical ideas beyond the scope of this course. You should not worry if you find them difficult.

Study skills. At the end of some sections you will find some well-meant advice about learning mathematics.

If you can do the compulsory questions on problem sheets, know the definitions and main results from lectures, and can prove the theorems whose proofs are marked as examinable, then you should do very well in the examination.

Part A: Sets, functions and complex numbers

1. Introduction: Sets of numbers

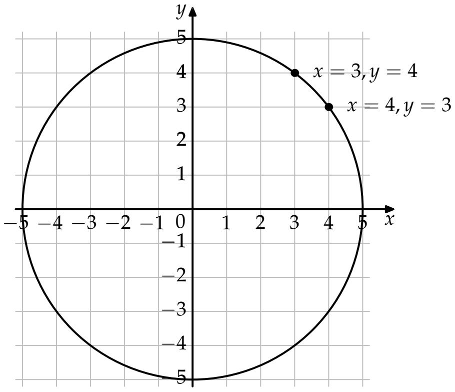
A MOTIVATING EXAMPLE. One of the unifying ideas in this course is solving equations. I hope we can all agree this is an useful and interesting thing to do. For example, consider the equation

$$x^2 + y^2 = 25.$$

How many solutions are there?

The answer depends on what sort of numbers x and y can be. If x and y can be any real numbers, then there are infinitely many solutions and they form a circle in the plane. But maybe we are only interested in solutions where x and y have to be one of the *natural numbers* $1, 2, 3, 4, \ldots$, Then the only solutions are $x = 3$, $y = 4$ and $x = 4$, $y = 3$.

Another important idea is mathematical proof. Imagine someone says to you 'Okay, I can see that $x = 3$, $y = 4$ and $x = 4$, $y = 3$ are two solutions using natural numbers. But why aren't there any more?'. How could you convince him or her, beyond any doubt, that you had found every solution?



Exercise 1.1. How many solutions to the equation $x^2 + y^2 = 25$ are there if x and y can be any integer? (The *integers* are the numbers $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$.)

SETS. A *set* is any collection of objects. These objects are called the *members* or *elements* of the set. One way to specify a set is to put a list of its elements inside a pair of curly braces. For example $\{2, 3, 5, 7, 11, 13\}$ is a set. Alternatively we may describe a set in words. For example,

the set of prime numbers that are less than or equal to 13

is another way to specify the set $\{2, 3, 5, 7, 11, 13\}$. (In Part B we will see why 1 is not considered to be a prime number.)

If X is a set and x is an element of X then we write $x \in X$. (This can be read as 'x is in X', or 'X contains x'.) If y is not an element of X then we write $y \notin X$. For example, $7 \in \{2,3,5,7,11,13\}$ and $8 \notin \{2,3,5,7,11,13\}$.

Exercise 1.2. True or false?

(i) 29 is a member of the set of prime numbers;

(ii) 87 is a member of the set of prime numbers;

(iii) $\{2,3,5,7,11\} = \{5,7,11,2,3\}$.

SETS OF NUMBERS. We write $\mathbb{N}$ for the set of natural numbers:

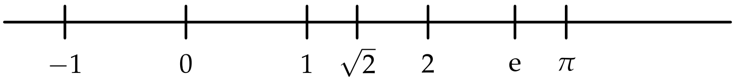
$$\mathbb{N} = \{1,2,3,4,\ldots\}.$$

We write $\mathbb{Z}$ for the set of integers:

$$\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}.$$

A number r/s with $r \in \mathbb{Z}$, $s \in \mathbb{Z}$ and $s \neq 0$ is said to be *rational*. We write $\mathbb{Q}$ for the set of rational numbers. Finally we write $\mathbb{R}$ for the set of real numbers.

It is not possible to write down all the elements of $\mathbb{R}$—if you are taking 110 *From Euclid to Mandelbrot* you will find out why, later in this course—but we can still visualize $\mathbb{R}$ as the real number line. Some important real numbers are marked below.



It is an important fact that there are real numbers that are not rational numbers. For example $\sqrt{2} \notin \mathbb{Q}$. We say that such numbers are *irrational*. We will prove that $\sqrt{2}$ is irrational later in this course. So what sort of numbers are rational?

Example 1.3. Let $x = 3.123123123\ldots$ where the three dots indicate that the repeated blocks of 123 continue forever. When a decimal number is multiplied by a power of 10, the decimal point moves to the right. So

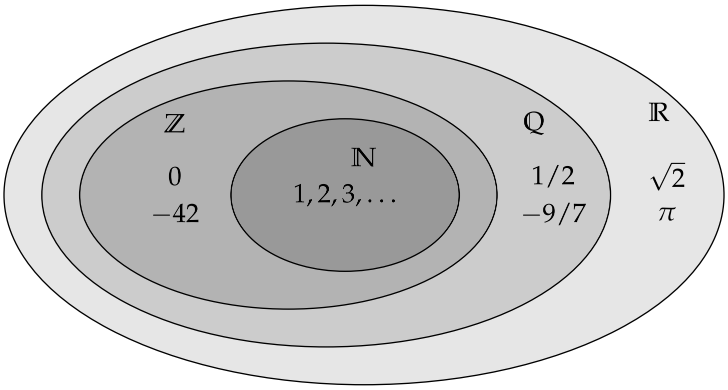
$$\begin{aligned} 1000x &= 3123.123123\ldots \\ x &= \quad\; 3.123123\ldots \\ \implies 999x &= 3120. \end{aligned}$$

Hence $x = 3120/999$ and so x is rational. Here '$\implies$' means 'implies'. It indicates that the third line is deduced from the first two, and makes the structure of the argument clearer.

In fact the rational numbers are exactly the real numbers whose decimal expansions are either finite, for example $3.123 = 3123/1000$, or recurring, as in Example 1.3.

Exercise 1.4. Find a simple expression for $1.153846153846\ldots$, where the repeated block is made of 6 digits.

The diagram below illustrates the sets of numbers we have seen so far. Note that a set contains all the numbers in the sets drawn inside it. For example, It is therefore entirely correct to say that 1 is a real number, or that -1 is a rational number.



CLOSURE AND EQUATION SOLVING. One important property of the natural numbers, which I hope you will agree is obviously true, is that if $m, n \in \mathbb{N}$ then $m + n \in \mathbb{N}$ and $mn \in \mathbb{N}$. To refer to these properties in a concise way we make the following definition.

Definition 1.5. Let X be a set of numbers. We say that X is

- *closed under addition* if $x + y \in X$ whenever $x \in X$ and $y \in X$;
- *closed under multiplication* if $xy \in X$ whenever $x \in X$ and $y \in X$;
- *closed under subtraction* if $x - y \in X$ whenever $x \in X$ and $y \in X$;
- *closed under division* if $x/y \in X$ whenever $x \in X$, $y \in X$ and $y \neq 0$.

Thus $\mathbb{N}$ is closed under addition and multiplication. But $\mathbb{N}$ is not closed under subtraction: for instance $1 \in \mathbb{N}$, $2 \in \mathbb{N}$, but $1 - 2 \notin \mathbb{N}$. There are of course many other examples you could take, but to show that $\mathbb{N}$ is not closed under subtraction, one is enough.

Exercise 1.6. Is $\mathbb{N}$ closed under division? Is $\mathbb{Z}$ closed under division? Is $\mathbb{Q}$ closed under (i) addition, (ii) division?

There is a connection between the closure properties of a set and the equations that can be solved using numbers from that set.

For example, $1-2 \notin \mathbb{N}$, and correspondingly, the equation $1 = 2 + x$ has no solution in $\mathbb{N}$. Going the other way, the equation $3x = 4$ has no solution in $\mathbb{Z}$, and correspondingly, $\mathbb{Z}$ is not closed under division.

A PROOF. To show that $\mathbb{Q}$ is closed under addition we must show that for *any* $x, y \in \mathbb{Q}$, we have $x + y \in \mathbb{Q}$. A rigorous proof must start from the definition of $\mathbb{Q}$.

Proof that $\mathbb{Q}$ *is closed under addition.* Let $x \in \mathbb{Q}$ and $y \in \mathbb{Q}$. Since $x \in \mathbb{Q}$ there exist $r, s \in \mathbb{Z}$ such that $s \neq 0$ and $x = r/s$. Since $y \in \mathbb{Q}$ there exist $t, u \in \mathbb{Z}$ such that $u \neq 0$ and $y = t/u$. Now

$$x + y = \frac{r}{s} + \frac{t}{u} = \frac{ru + st}{su}.$$

Hence $x + y = m/n$ where $m = ru + st$ and $n = su$. Since the integers are closed under addition and multiplication, we have $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$. Therefore $x + y$ is rational. □

Exercise 1.7. At the end this proof has one (easily fixed) gap. You might also object to it for other reasons. Come up with at least one objection.

SUBSETS. If X and Y are sets and every element of X is an element of Y, then we say that X is a *subset* of Y, and write $X \subseteq Y$. In symbols the condition $X \subseteq Y$ is

$$x \in X \implies x \in Y.$$

A natural number is a special kind of integer, an integer is a special sort of rational number, and a rational number is a special kind of real numbers. So $\mathbb{N}$ is a subset of $\mathbb{Z}$, and so on. In symbols:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

There is a special notation for defining subsets of a set. For example if Y is the set of prime numbers and

$$X = \{x \in Y : x \leq 13\}$$

then X is the set of prime numbers x such that $x \leq 13$. The set $\mathbb{Q}$ of rational numbers can be defined as

$$\mathbb{Q} = \{r/s : r \in \mathbb{Z}, s \in \mathbb{Z}, s \neq 0\}.$$

Example 1.8. Let

$$X = \{x \in \mathbb{R} : x \geq 2 + \sqrt{5}.\}$$
$$Y = \{x \in \mathbb{R} : x^2 - 4x + 1 \geq 2\}$$

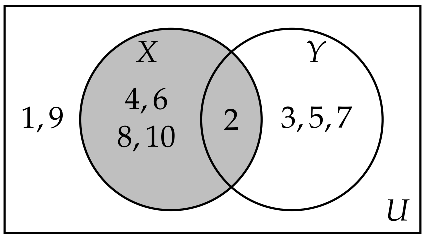
We will show that $X \subseteq Y$. Is it true that $X = Y$?

The symbol '$\Longleftrightarrow$' will be used in the proof: if A and B are mathematical statements then $A \iff B$ means that A implies B *and* B implies A. So A and B are either both true, or both false.

VENN DIAGRAMS. A *Venn diagram* is a diagram, like the one on page 6, that represents sets by regions of the plane. For example, the sets

$$U = \{1, 2, 3, \ldots, 9, 10\}$$
$$X = \{n \in U : n \text{ is even}\}$$
$$Y = \{n \in U : n \text{ is a prime number}\}$$

are shown in the Venn diagram below. The region representing X is shaded.



INTERSECTION, UNION, COMPLEMENT. Let X and Y be sets.

- The *intersection* of X and Y, written $X \cap Y$, is the set of elements that are in both X and Y.
- The *union* of X and Y, written $X \cup Y$, is the set of elements in at least one of X and Y.
- If X is a subset of a set U then we define the *complement of X in U* by $X' = \{y \in U : y \notin X\}$.

Exercise 1.9. Draw Venn diagrams representing $X \cap Y$, $X \cup Y$ and X'.

If X, Y and U are as above then $X \cap Y = \{2\}$ is the set of numbers in U that are both even and prime, and $X \cup Y = \{2, 3, 4, 5, 6, 7, 8, 10\}$ is the set of numbers in U that are either even or prime. The complement of X in U is $\{1, 3, 5, 7, 9\}$.

Intersections and unions of larger number of sets are defined as you would expect. For example, if $Z = \{1, 2, 8\}$ then $X \cap Y \cap Z = \{2\}$

and $X \cup Y \cup Z = \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$. More sets can be formed by combining these operations. For example $Z \cap (X \cup Y) = \{2, 8\}$.

The first problem sheet has a Venn diagram showing all the ways in which three sets can meet. It is possible to draw Venn diagrams for larger number of sets, but they become hard to use. So we need to be able to reason about sets without the help of Venn diagrams.

Claim 1.10 (De Morgan's Laws)**.** *Let X and Y be subsets of a set U. Then*

(i) $(X \cup Y)' = X' \cap Y'$,

(ii) $(X \cap Y)' = X' \cup Y'$.

The proof of (ii) will be left to you on the first problem sheet.

EXTRAS. According to the definition at the start of this section, a set is any collection of objects. This is a workable definition for this course, but it can lead to problems. For example, should we allow

$P =$ the collection of people apart from me in this room

to be a set? Probably not, since two people in the same room will disagree about what members P has. We could try to avoid these ambiguities by saying that a set is any collection of *mathematical* objects, but this might just lead to arguments about whether an object is mathematical or not.

Another feature that may seem unintuitive, particularly if we are not fussy about what objects can go into a set, concerns sets with no elements. For example, define

$A = \{x \in \mathbb{R} : x^2 = -1\}$

$B =$ the set of people who walked on the moon in 2012.

Two sets are equal if they have the same elements, and clearly A and B have the same elements (none, in both cases). So $A = B$. But can it really be sensible to equate a set of people who walked on the moon with a set of real numbers?

Still more alarmingly, in 1901 Bertrand Russell found what seemed at the time like a paradox in set theory. He argued as follows. Suppose that there is a set U which has *every* set as a member. Let R be the subset of U containing all those sets that are not members of themselves. In symbols

$$R = \{X \in U : X \notin X\}.$$

The definition of R takes a lot of thinking about. Most sets are not members of themselves. For example, the set $\mathbb{N}$ of natural numbers $\{1, 2, 3, \ldots\}$ does not have $\mathbb{N}$ as a member. But U does contain itself, since U is a set, and U is supposed to contain every set. So U is an example of a set that is not in R.

Russell asked: is R a member of itself? Either possibility leads to a contradiction! Suppose that $R \notin R$. Then R is a set that does not contain itself so, *by definition of* R, R is a member of R, a contradiction.

Exercise 1.11. Deduce similarly that if $R \in R$ then $R \notin R$.

All these problems are avoided in modern axiomatic set theory by restricting the ways in which new sets can be formed. In particular, there is no set U containing all other sets, and Russell's set R cannot be defined.

Modern axiomatic set theory works very well as a foundation for mathematics. From this point of view, *everything is a set*. For example, the number 0 is, by definition, the empty set, which we write $\varnothing$. The number 1 is, by definition, $\{\varnothing\}$; this is the set whose single element is the empty set. So $1 = \{0\}$. The number 2 is, by definition $\big\{\varnothing, \{\varnothing\}\big\}$. So $2 = \{0, 1\}$. And so on. Integers, rational numbers and real numbers are all encoded as special kinds of sets.

The Grelling–Nelson paradox is a version of Russell's paradox using English words rather than sets. You will be able to find out about it on the web.

STUDY SKILLS (LECTURES). To get the most out of your lectures, you need to put in some work. Even if printed notes are issued, you should make your own notes during the lecture. This helps to keep the mind focused, and makes sure that you get a complete record.

Don't just copy things down without thinking: you should try to follow each step in a proof or example as it appears on the board. Most people take in more when they see the mathematics created in front of them. Of course you learn most by actually using the mathematics to solve problems. So you need to do problem sheets as well as attend lectures.

You will find you pay more attention at some points than others. Lectures are demanding, and it is impossible to concentrate fully at all times. But please *do not* conduct non-mathematical conversations during a lecture. You will distract your colleagues who may be trying hard to follow the lecture.

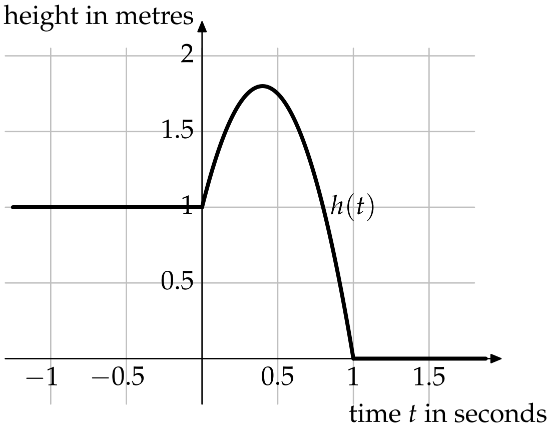
Please make the effort to participate in quizzes, and to ask questions in lectures. (Most lecturers will pause to invite questions.) In 181 lectures, you are welcome to ask questions at any time.

2. FUNCTIONS

MOTIVATION. For a long time, the common way to think of a function was as something expressed by a formula, such as $f(x) = x^2 - 4x + 1$. This is still a reasonable way to think, but it has its limitations. For example, suppose at time zero I throw a bean-bag (so no bounces) up in the air at four metres per second. Taking the convenient, but inaccurate, value of $10\,\mathrm{ms}^{-2}$ for the acceleration due to gravity, and making various other simplifying assumptions, its height $h(t)$ above the ground at time t will be

$$h(t) = \begin{cases} 1 & \text{if } t \leq 0 \\ 1 + 4t - 5t^2 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{if } t \geq 1, \end{cases}$$

as shown in the graph below.



This suggests that we should allow functions that are defined using 'split' definitions.

Note that $h(t)$ is not differentiable when $t = 0$ or when $t = 1$. Another function, which is far more badly behaved, but still of interest to mathematicians, is

$$g(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

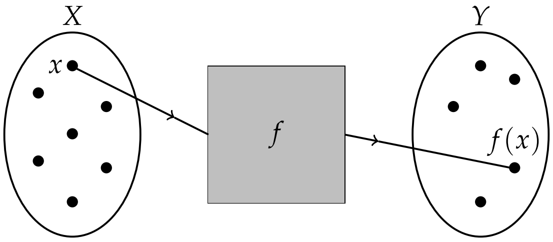
The functions f, g and h are defined for every real number. This is also an unnecessary restriction. For example, suppose you are given a weighted die, which is biased to land on 6 with probability $3/8$. We could describe the probability of rolling each number by

$$p(n) = \begin{cases} 3/8 & \text{if } n = 6 \\ 1/8 & \text{if } n \in \{1, 2, 3, 4, 5\}. \end{cases}$$

Then p sends the set $\{1, 2, 3, 4, 5, 6\}$ to $\mathbb{R}$. In fact, since the values taken by p are rational numbers, we could replace $\mathbb{R}$ with $\mathbb{Q}$, or even with $\{1/8, 3/8\}$.

DEFINITION OF FUNCTIONS. Eventually it was realized that it was best *not* to impose any special conditions on what a function can do.

Definition 2.1. Let X and Y be sets. A *function* from X to Y is a black box such that, when an element $x \in X$ is put in, an element $y \in Y$ comes out. If the function is called f, then we write $f : X \to Y$. The output for the input x is written $f(x)$.



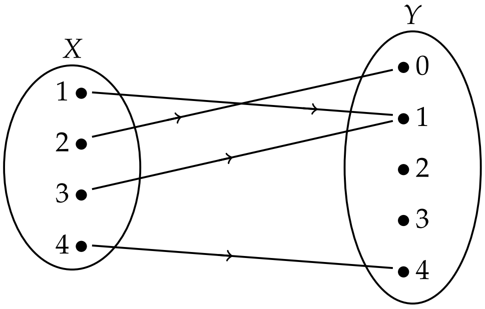
Let f and g be functions both with input set X and output set Y. If $f(x) = g(x)$ for all $x \in X$ then we say that f is equal to g, and write $f = g$. So f and g are equal if whenever we put in the same input, we get out the same output. We do not care at all how the mechanism in the black box produces its output.

Example 2.2. Let $f : \{1, 2, 3, 4\} \to \{0, 1, 2, 3, 4\}$ be the function defined by

$$f(1) = 1, \quad f(2) = 0, \quad f(3) = 1, \quad f(4) = 4.$$

Define $g : \{1, 2, 3, 4\} \to \{0, 1, 2, 3, 4\}$ by $g(x) = (x - 2)^2$. Then $f = g$, since $f(x) = g(x)$ for all $x \in \{1, 2, 3, 4\}$.

We have already seen that a graph can be one good way to visualize a function. Another is a diagram such as the one below, which shows the equal functions f and g from Example 2.2.



Definition 2.3. Let $f : X \to Y$ be a function. The set X of allowed inputs to f is called the *domain* of f. The set Y of allowed outputs is called the *codomain* of f. The set $\{f(x) : x \in X\}$ of all outputs that actually appear is called the *range* of f.

You will need to think carefully about this definition. In particular, note the distinction between the codomain and the range. For example, if g is the function in Example 2.2 then

- The domain of g is $\{1, 2, 3, 4\}$. This is the set of allowed inputs.
- The codomain of g is $\{0, 1, 2, 3, 4\}$. Every output of g is promised to be in this set. [**Misprinted in version issued 4th October.**]
- The range of g is $\{0, 1, 4\}$. These are the outputs that actually appear. [**Misprinted in version issued 4th October**.]

In general it might take quite a bit of work to define the range of a function $f : X \to Y$. But the domain and codomain are given for you, as part of its definition. It is maybe a bit of a stretch, but try imagining that written on the side of the black box are the sets X and Y, and the text

> 'If you input any $x \in X$ this machine promises faithfully
> to output $f(x) \in Y$'.

Exercise 2.4. Let $h(t)$ be the height function at the start of this section. Find $h'(t)$ when $t \neq 0, 1$. At $t = 0$ or 1 the derivative is not defined, because at these points the slope of the graph is different depending on whether we approach from the left or from the right. So the domain of $h'(t)$ is $\{x \in R : x \neq 0, 1\}$. What are the domain and range of $h''(t)$?

INJECTIVE, SURJECTIVE, BIJECTIVE.

Definition 2.5. Let X and Y be sets and let $f : X \to Y$ be a function.

(i) We say that f is *injective* if for all $x, x' \in X$,

$$f(x) = f(x') \implies x = x'.$$

(ii) We say that f is *surjective* if for all $y \in Y$ there exists $x \in X$ such that $f(x) = y$.

(iii) We say that f is *bijective* if f is injective and surjective.

The definition of surjective is the easiest to understand. It can be restated as follows

f is surjective $\iff$ for all $y \in Y$, the equation $f(x) = y$ has a solution.

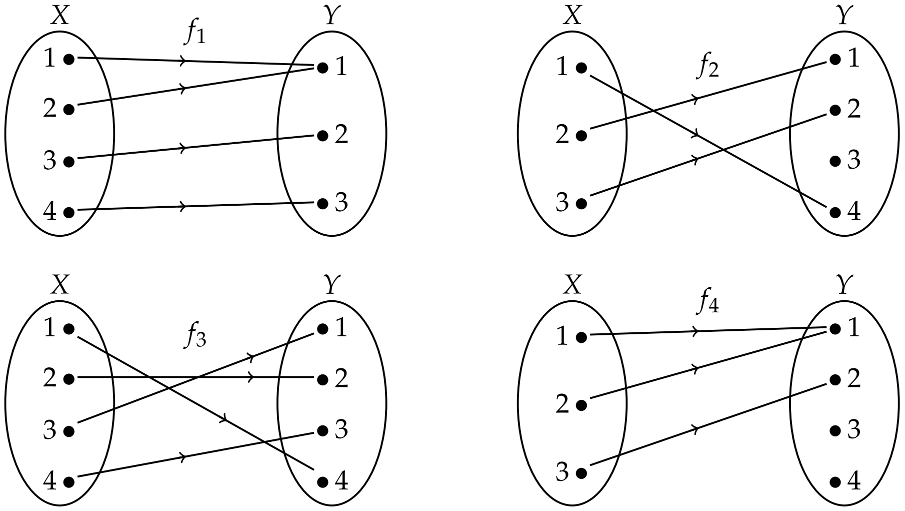
Injective is a bit harder, and people often get it wrong. Here are three equivalent restatements:

$$\begin{aligned} f \text{ is injective} &\iff \text{different inputs to } f \text{ always give different outputs} \\ &\iff \text{if } x \neq x' \text{ then } f(x) \neq f(x') \\ &\iff \begin{array}{l}\text{for all } y \in Y, \text{ the equation } f(x) = y \text{ has at most} \\ \text{one solution.}\end{array} \end{aligned}$$

Either directly, or by putting together the equivalent ways to state that a function is injective and surjective given above, you should be able to see that

f is bijective $\iff$ for all $y \in Y$ the equation $f(x) = y$ has a unique solution.

Example 2.6. For each of the functions f_1, f_2, f_3, f_4 drawn as a diagram below, we will determine which combination of the properties injective, surjective, bijective it has. (One function has none of these special properties.)



Exercise 2.7. Let $f : X \to Y$ be represented by a diagram like the ones above. Then

f is injective $\iff$ no element of the codomain Y has two (or more) arrows pointing to it.

Give a similar condition for f to be surjective. Give a similar condition for f to be bijective.

By looking at the graph of a function $f : X \to \mathbb{R}$, where $X \subseteq \mathbb{R}$, we can detect whether f is injective by looking at the horizontal lines going through $(0, y)$ for each $y \in \mathbb{R}$. This is called the *horizontal line test*:

f is injective $\iff$ each horizontal line hits the graph of f at most once.

Example 2.8. The function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^2 - 4x + 1$ is neither injective nor surjective. Let $X = \{x \in \mathbb{R} : x \geq 2\}$. If we define $g : X \to \mathbb{R}$ by $g(x) = x^2 - 4x + 1$ then g is injective. Note that g is a different function to f, because the domains of f and g are different.

A bijective function is also called a *bijection*.

Exercise 2.9. Let $X = \{x \in \mathbb{R} : x \geq 2\}$. What subset Y of $\mathbb{R}$ should you choose so that the function $h : X \to Y$ defined by $h(x) = x^2 - 4x + 1$ is a bijection?

INVERSE FUNCTIONS. Suppose that $f : X \to Y$ is a bijection. As remarked at the top of page 14, for each $y \in Y$ there exists a unique $x \in X$ such that $f(x) = y$.

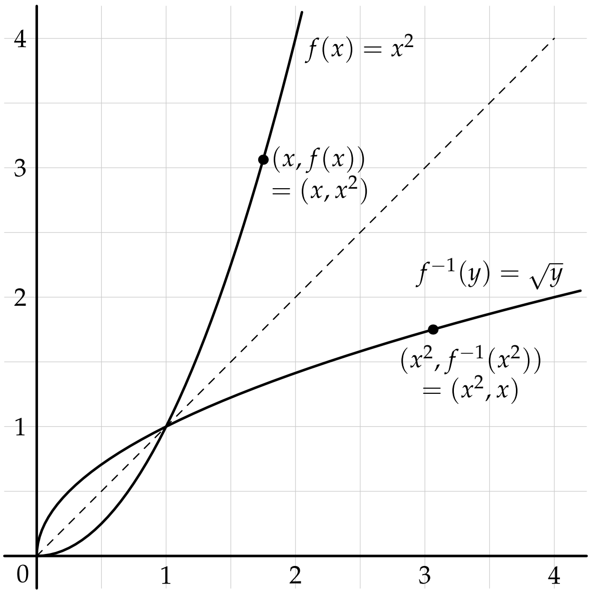
We define the *inverse function to* f to be the function $f^{-1} : Y \to X$ which sends $y \in Y$ to the unique $x \in X$ such that $f(x) = y$. In symbols

$$f^{-1}(y) = x \iff f(x) = y.$$

Exercise 2.10. Suppose that $f : X \to Y$ is represented by a diagram, as in Example 2.6. How can you obtain the diagram representing the inverse function $f^{-1} : Y \to X$? [*Hint: a complete answer can be given in four words.*]

It is also useful to think about how the graphs of a bijective function and its inverse are related.

Let $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$. The graph below shows the function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $f(x) = x^2$. The inverse function to f is $f^{-1}(y) = \sqrt{y}$. The marked point (x, x^2) is on the graph of f. Its reflection in the line $y = x$ is (x^2, x), which is on the graph of f^{-1}.



In general, (x, y) is on the graph of $f : X \to Y \iff (y, x)$ is on the graph of $f^{-1} : Y \to X$. So the graph of f^{-1} is obtained from the graph of f by reflecting it in the line $y = x$.

Here are two further examples on inverse functions. If you plot the graphs you should find they are related by the reflection just described. The first example gives one (of several) ways to do Question 4(a) on Sheet 2.

Example 2.11. Let $Y = \{y \in \mathbb{R} : 0 \leq y < 2\}$. Let $f : \mathbb{R}_{\geq 0} \to Y$ be the function defined by $h(x) = 2x/(1 + x)$. For $y \in Y$ we have

$$\frac{2x}{1+x} = y \iff y + xy = 2x \iff y = x(2-y) \iff \frac{y}{2-y} = x.$$

Hence $f(x) = y \iff x = y/(2 - y)$. Since $y \geq 0$ and $2 - y > 0$, the solution $x = y/(2 - y)$ is in the domain $\mathbb{R}_{\geq 0}$ of h. Therefore f is a bijection with inverse $f^{-1}(y) = y/(2 - y)$.

Example 2.12. Let $Y = \{y \in \mathbb{R} : -1 \leq y \leq 1\}$. Consider $\sin : \mathbb{R} \to Y$. This function is not bijective, because it is not injective. For example, $\sin 0 = \sin 2\pi = 1$. To find an inverse we must first restrict the domain. The usual choice is to take

$$X = \{x \in \mathbb{R} : -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}\}.$$

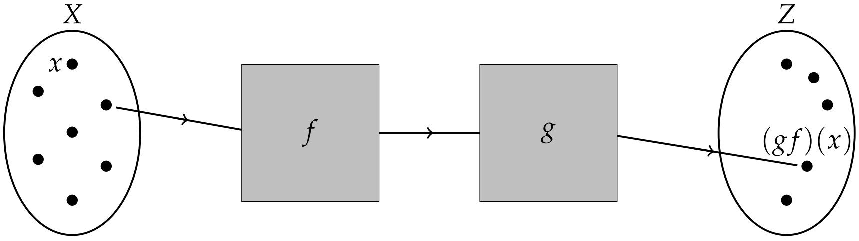
Then $\sin : X \to Y$ is a bijection, with inverse the usual $\sin^{-1}$ function on your calculator (provided it is set to radians). **Exercise:** Give a different domain Z, with $Z \neq X$, such that $\sin : Z \to Y$ is again a bijection.

COMPOSING FUNCTIONS. Let $f : X \to Y$ and $g : Y \to Z$ be functions. The *composition of f and g* is the function $gf : X \to Z$, defined by

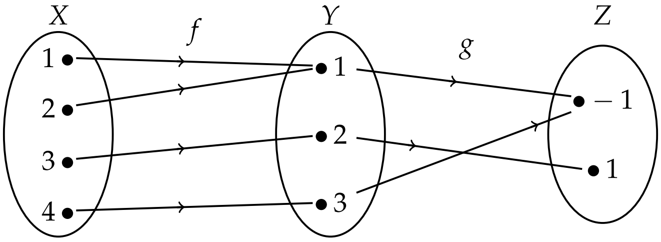
$$(gf)(x) = g\big(f(x)\big).$$

Note that gf means 'do f, then do g'. One has to get used to reading function compositions from right to left.

To create a black box for the function gf, connect the output of the black box for f to the input of the black box for g, as shown below. Since g expects inputs from the set Y (the domain of g) and f produces outputs in the set Y (the codomain of f) this is a sensible thing to do.



Example 2.13. Let $f : \{1, 2, 3, 4\} \to \{1, 2, 3\}$ be the function f_1 from Example 2.6. Let $g : \{1, 2, 3\} \to \{-1, 1\}$ be defined by $g(x) = (-1)^x$.



Note that f and g are both surjective. Clearly $(gf)(3) = 1$ and $(gf)(2) = -1$. (There are two other elements also sent to -1.) So for every $z \in Z$ there exists $x \in X$ such that $(gf)(x) = z$. Therefore gf is also surjective.

More generally the composition of *any* two surjective functions is surjective. The proof of this is left to you on Sheet 2, Question 5(a). The previous example should suggest how to prove it. In lectures we will prove (a) and (c) below.

Lemma 2.14 (Examinable). *Let $f : X \to Y$ and $g : Y \to Z$ be functions.*

(a) *If f and g are injective then gf is injective.*
(b) *If f and g are surjective then gf is surjective.*
(c) *If f and g are bijective then gf is bijective.*

By (c), if $f : X \to Y$ and $g : Y \to Z$ are bijections, then $gf : X \to Z$ is a bijection, and so it has an inverse function. To undo the composition $gf : X \to Z$ we must first undo $g : Y \to Z$, then undo $f : X \to Y$. Hence

$$(gf)^{-1} = f^{-1}g^{-1}.$$

This result can be useful when finding inverse functions.

Example 2.15. Let

$$f(x) = \sqrt{\frac{2x^2}{1 + x^2}}.$$

We can write f as a composition: $f = f_3 f_2 f_1$ where $f_1(x) = x^2$, $f_2(x) = 2x/(1 + x)$ and $f_3(x) = \sqrt{x}$. In the lecture we will sort out the domains and codomains of f and f_1, f_2, f_3, and hence find the inverse to f.

Strictly speaking it is not clear what $f_3 f_2 f_1$ means, since we have only defined the composition of two functions. We could interpret it as meaning either $(f_3 f_2) f_1$, or $f_3(f_2 f_1)$. Fortunately, these are exactly the same function.

The *associative property of composition* states that if $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$ are any functions then

$$(hg)f = h(gf) : X \to W.$$

This has a one-line proof.

We will see associativity again in §10 of the course on rings.

IDENTITY FUNCTIONS. Suppose $f : X \to Y$ is a bijection. We have seen that f has an inverse function $f^{-1} : Y \to X$. What happens when we compose f and f^{-1}? The defining property of f^{-1}, stated on page 15, is

$$f^{-1}(y) = x \iff f(x) = y.$$

It follows that $(f^{-1}f)(x) = x$ for all $x \in X$ and $(ff^{-1})(y) = y$ for all $y \in Y$.

The *identity* function on a set X is the function $\mathrm{id}_X : X \to X$ defined by $\mathrm{id}_X(x) = x$ for all $x \in X$.

Using identity functions, the results just mentioned on $f^{-1}f$ and ff^{-1}, become $f^{-1}f = \mathrm{id}_X$ and $ff^{-1} = \mathrm{id}_Y$.

EXTRAS. The following lemma gives yet another way to think about injective and surjective functions.

Lemma 2.16. *Let X and Y be non-empty sets and let $f : X \to Y$ be a function.*

(a) *f is injective $\iff$ there exists $g : Y \to X$ such that $gf = \mathrm{id}_X$.*

(b) *f is surjective $\iff$ there exists $h : Y \to X$ such that $fh = \mathrm{id}_Y$.*

Moreover, if $gf = \mathrm{id}_X$ and $fh = \mathrm{id}_Y$ then f is bijective and $g = h = f^{-1}$.

Exercise 2.17. Prove '$\Longleftarrow$' in (a) and (b). A good way to start in (a) is to write $f(x) = f(x') \implies g\big(f(x)\big) = g\big(f(x')\big)$ and use that $gf = \mathrm{id}_X$.

Exercise 2.18. Prove '$\Longrightarrow$' in (a) and (b). This is a bit harder, since it is now *your* job to define the functions g and h. Diagrams might be helpful.

To prove the final line, note that by (a) '$\Longleftarrow$', f is injective, and by (b) '$\Longleftarrow$', f is surjective. Hence f is bijective. Now using associativity, we have

$$g = g\,\mathrm{id}_Y = g(ff^{-1}) = (gf)f^{-1} = \mathrm{id}_X f^{-1} = f^{-1}$$

and similarly you can show that $h = f^{-1}$. So $g = h = f^{-1}$, as required.

3. Complex Numbers

Motivation and definition. If $x \in \mathbb{R}$ then $x^2 \geq 0$. So the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not surjective, and the equation $x^2 = -1$ has no solutions in $\mathbb{R}$.

We will create a larger number system, called the complex numbers, in which this equation has a solution. We start by introducing a symbol i, such that $i^2 = -1$. The number systems $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are closed under addition and multiplication (see Definition 1.5). This is a very useful property. So our new number system will contain $-3i$, $1 + \sqrt{3}i$, $3 - 2i$, and so on.

Definition 3.1. A *complex number* is defined to be a symbol of the form $a + bi$ where $a, b \in \mathbb{R}$. If $z = a + bi$ then we say that a is the *real part* of z, and b is the *imaginary part of* z, and write $\operatorname{Re} z = a$, $\operatorname{Im} z = b$. We write $\mathbb{C}$ for the set of all complex numbers.

Using the notation introduced at the bottom of page 7, we have

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

It is fine to write $a + ib$ rather than $a + bi$, and to write $a + 0i$ as a and $0 + bi$ as bi. A complex number written in any of these ways (with $a, b \in \mathbb{R}$) is said to be in *Cartesian form*.

In $\mathbb{C}$ we can find square roots of any real number. For example,

$$(\sqrt{3}\,i)^2 = \sqrt{3}^2\, i^2 = -3.$$

We will see at the end of this section that *any* polynomial equation has a solution in $\mathbb{C}$. Solving cubic equations, such as $x^3 - 6x - 40 = 0$, was an important historical motivation for complex numbers. See Example 3.12 below and the extras at the end of this section.

Please interpret the 'complex' in complex number as meaning 'made of more than one part', rather than 'difficult'. The word 'imaginary' is also standard—please do not be put off by it. I hope you will see that it is not hard to calculate with complex numbers, and there is no reason to be scared of them!

Exercise 3.2. Calculate $(1 + i)^3$.

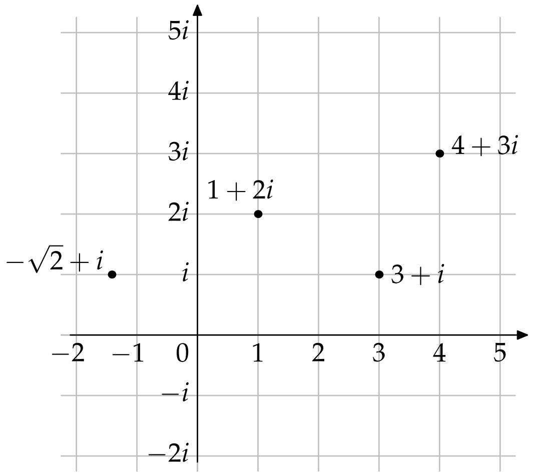
Addition, subtraction and multiplication. The rules for adding, multiplying and subtracting complex numbers follow from the property that $i^2 = -1$. If $a + bi$ and $c + di \in \mathbb{C}$ are complex numbers in

Cartesian form then

$$\begin{aligned}(a+bi)+(c+di)&=(a+c)+(b+d)i\\(a+bi)-(c+di)&=(a-c)+(b-d)i\\(a+bi)(c+di)&=(ac-bd)+(ad+bc)i.\end{aligned}$$

So the set $\mathbb{C}$ of complex numbers is closed under addition, subtraction and multiplication. To see that the complex numbers are also closed under division, it is useful to think about them geometrically.

ARGAND DIAGRAM, COMPLEX CONJUGATE AND MODULUS. We represent complex numbers by points in a plane (called an *Argand diagram*).



(When drawing these diagrams yourself, there is no need to put in gridlines or all the axis labels.)

Exercise 3.3. By the rules for addition above, $(1+3i)+(3+i)=4+3i$. What does this mean geometrically?

Let $z=a+bi\in\mathbb{C}$ be in Cartesian form. We define the *modulus* of z, written $|z|$, to be $\sqrt{a^2+b^2}$. We define the *complex conjugate* of z, written $\overline{z}$, to be $a-bi$.

We read $|z|$ as 'mod z' and $\overline{z}$ as 'z bar'. Note that $\overline{z}$ is the reflection of z in the real axis. By Pythagoras' Theorem, $|z|$ is the distance from 0 to z on the Argand diagram. For example $|4+3i|=\sqrt{4^2+3^2}=\sqrt{25}=5$.

Lemma 3.4 (Examinable). *Let $z\in\mathbb{C}$. Then*

(a) $|z|^2=z\overline{z}$.

(b) *If $z\neq 0$ then $1/z=\overline{z}/|z|^2$.*

(c) *The set $\mathbb{C}$ of complex numbers is closed under division.*

Part (b) of Lemma 3.4 is a practical way to find $1/z$ for any non-zero complex number z. This is shown in the next example, which also gives some examples of solving equations in $\mathbb{C}$.

Example 3.5.

(1) To solve the equation $(1-2i)z + (1-4i) = 2+2i$ we start by subtracting $1-4i$ from both sides, to get $(1-2i)z = 1+6i$. Now use the division trick in (b) to get
$$z = \frac{1+6i}{1-2i} = \frac{(1+6i)(1+2i)}{(1-2i)(1+2i)} = \frac{-11+8i}{5} = -\frac{11}{5} + \frac{8}{5}i.$$
Hence $z = -\frac{11}{5} + \frac{8}{5}i$ is the unique solution.

(2) Consider the simultaneous equations $|z| = 5$ and $z + \overline{z} = 8$. If $z = a + bi$ then
$$z + \overline{z} = (a+bi) + (a-bi) = 2a.$$
So $z + \overline{z} = 8 \iff z = 4 + bi$ for some $b \in \mathbb{R}$. Since
$$|4+bi|^2 = (4+bi)(4-bi) = 16 + b^2$$
we have
$$|4+bi| = 5 \iff b^2 = 9 \iff b = 3 \text{ or } b = -3.$$
So there are exactly two solutions to the simultaneous equations, namely $z = 4+3i$ and $z = 4-3i$.

POLAR FORM AND ARGUMENTS. Any complex number z can be written in the form
$$z = r(\cos\theta + i\sin\theta)$$
where $r \in \mathbb{R}_{\geq 0}$ and θ is an angle, measured in radians. This is called the *polar form* of z. Observe that $r = |z|$. We say that θ is an *argument* of z.

Example 3.6. Let $z = \frac{1}{2} + \frac{\sqrt{3}}{2}i$. Then $\frac{\pi}{3}$ is an argument of z and
$$z = \cos\frac{\pi}{3} + i\sin\frac{\pi}{3}.$$
The full set of arguments of z is
$$\{\ldots, -\frac{5\pi}{3}, \frac{\pi}{3}, \frac{7\pi}{3}, \frac{13\pi}{3}, \ldots\}.$$
Comparing real and imaginary parts in the two expression for z, we get $\cos\frac{\pi}{3} = \frac{1}{2}$ and $\sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}$. **Exercise:** show by drawing another triangle that $\frac{\pi}{4}$ is an argument of $1+i$ and
$$1 + i = \sqrt{2}(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}).$$
Deduce that $\cos\frac{\pi}{4} = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$.

More generally, let $z \in \mathbb{C}$ be non-zero. Suppose that θ is an argument of z. Since sin and cos are periodic with period 2π,

$$\phi \text{ is an argument of } z \iff \phi = \theta + 2n\pi \text{ for some } n \in \mathbb{Z}.$$

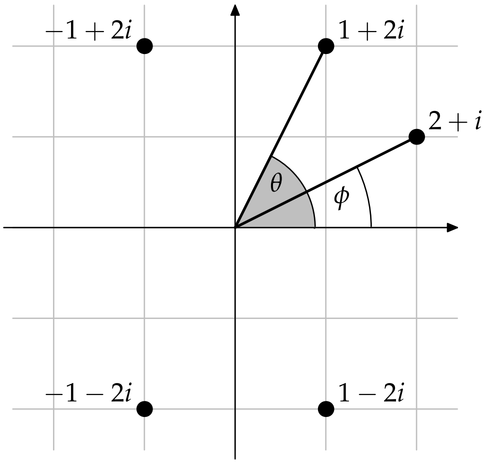
Hence the set of angles that are arguments of z is

$$\{\ldots, \theta - 4\pi, \theta - 2\pi, \theta, \theta + 2\pi, \theta + 4\pi, \ldots\}.$$

To get around the non-uniqueness of arguments, we make the following definition.

Definition 3.7. Let $z \in \mathbb{C}$ be non-zero. If $z = r(\cos\theta + i\sin\theta)$ where $-\pi < \theta \le \pi$, then we say that θ is the *principal argument* of z, and write $\theta = \operatorname{Arg}(z)$.

Example 3.8. We will find the principal argument of the complex numbers shown on the Argand diagram below in terms of the angles θ and ϕ.



There is an often misapplied 'rule' that $\operatorname{Arg}(a+bi) = \tan^{-1}(b/a)$. **This only works when $a > 0$.** It is better to draw a diagram and find the principal argument using a relevant triangle, as in this example.

Example 3.9. Let $z = r(\cos\theta + i\sin\theta)$ and $w = s(\cos\phi + i\sin\phi)$ be complex numbers in polar form. Using the formulae

$$\begin{aligned}\cos(\theta+\phi) &= \cos\theta\cos\phi - \sin\theta\sin\phi\\ \sin(\theta+\phi) &= \cos\theta\sin\phi + \sin\theta\cos\phi\end{aligned}$$

it follows that

$$zw = rs\big(\cos(\theta+\phi) + i\sin(\theta+\phi)\big).$$

In short: to multiply numbers in polar form, multiply the moduli and add the arguments.

Exercise 3.10. Let w and z be as in Example 3.10 and suppose that $z \neq 0$. Express w/z in polar form.

DE MOIVRE'S THEOREM AND A CUBIC EQUATION. If $\theta \in \mathbb{R}$ and $n \in \mathbb{N}$ then

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

Moivre's Theorem can be proved using mathematical induction and Example 3.10. Example 3.16(2) below gives a quicker proof, using the exponential function.

Example 3.11. Comparing real and imaginary parts in the case $n = 3$ of De Moivre's Theorem shows that

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$$

We proved an identity about the real cosine function, $\cos : \mathbb{R} \to \mathbb{R}$ using complex numbers. **Exercise:** find the analogous identity for $\sin 3\theta$.

This leads to a way to solve cubic equations. The example below gives a special case, but the method generalizes quite easily to any real cubic with three real roots: see the extras for this section.

Example 3.12. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3 - 12x - 8$. Substitute $x = 4\cos\theta$. Then

$$\begin{aligned} f(x) = 0 &\iff 64\cos^3\theta - 48\cos\theta - 8 = 0 \\ &\iff 16(4\cos^3\theta - 3\cos\theta) = 8 \\ &\iff 16\cos 3\theta = 8 \\ &\iff \cos 3\theta = 1/2. \end{aligned}$$

Exercise: by drawing the graph for cos, and using Example 3.6, show that $\cos 3\theta = 1/2 \iff 3\theta = \pm\pi/3 + 2n\pi$ for some $n \in \mathbb{Z}$. Deduce that the roots of f are

$$4\cos\tfrac{\pi}{9}, \quad 4\cos\tfrac{7\pi}{9}, \quad 4\cos\tfrac{13\pi}{9}.$$

Further Exercise: sketch the graph of f and label the roots correctly. Include the coordinates of the turning points.

MOTIVATION FOR THE EXPONENTIAL FUNCTION. (Non-examinable.) You will have seen the real exponential function, which sends $x \in \mathbb{R}$ to $\mathrm{e}^x \in \mathbb{R}$. You might also have seen its expression as a Taylor / Maclaurin series:

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots .$$

Using this power series, we can define e^z for any complex number z by

$$\mathrm{e}^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .$$

Now, if $z = bi$ is a purely imaginary number, something amazing happens:

$$\begin{aligned}\mathrm{e}^{bi} &= 1 + bi + \frac{(bi)^2}{2!} + \frac{(bi)^3}{3!} + \frac{(bi)^4}{4!} + \frac{(bi)^5}{5!} + \cdots \\ &= \left(1 - \frac{b^2}{2!} + \frac{b^4}{4!} - \cdots\right) + \left(b - \frac{b^3}{3!} + \frac{b^5}{5!} - \cdots\right)i \\ &= \cos b + i \sin b\end{aligned}$$

where the final step uses the power series for the cosine and sine functions.

A familiar property of the real exponential function is that $\mathrm{e}^{a+c} = \mathrm{e}^a\mathrm{e}^c$ for all $a, c \in \mathbb{R}$. We define the complex exponential function so that this property holds for all complex numbers.

The complex exponential function.

Definition 3.13. Let $z = a + bi \in \mathbb{C}$ be a complex number in Cartesian form. We define the *complex exponential function* $\exp : \mathbb{C} \to \mathbb{C}$ by

$$\exp(z) = \mathrm{e}^a(\cos b + i \sin b).$$

It is fine to write e^z for $\exp(z)$. The complex exponential function unites in one beautiful mathematical object the real valued exponential, cosine and sine functions.

Exercise 3.14. Show that $\exp(z + w) = \exp z \exp w$ for all complex numbers z and w. [*Hint:* write $z = a + bi$, $w = c + di$ and use Example 3.9.]

A complex number written as $r\mathrm{e}^{i\theta}$ where $r \in \mathbb{R}_{\geq 0}$ and $\theta \in \mathbb{R}$ is said to be in *exponential form*. It is easy to convert between polar and exponential form:

$$z = r(\cos\theta + i \sin\theta) \iff z = r\mathrm{e}^{i\theta}.$$

Example 3.15.

(1) Put $z = i\pi$ in the complex exponential function. We get $\mathrm{e}^{i\pi} = -1$, or equivalently,

$$\mathrm{e}^{i\pi} + 1 = 0.$$

This is *Euler's Identity*. It relates five fundamental mathematical constants: $0, 1, \mathrm{e}, \pi$ and i.

(2) Let $\theta \in \mathbb{R}$. Put $z = n\theta i$ in the complex exponential function to get

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

This proves De Moivre's Theorem.

We saw earlier that the polar form of a complex number is not unique. The exponential form has the same lack of uniqueness. In fact, if $re^{i\theta}$ and $se^{i\phi}$ are non-zero complex numbers in exponential form, then

$$(\star) \qquad re^{i\theta} = se^{i\phi} \iff r = s \text{ and } \phi = \theta + 2n\pi \text{ for some } n \in \mathbb{Z}.$$

Exponential form is very useful for finding roots of complex numbers. The next example shows the general strategy to use. To find all solutions, it is essential to remember the remark about uniqueness above.

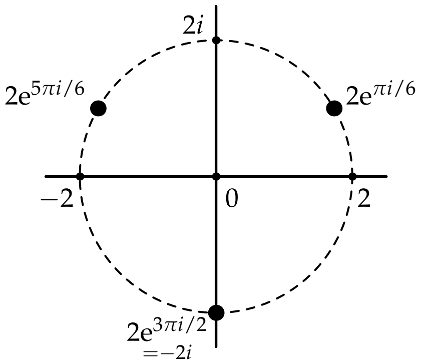
Example 3.16. We will solve the equation $z^3 = 8i$. The argument of $8i$ is $\pi/2$, so in exponential form we have $8i = 8e^{i\pi/2}$. If $z = re^{i\theta}$ then $z^3 = r^3e^{i3\theta}$. By $(\star)$ we have

$$\begin{aligned} r^3e^{i3\theta} = 8e^{i\pi/2} &\iff r^3 = 8 \text{ and } 3\theta = \frac{\pi}{2} + 2n\pi \text{ for some } n \in \mathbb{Z} \\ &\iff r = 2 \text{ and } \theta = \frac{\pi}{6} + \frac{2\pi}{3}n \text{ for some } n \in \mathbb{Z}. \end{aligned}$$

Taking $n = 0, 1, 2$ and noting that $\pi/6 + 2\pi/3 = 5\pi/6$ and $\pi/6 + 4\pi/3 = 9\pi/6 = 3\pi/2$ we get the three solutions

$$z = 2e^{i\pi/6}, \quad 2e^{i5\pi/6}, \quad 2e^{i3\pi/2}$$

shown below. These are all the solutions, since adding (or subtracting) $2\pi/3$ to the arguments above gives another solution already found.



You should draw an Argand diagram whenever you do a problem of this sort. The rotational symmetry helps to check that no root has been overlooked.

In general, a non-zero complex number has n distinct n-th roots.

LOGS OF A COMPLEX NUMBER. Let $z = re^{i\theta}$ be a complex number in exponential form. If $z = 0$ then there is no $w \in \mathbb{C}$ such that $\mathrm{e}^w = z$, since $|\mathrm{e}^{a+bi}| = \mathrm{e}^a$ and $e^a > 0$ for all $a \in \mathbb{R}$. If $z \neq 0$ then

$$e^w = z \iff w = \ln r + (\theta + 2\pi n)i \text{ for some } n \in \mathbb{Z}.$$

Any such number w is called a *logarithm* of z.

Example 3.17. In exponential form $2i = 2\mathrm{e}^{i\pi/2}$. So the set of logarithms of $2i$ is

$$\{\ln 2 + (\frac{\pi}{2} + 2n\pi)i \text{ for some } n \in \mathbb{Z}\}.$$

Exercise 3.18. Consider $\exp : \mathbb{C} \to \mathbb{C}$. What are the domain, codomain and range of exp? Is exp surjective? Is exp injective?

QUADRATIC EQUATIONS. You are probably familiar with how to solve quadratic equations over the real numbers. Essentially the same method works over $\mathbb{C}$. Exponential form might be useful for finding the square root needed.

Lemma 3.19 (Examinable). *Let $a, b, c \in \mathbb{C}$ and suppose that $a \neq 0$. The solutions to the quadratic equation $az^2 + bz + c = 0$ are*

$$z = \frac{-b \pm D}{2a}$$

where $D \in \mathbb{C}$ satisfies $D^2 = b^2 - 4ac$.

Example 3.20. Observe that $z^3 - 1 = (z-1)(z^2 + z + 1)$. So if z is a third root of unity other than 1 then z is a solution of $z^2 + z + 1 = 0$. Using Lemma 3.19 we get

$$z = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i.$$

Since the third roots of unity are 1, $\mathrm{e}^{2\pi i/3}$ and $\mathrm{e}^{4\pi i/3}$, this shows that $\cos\frac{2\pi}{3} = -\frac{1}{2}$ and $\sin\frac{2\pi}{3} = \frac{\sqrt{3}}{2}$. (You can check these values agree with Example 3.6.)

FUNDAMENTAL THEOREM OF ALGEBRA. We defined the complex numbers so we could solve the equation $x^2 = -1$, and we have just seen that any quadratic equation has a solution in $\mathbb{C}$. Remarkably, *any* polynomial equation has a solution in $\mathbb{C}$.

Theorem 3.21 (Fundamental Theorem of Algebra). *Let $n \in \mathbb{N}$ and let $a_0, a_1, \ldots, a_n \in \mathbb{C}$ with $a_n \neq 0$. Then the equation*

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0$$

has a solution in $\mathbb{C}$.

The conclusion of the Fundamental Theorem of Algebra is just that one solution exist. We will see in §10 that it follows from the theorem, as stated above, that there exist distinct $w_1, w_2, \ldots, w_r \in \mathbb{C}$ and $m_1, \ldots, m_r \in \mathbb{N}$ such that $m_1 + \cdots + m_r = n$ and

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = a_n (z - w_1)^{m_1} (z - w_2)^{m_2} \ldots (z - w_r)^{m_r}.$$

Hence a polynomial of degree n has exactly n roots in $\mathbb{C}$, provided we agree to say that $1 + i$ is a root of $z\big(z - (1 + i)\big)^2$ twice, and 0 is a root of $z^3(z + 1)$ three times, and so on. This is called *counting roots with multiplicity.*

Exercise 3.22. Find all solutions to the quartic equation $z^4 + 2z^3 + 3z^2 + 4z + 2 = 0$. (*Hint:* one solution is in $\mathbb{Z}$.)

EXTRAS: CUBIC EQUATIONS. This section is an extended exercise on solving cubic equations. By dividing through by the coefficient of z^3, we can reduce any cubic equation to one of the form $z^3 + bz^2 + cz + d = 0$. The next exercise gives a further useful reduction.

Exercise 3.23. Show that substituting $z = w - b/3$ into $z^3 + bz^2 + cz + d = 0$ gives a new equation $w^3 - 3Aw - 2B = 0$ for suitable A and B. How are the roots of these equations related?

We shall first solve the equation $w^3 - 3Aw - 2B = 0$ in the special case when A and B are real and the trigonometric solution in Example 3.12 works.

Exercise 3.24. Suppose that $A, B \in \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3 - 3Ax - 2B$.

(a) Show that f has three distinct real roots if and only if $A^3 > B^2$. [*Hint:* the real roots of a real polynomial are separated by zeros of its derivative. Try finding the coordinates of the turning points of f.]

(b) Suppose that $A^3 > B^2$. Then $A > 0$ and so $\sqrt{A} \in \mathbb{R}$. Use Example 3.11 to show that

$$f(2\sqrt{A}\cos\theta) = 0 \Longleftrightarrow \cos 3\theta = \frac{B}{A^{3/2}}.$$

Since $|B/A^{3/2}| < 1$, there exists $\phi \in \mathbb{R}$ such that $\cos\phi = B/A^{3/2}$. Deduce that the roots of f are

$$2\sqrt{A}\cos\frac{\phi}{3}, \quad 2\sqrt{A}\cos\big(\frac{\phi}{3}+\frac{2\pi}{3}\big), \quad 2\sqrt{A}\cos\big(\frac{\phi}{3}+\frac{4\pi}{3}\big).$$

An alternative method that works when A and B are any non-zero complex numbers is credited variously to the 16th Century Italian mathematicians del Ferro, Tartaglia and Cardano (with del Ferro having the strongest claim). Observe that

$$(u+v)^3 = 3uv(u+v) + u^3 + v^3.$$

Hence if we set $w = u + v$ then w will satisfy $w^3 = 3Aw + 2B$ provided u and v are chosen so that $A = uv$ and $2B = u^3 + v^3$.

Exercise 3.25. Suppose that $A \neq 0$. Let

$$D = -A^3 + B^2$$

and let $\Delta \in \mathbb{C}$ be such that $\Delta^2 = D$. Show that $uv = A$ and $u^3 + v^3 = B$ if and only if

$$u^3 = B \pm \Delta \quad \text{and} \quad v = \frac{A}{u}.$$

Deduce that if $u^3 = B + \Delta$ then the solutions of $w^3 = 3Aw + 2B$ are

$$u + \frac{A}{u}, \quad \zeta u + \frac{\zeta^2 A}{u}, \quad \zeta^2 u + \frac{\zeta A}{u}$$

where $\zeta = \exp(2\pi i/3)$ is a third root of unity.

When A and B are real and $A^3 > B^2$, we saw in Exercise 3.24 that $x^3 = 3Ax + 2B$ has three real roots. But Exercise 3.25 expresses the solutions in terms of cube roots of $B + \Delta$ where $\Delta = i\sqrt{A^3 - B^2}$ is a purely imaginary complex number. This shows that the del Ferro/Tartaglia/Cardano method may need cube roots of non-real complex numbers *even when all the three solutions are real!*

Exercise 3.26. Let $f(x) = x^3 - 12x - 8$ be the cubic in Example 3.12. Show that, if we follow the method in Exercise 3.25 and take $\Delta = 4i\sqrt{3}$, then $B + \Delta = 8\exp(\pi i/3)$. Hence show that the solutions to $f(x) = 0$ given by Example 3.12 and Exercise 3.25 are the same.

Exercise 3.27. Solve the following cubic equations.
(a) $w^3 - 3w + 4 = 0$,
(b) $z^3 - 3z^2 - 3z - 35 = 0$,
(c) $w^3 - 3w + 1 = 0$.

See Chapter 7 of Liebeck *A concise introduction to pure mathematics* ([2] in the list of recommended reading) for more about cubic equations.

Part B: Natural numbers and induction

4. INDUCTION

PROPOSITIONS. A *proposition* is a self-contained statement which is either true or false. For example the statement

There is a real number x such that $x^2 + 1 = 0$

is a false proposition. More briefly, we can write

P: The integers are closed under addition.

This defines P to be the true proposition that the integers are closed under addition. Some statements are too vague or subjective to be considered propositions. For instance:

Q: Houses in Englefield Green are too expensive.

MORE PROPOSITIONS. We often want to consider statements that depend on the value of a variable. For example, for each $x \in \mathbb{R}$, define

$P(x)$: $x^2 - 4x + 1 \geq 2$.

This defines an infinite collection of propositions, one proposition for each real number. Some of these propositions are true, and others are false. For example $P(6)$ and $P(2 + \sqrt{5})$ are true, and $P(1)$ is false.

Using these propositions, the set $Y = \{x \in \mathbb{R} : x^2 - 4x + 1 \geq 2\}$ seen in Example 1.8 can be defined by

$$Y = \{x \in \mathbb{R} : P(x)\}.$$

Note that it only makes sense to ask if $P(x)$ is true *when x is a specific real number*.

In this section we will consider propositions that depend on a natural number. Recall that the set of natural numbers $\mathbb{N}$ is $\{1, 2, 3, \ldots\}$.

Example 4.1. For $n \in \mathbb{N}$ define

$Q(n)$: $n^2 + n + 41$ is a prime number

So we have defined propositions

$Q(1)$: $1^2 + 1 + 41$ is a prime number

$Q(2)$: $2^2 + 2 + 41$ is a prime number

$Q(3)$: $3^2 + 3 + 41$ is a prime number

and so on. In this case $Q(1), Q(2), \ldots, Q(39)$ are all true propositions. But $Q(40)$ and $Q(41)$ are false.

Example 4.2. For $n \in \mathbb{N}$ define

$P(n)$: The sum of the odd numbers from 1 up to and including $2n - 1$ is equal to n^2.

So we have defined propositions

$$P(1):\ 1 = 1^2$$
$$P(2):\ 1 + 3 = 2^2$$
$$P(3):\ 1 + 3 + 5 = 3^2$$

and so on. If you look at a few more cases you will probably be convinced that $P(n)$ is true for every $n \in \mathbb{N}$. But Example 4.1 shows it can be dangerous to make conjectures on limited evidence!

The proposition $P(n)$ in Example 4.2 can also be written as

$$P(n):\ 1 + 3 + \cdots + (2n - 1) = n^2.$$

Here $\cdots$ indicates that you should continue the pattern by adding odd numbers until $2n - 1$ is reached. (We will later see Σ notation, which gives an alternative way to write sums like this.)

THE PRINCIPLE OF MATHEMATICAL INDUCTION. Suppose that $P(n)$ is a proposition for each $n \in \mathbb{N}$. The Principle of Mathematical Induction states that if

- $P(1)$ is true
- $P(n) \implies P(n+1)$ for each $n \in \mathbb{N}$,

then $P(n)$ is true for all $n \in \mathbb{N}$.

In this course we shall take it for granted that the Principle of Mathematical Induction is a legitimate way to prove things. This is justified in 194 Numbers and Functions using properties of the real numbers.

INFORMAL JUSTIFICATION. Here is a (non-examinable) informal argument for the Principle of Mathematical Induction.

Let $P(n)$ be a proposition for each $n \in \mathbb{N}$. Suppose the hypotheses for the Principle of Mathematical Induction hold, so $P(1)$ is true, and $P(n) \implies P(n+1)$ for each $n \in \mathbb{N}$. Imagine somehow challenges you to prove that $P(m)$ is true for a particular $m \in \mathbb{N}$. You can then say

Since $P(1)$ is true, and $P(1) \implies P(2)$, $P(2)$ is true.
Since $P(2)$ is true, and $P(2) \implies P(3)$, $P(3)$ is true.

and so on, eventually speaking the sentence

Since $P(m-1)$ is true, and $P(m-1) \implies P(m)$, $P(m)$ is true.

At this point your challenger must be convinced! This argument works for all $m \in \mathbb{N}$, so $P(m)$ is true for all $m \in \mathbb{N}$.

EXAMPLES OF INDUCTION.

Example 4.3. For all $n \in \mathbb{N}$ we have

$$1+2+\cdots+n = \frac{n(n+1)}{2}.$$

Proof. For $n \in \mathbb{N}$ let $P(n)$ be the proposition

$$P(n):\ 1+2+\cdots+n = \frac{n(n+1)}{2}.$$

The proposition $P(1)$ asserts that $1 = 1(1+1)/2$. Clearly this is true. Suppose, inductively, that $P(n)$ is true. Then

$$\begin{aligned} 1+2+\cdots+n+(n+1) &= (1+2+\cdots+n)+(n+1) \\ &= \frac{n(n+1)}{2}+(n+1) \end{aligned}$$

where we used $P(n)$ to replace $1+2+\cdots+n$ with $n(n+1)/2$. Factoring out $n+1$ we get

$$\begin{aligned} 1+2+\cdots+n+(n+1) &= (n+1)\Big(\frac{n}{2}+1\Big) \\ &= \frac{(n+1)(n+2)}{2}. \end{aligned}$$

Therefore $P(n) \implies P(n+1)$ for each $n \in \mathbb{N}$. Hence, by the Principle of Mathematical Induction, $P(n)$ is true for all $n \in \mathbb{N}$. □

You may prefer to abbreviate the final two sentences, and write something like 'Hence by induction, $P(n)$ is true for all n'.

Proofs by induction all fit into the same template.

(1) Formulate the statement you want to prove as a proposition $P(n)$, depending on a natural number n.

(2) Prove $P(1)$. This is called the *base case*.

(3) Prove that $P(n) \implies P(n+1)$ for each $n \in \mathbb{N}$. In other words: **assume** $P(n)$ **and use it to prove** $P(n+1)$. This is called the *inductive step*.

(4) Announce that you have finished!

Example 4.4. For $n \in \mathbb{N}$ define

$$P(n):\ 2^{2n}-1 \text{ is a multiple of } 3.$$

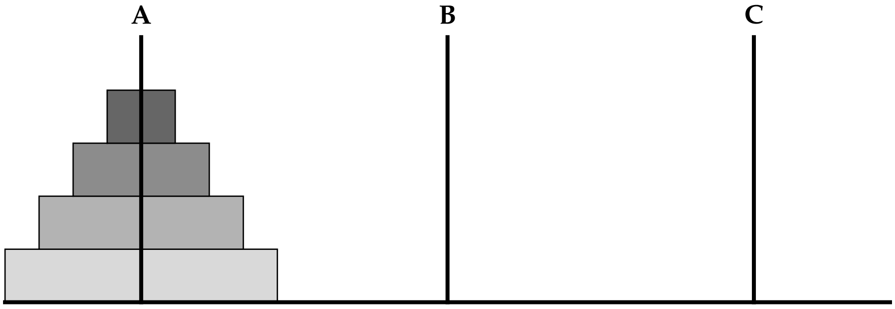
Then $P(n)$ is true for all $n \in \mathbb{N}$.

Sometimes we need to take the base case to be $P(b)$ for some $b > 1$.

Example 4.5. If $n \in \mathbb{N}$ and $n \geq 4$ then $2^n \geq 4n$.

Here is a more substantial example of induction.

Problem 4.6 (Towers of Hanoi)**.** You are given a board with three pegs. On peg **A** there are n discs of strictly increasing radius. The starting position for a four disc game is shown below.



A *move* consists of taking a single disc at the top of the pile on one peg, and moving it to another peg. **At no point may a larger disc be placed on top of a smaller disc.** Your aim is to transfer all the discs from peg **A** to one of the other pegs. How many moves are required?

Exercise 4.7. Prove by induction on n that no solution can use fewer moves than the solution found in lectures.

Sigma notation. Let $m, n \in \mathbb{Z}$ with $m \leq n$. If $a_m, a_{m+1}, \ldots, a_n$ are complex numbers then we write their sum $a_m + a_{m+1} + \cdots + a_n$ as

$$\sum_{k=m}^{n} a_k.$$

This is read as 'the sum of a_k for k from m to n', or 'sigma a_k for k from m to n'. We say that k is the *summation variable*, m is the *lower limit* and n is the *upper limit*.

Using Σ notation we can rewrite the sums seen earlier using the $\cdots$ notation. For example, $1 + 2 + \cdots + n = \sum_{k=1}^{n} k$. We have

$$\sum_{k=1}^{1} k = 1, \quad \sum_{k=1}^{2} k = 1 + 2 = 3, \quad \sum_{k=1}^{3} k = 1 + 2 + 3 = 6, \quad \text{and so on.}$$

There is no mathematical difference between $\displaystyle\sum_{k=1}^{n} a_k$ and $\sum_{k=1}^{n} a_k$.

Exercise 4.8.

(i) Express the sums $1+3+\cdots+(2n-1)$ and $1+2+2^2+\cdots+2^n$ using Σ notation.

(ii) Calculate $\sum_{m=-2}^{3} m^2$.

At A-level you might have written $\sum_1^n a_k$ rather than $\sum_{k=1}^n a_k$. Omitting the summation variable k from the limits of the sum can lead to ambiguities. For example, consider

$$1^m + 2^m + \cdots + n^m = \sum_{k=1}^{n} k^m.$$

If we write $\sum_1^n k^m$ for this sum, then it is no longer clear that k should vary while m is fixed.

Example 4.9. Let z be a complex number. Then

(i) $\sum_{k=1}^n z = nz$,

(ii) $\sum_{k=1}^n k = n(n+1)/2$,

(iii) $\sum_{k=1}^n n = n^2$.

RULES FOR MANIPULATING SIGMA NOTATION.

(1) The summation variable can be renamed:

$$\sum_{k=0}^{n} 2^k = \sum_{j=0}^{n} 2^j.$$

A similar renaming is possible for sets: $\{x \in \mathbb{R} : x^2 \geq 2\}$ is exactly the same set as $\{y \in \mathbb{R} : y^2 \geq 2\}$.

(2) In a product, expressions not involving the summation variable can be taken outside the sum:

$$\sum_{j=0}^{n} 5(j+1)^2 = 5\sum_{j=0}^{n}(j+1)^2$$

and

$$\sum_{j=0}^{n} 5m(j+m)^2 = 5m\sum_{j=0}^{n}(j+m)^2.$$

(3) Sums can be split up:

$$\sum_{j=0}^{n}(2^j + j^2) = \sum_{j=0}^{n} 2^j + \sum_{j=0}^{n} j^2,$$

and terms taken out: $\sum_{k=0}^n a_k = a_0 + \sum_{k=1}^n a_k = \sum_{k=0}^{n-1} a_k + a_n$. The latter is often useful in proofs by induction.

(4) The limits can be shifted. For example, if $x \in \mathbb{R}$ then

$$\sum_{k=0}^{n-1} (k+1)x^k = \sum_{r=1}^{n} rx^{r-1}.$$

We replaced k with $r-1$. The original sum had k varying from 0 to $n-1$. Hence $r-1$ should also vary from 0 to $n-1$, and so r should vary from 1 to n.

It is no coincidence that integrals can be manipulated in similar ways.

An example combining Sigma notation and induction is given below. The inductive part of the proof is very similar to Example 4.3.

Example 4.10. Define

$$P(n) \colon \sum_{k=1}^{n} k^2 = \tfrac{1}{6}n(n+1)(2n+1).$$

The proposition $P(1)$ asserts that

$$\textstyle\sum_{k=1}^{1} k^2 = \tfrac{1}{6}1(1+1)(2 \times 1 + 1).$$

This is true, because the left-hand side is $1^2 = 1$ and the right-hand side is $\frac{1}{6}(1 \times 2 \times 3) = 1$.

Now consider $\sum_{k=1}^{n+1} k^2$. Split off the final summand using rule (3), and then use the inductive assumption $P(n)$ to get

$$\sum_{k=1}^{n+1} k^2 = \sum_{k=1}^{n} k^2 + (n+1)^2 = \tfrac{1}{6}n(n+1)(2n+1) + (n+1)^2.$$

Routine algebraic manipulations give

$$\begin{aligned}\sum_{k=1}^{n+1} k^2 &= \tfrac{1}{6}(n+1)\bigl(n(2n+1) + 6(n+1)\bigr)\\ &= \tfrac{1}{6}(n+1)\bigl(2n^2+7n+6\bigr)\\ &= \tfrac{1}{6}(n+1)(n+2)(2n+3).\end{aligned}$$

Hence $P(n+1)$ is true. Therefore $P(n) \implies P(n+1)$. By induction $P(n)$ is true for all $n \in \mathbb{N}$.

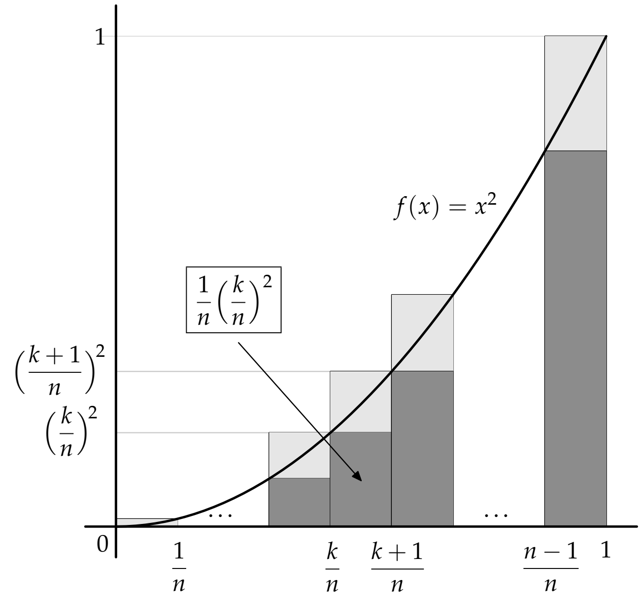
An application of induction and Sigma notation.

Example 4.11. Consider the graph opposite. From the dark boxes we get

$$\sum_{k=1}^{n-1} \frac{1}{n}\Bigl(\frac{k}{n}\Bigr)^2 \le \int_0^1 x^2 \, \mathrm{d}x$$

and from the light boxes (which are just the dark boxes shifted to the left, with one extra box of area $1/n$ at the far right), we get

$$\sum_{k=1}^{n-1} \frac{1}{n}\Bigl(\frac{k}{n}\Bigr)^2 + \frac{1}{n} \ge \int_0^1 x^2 \, \mathrm{d}x.$$


1
$f(x) = x^2$
$\frac{1}{n}\left(\frac{k}{n}\right)^2$
$\left(\frac{k+1}{n}\right)^2$
$\left(\frac{k}{n}\right)^2$
0
$\frac{1}{n}$
$\frac{k}{n}$
$\frac{k+1}{n}$
$\frac{n-1}{n}$
1
$\cdots$
$\cdots$

By Example 4.10 we have

$$\begin{aligned}\sum_{k=1}^{n-1}\frac{1}{n}\left(\frac{k}{n}\right)^2 &= \frac{1}{n^3}\sum_{k=1}^{n-1}k^2\\ &= \frac{(n-1)n(2n-1)}{6n^3}\\ &= \frac{1}{3}\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right).\end{aligned}$$

Hence

$$\frac{1}{3}\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right) \le \int_0^1 x^2\,\mathrm{d}x \le \frac{1}{3}\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)+\frac{1}{n}.$$

Now $1/n$ converges to 0 as n tends to infinity, and so the left-hand side and right-hand side both converge to $1/3$. (If you are doing 194 Numbers and Functions you could prove this using the Algebraic Limit Theorem.) Hence $\int_0^1 x^2\,\mathrm{d}x$ is sandwiched between two sequences that converge to $1/3$. Therefore

$$\int_0^1 x^2\,\mathrm{d}x = \frac{1}{3}.$$

Exercise 4.12. The light box containing the dark box marked in the diagram has area $\frac{1}{n}\left(\frac{k+1}{n}\right)^2$. Calculating the area of the light boxes this way gives

$$\sum_{k=0}^{n-1}\frac{1}{n}\left(\frac{k+1}{n}\right)^2$$

Use Rules (3) and (4) to show that this is equal to $\sum_{k=1}^{n-1}\frac{1}{n}\left(\frac{k}{n}\right)^2+\frac{1}{n}$, which is the expression used above.

5. Prime numbers

In this section we will look at prime numbers and prime factorizations. One highlight will be Euclid's proof (from approx 300 BCE) that there are infinitely many primes. We will also see a quick way to prove that $\sqrt{2}$, $\sqrt{3}$, $\sqrt[3]{5}$, and many other similar numbers, are all irrational.

Integer division. Division with remainder should be familiar from school. It is stated formally in the next theorem.

Theorem 5.1 (Examinable). *Let $n \in \mathbb{Z}$ and let $m \in \mathbb{N}$. There exist unique integers q and r such that $n = qm + r$ and $0 \leq r < m$.*

The proof shows that $q = \lfloor n/m \rfloor$ where $\lfloor x \rfloor$ is the floor function, seen in Question 3 of Sheet 2. So the existence part of the proof gives an effective way to find q.

We say that q is the *quotient*, and r is the *remainder* when n is divided by m. If $r = 0$ then we say that *m divides n*, or that n is a *multiple* of m.

Example 5.2.

(i) Let $n = 44$ and $m = 6$. Then $44/6 = 7\frac{2}{6}$ and so, when 44 is divided by 6, the quotient is 7 and the remainder is 2. Note that for this calculation it is better to leave the fractional part as $\frac{2}{6}$ than to simplify it to $\frac{1}{3}$.

(ii) Let $n = 63$ and $m = 7$. Then $63/7 = 9$ so 7 divides 63. The quotient is 9 and the remainder is 0.

(iii) Since $-13 = -3 \times 6 + 5$, when -13 is divided by 6 the quotient is -3 and the remainder is 5.

Exercise 5.3. Find the quotient q and the remainder r when n is divided by m in each of these cases:

(i) $n = 20, m = 7$, (ii) $n = 21, m = 7$, (iii) $n = 22, m = 7$
(iv) $n = 7, m = 22$, (v) $m = -10, m = 7$, (vi) $n = 0, m = 1$.

The answers are available from the Part B Slides on Moodle.

Factorization into primes.

Definition 5.4. Let $n \in \mathbb{N}$ and suppose that $n > 1$.

(i) We say that n is *prime* if the only natural numbers that divide n are 1 and n.

(ii) We say that n is *composite* if it is not prime.

The first few prime numbers are

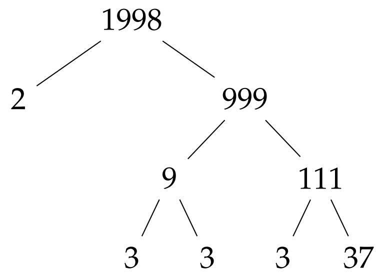
$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 37,\ 41,\ 43,\ 47,\ \ldots.$$

By Definition 5.4, the number 1 is neither prime nor composite.

Example 5.5. Take $n = 1998$. We might spot that $n = 2 \times 999$ and that $999 = 9 \times 111$. Then $9 = 3 \times 3$, and $111 = 3 \times 37$, so

$$1998 = 2 \times 3 \times 3 \times 3 \times 37 = 2 \times 3^3 \times 37.$$

The tree below records these steps. (For some reason mathematical trees usually grow downwards.)



More generally, let $n > 1$ be a natural number

- If n is prime then n is equal to a product of primes (since $n = n$).
- If n is composite then n is divisible by some $m \in \mathbb{N}$ such that $1 < m < n$. So $n/m \in \mathbb{N}$. Let $m' = n/m$. Repeating this argument with m and m' shows that n is equal to a product of prime numbers.

Therefore any number has a *factorization* as a product of prime numbers. For example, $1998 = 2 \times 3^3 \times 37$ and $31460 = 2^2 \times 5 \times 11^2 \times 13$.

INFINITELY MANY PRIMES. The next theorem, due to Euclid, needs only the existence of prime factorizations, proved above.

Theorem 5.6 (Examinable). *There are infinitely many prime numbers.*

How much should one trust a proof? Euclid's proof is a mathematical gem that has been understood and enjoyed by mathematicians since 300 BCE. Can any reasonable person doubt that there are infinitely many primes?

Exercise 5.7. The first five prime numbers are $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, $p_5 = 11$, $p_6 = 13$. Show that $p_1 + 1$, $p_1p_2 + 1$, $p_1p_2p_3 + 1$, $p_1p_2p_3p_4 + 1$ and $p_1p_2p_3p_4p_5 + 1$ are all prime, but

$$\begin{aligned} p_1p_2p_3p_4p_5p_6 + 1 &= 2 \times 3 \times 5 \times 7 \times 11 \times 13 + 1 \\ &= 59 \times 509. \end{aligned}$$

This shows that the number N in Euclid's proof is not always prime.

UNIQUE FACTORIZATION. Let $\mathbb{N}_0$ be the set $\{0, 1, 2, 3, \ldots\}$ of the natural numbers *together with* 0.

Theorem 5.8 (Fundamental Theorem of Arithmetic)**.** *Let $n \in \mathbb{N}$. Let $p_1, p_2, p_3, \ldots$ be the primes in increasing order. There exists unique $e_i \in \mathbb{N}_0$ such that*

$$n = p_1^{e_1} p_2^{e_2} p_3^{e_3} \ldots.$$

We have already proved the existence part of the following theorem. The proof of uniqueness is not examinable, **but you need to understand its statement**.

Writing out prime factorizations in the form in this theorem is a bit long-winded. For example

$$31460 = 2^2 \times 3^0 \times 5^1 \times 7^0 \times 11^2 \times 13^1 \times 17^0 \times 19^0 \ldots,$$

where all the exponents of the primes 17 or more are zero. But thinking about prime factorizations in this way is useful in proofs.

The reason why 1 is not defined to be a prime number is because this would destroy unique factorization. For instance,

$$20 = 2^2 \times 5 = 1 \times 2^2 \times 5 = 1^2 \times 2^2 \times 5 = \ldots$$

would all be different prime factorizations of 20.

The uniqueness property is very powerful: it is proved in the extras for §8. The extras for this section might convince you that it is not obvious. (As ever, all the extras are optional and non-examinable.)

IRRATIONAL NUMBERS. As an example of how to use unique factorization we will prove that $\sqrt{3}$ is irrational. The exercise below shows the key idea.

Example 5.9. A manufacturer of cheap calculators claims to you that $\sqrt{3} = \frac{2148105}{1240209}$. Calculate the prime factorizations of 2148105 and 1240209 (in principle you could do this by repeated division, even using one of his cheapest calculators). Hence show that he is wrong.

Note the critical role played by the powers of 3 dividing 2148105 and 1240209.

Claim 5.10. *$\sqrt{3}$ is an irrational number.*

We will prove Claim 5.10 using *proof by contradiction*. While related, this is not the same proof by contradiction you will have seen if you are doing 194 Numbers and Functions.

The same argument goes through with minor changes to prove that $\sqrt{5}$, $\sqrt{8}$, $\sqrt[3]{5}$, and so on, are all irrational. For instance, to prove that $\sqrt[3]{5}$ is irrational, use the argument above, but replace 5 with 3 and cube instead of square.

BINARY AND OTHER BASES. We will now see another application of the integer division defined at the start of this section.

Example 5.11. To write 145 in base 3:

Divide 145 by 3:	$145 = 48 \times 3 + \mathbf{1}$
Divide the quotient 48 by 3:	$48 = 16 \times 3 + \mathbf{0}$
Divide the quotient 16 by 3:	$16 = 5 \times 3 + \mathbf{1}$
Divide the quotient 5 by 3:	$5 = 1 \times 3 + \mathbf{2}$
Divide the quotient 1 by 3:	$1 = 0 \times 3 + \mathbf{1}$

We now stop, because the last quotient was 0. Reading the list of remainders from bottom to top we get

$$145 = 1 \times 3^4 + 2 \times 3^3 + 1 \times 3^2 + 0 \times 3^1 + 1 \times 3^0.$$

Hence 145 is 12101 in base 3. We write this as $145 = 12101_3$.

Our usual way of writing numbers uses base 10. If no base is specified, as is usually the case, then base 10 is intended.

The example above should suggest a general algorithm.

Algorithm 5.12. *Let $n \in \mathbb{N}$ and let $b \in \mathbb{N}$. To write n in base b, divide n by b, then divide the quotient by b, and so on, until the quotient is 0. If $r_0, r_1, r_2, \ldots, r_k$ is the sequence of remainders then*

$$n = r_k b^k + r_{k-1} b^{k-1} + \cdots + r_1 b + r_0$$

and $n = (r_k r_{k-1} \ldots r_1 r_0)_b$.

In Example 5.11, the base was 3 and the sequence of remainders was $r_0 = 1, r_1 = 0, r_2 = 1, r_3 = 2$ and $r_4 = 1$.

If time permits we will prove that the algorithm is correct by induction on k, taking as the base case $k = 0$.

Base 2 is known as *binary*. Binary is particularly important because computers store and process data as sequences of the *bi*nary digi*ts*, or *bits*, 0 and 1. For a nice introduction to programming at the level of bits, see `pleasingfungus.com/Manufactoria/`.

Exercise 5.13. Show that $21 = 10101_2$ and write 63, 64 and 65 in binary.

Exercise 5.14. Let $n = r_k r_{k-1} \dots r_1 r_0$ be a binary number. Describe, in terms of operations on the string of bits $r_k r_{k-1} \dots r_1 r_0$, how to

(i) Multiply n by 2,
(ii) Add 1 to n,
(iii) Subtract 1 from n,
(iv) Find the quotient and remainder when n is divided by 2.

[*Hint:* for base 10, you probably learned how to do these at school. The MATHEMATICA command `BaseForm[n,2]` will write $n \in \mathbb{N}_0$ in binary.]

EXTRAS: UNIQUE FACTORIZATION. Let S be the set of all $n \in \mathbb{N}$ such that the remainder when n is divided by 4 is 1. So

$$S = \{1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, \dots\}.$$

Say, just for this section, that $n \in S$ is *S-prime* if $n \neq 1$ and the only elements of S that divide n are 1 and n.

Exercise 5.15. Show that 5 is S-prime. Show, more surprisingly, that 9, 21 and 49 are all S-primes. Give an example of a number in S, apart from 1, that is not S-prime.

Now observe that $9 \times 49 = 21 \times 21$. The three numbers 9, 21, 49, are all S-primes. So although S-primes are defined very like normal primes, in the world of S-primes, unique factorization does not hold.

EXTRAS: STRONG INDUCTION. In the proof of the existence part of the Fundamental Theorem of Arithmetic, we said 'Repeating this argument with m and m' ...'. This can be made more formal using *strong induction*. In a strong inductive proof, you may assume all of $P(1), \dots, P(n)$ when proving $P(n+1)$.

Exercise 5.16. For $n \in \mathbb{N}$ define

$P(n)$: Either $n = 1$ or n is equal to a product of prime numbers.

Prove that $P(n)$ is true for all $n \in \mathbb{N}$ by strong induction.

For more on strong induction see Chapter 8 of *A concise introduction to pure mathematics* by Martin Liebeck, [2] in the recommended reading list on page 2. (Or page 177 of [1] also has a very brief account.)

EXTRAS: MORE ON PRIMES. Prime numbers are an important area of mathematical research. To learn more, try searching on the web for: 'Prime Number Theorem', 'Green–Tao Theorem', or 'Goldbach's Conjecture'. The Twin Primes Conjecture states there are infinitely many primes p such that both p and $p + 2$ are primes. In April 2013 Yitang Zhang made a huge step towards proving this conjecture, by showing that there are infinitely many pairs of prime numbers that differ by less than 70 million.

Part C: Logic and sets

6. LOGIC

MOTIVATION: AMBIGUOUS SENTENCES. Logical reasoning is at the heart of mathematics. In pairs discuss the meaning of the following sentences. Each has two interpretations that are logically reasonable.

(a) The picture of the woman in the museum.
(b) The lady hit the man with an umbrella.
(c) Nurses help dog bite victim.
(d) Did you see the girl with the telescope?
(e) Walk to Windsor or swim the Channel and climb the Matterhorn.

The ambiguities in everyday language are often resolved, either from the context, or because we are conditioned to expect one meaning.

In mathematics we instead try to avoid ambiguity by careful use of mathematical language and symbols. Mathematical language has some usages that may seem strange at first.

One such usage is word 'let'. For example, the proof that $\mathbb{Q}$ is closed under addition (before Exercise 1.7), started 'Let $x \in \mathbb{Q}$ and let $y \in \mathbb{Q}$.' This means that x and y are arbitrary rational numbers: the proof has to work for all of them.

'AND', 'OR' AND 'NOT'. Another word that is used in mathematics in a way that may seem non-standard is 'or'. Let P and Q be propositions. (Remember, this means P and Q can be any propositions!)

(i) *P or Q*, written $P \vee Q$, means at least one of P and Q is true.
(ii) *P and Q*, written $P \wedge Q$, means P and Q are both true.
(iii) *not P*, written $\neg P$, means that P is false.

Parts (b) and (c) in the next example show that there is a correspondence between the logical operations $\wedge$, $\vee$ and $\neg$ and the set operations $\cap$, $\cup$ and set complement.

Example 6.1. Consider the following propositions, depending on a natural number n.

$$\begin{aligned} P(n)&: \text{ n is even} \\ Q(n)&: \text{ n is a multiple of 3} \end{aligned}$$

(a) In words, $\neg P(n) \wedge Q(n)$ states 'n is an odd number and n is a multiple of 3'.

(b) A short way to state $P(n) \wedge Q(n)$ is 'n is divisible by 6'. Correspondingly

$$\{n \in \mathbb{N} : P(n)\} \cap \{n \in \mathbb{N} : Q(n)\} = \{n \in \mathbb{N} : P(n) \wedge Q(n)\}.$$

(c) The negation of $P(n)$ is 'n is odd', and correspondingly

$$\{n \in \mathbb{N} : n \text{ is even}\}' = \{n \in \mathbb{N} : n \text{ is odd}\}.$$

To avoid ambiguities such as those in sentence (e) above, the parentheses '(' and ')' are used.

Exercise 6.2. Let $R(n) = \big(P(n) \vee Q(n)\big) \wedge \neg\big(P(n) \wedge Q(n)\big)$ where $P(n)$ and $Q(n)$ are as in Example 6.1. **[Corrected 14th November.]**

(a) State $R(n)$ in words.

(b) Draw a Venn diagram representing the sets $\{n \in \mathbb{N} : P(n)\}$, $\{n \in \mathbb{N} : Q(n)\}$, $\{n \in \mathbb{N} : R(n)\}$ and $\{n \in \mathbb{N} : \neg P(n) \wedge Q(n)\}$.

TRUTH TABLES. A concise way to specify a logical operation such as $\vee$, $\wedge$ or $\neg$ is by a *truth table*, such as the one below for $\vee$.

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

Truth tables can be used to prove logical identities. The next result is the analogue of Claim 1.10 for propositions.

Claim 6.3 (De Morgan's Laws for propositions). *Let P and Q be propositions. Then the following are true:*

(i) $\neg(P \vee Q) \iff \neg P \wedge \neg Q$,

(ii) $\neg(P \wedge Q) \iff \neg P \vee \neg Q$.

The proof of (ii) is left to you in Question 2(b) on Sheet 7. The truth table method is easy and systematic. But it is sometimes more illuminating to argue directly.

IMPLICATION, LOGICAL EQUIVALENCE AND TAUTOLOGIES. We have already used $\Longrightarrow$ 'implies' and $\iff$ 'if and only if' many times. Let P and Q be propositions. Stated formally:

- $P \Longrightarrow Q$ means that if P is true then Q is true.
- $P \iff Q$ means that $P \Longrightarrow Q$ and $Q \Longrightarrow P$.

If $P \iff Q$ is true then we say that P and Q are *logically equivalent*. For example, by Claim 6.3(i), the propositions $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$ are logically equivalent.

If a proposition is always true, then it is said to be a *tautology*. For instance

$$(P \iff Q) \iff \big((P \implies Q) \wedge (Q \implies P)\big)$$

is a tautology, and so is $P \iff \neg(\neg P)$.

See the extras for some of the equivalent ways to state implications: in this course $\implies$, $\iff$ and 'if P then Q' will be enough. But other formulations can be more expressive, and will be met in other courses.

TRUTH TABLE FOR $\implies$. By definition, $P \implies Q$ means 'if P is true then Q is true'. If P is false then this statement makes no claim about Q. Therefore if P is false then $P \implies Q$ is true.

This may seem surprising to you. But it is consistent with how we use implication. Think of $P \implies Q$ as a promise: if P is true, then Q is true. If P is false, then it does not matter whether Q is true or not: the promise is still kept.

The truth table for $P \implies Q$ is shown below.

P	Q	$P \implies Q$
T	T	T
T	F	F
F	T	T
F	F	T

Notice that there is only one false in the column for $P \implies Q$.

$P \implies Q$ is false if and only if P is true and Q is false

This might be useful for Sheet 7, particularly Question 3. See the extras for some discussion of the use of implication in everyday language.

Example 6.4. Logically $P \implies Q$ says nothing about $Q \implies P$. For example, consider these two propositions concerning a real number x.

$$P(x):\ x \geq 2 + \sqrt{5}$$
$$Q(x):\ x^2 - 4x + 1 \geq 2$$

Then $P(x) \implies Q(x)$ for every $x \in \mathbb{R}$. But $Q(-10)$ is true and $P(-10)$ is false, so $Q(-10) \not\Longrightarrow P(-10)$. Correspondingly, as seen in Example 1.8,

$$\{x \in \mathbb{R} : P(x)\} \subseteq \{x \in \mathbb{R} : Q(x)\},$$

and -10 is in the right-hand set, but not in the left-hand set.

Exercise 6.5. Let x and y be real numbers. Which of the following propositions are true?

(a) $x \geq 4 \implies x \geq 3$,
(b) $x \geq 3 \implies x \geq 4$,
(c) $x^2 - 2x - 3 = 0 \implies x = -1, x = 3$ or $x = 37$,
(d) $x \geq 0$ and $x^2 - 2x - 3 = 0 \implies x = 3$,
(e) $x^2 = y^2 \implies x = y$,
(f) $x^3 = y^3 \implies x = y$.

When can $\implies$ be replaced with $\iff$?

PROOF BY CONTRAPOSITIVE. The next claim can be proved easily using a truth table (see Part C Slides). But it is may be more illuminating to argue directly.

Claim 6.6. *Let P and Q be propositions. Then $P \implies Q$ and $\neg Q \implies \neg P$ are logically equivalent.*

Switching to the contrapositive can be useful first step in a proof, particularly when statements appear in negated form.

Claim 6.7. *Let $a \in \mathbb{Q}$ and let $x \in \mathbb{R}$. If $x \notin \mathbb{Q}$ then $x + a \notin \mathbb{Q}$.*

USING IMPLICATION TO CLARIFY PROOFS. It is often tempting to start with the statement we are trying to prove, and manipulate it until it becomes obviously true. **But this is only valid if every step is reversible**.

Example 6.8. Suppose we want to find all $x \in \mathbb{R}$ such that

$$\sqrt{x+3} = x + 1.$$

We might argue as follows:

$$\begin{aligned}
\sqrt{x+3} = x + 1 &\implies (x+3) = (x+1)^2 \\
&\implies x + 3 = x^2 + 2x + 1 \\
&\implies x^2 + x - 2 = 0 \\
&\implies (x+2)(x-1) = 0 \\
&\implies x = -2 \text{ or } x = 1.
\end{aligned}$$

There is nothing logically incorrect so far. But note that $\sqrt{-2+3} = \sqrt{1} \neq -2 + 1$. So it would be wrong to deduce that -2 and 1 are both solutions.

Without any implication signs at all, proofs can become frustratingly difficult to read. It is not enough to have a clear intention in mind: you must communicate it to the reader. Here is an example from Question 5(b) on Sheet 5.

Exercise 6.9. Criticize and improve the following proof that $2^n \geq 6n$ for all n such that $n \geq 5$.

> $P(n) = 2^n \geq 6n$ where $n \geq 5$.
>
> $P(5) = 2^5 \geq 6 \times 5$. True.
>
> $P(n+1)$ where $n \in \mathbb{N}$, $n \geq 5$.
>
> > $2^{n+1} \geq 6(n+1)$
> >
> > $2^n \geq 3n + 3$
> >
> > $6n \geq 3n + 3$
> >
> > $3n \geq 3$
>
> Hence by the Principle of Mathematical Induction, $P(n)$ is true for all $n \in \mathbb{N}$ when $n \geq 5$.

PROVING THAT TWO SETS ARE EQUAL. Question 2(c) of Sheet 7 shows that $(P \vee Q) \wedge R$ and $(P \wedge R) \vee (Q \wedge R)$ are logically equivalent. The corresponding set theory identity is (i) below.

Claim 6.10. *Let X, Y and Z be sets. Then*

(i) $(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z)$,
(ii) $(X \cap Y) \cup Z = (X \cup Z) \cap (Y \cup Z)$.

The proof will use the following principle: if A and B are sets then

$$A = B \iff A \subseteq B \text{ and } B \subseteq A.$$

This is often a good way to show that two sets are equal.

'FOR ALL' AND 'EXISTS'. Let $P(x)$ be a propositions depending on an element x of a set X.

- If $P(x)$ is true for all $x \in X$, then we write $(\forall x \in X)\, P(x)$.
- If there exists an element $x \in X$ such that $P(x)$ is true, then we write $(\exists x \in X)\, P(x)$.

The parentheses around $\forall x \in X$ and $\exists x \in X$ are often omitted.

The negation of

- $(\forall x \in X)\, P(x)$ is $(\exists x \in X)\, \neg P(x)$.
- $(\exists x \in X)\, P(x)$ is $(\forall x \in X)\, \neg P(x)$.

Once you have understood the meaning of the $\forall$ and $\exists$ symbols, these rules should seem fairly obvious to you. Negating long compound statements becomes routine if you apply these rules step-by-step.

Exercise 6.11. Sometimes the set X in $\forall x \in X$ is indicated by inequalities. For example,

$(\forall \epsilon > 0)\, Q(\epsilon)$	means that $Q(\epsilon)$ is true for all ϵ in the set of positive real numbers,
$(\forall n \geq N)\, S(n)$	means that $S(n)$ is true for all $n \in \mathbb{N}$ such that $n \geq N$.

Let $a_1, a_2, a_3, \ldots$ be real numbers. Write down the negation of

$$(\exists \ell \in \mathbb{R})(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall n \geq N)\ |a_n - \ell| < \epsilon.$$

If you are doing MT194 Numbers and Functions then you should have noticed that a logically equivalent statement is 'the sequence (a_n) converges'.

DEFINITIONS. Definitions are a frequent source of confusion. Here is the definition of prime from Definition 5.4.

> Let $n \in \mathbb{N}$ and suppose that $n > 1$. We say that n is *prime* if the only natural numbers that divide n are 1 and n.

By convention, when 'if' is written in a definition, it means 'if and only if'. (Question 1 on Sheet 7 asks you to read Chapter 15 of *How to think like a mathematician*; this convention is discussed on page 104.) So the definition of prime for a natural number n can be restated as

$$n > 1 \text{ and } n \text{ is only divisible by 1 and } n \iff n \text{ is prime.}$$

It is important to realise that this definition of prime is complete as it stands. All the other properties of prime numbers, for example, the Fundamental Theorem of Arithmetic, Euclid's Theorem that there are infinitely many primes, and so on, all follow from this basic definition.

Exercise 6.12. In examinations you may be asked to give some definitions. Here is a typical question and a sadly not atypical answer.

> 'What does it mean to say that $f : X \to Y$ is injective'?
>
> **Answer:** $f(x) \neq f(x')$. So f is injective if there aren't two arrows in the same dot, like $f(x) = x + 1$.

Criticize this answer.

EXTRAS: ONLY IF, NECESSARY, SUFFICIENT. As usual, let P and Q be propositions. The following are all different ways to state $P \implies Q$:

- if P then Q,
- Q if P,
- P only if Q,
- P is sufficient for Q,
- Q is necessary for P.

Note that reading $P \iff Q$ as 'P if and only if Q' is consistent with the meaning of 'only if' just stated: 'P if Q' means $Q \implies P$ and 'P only if Q' means $P \implies Q$.

Exercise 6.13. Please assume that the following statements are true. (Arguably these are not propositions, since they are not particularly mathematical, but they are convenient for this example.)

P: If it is raining then the sky is cloudy.
Q: If it rains in the morning then Prof. Z carries her umbrella all day.
R: People who carry umbrellas never get soaked.

Which of the following statements can be deduced from P, Q and R?

A: A cloudy sky is a necessary condition for rain.
B: A cloudy sky is a sufficient condition for rain.
C: It is raining only if the sky is cloudy.
D: Rain in the morning is a necessary condition for Prof. Z to carry her umbrella.
E: Rain in the morning is a sufficient condition for Prof. Z to carry her umbrella.
F: Rain falling from the sky implies that the sky is cloudy.
G: The sky is cloudy implies that rain is falling.
H: If Prof. Z is soaked then it did not rain this morning.

The answers are on the Part C Slides.

EXTRAS: IMPLICATION IN EVERYDAY LANGUAGE. Does the truth table for implication on page 43 reflect the way implication is used in everyday language? Here are two examples.

(a) Every general election someone, call him Y, will make an announcement of the form:

'If party X wins the election then I will leave the country'.

If party X does not win, then Person Y is not committing himself either way: he might stay or he might leave. Either way his announcement will be true. So it seems that Y's announcement is the same as

$$\text{Party } X \text{ wins the election} \implies \text{Person } Y \text{ will leave the country.}$$

(b) However there are other cases which are less clear. For example, consider

'If you don't do your homework, you can't go to the cinema'.

The natural interpretation is that if we do our homework, we can go to the cinema. But this does not logically follow! Let P be 'Did homework', and let Q be 'Went to cinema'. The original statement can reasonably be interpreted as

$$\neg P \implies \neg Q.$$

This does not imply $P \implies Q$. It could be that everyone always does their homework, and no-one ever goes to the cinema.

Exercise 6.14. State the contrapositive of $\neg P \implies \neg Q$ in English. By Claim 6.6, this *does* follow logically from $\neg P \implies \neg Q$.

EXTRAS: MORE LOGIC. Logic is a much deeper subject than this introduction might suggest. It underpins Gödel's incompleteness theorem on the limits of formal mathematical proofs and Turing's work on the relationship between mathematical truth and computability.

There are many good books on these subjects written for the non-expert. For example, *Gödel's Proof* by Ernest Nagel and James Newman, NYU Press (2001).

STUDY SKILLS (PROOFS, THEOREMS AND DEFINITIONS). In MT181 you need to know the proofs of the theorems marked as 'examinable' in the printed notes and proved carefully in lectures. For the other theorems, such as the Fundamental Theorem of Arithmetic, you need to understand the statements. But you are not expected to be able to prove them.

You also need to know all the definitions, both so that you can state them accurately in examinations, and so you can apply them in theorems and examples.

It is *not* necessary to memorise definitions, theorems or proofs verbatim. Any logically correct version will get full marks. For example, 'A function $f : X \to Y$ is *injective* if and only if for all $y \in Y$ the equation $f(x) = y$ has at most one solution', is a perfect alternative to the formulation in Definition 2.5(i).

Learning proofs is hard work, but it pays off in fixing ideas in your head. The best strategy is to break down the proof into key steps. Think about what hypotheses (or background knowledge) each step uses, and how it leads toward the conclusion. If you can remember the key steps, you will be able to fill in the details for yourself.

7. Sets and counting

In this section we will look again at sets and subsets (see pages 4 and 7) and solve some counting problems. We will need the set with no elements. This set is called the *empty set*, and is denoted $\varnothing$.

Sizes of sets.

Definition 7.1. Let X be a set. We say that X is *finite* if it has finitely many elements, and *infinite* otherwise. The *size* of a finite set X is its number of elements. We denote the size of X by $|X|$,

Note that $|X|$ is read as 'mod X'.

Exercise 7.2. State the truth value (true or false) of each of the propositions below.

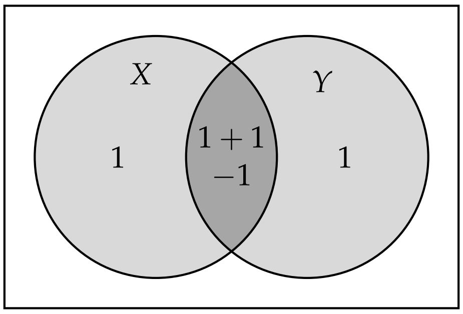
(a) 1 is an element of $\mathbb{N}$.
(b) $\{1\}$ is an element of $\mathbb{N}$.
(c) $| x \in \mathbb{R} : x^2 = -1\}| = 0$.
(d) $|\{z \in \mathbb{C} : z^3 = 1\}| = 3$.
(e) $\big|\{\mathbb{N}, \mathbb{Q}, \{0,1\}\}\big| = 3$.
(f) The set of natural numbers is infinite.
(g) The empty set is a subset of every set;
(h) The empty set is an element of every set;

If X is any set then $\varnothing \subseteq X$ (there is nothing to check), and $X \subseteq X$, since $x \in X \Longrightarrow x \in X$.

Inclusion and Exclusion. Let X and Y be finite sets. In the sum $|X| + |Y|$ we count each element of X once, and each element of Y once. So the elements of $X \cap Y$ are counted twice, once as elements of X, and once as elements of Y. If we subtract $|X \cap Y|$ to correct for this overcounting, we get

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

This is illustrated on the Venn diagram below.



For example, if $z \in X \cap Y$ then z is counted in $|X|$, $|Y|$ and in $|X \cap Y|$, for a total contribution of $1 + 1 - 1 = 1$.

If X and Y are contained in a universe set U then, since

$$|(X \cup Y)'| = |U| - |X \cup Y|$$

we have

$$|(X \cup Y)'| = |U| - |X| - |Y| + |X \cap Y|.$$

Exercise 7.3. At the University of Erewhon, there are 100 students. At each algebra lecture there are 65 students and at each analysis lecture that are 70 students. Let b be the number of students doing both algebra and analysis.

(i) If $b = 50$, how many students are doing neither algebra nor analysis?
(ii) What is the greatest possible value of b?
(iii) What is the least possible value of b?

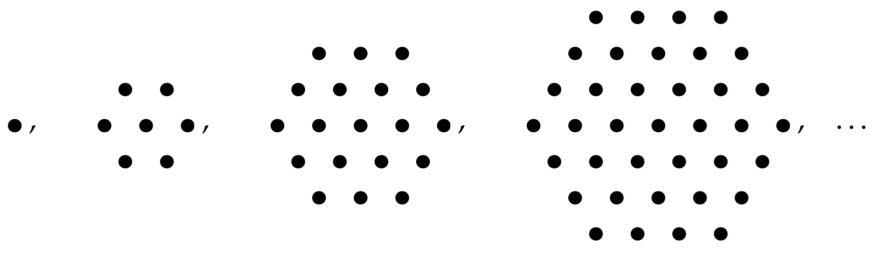
The claim below is the Principle of Inclusion and Exclusion for three sets.

Claim 7.4. *If X, Y and Z are finite sets then*

$$|X \cup Y \cup Z| = |X| + |Y| + |Z| - |X \cap Y| - |Y \cap Z| - |Z \cap X| + |X \cap Y \cap Z|.$$

Exercise 7.5. Suppose that X, Y, Z are subsets of a finite universe set U. Use Claim 7.4 **[corrected from 7.5 on 22nd November]** to write down a formula for the size of $\big|(X \cup Y \cup Z)'\big|$. (This formula will be useful for Question 5(b) on Sheet 8.)

Example 7.6. Let a_n be the number of dots in the nth diagram below. So $a_1 = 1$, $a_2 = 7$, $a_3 = 19$, $a_4 = 37$, and so on.



We will use Claim 7.4 **[corrected from 7.5 on 22nd November]** to find a formula for a_n.

Exercise 7.7. An alternative approach would to be to count dots in rows from top to bottom. This gives

$$a_n = n + \cdots + (2n-2) + (2n-1) + (2n-2) + \cdots + n.$$

Using Sigma notation, an equivalent form is

$$a_n = (2n-1) + 2\sum_{k=n}^{2n-2} k.$$

By expressing the sum as $\sum_{k=1}^{2n-2} k - \sum_{k=1}^{n-1} k$ and using Example 4.3, give an alternative proof of the formula for a_n.

COUNTING SUBSETS. There are four subsets of $\{1,2\}$, namely $\varnothing, \{1\}, \{2\}, \{1,2\}$. **Exercise:** Write down all the subsets of $\{1\}$ and $\{1,2,3\}$. How many are there in each case?

To make a subset X of $\{1,2,3\}$, we must decide for each element of $\{1,2,3\}$ whether or not to put it in X. We have three independent yes/no choices, so there are $2 \times 2 \times 2 = 8$ subsets of $\{1,2,3\}$. Generalizing this argument shows that there are 2^n subsets of $\{1,2,\ldots,n\}$.

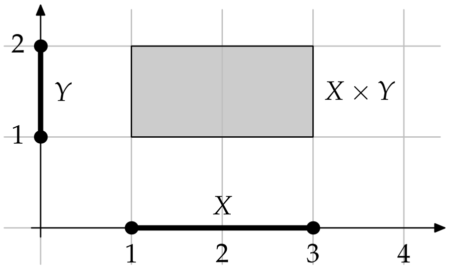
This principle, that numbers of independent choices can be multiplied to find the size of a set, is very useful when solving combinatorial problems.

CARTESIAN PRODUCTS. If X and Y are sets then we denote by $X \times Y$ the set of all *ordered pairs* (x,y) with $x \in X$ and $y \in Y$. It is usual to write X^2 for $X \times X$. Thus the plane is the set $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Exercise 7.8. Let

$$X = \{x \in \mathbb{R} : 1 \le x \le 3\}$$
$$Y = \{y \in \mathbb{R} : 1 \le y \le 2\}.$$

The set $X \times Y$ is drawn below.



Decide on the truth value of the following propositions.

(a) $(1,2) = (2,1)$, (b) $\{1,2\} = \{2,1\}$,
(c) $(5/2, 3/2) \in X \times Y$, (d) $(3/2, 5/2) \in X \times Y$,
(e) $Y \times Y \subseteq X \times Y$, (f) $X \subseteq Y$,
(g) $\varnothing \times X \subseteq \varnothing \times Y$.

Suppose that X and Y are finite sets. The number of ordered pairs (x, y) with $x \in X$ and $y \in Y$ is $|X||Y|$, since we have $|X|$ choices for x and $|Y|$ independent choices for y. Hence $|X \times Y| = |X||Y|$.

Example 7.9. A menu offers a choice of 3 starters, 5 main courses, and 2 desserts.

(a) How many three course meals can be ordered?
(b) How many two course meals can be ordered?

EXTRAS: COUNTING FUNCTIONS. Let X and Y be finite sets. How many functions are there with domain X and codomain Y? To define a function $f : X \to Y$, we have to pick, for each $x \in X$, its image $f(x) \in Y$. (See Definition 2.1.) So for each $x \in X$ we have $|Y|$ choices. Multiplying independent choices shows that the number of functions is

$$|Y| \times |Y| \times \overset{|X| \text{ factors}}{\cdots} \times |Y| = |Y|^{|X|}.$$

Suppose instead we want to count the number of injective functions. If $|Y| < |X|$ then there are no injective functions. (This is the pigeonhole principle: see Question 7 of Sheet 8.)

Exercise 7.10. Let $X = \{1, 2, \ldots, m\}$ and let $Y = \{1, 2, \ldots, n\}$ where $m, n \in \mathbb{N}$ and $n \geq m$. Show that the number of injective functions $f : X \to Y$ is

$$n(n-1)\ldots(n-m+1) = \frac{n!}{(n-m)!}.$$

[*Hint:* construct f step-by-step. How many choices are for $f(1)$? How many choices are there for $f(2)$? Repeat until you get to $f(m)$, then multiply numbers of choices.]

EXTRAS: SET DUALITY. Given any identity involving subsets of a universe set U, the principle of *duality* says that if $\cup$ and $\cap$ are swapped, and every set is replaced with its complement in U, the new identity still holds.

For example, the first de Morgan's law for sets, seen in Claim 1.10(i), states that $(X \cup Y)' = X' \cap Y'$. The dual identity is $(X' \cap Y')' = X \cup Y$. **Exercise:** deduce Claim 6.10(ii) from Claim 6.10(i) using duality.

Part D: Integers and rings

8. Euclid's Algorithm and congruences

Recall from Theorem 5.1 that if $n, m \in \mathbb{Z}$ and $m \neq 0$ then there exists a unique quotient $q \in \mathbb{Z}$ and remainder $r \in \mathbb{Z}$ such that $n = qm + r$ and $0 \leq r < m$.

In this section we will look at some more consequences of integer division, and see applications in coding theory (ISBNs) and cryptography (Diffie–Hellman key exchange, non-examinable).

Greatest common divisors.

Definition 8.1. Let $m, n \in \mathbb{N}$. We say that $d \in \mathbb{N}$ is the *greatest common divisor* of m and n, and write $\gcd(m, n) = d$, if d is the greatest natural number that divides both m and n.

One way to find $\gcd(m, n)$ is to find the prime factorizations of m and n. If p^a is the highest power of p dividing m and p^b is the highest power of p dividing n then $p^{\min(a,b)}$ is the highest power of p dividing $\gcd(m, n)$. This determines the prime factorization of $\gcd(m, n)$.

Example 8.2. Since

$$5848 = 2^3 \times 17 \times 43$$
$$2652 = 2^2 \times 3 \times 13 \times 17$$

we have $\gcd(5848, 2652) = 2^2 \times 17 = 68$.

Exercise 8.3. Find $\gcd(m, n)$ in each of these cases:

(i) $m = 310, n = 42$,
(ii) $m = 23, n = 46$,
(iii) $m = 31460, n = 41\,991\,752$.

Hint: on page 38 we saw that $31460 = 2^2 \times 5 \times 11^2 \times 13$. You do not need to factor m completely to find the gcd.

Euclid's Algorithm. There is a fast algorithm for finding greatest common divisors that is usually attributed to Euclid. The key idea is given in the following lemma.

Lemma 8.4 (Examinable). *Let $m, n \in \mathbb{N}$. Let $n = qm + r$. Then*

$$\gcd(n, m) = \gcd(m, r).$$

For example, $5848 = 2 \times 2652 + 544$, so Lemma 8.4 implies that $\gcd(5848, 2652) = \gcd(2652, 544)$. We could now factorize the smaller number 544 to find the gcd. This already uses less calculation than Example 8.2. But a much better idea is to keep on applying Lemma 8.4.

Algorithm 8.5 (Euclid's Algorithm)**.** *Let n, $m \in \mathbb{N}$. Find the quotient q and the remainder r when n is divided by m.*

- *If $r = 0$ then m divides n and $\gcd(n, m) = m$.*
- *Otherwise repeat from the start with m and r.*

Euclid's Algorithm always finishes because the remainder get smaller at each step. Lemma 8.4 implies that the final output of the algorithm is $\gcd(m, n)$. So Euclid's Algorithm has the two key properties of a good algorithm: it always finishes, and it always finishes with the right answer.

Example 8.6. Let $n = 3933$ and let $m = 389$. The equations below show the quotient and remainder at each step of Euclid's Algorithm:

$$\begin{aligned} 3933 &= 10 \times 389 + 43 \\ 389 &= 9 \times 43 + 2 \\ 43 &= 21 \times 2 + 1 \\ 2 &= 2 \times 1. \end{aligned}$$

Hence $\gcd(3933, 389) = 1$.

Note that we got this answer without computing any prime factorizations. For the record, $3933 = 3^2 \times 19 \times 23$ and 389 is prime (so 389 is its own prime factorization).

By working backwards through the steps in Euclid's Algorithm it is possible to find $s, t \in \mathbb{Z}$ such that $sm + tn = \gcd(m, n)$. We will see why this is useful shortly.

Example 8.7. By the penultimate line of Example 8.6 we have $\mathbf{1} = \mathbf{43} - 21 \times \mathbf{2}$, where bold type indicates the numbers that were remainders in Euclid's algorithm. Working back by finding the rows where **43** and **2** first appeared, we get

$$\begin{aligned} 1 &= \mathbf{43} - 21 \times \mathbf{2} \\ &= \mathbf{43} - 21 \times (\mathbf{389} - 9 \times \mathbf{43}) \\ &= 190 \times \mathbf{43} - 21 \times \mathbf{389} \\ &= 190 \times (\mathbf{3933} - 10 \times \mathbf{389}) - 21 \times \mathbf{389} \\ &= 190 \times \mathbf{3933} - 1921 \times \mathbf{389}. \end{aligned}$$

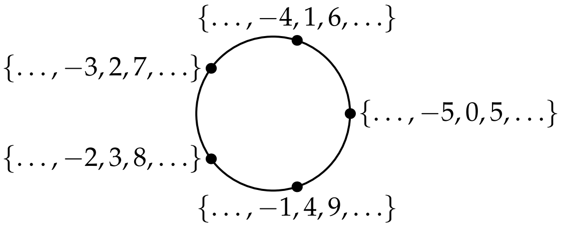
CONGRUENCES.

Definition 8.8. Let $m \in \mathbb{N}$. Let $n, n' \in \mathbb{Z}$. If $n - n'$ is divisible by m then we say that n is *congruent to n' modulo m*, and write

$$n \equiv n' \bmod m \textbf{ [typo } n \textbf{ corrected 29th November]}.$$

If $n = qm + r$ then $n - r$ is divisible by m. Hence if the remainder when n is divided by m is r, then $n \equiv r \bmod m$. This shows that any integer is congruent to one of $\{0, 1, \ldots, m-1\}$ modulo m.

The diagram below shows some of the numbers congruent to 0, 1, 2, 3 and 4 modulo 5.


$\{\ldots, -4, 1, 6, \ldots\}$
$\{\ldots, -3, 2, 7, \ldots\}$
$\{\ldots, -5, 0, 5, \ldots\}$
$\{\ldots, -2, 3, 8, \ldots\}$
$\{\ldots, -1, 4, 9, \ldots\}$

You might have seen congruences before as 'clock arithmetic'. For example, working in the twenty-four hour clock, seven hours after 23:00 is 6:00. Correspondingly, $23 + 7 = 30 \equiv 6 \bmod 24$.

Example 8.9.

(a) Since $5848 = 2 \times 2652 + 544$, we have $5848 \equiv 544 \bmod 2652$.
(b) $-7 \equiv 10 \bmod 17$ since $-7 - 10$ is divisible by 17.
(c) $27 \times 33 \equiv 17 \times 33 \equiv 7 \times 33 \equiv 7 \times 3 \equiv 1 \bmod 10$. **Exercise:** make up a similar example working modulo 5.

Part (c) of the example suggests the following general result.

Lemma 8.10 (Examinable). *Let $m \in \mathbb{N}$ and let $r, r', s, s' \in \mathbb{Z}$. If $r \equiv r'$ mod m and $s \equiv s'$ mod m then*

(i) *$r + s \equiv r' + s'$ mod m,*
(ii) *$rs \equiv r's'$ mod m.*

Lemma 8.10 justifies many manipulations with congruences.

For example, suppose we know that $2x \equiv 1 \bmod 5$. Let $r = r' = 3$ and let $s = 2x$, $s' = 1$. Then by Lemma 8.11 we have $3 \times 2x \equiv 3 \times 1 \bmod 5$. This simplifies to $x \equiv 3 \bmod 5$.

SOLVING CONGRUENCE EQUATIONS. Solving equations such at $x+9 \equiv 7 \bmod 12$ is straightforward. For equations like $3x \equiv 2 \bmod m$, a new idea is required. If the modulus m is small, you can just try $x = 0, 1, \ldots, m-1$ in turn: if a solution exists, this will find one.

Exercise 8.11.

(a) Find $x \in \mathbb{Z}$ such that $0 \leq x < 11$ and $x + 9 \equiv 7 \bmod 12$.
(b) Find an $x \in \mathbb{Z}$ such that $3x \equiv 2 \bmod 5$.
(c) Find *all* $x \in \mathbb{Z}$ such that $3x \equiv 2 \bmod 5$.

The answers are on the Part D slides available from Moodle.

When the modulus is larger, Euclid's algorithm can be used.

Example 8.12. To solve the congruence $389x \equiv 103 \bmod 3933$ we use Euclid's algorithm to express $1 = \gcd(389, 3933)$ as $389s + 3933t$ where $s, t \in \mathbb{Z}$. By Example 8.7,

$$1 = -1921 \times 389 + 190 \times 3933.$$

Note this is equivalent to the congruence

$$1 \equiv -1921 \times 389 \bmod 3933.$$

Hence

$$\begin{aligned} 389x \equiv 103 \bmod 3933 &\iff -1921 \times 389x \equiv -1921 \times 103 \bmod 3933 \\ &\iff x \equiv -1921 \times 103 \bmod 3933 \\ &\iff x \equiv 2720 \bmod 3933. \end{aligned}$$

Therefore $389x \equiv 103 \bmod 3933$ if and only if $x = 2720 + 3933q$ for some $q \in \mathbb{Z}$.

Not all congruences can be solved. For example $2x \equiv 3 \bmod 4$ has no solution, because $2x$ is always even, but any number congruent to 3 modulo 4 is odd.

THE ISBN CODE. All recent books have an International Standard Book Number (ISBN) assigned by the publisher. In the system used before 2007, each book is given a sequence of length 10 with entries from

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, \mathrm{X}\}.$$

For example *A Concise Introduction to Pure Mathematics*, number [2] in the recommended reading list, has ISBN

1-4398-3598-5

The dashed are put into to improve readability, but are not formally part of the code. For background only,

- 1 identifies the country of publication,

- 4398 identifies the publisher (CRC Press)
- 3598 is the item number assigned by the publisher.
- 5 is the *check digit*.

The check digit is chosen so that if $u_1u_2u_3u_4u_5u_6u_7u_8u_9u_{10}$ is an ISBN then

$$\sum_{j=1}^{10}(11-j)u_j \equiv 0 \bmod 11.$$

The sum on the left-hand side is called the *check sum* It might be necessary to take 10 as a check digit. In this case the letter X is used to stand for 10; X never appears in the first nine positions of an ISBN.

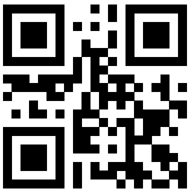
Exercise 8.13.

(a) Suppose that an error is made in position 8, and the ISBN is miscopied as 1-4398-3568-5. Show that the error will be detected because the check sum is not divisible by 11.

(b) Suppose that the digits in positions 8 and 9 are swapped, and so 1-4398-3589-5 is written down. Show again that the error will be detected.

In fact the ISBN code detects all errors of types (a) and (b): see Question 5 on Sheet 9.

Modern codes can even correct errors, by using redundant information stored in the codewords to deduce which positions have been corrupted. Error correcting codes are used in compact discs, Bluray discs, and QR codes, such as the one in the margin.



EXTRAS: PUBLIC KEY CRYPTOGRAPHY. Suppose you are Alice, and you want to send a message to your friend Bob, so that only Bob can read it. In the traditional approach, you first meet up to exchange a secret key, and later encrypt your message using a cipher with this secret key.

For online banking and retail, this is obviously not a workable approach. It is not realistic to expect a potential customer of your business to arrange a secret meeting before placing an order.

Instead a number of remarkable mathematical ideas are used. Here is the Diffie–Hellman key exchange scheme. Alice starts by picking a large prime p, and an element

$$g \in \{2, \ldots, p-1\}$$

and sends both p and g to Bob. Then

- Alice chooses some $a \in \mathbb{N}$ and sends g^a mod p to Bob.
- Bob chooses some $b \in \mathbb{N}$ and sends g^b mod p to Alice.

Now Alice knows a and g^b, so she can calculate $(g^b)^a$ mod p. Similarly Bob knows b and g^a, so he can calculate $(g^a)^b$ mod p. Hence Alice and Bob can both calculate g^{ab} mod p. An eavesdropper knows p, g and g^a, g^b mod p. However, there is no obvious way to determine g^{ab} mod p given these numbers, and it is believed that this is a hard problem.

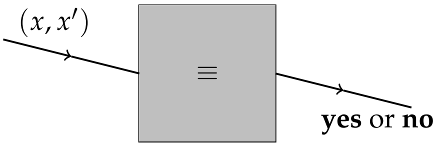
Therefore Alice and Bob have succeeded in sharing the secret key g^{ab} mod p using only public communications. If you do not find this surprising, you have not fully grasped the problem public key cryptography solves.

The account above is oversimplified in many ways but shows the basic idea. Try searching for 'Diffie–Hellman' or 'RSA' on the web for more details. Currently we offer third or fourth year courses on coding theory, cipher systems, and public key cryptography.

9. Relations and the integers modulo m

The following definition generalizes the congruence relation.

Definition 9.1. Let X be a set. A *relation* on X is a black box which, given an ordered pair (x, x') where $x, x' \in X$, outputs either **yes** or **no**. A **yes** means x is related to x', and a **no** means x is not related to x'.



Two relations on a set X are equal if they agree on all ordered pairs (x, x'). As for functions, it is irrelevant how the black box arrives at its answer.

Example 9.2.

(i) Fix $m \in \mathbb{N}$. Let $n, n' \in \mathbb{Z}$. For the input (n, n'), let the black box output **yes** if $n \equiv n'$ mod m and **no** otherwise. This defines the congruence modulo m relation on $\mathbb{Z}$.

(ii) Let P be the set of all subsets of $\{1, 2, 3\}$. Given an ordered pair (X, Y) of elements of P, let the black box output **yes** if $X \subseteq Y$ and **no** otherwise.

Relations can be defined more briefly. For example, suppose that $X = \{1, 2, 3, 4, 5, 6\}$. Then

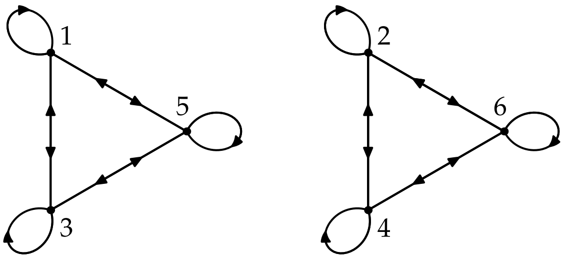
$$x \text{ relates to } y \iff x < y$$

defines the relation 'strictly less than' on X. An analogous relation can be defined replacing X with any other subset of $\mathbb{R}$.

The symbol $\sim$ (pronounced 'tilde', or perhaps 'twiddle') is often used for relations. We will also use $\equiv$, but only for congruence relations.

DIAGRAMS. Let X be a set and let $\sim$ be a relation defined on X. To represent $\sim$ on a diagram, draw a dot for each element of X. Then for each $x, y \in X$ such that $x \sim y$, draw an arrow *from x to y*. If $x \sim x$ draw a loop from x to itself.

Example 9.3. Let $X = \{1, 2, 3, 4, 5, 6\}$. The relation $x \equiv y$ mod 2 on X is shown in the diagram below.



Exercise: Draw a similar diagram for the relation on $\{1, 2, 3, 4, 5, 6\}$ defined by

$$x \sim y \iff x - y \text{ is even and } x > y.$$

Diagrams become impractical for relations on bigger sets. But they are still useful for thinking about the properties below.

PROPERTIES OF RELATIONS.

Definition 9.4. Let $\sim$ be a relation on a set X. We say that $\sim$ is

(i) *reflexive* if $x \sim x$ for all $x \in X$;

(ii) *symmetric* if for all $x, y \in X$,
$$x \sim y \Longrightarrow y \sim x;$$

(iii) *transitive* if for all $x, y, z \in X$,
$$x \sim y \text{ and } y \sim z \Longrightarrow x \sim z.$$

A relation that is reflexive, symmetric and transitive is said to be an *equivalence relation*.

Example 9.5. Fix $m \in \mathbb{N}$. The congruence relation $n \equiv n'$ mod m is an equivalence relation on $\mathbb{Z}$.

Most relations that are important in mathematics are either equivalence relations or order relations, such as $x < y$ on $\{1, 2, 3, 4, 5, 6\}$ or $X \subseteq Y$ on the set of subsets of $\{1, 2, 3\}$. These relations are transitive, but never symmetric (except in trivial cases).

Exercise 9.6. Decide which of the two relations $<$ and $\subseteq$ just mentioned is reflexive.

In general a relation can have any combination of the properties reflexive, symmetric and transitive. See Question 2 of Sheet 10.

Exercise 9.7. Let X be the set of people sitting in a full lecture room. For each of the following relations, decide whether it is (i) reflexive, (ii) symmetric and (iii) transitive.

(a) $x \sim y$ if x and y are sitting in the same row,
(b) $x \sim y$ if x is sitting in a strictly higher row than y,
(c) $x \sim y$ if x and y are friends.

EQUIVALENCE CLASSES. Suppose that $\sim$ is an equivalence relation on a set X. For $x \in X$, we define the *equivalence class of* x to be the set of all elements of X that relate to x. In symbols

$$[x] = \{z \in X : z \sim x\}.$$

For example, the equivalence classes for the relation $x \equiv y \bmod 2$ on the set $\{1, 2, 3, 4, 5, 6\}$ are

$$\begin{aligned} [0] = [2] = [4] = \{0, 2, 4\} \\ [1] = [3] = [5] = \{1, 3, 5\} \end{aligned}$$

Notice how the equivalence classes can be read off from the diagram on page 59.

More generally there is the following result.

Theorem 9.8. *Let $\sim$ be an equivalence relation on a set X. Let $x, y \in X$.*

(i) $x \in [x]$,
(ii) $x \sim y \iff [x] = [y]$,
(ii) $x \not\sim y \iff [x] \cap [y] = \varnothing$.

Thus, by (i), every element of X lies in an equivalence class, and by (ii) and (iii), X is a disjoint union of the distinct equivalence classes.

The proof of Theorem 9.8 is non-examinable and will be skipped if time is pressing. See Theorem 31.13 in *How to think like a mathematician* for a careful (and exhaustively analysed) proof.

THE NUMBER SYSTEM $\mathbb{Z}_m$ OF INTEGERS MODULO m. Fix $m \in \mathbb{N}$. Let

$$\mathbb{Z}_m = \{[n] : n \in \mathbb{Z}\}$$

be the set of equivalence classes for congruence modulo m.

This definition is liked by pure mathematicians, because it avoids having to make an explicit choice of 'representative' elements for each equivalence class. But you might prefer the equivalent definition

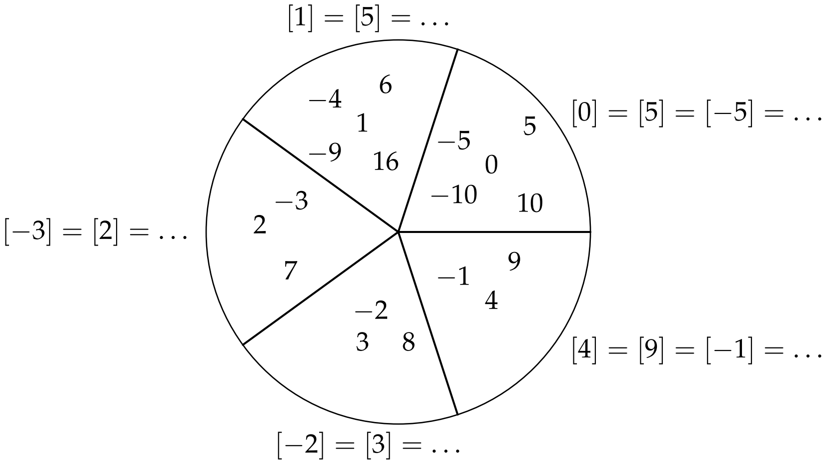
$$\mathbb{Z}_m = \{[0], [1], \ldots, [m-1]\}.$$

Here

$$[r] = \{r + qm : q \in \mathbb{Z}\}$$

is the set of integers that leave remainder r on division by q.

The diagram below shows some elements of each of the equivalence classes for the relation of congruence modulo 5 on $\mathbb{Z}$.



We turn the set $\mathbb{Z}_m$ of equivalence classes into a number system by defining addition and multiplication as follows.

Definition 9.9. Fix $m \in \mathbb{N}$. Given $[r], [s] \in \mathbb{Z}_m$ we define $[r] + [s] = [r + s]$ and $[r][s] = [rs]$.

There is a subtle point here: we need to check that this definition is *well defined*, that is, $[r] + [s]$ depends only on the equivalence classes $[r]$ and $[s]$, and not on the particular representatives r and s. Suppose that $[r] = [r']$ and $[s] = [s']$. By Theorem 9.8(ii),

$$[r] = [r'] \iff r \equiv r' \bmod m$$
$$[s] = [s'] \iff s \equiv s' \bmod m.$$

Hence, by Lemma 8.10, we have $rs \equiv r's' \bmod m$ and $r + s \equiv r' + s' \bmod m$. Now Theorem 9.8(ii) implies that $[rs] = [r's']$ and $[r + s] = [r' + s']$, as required.

Exercise 9.10. Recall that a square number is a number of the form n^2 where $n \in \mathbb{N}$.

(i) Calculate $1^2, 2^2, 3^2, 4^2, 5^2, 6^2, 7^2, \ldots$ modulo 4. State and prove a conjecture on the pattern you observe.

(ii) Is 2015 the sum of two square numbers?

Example 9.11. The addition and multiplication tables for $\mathbb{Z}_5$ are shown below. For example, the entry in the addition table in the row for [4] and the column for [3] is

$$[4] + [3] = [2]$$

since $4 + 3 = 7$ and $7 \equiv 2 \bmod 5$.

+	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

×	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]
[2]	[0]	[2]	[4]	[1]	[3]
[3]	[0]	[3]	[1]	[4]	[2]
[4]	[0]	[4]	[3]	[2]	[1]

You may omit [0] from the multiplication table if you prefer.

10. Rings

The addition and multiplication operations on $\mathbb{Z}_m$ have all the properties you would expect. Formally this is expressed by saying that $\mathbb{Z}_m$ is a ring, as defined in the next definition.

Definition 10.1. Suppose that R is a set on which addition and multiplication are defined, so that given any two elements $x, y \in R$, their sum $x + y$ and product xy are elements of R. We say that R is a *ring* if the following properties hold:

(1) (*Commutative law of addition*) $x + y = y + x$ for all $x, y \in R$,

(2) (*Existence of zero*) There is an element $0 \in R$ such that $0 + x = x$ for all $x \in R$,

(3) (*Existence of additive inverses*) For each $x \in R$ there exists an element $-x \in R$ such that $-x + x = 0$, where 0 is the element in property (2),

(4) (*Associative law of addition*) $(x + y) + z = x + (y + z)$ for all $x, y, z \in R$,

(5) (*Existence of one*) There exists an element $1 \in R$ such that $1x = x1 = x$ for all $x \in R$,

(6) (*Associative law of multiplication*) $(xy)z = x(yz)$ for all $x, y, z \in R$,

(7) (*Distributivity*) $x(y+z) = xy + xz$ and $(x+y)z = xz + yz$ for all $x, y, z \in R$.

The number systems $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{C}$ and $\mathbb{Z}_m$ for $m \in \mathbb{N}$ are rings.

Definition 10.2.

(i) A ring R is *commutative* if $xy = yx$ for all $x, y \in R$.

(ii) A commutative ring R is a *field* if for all non-zero $x \in R$ there exists an element $y \in R$ such that $xy = yx = 1$, where 1 is the one element in property (5). We say that y is the *inverse* of x and write $y = x^{-1}$.

Note that 'inverse' refers to the multiplicative structure on R. The element $-x$ in property (3) is called the *additive inverse* of x.

Some familiar examples of fields are $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. More interestingly, $\mathbb{Z}_5$ is a field. The inverses of the non-zero elements can be found from the multiplication table in Example 9.11. They are

$$[1]^{-1} = [1], \quad [2]^{-1} = [3], \quad [3]^{-1} = [2], \quad [4]^{-1} = [4].$$

If R is a field then we can define division in R by $x/y = xy^{-1}$, for $x, y \in R$ with y non-zero. Thus fields are closed under division in the sense of Definition 1.5.

Theorem 10.3 (Examinable). *If p is prime then $\mathbb{Z}_p$ is a field.*

The proof of this theorem gives an effective way to find the inverse of a non-zero element $[x] \in \mathbb{Z}_p$: use Euclid's Algorithm to find $s, t \in \mathbb{Z}$ such that

$$sx + tp = 1.$$

Then $sx \equiv 1 \bmod p$, and so $[s][x] = [1]$ and $[x]^{-1} = [s]$. The method is illustrated in the example below.

Example 10.4. Let $x = [7] \in \mathbb{Z}_{23}$. Using Euclid's algorithm on 23 and 7 we get

$$\begin{aligned} 23 &= 3 \times 7 + 2 \\ 7 &= 3 \times 2 + 1 \\ 2 &= 2 \times 1 \end{aligned}$$

so, as expected, the greatest commmon divisor is 1. Working back gives

$$1 = 7 - 3 \times 2 = 7 - 3 \times (23 - 3 \times 7) = 10 \times 7 - 3 \times 23.$$

Therefore

$$10 \times 7 \equiv 1 \bmod 23$$

and $[7]^{-1} = [10]$.

A more surprising example of a field is

$$S = \{a + bi\sqrt{3} : a, b \in \mathbb{Q}\}.$$

Question 3 on Sheet 4 shows that S is closed under multiplication and if $z \in S$ and $z \neq 0$ then $1/z \in S$. It is easily seen that S is closed under addition and subtraction. So S is a field.

Here are some properties that hold for all rings. Some of the proofs are left to you on Question 8 of Problem Sheet 10.

Lemma 10.5. *Let R be a ring.*

(i) *There is a unique zero element in R satisfying property* (2).

(ii) *There is a unique one element in R satisfying property* (5).

(iii) *For each $x \in R$ there exists a unique $y \in R$ such that*

$$y + x = x + y = 0.$$

(iv) *If $x, z \in R$ and $x + z = x$ then $z = 0$.*

(v) *We have $0x = 0 = x0$ for all $x \in R$.*

(vi) *We have $-x = (-1)x = x(-1)$ for all $x \in R$.*

(vii) *For all $x \in R$ we have $-(-x) = x$.*

(viii) *For all $x, y \in R$ we have*

$$-(xy) = (-x)y = y(-x) \text{ and } (-x)(-y) = xy.$$

(ix) $0 = 1$ *if and only if $R = \{0\}$.*

In (vi), (vii), (viii) you should bear in mind that $-x$ means the additive inverse of x given by property (3). So the only thing we can use about $-x$ is that $-x + x = 0$. It is a non-trivial result that $-x = (-1)x$.

POLYNOMIAL RINGS. The course ends with rings of polynomials. We define polynomial rings over an arbitrary field: the main examples to bear in mind are $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}_p$ for prime p.

Definition 10.6. Let F be a field. Let $F[x]$ denote the set of all polynomials

$$f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$$

where $d \in \mathbb{N}_0$ and $a_0, a_1, a_2, \ldots, a_d \in F$. If $d = 0$, so $f(x) = a_0$, then $f(x)$ is a *constant polynomial*.

When writing polynomials we usually omit coefficients of 1, and do not include powers of x whose coefficient is 0. For example, in $\mathbb{Z}_2[x]$, we write $x^2 + [1]$ rather than $[1]x^2 + [0]x + [1]$.

The x in $f(x)$ is called an *indeterminate*. You can think of it as standing for an unspecified element of F.

Polynomials are added and multiplied in the expected way.

Example 10.7. In $\mathbb{Z}_3[x]$, we have

$$\begin{aligned}(x^4 + [2]x^3 + [1]) + ([2]x^4 + x^2 + [1])& \\ &= ([1] + [2])x^4 + [2]x^3 + x^2 + ([1] + [1]) \\ &= [2]x^3 + x^2 + [2]\end{aligned}$$

and

$$(x + [1])(x + [2]) = x^2 + ([1] + [2])x + [1][2] = x^2 + [2].$$

It is routine to check that $F[x]$ has all the properties to be a ring. The zero element is the constant polynomial $0 \in F$ and the one element is the constant polynomial $1 \in F$.

There is a remarkable analogy between the ring of integers $\mathbb{Z}$ and polynomial rings. For example, polynomials can be divided with remainder in a similar way to integers (see Theorem 5.1).

Definition 10.8. Let $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_2x^2 + a_1x + a_0$ where $a_d \neq 0$.

(i) We say that d is the *degree* of $f(x)$ and write $\deg f(x) = d$.

(ii) The *leading coefficient* of $f(x)$ is a_d. If $a_d = 1$ we say that $f(x)$ is *monic*.

The degree of zero polynomial $f(x) = 0$ is undefined.

Theorem 10.9. *Let F be a field, let $f(x) \in F[x]$ be a non-zero polynomial and let $g(x) \in F[x]$. There exist polynomials $q(x), r(x) \in F[x]$ such that*

$$g(x) = q(x)f(x) + r(x)$$

and either $r(x) = 0$ or $\deg r(x) < \deg f(x)$.

By analogy with integer division, we say that $q(x)$ is the *quotient* and $r(x)$ is the *remainder* when $g(x)$ is divided by $f(x)$.

Example 10.10.

(1) Working in $\mathbb{Q}[x]$, let $g(x) = 3x^2 + 2x - 1$ and let $f(x) = 2x + 1$. Then

$$g(x) = (\tfrac{3}{2}x + \tfrac{1}{4})f(x) - \tfrac{5}{4}$$

so the quotient is $q(x) = \frac{3}{2}x + \frac{1}{4}$ and the remainder is $r(x) = -\frac{5}{4}$. If instead we take $h(x) = x + 1$ then

$$g(x) = (3x - 1)h(x).$$

So when $g(x)$ is divided by $h(x)$ the quotient is $3x - 1$ and the remainder is 0.

(2) Working in $\mathbb{Z}_3[x]$, let $g(x) = x^4 + x^3 + [2]x^2 + x + 1$ and let $f(x) = x^2 + x$. Then

$$g(x) = (x^2 + [2]x)f(x) + 2[x] + 1.$$

So the quotient when $g(x)$ is divided by $f(x)$ is $x^2 + [2]x$ and the remainder is $2[x] + 1$.

There is a MATHEMATICA notebook on Moodle you can use to check calculations with polynomials.

The next theorem is known as the Remainder Theorem, or Factor Theorem.

Theorem 10.11. *Let F be a field and let $f(x) \in F[x]$ be a polynomial. Let $c \in F$. Then*

$$f(x) = q(x)(x - c) + r$$

for some polynomial $q(x) \in F[x]$ and some $r \in \mathbb{F}$. Moreover $f(c) = 0$ if and only if $r = 0$.

This theorem is very useful for factorizing polynomials and solving polynomial equations. It was used earlier in Example 3.22. Here is another example.

Example 10.12. Working in $\mathbb{Z}_3[x]$, let $g(x) = x^4 + x^3 + [2]x^2 + x + [1]$ as in Example 10.10(2). Since

$$g([1]) = [1] + [1] + [2] + [1] + [1] = [6] = [0],$$

the Factor Theorem says that $x - [1]$ divides $g(x)$. Division gives

$$g(x) = (x - [1])(x^3 + [2]x^2 + x + [2]).$$

The cubic $x^3 + [2]x^2 + x + [2]$ also has $[1]$ as a root. Dividing it by $x - [1]$ gives

$$g(x) = (x - [1])^2(x^2 + [1]).$$

Therefore $g(x)$ has $[1]$ as a root with multiplicity 2, and no other roots in $\mathbb{Z}_3$.

We end with a corollary of Theorem 10.9 that gives a stronger version of the Fundamental Theorem of Algebra (Theorem 3.21).

Corollary 10.13. *Let $f(x) \in \mathbb{C}[x]$ be a polynomial of degree n. There exist distinct $w_1, w_2, \ldots, w_r \in \mathbb{C}$ and $m_1, \ldots, m_r \in \mathbb{N}$ such that $m_1 + \cdots + m_r = n$ and*

$$a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = a_n(z - w_1)^{m_1}(z - w_2)^{m_2} \ldots (z - w_r)^{m_r}.$$

Hence a polynomial in $\mathbb{C}[x]$ of degree n has exactly n roots, counted with multiplicities.

EXTRAS: EUCLID'S ALGORITHM FOR POLYNOMIALS. Given any two polynomials $f(x)$, $g(x)$, there is a unique monic polynomial $h(x)$ such that $h(x)$ divides $f(x)$ and $g(x)$ and $\deg h(x)$ is as large as possible. We call this polynomial the *greatest common divisor* of $f(x)$ and $g(x)$. It can be found by Euclid's Algorithm.

For example, working in $\mathbb{Z}_3[x]$, let $f(x) = x^3 + x^2 + x + [1]$ and $g(x) = x^2 - [1]$. Then

$$\begin{aligned} x^3 + x^2 + x + [1] &= (x + [1])(x^2 - [1]) + [2]x + [2] \\ x^2 - [1] &= ([2]x + [1])([2]x + [2]). \end{aligned}$$

So the algorithm terminates with final non-zero remainder $[2]x + [2]$. We now multiply $[2]x + [2]$ by $[2]^{-1} = [2]$ to get $x + [1]$. Hence

$$x + [1] = \gcd\bigl(f(x), g(x)\bigr).$$

Applications of this will be seen in MT283 Rings and Factorisation.

EXTRAS: MATRIX RINGS. Let R be a ring. If $a, b, c, d \in R$ then we say that

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

is a 2×2-*matrix* over R. The set of all such matrices forms a ring, with addition and multiplication defined by

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

and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}$$

The zero element is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and the one element is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

One interesting property of the ring of 2×2-matrices is that, unlike all the rings seen so far, multiplication is not commutative.

Exercise 10.14. Compute the matrix products

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and deduce that multiplication of matrices is not commutative.

EXTRAS: INTEGRAL DOMAINS. An important property of the integers $\mathbb{Z}$ is that the product of two non-zero integers is always non-zero.

Definition 10.15. Let R be a ring and suppose that $x, y \in R$. If

$$xy = 0 \Longrightarrow x = 0 \text{ or } y = 0$$

then we say that R is an *integral domain*.

Thus $\mathbb{Z}$ is an integral domain. However, the ring of 2×2-matrices is not an integral domain. For example, the second product in Exercise 10.4 is the zero matrix.

Theorem 10.16. *If F is a field then F is an integral domain*

Proof. Suppose that $x, y \in F$ are such that $xy = 0$. If $x \neq 0$ then x has an inverse, $x^{-1} \in F$. Multiplying by x^{-1} we get

$$0 = x^{-1}0 = x^{-1}(xy) = (x^{-1}x)y = 1y = y$$

using Lemma 10.5(v) and the ring properties in Definition 10.1. Hence $x \neq 0 \Longrightarrow y = 0$, and so either $x = 0$ or $y = 0$. □

In particular, it follows that $\mathbb{Z}_p$ is an integral domain whenever p is prime. If $n = ab$ is composite then $\mathbb{Z}_n$ is not an integral domain, because $[a][b] = [n] = [0]$.

Question 9 on Sheet 10 outlines a proof of the interesting result that any finite integral domain is a field.

EXTRAS: RINGS OF SUBSETS. Let X be a set and let R be the set of all subsets of X. There is a remarkable way to make R a ring. For $A, B \in R$, we define

$$A + B = (A \cap B') \cup (A' \cap B).$$

So $A + B$ is the set of elements of X that are in *exactly* one of A and B. We define

$$AB = A \cap B.$$

One can check that with these definitions, R satisfies all the conditions to be a commutative ring. Since

$$A + \varnothing = (A \cap \varnothing') \cup (A' \cap \varnothing) = (A \cap X) \cup (A' \cap \varnothing) = A \cup \varnothing = A$$

and $AX = A \cap X = A$ for all $A \in R$, the zero element is $\varnothing$ and the one element is X.

Exercise 10.17.

(i) Check that the distributivity law holds for R.
(ii) Show that $A + A = 0$ for all $A \in R$.
(iii) Use the ring structure of R to show that if A, B, C, D are subsets of X then $A + B = C + D$ if and only if $A + D = B + C$.