MT1003 Pure Mathematics

MQ

Contents

Introduction

The following is an excerpt from the *Student Handbook* relating to this course:

> **Aims:** The aim of this module is to provide students with a taste of (both) pure (and applied) mathematics, to give them insight into areas available for study in later years and to provide them with the opportunity to broaden their mathematical experience.
>
> **Objectives:** By the end of the course students should have gained facility with each of the topic mentioned below and should be able to perform calculations and prove results....

With these in mind, my goals in my half of this lecture course will be roughly as follows:

- to introduce some aspects (and methods) of pure mathematics so as to provoke interest and enthusiasm for future courses;
- to introduce the concept of proof in contexts which are not too unfamiliar.

Textbooks

This is a rather broad course and so it is rather difficult to find textbooks that correspond very well to it. There are many books that between them cover the material in the course, but usually they go much further than is required. The following are possibilities for consulting:

- D. M. Burton, *Elementary Number Theory* (Allyn & Bacon, 1976) [Short Loan, QA241.B8]
- R. P. Grimaldi, *Discrete and Combinatorial Mathematics: An Applied Introduction* (Addison-Wesley, 1989) [QA39.2G85F89]
- T. S. Blyth & E. F. Robertson, *Essential Student Algebra, Vol. I* (Chapman & Hall, 1986) [Short Loan, QA155.B6R8]

Those that are still in print are probably too expensive to consider buying, but they are worth consulting every so often for additional background reading. These three are all available in the Mathematics and Physics Library (the first and third are in short loan). Equally most introductory texts on (elementary) number theory, discrete mathematics and combinatorics, and algebra would be of help.

What is Pure Mathematics?

To introduce this course, it is worth considering what pure mathematics actually is. When undergoing sixth-form studies at school, one may well believe that pure mathematics is about the development of mathematical techniques for application in mathematical problems. However, these techniques are precisely what is taught in the MT1002 course and yet it is this current course that has "Pure Mathematics" in its title. So what is the real meaning of "Pure Mathematics" and how does it differ from "Mathematical Methods"?

The study of Pure Mathematics typically has the following format:

- an abstract definition is made of some mathematical object;
- investigation is made of such mathematical objects using arguments that only exploit the details present within the abstract definition;
- the "theorems" established in the aforementioned investigation are applied to any example that satisfies the original abstract definition.

This will be the pattern of the course. We shall make various definitions and then establish theorems to develop the theory and apply the theory to examples to illustrate the progress that has been made. Initially the level of abstraction involved may take some getting used to, but in the end this abstraction is both the power and the beauty behind pure mathematics.

The one skill that will need careful development, and which should evolve over the course with practice, is the writing of proofs. Many of the questions on the problem sheets will call for a proof. Initially these may be hard to generate, but the key is usually to make careful logical deductions from the given hypotheses and *continued practice is essential*. The proofs given for theorems in lectures will illustrate this. The most common obstruction to producing a proof is not knowing what the terms involved actually mean. It is important to learn definitions, since otherwise statements of theorems and questions cannot be understood.

Course content

The structure of the course will be roughly as follows:

- Elementary Number Theory: integers, divisibility, greatest common divisor and Euclidean algorithm, factorisation and primes, linear Diophantine equations, congruences
- Functions and Relations: equivalence relations, application to congruences
- Higher Diophantine Equations: Pythagorean triples
- Graphs: examples and properties
- Groups: permutations, introduction to group theory

Section 1

Divisibility of Integers

Number Theory can be described as the study of the integers (and their generalisations). Accordingly we are interested in properties of

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

We may perform three of the basic arithmetical operations within $\mathbb{Z}$:

- Addition; e.g., $4 + 5 = 9$
- Multiplication; e.g., $(-3) \times 7 = -21$
- Subtraction; e.g., $3 - 8 = -5$.

Division, however, is *not* always defined for any pair of integers. For example, we cannot divide 3 by 2 to obtain an integer (although $3/2$ is defined in the set of all real numbers, it is not defined in $\mathbb{Z}$). What we can do in the integers is to divide and obtain a quotient and remainder.

Basic Fact 1.1 *For every two integers a and b with $b > 0$ there exist unique integers q and r such that*

$$a = bq + r \qquad \text{and} \qquad 0 \leqslant r < b.$$

We call q the quotient *and r the* remainder.

This fact should actually be familiar to many, though possibly its significance may not be so apparent. It will actually be key to most of what we do in the Number Theory part of this course.

Example 1.2 (i) Dividing 17 by 5 gives: $17 = 5 \cdot 3 + 2$.

(ii) Dividing 20 by 4 gives: $20 = 4 \cdot 5 + 0$.

(iii) The Basic Fact is also what underpins "long division" which was probably met at school. For example, if we attempt to divide 417 by 13 we proceed as follows:

$$\begin{array}{r} 32 \\ 13 \overline{)\,417} \\ \underline{390} \\ 27 \\ \underline{26} \\ 1 \end{array}$$

So $417 = 13 \cdot 32 + 1$.

We shall take the Basic Fact for granted. Essentially we shall be treating it as though it is a defining property of the integers (an 'axiom') and then proceed to deduce other information and theorems from it.

There is an alternatve: we could construct the integers by some method and then prove that the Basic Fact holds. Indeed this can be done without too much difficulty (at least, in the grand scheme of mathematics it is not too difficult), but doing so would go well beyond what is expected for this course. Also doing so might well be unhelpful and discouraging for students meeting pure mathematics for the first time.

We shall actually be able to make considerable progress using only the Basic Fact, and we begin with the following definition.

Definition 1.3 For two integers a and b with $b \neq 0$, we say that b *divides* a (or that a *is divisible by* b) if $a = bq$ for some integer q. We denote this by $b \mid a$.

So when we say "b divides a" what we mean is that we get zero remainder when we attempt to divide a by b (as in the Basic Fact).

To give a basic flavour of the sort of thing we shall encounter in the number theory part of these lectures, we are already in a position where we can prove the following:

Theorem 1.4 *The square of an integer is either divisible by* 4*, or else it gives remainder* 1 *when divided by* 8*.*

PROOF: Let a be any integer. When we attempt to divide a by 4, the possible remainders are 0, 1, 2 or 3. We consider each of these possibilities in turn.

- If $a = 4k$, then
$$a^2 = (4k)^2 = 16k^2 = 4(4k^2),$$
which is divisible by 4.

- If $a = 4k + 1$, then

$$a^2 = (4k + 1)^2 = 16k^2 + 8k + 1 = 8(2k^2 + k) + 1,$$

  so a^2 has remainder 1 when we divide by 8.
- If $a = 4k + 2$, then

$$a^2 = (4k + 2)^2 = 16k^2 + 16k + 4 = 4(4k^2 + 4k + 1),$$

  which is divisible by 4.
- Finally, if $a = 4k + 3$, then

$$a^2 = (4k + 3)^2 = 16k^2 + 24k + 9 = 8(2k^2 + 3k + 1) + 1,$$

  so a^2 has remainder 1 when we divide by 8.

Hence the theorem is proved. □

It is sensible to review this proof and wonder why it is the correct way (or at least a sensible way) to proceed. After the event one can see it does indeed do what it should, but how does one come up with the proof? Such a question is important to address when as a student one will need to do something similar for a question on a problem sheet or in an exam.

We knew that in the end we would be interested in the result when we attempted to divide by 8. Accordingly it is sensible to consider the result obtained when our original integer a is divided by something. It would seem sensible to try to divide a by 2, 4 or 8. If one tries looking at the remainder when we divide a by 2, then the proof does not quite work. (One instead ends up proving the weaker result that a^2 is either divisible by 4 or has remainder 1 upon dividing by 4.) If instead one examines the remainder upon dividing by 8, then the proof actually does work but turns out to be twice as long as the one presented here. So looking at the remainder when dividing by 4 turns out to be the best choice: it provides both a proof that works and is not too long.

It is worth considering this sort of review after every proof and solution to a question. (Hopefully most of the time it can be done quite quickly.)

One of the standard things that is done in pure mathematics is that certain concepts are introduced and then examined in great detail. We have introduced the concept of "dividing" and the sensible thing to do is decide what properties it has. Accordingly, in the next lecture we shall prove the following result:

Theorem 1.5 (Basic Properties of |) *Let a, b, c, d, x, y be any integers.*

(i) $a \mid 0$, $1 \mid a$, $a \mid a$.

(ii) $a \mid 1$ *if and only if* $a = \pm 1$.

(iii) *If* $a \mid b$ *and* $c \mid d$, *then* $ac \mid bd$.

(iv) *If* $a \mid b$ *and* $b \mid c$, *then* $a \mid c$.

(v) *If* $a \mid b$ *and* $b \mid a$, *then* $a = \pm b$.

(vi) *If* $a \mid b$ *and* $a \mid c$, *then* $a \mid (bx + cy)$.

PROOF OF THEOREM 1.5: This turns out to be reasonably easy — provided one keeps the definition of "divides" clear in one's head. Two of the parts (parts (iv) and (vi)) will be omitted and instead appear as exercises on Tutorial Sheet I.

(i) We need to express 0 as a product of a and something else:

$$0 = a \cdot 0$$

so

$$a \mid 0.$$

Similarly $a = a \cdot 1$, so

$$1 \mid a \qquad \text{and} \qquad a \mid a.$$

(ii) The 'if and only if' means that we have two things to do. We need on the one hand to show that if $a \mid 1$ then $a = \pm 1$; this is the 'only if' part. On the other hand, we need to check that if $a = \pm 1$ then $a \mid 1$. This is the 'if' part and is in this case probably the easiest part.

To prove the 'if' part we need to show that both 1 and -1 do divide 1; that is, we write 1 as a product of 1 and something and also write 1 as a product of -1 and something.

Now $1 = 1 \cdot 1$ and $1 = (-1) \cdot (-1)$, so $1 \mid 1$ and $-1 \mid 1$. Therefore if $a = \pm 1$, then $a \mid 1$.

Conversely, if $a \mid 1$, we have $1 = aq$ and so $1 = (-a)(-q)$. Therefore either a or $-a$ is a positive divides of 1. But no integer greater than 1 can divide 1, so either $a = 1$ or $-a = 1$. Thus $a = \pm 1$.

(iii) Suppose $a \mid b$ and $c \mid d$. This means that there are integers q and r such that

$$b = aq \qquad \text{and} \qquad d = cr.$$

So

$$bd = aq \cdot cr = (ac)(qr).$$

Hence

$$ac \mid bd.$$

(The proofs of (iv) and (vi) are similar.)

(v) Suppose $a \mid b$ and $b \mid a$. This means that there exist integers q and r such that

$$b = aq \qquad \text{and} \qquad a = br.$$

Substitute the former into the latter:

$$a = aqr$$

Therefore

$$a(1 - qr) = 0.$$

Hence either $a = 0$ or $1 - qr = 0$.

If $a = 0$, then $b = aq = 0$ and so $a = b$.

If $1 - qr = 0$, then $qr = 1$. Thus $q \mid 1$, so $q = \pm 1$ by (ii). Hence $b = aq = \pm a$.

Therefore in either case $a = \pm b$. □

Positional Notation

Finally for this section, we shall note that the Basic Fact also has an impact on how we write a number down.

Recall first that the number 1234 means

$$1 \cdot 1000 + 2 \cdot 100 + 3 \cdot 10 + 4 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0.$$

Generalising this idea leads us to:

Definition 1.6 Fix a positive integer $b > 1$. If a is a positive integer, we write

$$a = \overline{(d_n d_{n-1} \dots d_1 d_0)}_b$$

to denote the expression of a in *base* b. The d_i are called the *digits* and the notation means that

$$a = d_n b^n + d_{n-1} b^{n-1} + \cdots + d_1 b + d_0,$$

where $0 \leqslant d_i < b$ for all $i = 0, 1, \dots, n$.

To write a positive integer a in base b, follow the following method:

Method:

- Divide a by b: $\quad a = bq + r$.
- Then r is the last digit: $d_0 = r$.
- Apply the method with a replaced by q to find the successive digits $d_1, d_2, \dots, d_n$.

Example 1.7 Let us write 37 in base 5.

$$\begin{aligned} 37 &= 5 \cdot 7 + 2 && \Longrightarrow \quad \text{Last digit is 2} \\ 7 &= 5 \cdot 1 + 2 && \Longrightarrow \quad \text{Next digit is 2} \\ 1 &= 5 \cdot 0 + 1 && \Longrightarrow \quad \text{Next digit is 1} \end{aligned}$$

Thus

$$37 = \overline{122}_5.$$

Indeed

$$37 = 5^2 + 2 \cdot 5 + 2.$$

Now let us write 37 in base 3.

$$\begin{aligned} 37 &= 3 \cdot 12 + 1 \\ 12 &= 3 \cdot 4 + 0 \\ 4 &= 3 \cdot 1 + 1 \\ 1 &= 3 \cdot 0 + 1 \end{aligned}$$

So

$$37 = \overline{1101}_3 = 3^3 + 3^2 + 1.$$

We may apply the idea used to calculate the expression to prove that the expression for an integer in base b both exists and is unique.

Theorem 1.8 (Positional Notation) *Let $b > 1$ be a fixed integer. Every positive integer a can be written as*

$$a = d_n b^n + d_{n-1} b^{n-1} + \cdots + d_1 b + d_0$$

where $n \geqslant 0$ and $0 \leqslant d_i < b$ for all $i = 0, 1, \ldots, n$.

Moreover, if d_n is required to be non-zero, then n and all the d_i are uniquely determined.

PROOF: We prove the existence part of the theorem by induction on a. If $a = 1$, we take $n = 0$ and $d_0 = 1$, and the result holds. Suppose then that $a > 1$ and that the result holds for all integers smaller than a. First divide a by b:

$$a = bq + d_0$$

where $0 \leqslant d_0 < b$. Now $q = (a - d_0)/b < a$ (since $b > 1$), so by induction

$$q = d_n b^{n-1} + d_{n-1} b^{n-2} + \cdots + d_1$$

where $0 \leqslant d_i < b$ for all i. Thus

$$a = qb + d_0 = d_n b^n + d_{n-1} b^{n-1} + \cdots + d_1 b + d_0.$$

This completes the induction step and the existence part of the proof is established.

Now we turn to uniqueness. Suppose we have the above formula for a and also another formula

$$a = e_m b^m + e_{m-1} b^{m-1} + \cdots + e_1 b + e_0$$

with $e_m \neq 0$ and $d_n \neq 0$. Then

$$a = b(e_m b^{m-1} + \cdots + e_1) + e_0.$$

Since the quotient and the remainder in the Basic Fact are unique we must have

$$e_0 = d_0$$

and

$$e_m b^{m-1} + \cdots + e_1 = q = d_n b^{n-1} + \cdots + d_1.$$

Now applying induction to q, we see that $m = n$ and $d_i = e_i$ for all i. This establishes uniqueness. □

Section 2

Greatest Common Divisors and the Euclidean Algorithm

We have already met what it means for one integers to divide another. The greatest common divisor of two integers is the largest integer that divides both of them.

Definition 2.1 Let a and b be two integers (at least one of which is non-zero). The *greatest common divisor* $\gcd(a, b)$ is the largest integer d which divides both a and b.

Thus $d = \gcd(a, b)$ is defined by the following two properties:

- $d \mid a$ and $d \mid b$
- If $c \mid a$ and $c \mid b$, then $c \leqslant d$.

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

For rather small integers, it is fairly easy to calculate the greatest common divisor by inspection.

Example 2.2 (i) $\gcd(10, 15) = 5$

(ii) $\gcd(16, 20) = 4$

(iii) $\gcd(14, 25) = 1$

For much larger pairs of integers we shall need to use a more advanced tool to find the greatest common divisor. This is the Euclidean Algorithm which we shall now describe.

The first thing to note is that the positive divisors of an integer a coincide with those of $-a$. Accordingly we shall only describe the algorithm

for calculating the greatest common divisor for positive integers. Since $\gcd(a, b) = \gcd(b, a)$, there is also no loss of generality in assuming that the integers a and b involved satisfy $a \geqslant b$.

Algorithm 2.3 (Euclidean Algorithm)

Input: Two positive integers a and b with $a \geqslant b$.

Output: The greatest common divisor $\gcd(a, b)$.

Method:

- *Step 1:* Define $a_1 = a, \ \ b_1 = b$.
  Divide a_1 by b_1: $\quad a_1 = b_1 q_1 + r_1$.
- *Step n:* Define $a_n = b_{n-1}, \ \ b_n = r_{n-1}$.
  Divide a_n by b_n: $\quad a_n = b_n q_n + r_n$.
- Repeat until $r_k = 0$.
- The last non-zero remainder r_{k-1} is $\gcd(a, b)$.

Example 2.4 Calculate $\gcd(143, 559)$.

- *Step 1:* $a_1 = 559, \ \ b_1 = 143$.

$$559 = 143 \cdot 3 + 130$$

- *Step 2:* $a_2 = 143, \ \ b_2 = 130$.

$$143 = 130 \cdot 1 + 13$$

- *Step 3:* $a_3 = 130, \ \ b_3 = 13$.

$$130 = 13 \cdot 10 + 0$$

Here $r_3 = 0$ and so the last non-zero remainder is $r_2 = 13$. Thus

$$\gcd(143, 559) = 13.$$

Having a supposed algorithm is all well and good, but what reason do we have to believe the algorithm really does what it is supposed to? On the face of it the Euclidean Algorithm just returns some positive integer to us. We must prove that this positive integer actually is the greatest common divisor of the two input numbers. The first step in doing this is the following.

Lemma 2.5 *If a, b, q and r are integers satisfying $a = bq + r$, then*

$$\gcd(a, b) = \gcd(b, r).$$

PROOF: Let $d_1 = \gcd(a, b)$ and $d_2 = \gcd(b, r)$. By definition

$$d_1 \mid a \qquad \text{and} \qquad d_1 \mid b.$$

Therefore

$$d_1 \mid (a - qb) = r.$$

From $d_1 \mid b$ and $d_1 \mid r$, we deduce $d_1 \leqslant d_2$ (since d_2 is the *greatest* common divisor of b and r).

Similarly $d_2 \mid b$ and $d_2 \mid r$, so $d_2 \mid (bq + r) = a$. Hence d_2 divides both a and b, so $d_2 \leqslant d_1$ (as d_1 is the greatest common divisor of a and b).

Therefore $d_1 = d_2$, as claimed. □

Let us now return to the Euclidean Algorithm.

Suppose a and b are positive integers with $a \geqslant b$ to which we apply the Algorithm. This generates for us a sequence of pairs of non-negative integers:

$$a_1, b_1, \quad a_2, b_2, \quad \ldots, \quad a_k, b_k$$

defined by

$$a_1 = a, \qquad b_1 = b$$

and for $n \geqslant 2$ by

$$a_n = b_{n-1} \quad \text{and} \quad b_n = r_{n-1} \qquad \text{where } a_{n-1} = b_{n-1}q_{n-1} + r_{n-1}.$$

First note that since the remainder satisfies $r_{n-1} < b_{n-1}$ when we divide, we have

$$b_1 > b_2 > \cdots > b_k.$$

This tells us immediately that the process must stop: we do eventually hit a point where the remainder is zero.

Furthermore Lemma 2.5 tells us that

$$\gcd(a_{n-1}, b_{n-1}) = \gcd(b_{n-1}, r_{n-1}) = \gcd(a_n, b_n)$$

so

$$\gcd(a_1, b_1) = \gcd(a_2, b_2) = \cdots = \gcd(a_k, b_k).$$

At the last stage we have $r_k = 0$, so $a_k = b_k q_k$; that is, $b_k \mid a_k$. Hence

$$\gcd(a_k, b_k) = b_k = r_{k-1}.$$

Therefore

$$\gcd(a, b) = r_{k-1},$$

which is precisely what the Euclidean Algorithm gives us.

This means we have proved the first part of the following:

Theorem 2.6 (i) *The Euclidean Algorithm works: given positive integers a and b with $a \geqslant b$, applying the Euclidean Algorithm calculates* $\gcd(a, b)$.

(ii) *Given integers a and b (one of which is non-zero) there exist integers u and v such that*

$$\gcd(a, b) = ua + vb.$$

PROOF OF (ii): First note that $\gcd(a, 0) = a = 1 \cdot a + 1 \cdot 0$, so the result holds if one of a or b is zero.

To complete the proof of (ii) we may assume that a and b are both positive, since the greatest common divisor is unchanged if we alter the sign of a or b and to complete the proof we will simply need to change the sign of the corresponding u or v as appropriate.

Without loss of generality assume $a \geqslant b$. Apply the Euclidean Algorithm to a and b to generate a sequence of pairs of integers

$$a_1, b_1, \quad a_2, b_2, \quad \ldots, \quad a_k, b_k.$$

Claim: For all values of n, a_n and b_n both have the form $ua + vb$ (for various u and v).

We prove the claim by induction on n. Firstly

$$a_1 = a = 1 \cdot a + 0 \cdot b, \qquad b_1 = b = 0 \cdot a + 1 \cdot b,$$

so the result holds for $n = 1$.

Suppose now that $n > 1$ and that

$$a_{n-1} = ua + vb, \qquad b_{n-1} = u'a + v'b$$

for some integers u, v, u' and v'. Now the steps in the Euclidean Algorithm tell us that

$$a_n = b_{n-1} = u'a + v'b$$

while

$$\begin{aligned} b_n = r_{n-1} &= a_{n-1} - b_{n-1}q_{n-1} \\ &= ua + vb - q_{n-1}(u'a + v'b) \\ &= (u - q_{n-1}u')a + (v - q_{n-1}v')b. \end{aligned}$$

Hence a_n and b_n also have the required form. Therefore the claim holds by induction.

Finally $\gcd(a, b) = r_{k-1} = b_k$ (the last non-zero remainder). From the claim this has the required form. □

Example 2.7 Take $a = 776,\ \ b = 544$. Apply the Euclidean Algorithm:

- Step 1: $a_1 = 776,\ \ b_1 = 544.$

$$776 = 544 \cdot 1 + 232$$

- Step 2: $a_2 = 544,\ \ b_2 = 232.$

$$544 = 232 \cdot 2 + 80$$

- Step 3: $a_3 = 232,\ \ b_3 = 80.$

$$232 = 80 \cdot 2 + 72$$

- Step 4: $a_4 = 80,\ \ b_4 = 72.$

$$80 = 72 \cdot 1 + 8$$

- Step 5: $a_5 = 72,\ \ b_5 = 8.$

$$72 = 8 \cdot 9 + 0$$

So

$$\gcd(776, 544) = 8.$$

Reversing the steps enables us to write the greatest common divisor as a multiple of 776 added to a multiple of 544:

$$\begin{aligned} 8 &= 80 - 72 \\ &= 80 - (232 - 2 \cdot 80) \\ &= 3 \cdot 80 - 232 \\ &= 3(544 - 2 \cdot 232) - 232 \\ &= 3 \cdot 544 - 7 \cdot 232 \\ &= 3 \cdot 544 - 7(776 - 544) \\ &= 10 \cdot 544 - 7 \cdot 776. \end{aligned}$$

So

$$\gcd(776, 544) = 8 = (-7) \cdot 776 + 10 \cdot 544$$

and we need take $u = -7$ and $v = 10$.

Section 3

Prime Numbers and Prime Factorisation

Definition 3.1 A *prime number* is an integer $p > 1$ whose only positive divisors are 1 and p.

Example 3.2 The first few prime numbers are:

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

See the *Prime Pages* (`http://www.utm.edu/research/primes/`) for longer lists of primes and much other interesting information.

The primes are useful since they form the 'building blocks' from which all other integers are made.

Theorem 3.3 (Fundamental Theorem of Arithmetic) *Any integer n with $n > 1$ can be written uniquely in the form*

$$n = p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r}$$

where the p_i are prime numbers with $p_1 < p_2 < \cdots < p_r$ and the k_i are positive integers.

Example 3.4 $180 = 2^2 \cdot 3^2 \cdot 5$.

Proof of Theorem 3.3: We shall need some auxiliary results before we can establish the uniqueness part, but we can establish the existence part straight away.

We proceed by induction on n. If n is a prime (including $n = 2$, the base case in the induction), then already n is a product of prime powers (namely a single prime) so there is nothing to prove. Assume then that $n > 2$ and

that n factorises, say $n = st$ where $1 < s, t < n$. By induction, both s and t can be written as a product of prime powers. Hence, upon multiplying these expressions together, we see that $n = st$ is also a product of prime powers.

This establishes existence of a prime decomposition. □

For the uniqueness part we prove:

Lemma 3.5 *Let p be a prime number.*

(i) *If $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

(ii) *If $p \mid a_1 a_2 \dots a_s$, then $p \mid a_i$ for some i.*

(iii) *If $p \mid q_1 q_2 \dots q_t$ where each q_i is a prime number, then $p = q_j$ for some j.*

PROOF: (i) Assume $p \mid ab$. If $p \mid a$ then the result holds. So assume that p does not divide a. Then $\gcd(p, a) = 1$ (since the only divisors of p are 1 and p and the latter does not divide a). Now part (ii) of Theorem 2.6 tells us that

$$1 = up + va$$

for some $u, v \in \mathbb{Z}$. Hence

$$b = ubp + vab.$$

Now $p \mid ab$ (by assumption), so we deduce $p \mid (ubp + vab)$; that is, $p \mid b$, as required.

(ii) Proceed by induction on s. If $s = 1$, then $p \mid a_1$ and there is nothing to prove. Assume then that $s > 1$. We have $p \mid ba_s$ where $b = a_1 a_2 \dots a_{s-1}$. Hence, by (i), either $p \mid b$ or $p \mid a_s$. If the first holds, that is, $p \mid b$, then by induction p divides one of a_1, a_2, $\dots$, a_{s-1}. Hence we deduce $p \mid a_i$ for some i, completing the inductive step.

(iii) Apply (ii). We deduce $p \mid q_j$ for some j. But as q_j is prime, its only divisors are 1 and q_j. Hence $p = q_j$. □

We now return to the uniqueness part of Theorem 3.3. Assume that we have expressed n in the required form in two different ways:

$$n = p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} = q_1^{l_1} q_2^{l_2} \dots q_t^{l_t}. \tag{3.1}$$

Note that $p_i \mid n = q_1^{l_1} q_2^{l_2} \dots q_t^{l_t}$. Hence by Lemma 3.5 we have $p_i = q_j$ for some j. By the same argument, each q_l is equal to some p_m. We conclude that

$$r = t,\ p_1 = q_1,\ p_2 = q_2, \dots, p_r = q_r.$$

Then Equation (3.1) becomes

$$p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} = p_1^{l_1} p_2^{l_2} \dots p_r^{l_r}.$$

Assume $k_i \neq l_i$ for some i. We may assume without loss of generality that $k_i > l_i$. Then dividing through by $p_i^{l_i}$ gives

$$p_1^{k_1} \dots p_{i-1}^{k_{i-1}} p_i^{k_i - l_i} p_{i+1}^{k_{i+1}} \dots p_r^{k_r} = p_1^{l_1} \dots p_{i-1}^{l_{i-1}} p_{i+1}^{l_{i+1}} \dots p_r^{l_r}.$$

Hence p_i divides $p_1^{l_1} \dots p_{i-1}^{l_{i-1}} p_{i+1}^{l_{i+1}} \dots p_r^{l_r}$. This implies that $p_i = p_j$ (with $i \neq j$) by Lemma 3.5(iii). This is a contradiction. Thus $k_i = l_i$ for all i and we have established the required uniqueness. □

We have established that a positive integer can be factorised uniquely as a product of prime numbers. This decomposition can also be used to find all the divisors. Suppose

$$n = p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

where the p_i are prime numbers with $p_1 < p_2 < \dots < p_r$ and $k_i \in \mathbb{N}$ for all i. Let $m \mid n$ with m positive, so write

$$n = mu$$

for some integer u. We can decompose m as a product of prime powers

$$m = q_1^{l_1} q_2^{l_2} \dots q_s^{l_s}.$$

Then each $q_i \mid m$, so $q_i \mid n$ and Lemma 3.5 gives

$$q_i = p_j \qquad \text{for some } j.$$

Thus m is a product of powers of some of the primes occurring in the prime factorisation of n. The same argument can be applied to u: it too is a product of powers of some of the p_i. Let

$$m = p_1^{s_1} p_2^{s_2} \dots p_r^{s_r}, \qquad u = p_1^{t_1} p_2^{t_2} \dots p_r^{t_r}$$

where $s_i \geqslant 0$ and $t_i \geqslant 0$ for all i. Then

$$n = mu = p_1^{s_1+t_1} p_2^{s_2+t_2} \dots p_r^{s_r+t_r}.$$

The uniqueness of the expression implies

$$s_i + t_i = k_i \qquad \text{for all } i$$

and hence

$$0 \leqslant s_i = k_i - t_i \leqslant k_i \qquad \text{for all } i.$$

Theorem 3.6 *Let $n = p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$ be a positive integer expressed as a product of prime powers (so each p_i is a prime number and each k_i is a positive integer). The positive divisors of n are precisely the numbers of the form*

$$p_1^{s_1} p_2^{s_2} \dots p_r^{s_r}$$

where $0 \leqslant s_i \leqslant k_i$ for $i = 1, 2, \dots, r$. □

Example 3.7 The divisors of 180 are:

$$\begin{array}{lll}
1 = 2^0 \cdot 3^0 \cdot 5^0 & 9 = 2^0 \cdot 3^2 \cdot 5^0 & 15 = 2^0 \cdot 3^1 \cdot 5^1 \\
2 = 2^1 \cdot 3^0 \cdot 5^0 & 18 = 2^1 \cdot 3^2 \cdot 5^0 & 30 = 2^1 \cdot 3^1 \cdot 5^1 \\
4 = 2^2 \cdot 3^0 \cdot 5^0 & 36 = 2^2 \cdot 3^2 \cdot 5^0 & 60 = 2^2 \cdot 3^1 \cdot 5^1 \\
3 = 2^0 \cdot 3^1 \cdot 5^0 & 5 = 2^0 \cdot 3^0 \cdot 5^1 & 45 = 2^0 \cdot 3^2 \cdot 5^1 \\
6 = 2^1 \cdot 3^1 \cdot 5^0 & 10 = 2^1 \cdot 3^0 \cdot 5^1 & 90 = 2^1 \cdot 3^2 \cdot 5^1 \\
12 = 2^2 \cdot 3^1 \cdot 5^0 & 20 = 2^2 \cdot 3^0 \cdot 5^1 & 180 = 2^2 \cdot 3^2 \cdot 5^1
\end{array}$$

We now turn to consider the properties of the prime numbers. We have already observed that they are the basic building blocks from which all other positive integers are constructed via multiplication. Furthermore, they have been recognised as important and have been studied since ancient times. Many deep theorems have been proved about them, often involving startling and surprising methods. On the other hand, there are many questions about them which are easy to state but which have still not been answered. For the rest of this lecture I shall begin to examine some of these theorems and questions. We begin with an important fact which was known by Euclid.

Theorem 3.8 *There are infinitely many prime numbers.*

PROOF: Assume for a contradiction that there are only finitely many primes, say $p_1, p_2, \ldots, p_n$. Consider

$$N = p_1 p_2 \ldots p_n + 1.$$

This number N must be divisible by a prime, say p_i divides N. Then

$$p_i \mid (N - p_1 p_2 \ldots p_n) = 1,$$

which is a contradiction. □

The argument can be modified in a number of ways to show that there are infinitely many primes of certain forms. First note that as 2 is the only even prime, it follows that all primes greater than 2 either have the form $4k + 1$ or $4k + 3$. We can prove:

Theorem 3.9 *There are infinitely many prime numbers of the form* $4k+3$.

PROOF: Suppose that $p_1, p_2, \ldots, p_n$ are all the primes of the form $4k + 3$. Consider

$$N = 4p_1 p_2 \ldots p_n - 1 = 4(p_1 p_2 \ldots p_n - 1) + 3.$$

Then N is a product of prime numbers, all of which must be odd. Note that

$$(4k + 1)(4l + 1) = 4kl + 4k + 4l + 1 = 4(kl + k + l) + 1.$$

Hence if N were a product of primes all of which had the form $4k+1$, then N would also have this form. Therefore one of the prime divisors of N must have the form $4k+3$, so is p_i for some i. Then

$$p_i \mid (4p_1p_2 \dots p_n - N) = 1,$$

a contradiction. This completes the proof. □

Similarly it is possible to show that every prime greater than 3 has the form $6k+1$ or $6k+5$ and modifying the above argument will show that there are infinitely many primes of the form $6k+5$. In fact, there are also infinitely many primes of the form $4k+1$ and infinitely many primes of the form $6k+1$, but this is much harder to prove. These are all special cases of the following theorem proved by Dirichlet:

Theorem 3.10 (Dirichlet 1837) *If a and b are coprime positive integers then there are infinitely many prime numbers of the form $ak+b$ ($k = 0, 1, 2, \dots$).*

The proof of this theorem is well beyond what we can hope to prove in this course, but it does at least give a flavour of the development of mathematics.

Consider the sequence of numbers of the form $4k+3$:

$$\underline{3},\ \underline{7},\ \underline{11},\ 15,\ \underline{19},\ \underline{23},\ 27,\ \underline{31},\ 35,\ \dots.$$

We are not claiming that they are all prime, but just that infinitely many of them are: primes will contine to occur in this list no matter how far we go along it.

In fact there is no known simple formula which yields only prime numbers whatever we substitute into it. For some time mathematicians believed that

$$f(n) = n^2 + n + 41$$

was such a formula having checked that $f(n)$ is prime for $n = 1, 2, 3, \dots, 39$. (Here $f(39) = 1601$.) However this is, of course, not good enough and we find:

$$f(40) = 40^2 + 40 + 41 = 40(40+1) + 41 = 40 \cdot 41 + 41 = 41^2,$$

which is not prime. In fact we can prove that there is no polynomial formula which generates primes.

Theorem 3.11 *There is no polynomial $f(n)$ with integer coefficients which is not constant and which takes only prime values for all non-negative integers n.*

PROOF: Assume that

$$f(n) = a_k n^k + a_{k-1} n^{k-1} + \cdots + a_1 n + a_0$$

is such a polynomial. Then $f(0) = a_0$ is prime, and so is $f(ta_0)$ for all choices of $t = 1, 2, \ldots$. But

$$f(ta_0) = a_k t^k a_0^k + a_{k-1} t^{k-1} a_0^{k-1} + \cdots + a_1 t a_0 + a_0$$

and therefore $a_0 \mid f(ta_0)$ for all t. Since $f(ta_0)$ is prime this forces

$$f(ta_0) = a_0 \qquad \text{for } t = 1, 2, \ldots.$$

Hence the polynomial $f(n)$ takes the value a_0 infinitely many times and therefore $f(n)$ must be the constant polynomial taking value a_0. □

The other thing that we might ask about is the distribution of primes. Although we have observed it is difficult to generate primes, one can actually say quite a lot about their distribution.

Theorem 3.12 *Let p_n denote the nth prime number. Then*

$$p_n \leqslant 2^{2^{n-1}}.$$

This result is pointing to the fact that primes do occur reasonably often.

PROOF: We proceed by induction on n. To start with we have

$$p_1 = 2 = 2^{2^{1-1}}.$$

The inductive step relies on the method of proof of Theorem 3.8. We know that there is a prime number q which divides

$$p_1 p_2 \ldots p_n + 1$$

and this cannot be one of the first n prime numbers. Hence

$$\begin{aligned} p_{n+1} \leqslant q &\leqslant p_1 p_2 \ldots p_n + 1 \\ &\leqslant 2 \cdot 2^2 \cdot \ldots \cdot 2^{2^{n-1}} + 1 \\ &= 2^{1+2+\cdots+2^{n-1}} + 1 \\ &= 2^{2^n - 1} + 1 \\ &= 2^{2^n - 1} + 2^{2^n - 1} = 2^{2^n}. \end{aligned}$$

This completes the proof. □

A much stronger (and far harder to prove) result is that

$$\lim_{n\to\infty} \frac{n \log n}{p_n} = 1.$$

This is one of the equivalent formulations of the famous **Prime Number Theorem** (proved in 1896 by Hadamard and de la Vallée Poussin). It can be interpreted as saying that for large values of n, the nth prime is roughly nearby the number $n \log n$. (This being the *natural logarithm*.)

It is certainly worth mentioning a few open questions that mathematicians have still yet to solve:

Goldbach's Conjecture: Is it true that every even number greater than 2 can be written as the sum of two prime numbers?

Twin Primes Conjecture: Is it true that there are infinitely many prime numbers p such that $p + 2$ is also prime?

We can interpret the Prime Number Theorem as saying that there are many prime numbers and that they occur rather regularly. The above conjectures also assert similar things. However, we finish our discussion of primes by showing that there are large gaps where no primes occur. Namely we prove:

Theorem 3.13 *For every positive integer n, there is a sequence of n consecutive composite numbers.*

('Composite' means a number that is not prime; i.e., composed as a product of more than one prime number.)

PROOF: Consider the following numbers:

$$\begin{array}{ll} (n+1)!+2 & \text{divisible by } 2 \\ (n+1)!+3 & \text{divisible by } 3 \\ \quad\vdots & \\ (n+1)!+(n+1) & \text{divisible by } n+1. \end{array}$$

These n consecutive numbers are all composite. □

Section 4

Linear Diophantine Equations

A *Diophantine equation* is an equation involving a number of variables all of whose coefficients are integers and to which we seek solutions which are integers.

Diophantine Equations with One Variable: These are essentially uninteresting: one simply attempts to solve them as ordinary equations by any method possible and then examines whether the solutions obtained are integers or not.

The behaviour becomes much more interesting if we consider an equation involving two variables.

Example 4.1 Consider

$$x + y = 1.$$

For every choice of x there is a unique solution for y, namely $y = 1 - x$. Thus the equation has infinitely many solutions, all of the form $(x, 1 - x)$ for $x \in \mathbb{Z}$.

Example 4.2 Consider the equation

$$x + 2y = 1.$$

This time we see that a solution for x cannot be arbitrary: it must also be an odd number. On the other hand, given any y there is always a solution for x, namely $x = 1 - 2y$. Thus this equation also has infinitely many solutions, all of the form $(1 - 2y, y)$ for $y \in \mathbb{Z}$.

Example 4.3 Consider the equation

$$3x + 6y = 1.$$

This equation has no solutions: the left-hand side is always a multiple of 3 no matter what choice is made for x and y, while the right-hand side is not a multiple of 3.

Let us now move on to consider the general situation. We start by defining the object of concern.

Definition 4.4 A *linear Diophantine equation* (in two variables) is an equation of the form

$$ax + by = c$$

where a, b and c are integers.

In view of our previous discussion, we have the following natural questions to consider:

- Under what conditions does the above equation have integer solutions?
- If the equation does have solutions, how many solutions does it have?
- How can we find all the solutions?

In view of the last example, it should be unsurprising that the common divisors of a and b are of relevance. We shall address each of these questions in turn.

Existence of Solutions

Consider the general linear Diophantine equation

$$ax + by = c \tag{4.1}$$

where a, b and c are integers. Assume that a and b are both non-zero (so the equation genuinely involves two variables). Let

$$d = \gcd(a, b).$$

Then d divides both a and b so we may write

$$a = da_1 \qquad \text{and} \qquad b = db_1$$

for some integers a_1 and b_1.

Suppose that we do have a solution (x_0, y_0) to the equation. This means $ax_0 + by_0 = c$. Now since d divides a and b, we deduce $d \mid (ax_0 + by_0)$; that is, $d \mid c$.

Conversely suppose $d \mid c$. Write $c = dc_1$. We make use of part (ii) of Theorem 2.6. It tells us that there exist integers u and v such that

$$d = ua + vb.$$

Hence upon multiplying c_1 we obtain

$$uac_1 + vbc_1 = dc_1;$$

that is,

$$a(uc_1) + b(vc_1) = c.$$

Therefore (uc_1, vc_1) is a solution of the equation.

Conclusion: The equation has a solution if and only if $d \mid c$.

Number of Solutions

Suppose that we do have a solution (x_0, y_0) to Equation (4.1). We can find other solutions by taking

$$x = x_0 + b_1 t, \qquad y = y_0 - a_1 t$$

for any integer t. Indeed

$$\begin{aligned} ax + by &= ax_0 + ab_1 t + by_0 - ba_1 t \\ &= (ax_0 + by_0) + (da_1 b_1 t - db_1 a_1 t) \\ &= c + 0 = c. \end{aligned}$$

Since t can be any integer we deduce that our equation has infinitely many solutions.

Finding all Solutions

We have (under the condition $d \mid c$) infinitely many solutions to our linear Diophantine equation. But could there be others about which we are currently unaware?

We shall need the following result in the course of our discussion.

Lemma 4.5 *Let r, s and t be integers and assume that r and s are coprime. If $r \mid st$, then $r \mid t$.*

Recall that to say r and s are coprime is to say that their greatest common divisor is 1.

PROOF: $\gcd(r, s) = 1$, so by part (ii) of Theorem 2.6, there exist integers u and v such that

$$ur + vs = 1.$$

Therefore

$$t = t(ur + vs) = utr + vst.$$

Now $r \mid st$ by assumption, while clearly r divides utr. Hence $r \mid (utr + vst)$, so $r \mid t$, as claimed. □

Now let us return to our linear Diophantine equation (4.1). Suppose we have fixed one solution (x_0, y_0) to (4.1). Let (x, y) be any other solution. So we have

$$ax + by = c = ax_0 + by_0.$$

Hence

$$a_1 d(x - x_0) = b_1 d(y - y_0).$$

Dividing by d gives

$$a_1(x - x_0) = b_1(y - y_0).$$

Now $a_1 = a/d$ and $b_1 = b/d$, so we have $\gcd(a_1, b_1) = 1$ (see Question 3 on Tutorial Sheet II). Hence a_1 and b_1 are coprime, while the above equation tells us

$$a_1 \mid b_1(y - y_0).$$

Hence Lemma 4.5 tells us that

$$a_1 \mid (y_0 - y).$$

This means that $y_0 - y = a_1 t$ for some $t \in \mathbb{Z}$. Substituting into the above equation gives

$$a_1(x - x_0) = b_1 a_1 t.$$

Therefore

$$x - x_0 = b_1 t.$$

Hence $x = x_0 + b_1 t$ and $y = y_0 - a_1 t$.

So we have shown that all solutions to (4.1) arise in the form we previously presented.

We summarise our finding as follows:

Theorem 4.6 *Let a, b and c be integers with a and b not both zero.*

(i) *The linear Diophantine equation*

$$ax + by = c$$

has a solution if and only if $d = \gcd(a, b)$ divides c.

(ii) *If $d \mid c$, then one solution may be found by determining u and v such that $d = ua + vb$ and then setting*

$$x_0 = uc/d \qquad \text{and} \qquad y_0 = vc/d.$$

All other solutions are given by

$$x = x_0 + (b/d)t, \qquad y = y_0 - (a/d)t$$

for $t \in \mathbb{Z}$.

Example 4.7 We shall find all solutions of

$$77x + 42y = 35.$$

First we calculate $\gcd(77, 42)$ using the Euclidean Algorithm:

$$\begin{aligned} 77 &= 42 \cdot 1 + 35 \\ 42 &= 35 \cdot 1 + 7 \\ 35 &= 7 \cdot 5 + 0 \end{aligned}$$

So

$$\gcd(77, 42) = 7.$$

Since 7 does divide 35, this means that the linear Diophantine equation does have integer solutions. To actually find the solutions we first reverse the steps in the Euclidean Algorithm:

$$\begin{aligned} 7 &= 42 - 35 \\ &= 42 - (77 - 42) \\ &= (-1) \cdot 77 + 2 \cdot 42. \end{aligned}$$

So we take $u = -1$ and $v = 2$. One solution is then

$$x_0 = (-1) \cdot 35/7 = -5, \qquad y_0 = 2 \cdot 35/7 = 10.$$

All the solutions are given by

$$\begin{aligned} x &= x_0 + (42/7)t = -5 + 6t \\ y &= y_0 - (77/7)t = 10 - 11t \end{aligned}$$

where $t \in \mathbb{Z}$.

We can also apply these techniques to other types of problem, for example:

Example 4.8 *A customer bought some apples and some oranges, 12 pieces of fruit in total, and they cost him £1.32. If an apple costs 3p more than an orange, and if more apples than oranges were purchased, how many pieces of each fruit were bought?*

Solution: Let x be the number of apples bought. Then $12 - x$ is the number of oranges bought. Let y be the cost of an apple. Then $y - 3$ is the cost of an orange. We obtain the following equation

$$xy + (12 - x)(y - 3) = 132.$$

Therefore

$$xy + 12y - 36 - xy + 3x = 132$$

$$3x + 12y = 168$$

$$x + 4y = 56$$

We can solve this equation by inspection:

$$x = 56 - 4t, \quad y = t \qquad (\text{for } t \in \mathbb{Z}).$$

But we have further requirements: $6 < x < 12$, so

$$6 < 56 - 4t < 12.$$

Therefore

$$44 < 4t < 50$$

$$11 < t < 12\tfrac{1}{2}.$$

Hence $t = 12$. We deduce that

$$x = 8, \qquad y = 12.$$

So the customer bought 8 apples at 12p each and 4 oranges at 9p each.

(Finally check our working: $8 \cdot 12 + 4 \cdot 9 = 132$.)

Example 4.9 *Suppose that we have available postage stamps in two denominations: 5p and 7p. What values can one make using combinations of stamps?*

(E.g., $10 = 5 + 5$, $12 = 5 + 7$, etc.)

Solution: We are asking for what values of c does

$$5x + 7y = c$$

have (non-negative) solutions? Now $\gcd(5, 7) = 1$, so our theory tells us that the equation does always have solutions (but possibly they are negative and one cannot put a negative number of stamps on a parcel!)

Let us instead follow the standard method and adjust at the appropriate point to ensure we are getting non-negative solutions. First apply the Euclidean Algorithm:

$$7 = 5 \cdot 1 + 2$$
$$5 = 2 \cdot 2 + 1$$
$$2 = 1 \cdot 2 + 0.$$

(So the greatest common divisor is indeed 1.) Reversing these steps:

$$\begin{aligned} 1 &= 5 - 2 \cdot 2 \\ &= 5 - 2(7 - 5) \\ &= 3 \cdot 5 + (-2) \cdot 7. \end{aligned}$$

So take $u = 3,\ \ v = -2$. One solution to the linear Diophantine equation is then:

$$x_0 = 3c, \qquad y_0 = -2c.$$

The general solution to the problem is then

$$x = 3c - 7t, \qquad y = -2c + 5t.$$

To achieve non-negative solutions we require

$$3c - 7t \geqslant 0, \qquad \text{i.e., } t \leqslant 3c/7$$

and

$$-2c + 5t \geqslant 0, \qquad \text{i.e., } t \geqslant 2c/5.$$

Hence we require that the integer t lie between $2c/5$ and $3c/7$; that is, that there is at least one integer between these numbers. How far apart are they?

$$3c/7 - 2c/5 = (15c - 14c)/35 = c/35.$$

Hence if $c \geqslant 35$, this gap is $\geqslant 1$ and there definitely will be an integer in the region we want. Thus for $c \geqslant 35$, non-negative solutions exist.

Conclusion: Any value of 35p or greater can be achieved using 5p and 7p stamps.

(Values smaller than 35p will have to be checked by hand.)

In fact, it turns out that the crucial point here is that the a and b we are considering here (5 and 7) are coprime. Provided we know this there will always be some point beyond which all integers can be achieved using a combination of multiples of a and b.

Theorem 4.10 *Let a and b be coprime positive integers. Then every number $c \geqslant ab$ can be expressed as $\lambda a + \mu b$ with λ and μ non-negative integers.*

The proof is omitted, but essentially it is the same argument as supplied to solve the above problem.

Section 5

Congruences

Definition 5.1 Let m be an integer with $m > 1$. We say that two integers a and b are *congruent modulo* m if $a - b$ is divisible by m. We denote this by $a \equiv b \pmod{m}$.

Thus

$$a \equiv b \pmod{m} \qquad \text{if and only if} \qquad m \mid (a - b).$$

The concept of congruence links nicely back to our "basic fact" about division. If a is an integer and we attempt to divide by m, we obtain quotient q and remainder r:

$$a = mq + r.$$

Then $a - r = mq$, so $m \mid (a - r)$. Then, by definition of congruence, $a \equiv r \pmod{m}$.

Example 5.2 $3 \equiv 24 \pmod{7}$
$-31 \equiv 11 \pmod{7}$
$20 \not\equiv 100 \pmod{7}$

Example 5.3 $a \equiv 0 \pmod{m}$ if and only if $m \mid a$.

Example 5.4 Two integers are congruent modulo 10 if and only if they end with the same digit.

We shall begin by developing some properties of congruence arithmetic.

Theorem 5.5 *Let m be an integer with $m > 1$ and let a, b and c be integers.*

(i) $a \equiv a \pmod{m}$.

(ii) *If* $a \equiv b \pmod{m}$, *then* $b \equiv a \pmod{m}$.

(iii) *If* $a \equiv b \pmod{m}$ *and* $b \equiv c \pmod{m}$, *then* $a \equiv c \pmod{m}$.

We shall meet names for the properties presented in this theorem in the next section. For the moment we note that congruence modulo m has some similarity to equality: we can view integers that are congruent modulo m as being 'somehow the same.'

PROOF: (i) $a - a = 0$, which is divisible by m, so $a \equiv a \pmod{m}$.

(ii) Assume $a \equiv b \pmod{m}$. This means $m \mid (a - b)$, so $a - b = mq$ for some integer q. Therefore $b - a = -mq = m(-q)$, so $m \mid (b - a)$. Hence $b \equiv a \pmod{m}$.

(iii) Assume $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. This means $m \mid (a-b)$ and $m \mid (b - c)$, so there exist integers q and r such that $a - b = mq$ and $b - c = mr$. Therefore

$$a - c = (a - b) + (b - c) = mq + mr = m(q + r)$$

and we deduce $m \mid (a - c)$; that is, $a \equiv c \pmod{m}$. □

Theorem 5.6 (Congruence Arithmetic) *Suppose that* m *is an integer with* $m > 1$ *and let* a, b, c, d *and* k *be integers with* $k \geqslant 0$.

(i) *If* $a \equiv b \pmod{m}$ *and* $c \equiv d \pmod{m}$, *then*

$$a + c \equiv b + d \pmod{m}$$

and

$$ac \equiv bd \pmod{m}.$$

(ii) *If* $a \equiv b \pmod{m}$, *then*

$$a + c \equiv b + c \pmod{m}$$
$$ac \equiv bc \pmod{m}$$

and

$$a^k \equiv b^k \pmod{m}.$$

PROOF: (i) By assumption $m \mid (a-b)$ and $m \mid (c-d)$. Therefore there exist integers q and r such that

$$a - b = mq \qquad \text{and} \qquad c - d = mr.$$

Then

$$\begin{aligned}(a + c) - (b + d) &= (a - b) + (c - d)\\ &= mq + mr\\ &= m(q + r).\end{aligned}$$

So $m \mid \big((a+c)-(b+d)\big)$; that is

$$a + c \equiv b + d \pmod{m}.$$

Also

$$\begin{aligned} ac - bd &= ac - bc + bc - bd \\ &= (a-b)c + b(c-d) \\ &= mqc + bmr \\ &= m(qc + br), \end{aligned}$$

so $m \mid (ac - bd)$; that is,

$$ac \equiv bd \pmod{m}.$$

(ii) Since $c \equiv c \pmod{m}$, we can take $c = d$ in (i) to give

$$a + c \equiv b + c \pmod{m} \qquad \text{and} \qquad ac \equiv bc \pmod{m}.$$

Equally if we assume by induction that $a^{k-1} \equiv b^{k-1} \pmod{m}$, then part (i) gives

$$a^k = a \cdot a^{k-1} \equiv b \cdot b^{k-1} = b^k \pmod{m}.$$

□

Example 5.7 The number $2^{70} + 3^{70}$ is divisible by 13.

Indeed

$$2^6 = 64 \equiv -1 \pmod{13},$$

so

$$2^{66} = (2^6)^{11} \equiv (-1)^{11} = -1 \pmod{13},$$

so

$$2^{70} = 2^{66} \cdot 2^4 \equiv (-1) \cdot 16 \equiv 10 \pmod{13}.$$

Similarly

$$3^3 = 27 \equiv 1 \pmod{13},$$

so

$$3^{69} = (3^3)^{23} \equiv 1^{23} = 1 \pmod{13},$$

so

$$3^{70} = 3^{69} \cdot 3 \equiv 1 \cdot 3 = 3 \pmod{13}.$$

Hence

$$2^{70} + 3^{70} \equiv 10 + 3 = 13 \equiv 0 \pmod{13},$$

so

$$13 \mid (2^{70} + 3^{70}).$$

Example 5.8 Let us find the last digit of 2^{1003}. We work modulo 10:

$$\begin{aligned} 2^1 &= 2 \\ 2^2 &= 4 \\ 2^3 &= 8 \\ 2^4 &= 16 \equiv 6 \pmod{10}. \end{aligned}$$

Now note

$$6^2 = 36 \equiv 6 \pmod{10}.$$

Assume, as an inductive hypothesis, that $6^{k-1} \equiv 6 \pmod{10}$ (where $k \geqslant 2$). Then

$$6^k = 6^{k-1} \cdot 6 \equiv 6 \cdot 6 \equiv 6 \pmod{10}.$$

Hence $6^k \equiv 6 \pmod{10}$ for all positive integers k. So

$$\begin{aligned} 2^{1003} = (2^4)^{1003} \cdot 2^3 &\equiv 6^{250} \cdot 8 \pmod{10} \\ &\equiv 6 \cdot 8 \pmod{10} \\ &= 48 \\ &\equiv 8 \pmod{10}. \end{aligned}$$

So the last digit of 2^{1003} is 8.

Theorem 5.9 *Every positive integer is congruent to the sum of its digits modulo* 3 *and also modulo* 9.

PROOF: First note that $10 \equiv 1 \pmod{9}$ and $10 \equiv 1 \pmod{3}$. Therefore

$$10^k \equiv 1^k = 1 \pmod{9}.$$

Write $a = \overline{(d_n d_{n-1} \ldots d_1 d_0)}_{10}$, so

$$\begin{aligned} a &= d_n \cdot 10^n + d_{n-1} \cdot 10^{n-1} + \cdots + d_1 \cdot 10 + d_0 \\ &\equiv d_n \cdot 1 + d_{n-1} \cdot 1 + \cdots + d_1 \cdot 1 + d_0 \\ &= d_n + d_{n-1} + \cdots + d_1 + d_0. \end{aligned}$$

The same argument applies if we work modulo 3. □

Corollary 5.10 *A positive integer is divisible by* 9 *if and only if the sum of its digits is divisible by* 9.

A positive integer is divisible by 3 *if and only if the sum of its digits is divisible by* 3. □

Section 6

Functions and Relations

In this section we discuss two useful tools in Pure Mathematics. They will be applied afterwards to the concepts we have been discussing previously and also to further material we shall introduce.

Functions

You will have met functions in other courses, but probably were more interested in differentiating them. Here we shall be more interested in more abstract properties. We begin with a definition:

Definition 6.1 Let X and Y be sets. A *function* $f\colon X \to Y$ is a "rule" which associates to each $x \in X$ some element in Y. We denote this rule by

$$f\colon x \mapsto f(x) \qquad (\text{for } x \in X).$$

We then say that f *maps the element* x *to the element* $f(x)$.

We call X the *domain* of the function f and Y the *codomain* (or *range*) of f.

Example 6.2 (i) We can define a function

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

which maps each real number to its square. This is the sort of familiar function considered in MT1002, for example.

(ii) Define a function $f\colon \mathbb{Z} \to \mathbb{N}$ by

$$f(x) = \begin{cases} \text{the smallest prime dividing } x & \text{if this exists (i.e,., } x \neq \pm 1) \\ 1 & \text{otherwise.} \end{cases}$$

Although not given by closed formula as the first example is, this is a perfectly valid function. It does give a *unique* value to $f(x)$ for each $x \in \mathbb{Z}$. For example,

$$f(0) = 2, \quad f(1) = 1, \quad f(2) = 2, \quad f(3) = 3, \quad f(4) = 2, \\ f(5) = 5, \quad f(6) = 2, \quad f(7) = 7, \ldots.$$

(iii) Take $P = \{\, x \in \mathbb{R} \mid x > 0 \,\}$. The following is not a valid definition of a function $f\colon P \to \mathbb{R}$:

$$f(x) = y \text{ satisfying } y^2 = x.$$

The problem here is that there is no uniquely determined y for all x. For example, for $f(2)$ should we choose $\sqrt{2}$ or $-\sqrt{2}$?

In much of Pure Mathematics (especially algebra), we are interested in functions with the following properties:

Definition 6.3 Let X and Y be sets and let $f\colon X \to Y$ be a function.

(i) We say f is *one-one* (or *injective*) if different elements in X always map to different elements in Y: that is,

$$f(x_1) = f(x_2) \qquad \Longrightarrow \qquad x_1 = x_2$$

for $x_1, x_2 \in X$.

[PICTURE HERE – CONVERT TO TIKZ]

(ii) We say f is *onto* (or *surjective*) if every element in Y is the image of some element in X: that is,

$$\text{For all } y \in Y, \text{ there is some } x \in X \text{ with } f(x) = y.$$

PICTURE HERE – CONVERT TO TIKZ

(iii) We say f is *bijective* if it is both one-one and onto.

Example 6.4 Define $f\colon \mathbb{Z} \to \{0, 1\}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is even} \\ 1 & \text{if } x \text{ is odd.} \end{cases}$$

Then f is onto, since $f(0) = 0$ and $f(1) = 1$. If we were to choose a larger codomain (but the same definition of the function) then it would cease to be onto.

This f is not one-one: $f(1) = f(3)$, so we have two distinct points in the domain mapping to the same image.

Example 6.5 This example illustrates that the behaviour for the same formula can altered by the choice of domain and codomain.

Define $f\colon \mathbb{Z} \to \mathbb{Z}$ by

$$f(x) = x + 1.$$

***f* is one-one** Let $x, y \in \mathbb{Z}$ and suppose $f(x) = f(y)$. This means $x + 1 = y + 1$ and (subtracting 1 from both sides) we deduce $x = y$. Hence f is one-one.

***f* is onto** Let $b \in \mathbb{Z}$. Define $x = b - 1 \in \mathbb{Z}$. Then $f(x) = x + 1 = b$. Hence every $b \in \mathbb{Z}$ is the image of some element in $\mathbb{Z}$ under f, so f is onto.

Now define a function $g\colon \mathbb{N} \to \mathbb{N}$ (where $\mathbb{N} = \{1, 2, 3, \dots\}$) by

$$g(x) = x + 1.$$

This is a new function: although it has the same "rule" as f it has a different domain and codomain, so is different. This function g is still one-one (by the same argument). But

$$g(x) = x + 1 \geqslant 2 \qquad \text{for all } x \in \mathbb{N}$$

so $g(x) \neq 1$ always. Hence g is not onto.

We shall finish our discussion about functions by considering bijective functions in a little more detail. Suppose $f\colon X \to Y$ is a bijective function. Then for each $y \in Y$, there exists some $x \in X$ such that $f(x) = y$ (because f is *surjective*). On the other hand, since f is injective there can only be one such x because f is injective. Hence each $y \in Y$ has the form $y = f(x)$ for one *and only one* element $x \in X$. Thus the function f yields an exact correspondence between elements of X and elements of Y. For this reason bijections are often called *one-to-one correspondences*.

PICTURE HERE – CONVERT TO TIKZ

We shall consider these properties of functions again later, particularly in the context of graphs and of permutations. Let us now move on to consider "relations".

Definition 6.6 A *relation* on a set X is any set R of ordered pairs of elements of X. If $(x, y) \in R$, we usually denote this by xRy.

Strictly speaking this is the definition of a *binary* relation since it involves just two elements of X. The idea behind the concept of a relation is that it 'links' the two elements x and y.

Example 6.7 $\leqslant$ is a relation on $\mathbb{Z}$. As a set of ordered pairs, this relation corresponds to $\{ (x, y) \mid x \leqslant y \} \subseteq \mathbb{Z} \times \mathbb{Z}$.

$<$ is a relation on $\mathbb{Z}$.

$=$ is a relation on $\mathbb{Z}$.

$|$ is a relation on $\mathbb{Z}$.

$\subseteq$ is a relation on the set of all subsets of X.

The concept of a relation generalised all these ideas (they are all important and archetypal examples of relations) and it is for this reason that we typically write xRy to say that x and y are related under R.

If R is a relation on a finite set X, then we can represent it using a particular type of diagram. We draw a node (a blob) to represent each element of X and draw an arrow from x to y if xRy (that is, if $(x, y) \in R$). [Such a diagram will be referred to as a "graph" later in the course.]

Example 6.8 Consider the relation $|$ on the set $X = \{1, 2, 3, 4, 5, 6\}$. It is represented by the following diagram:

PICTURE HERE – CONVERT TO TIKZ

Example 6.9 Define R on $X = \{1, 2, 3, \ldots, 12\}$ by

xRy if and only if 2 appears the same number of times in the factorisation of x and y.

PICTURE HERE – CONVERT TO TIKZ

As a further example of a relation on $\mathbb{Z}$, we have congruence modulo m ($\equiv \pmod m$) where $m > 1$. We discussed this relation considerably previously. In particular, Theorem 5.5 observed that congruence (modulo m) enjoyed three properties to which we shall now give names.

Definition 6.10 Let R be a relation on a set X.

(i) R is *reflexive* (**R**) if xRx for all $x \in X$;

(ii) R is *symmetric* (**S**) if xRy implies yRx for all $x, y \in X$;

(iii) R is *anti-symmetric* (**AS**) if xRy and yRx imply $x = y$ for all $x, y \in X$;

(iv) R is *transitive* (**T**) if xRy and yRz imply xRz for all $x, y, z \in X$.

Example 6.11 (i) $\equiv \pmod m$ is reflexive, symmetric and transitive, but not anti-symmetric. (The first three are precisely what Theorem 5.5 says.)

(ii) $\leqslant$ (on $\mathbb{Z}$) is reflexive, anti-symmetric and transitive, but not symmetric.

(iii) $<$ (on $\mathbb{Z}$) is anti-symmetric and transitive, but not reflexive or symmetric.

(iv) $=$ (on any set) is reflexive, symmetric, and transitive.

(v) $|$ (on $\mathbb{N}$) is reflexive, anti-symmetric and transitive (see Theorem 1.5).

(vi) The relation R of Example 6.9 is reflexive, symmetric and transitive.

The properties R, S and T are *independent*: we can find a relation which satisfies any collection of them but not the others.

Example 6.12 The following diagram defines a relation which is reflexive and symmetric but not transitive.

PICTURE HERE – CONVERT TO TIKZ

We give a special name to relations that are reflexive, symmetric and transitive.

Definition 6.13 An *equivalence relation* is a relation that is reflexive, symmetric and transitive.

Example 6.14 $\equiv \pmod m$ on $\mathbb{Z}$, $=$ on any set, and the relation R of Example 6.9 are all equivalence relations.

$\leqslant$ (on $\mathbb{Z}$) and $|$ (on $\mathbb{N}$) are *not* equivalence relations — they are anti-symmetric *not* symmetric.

It is worth noting that relations that are reflexive, transitive and anti-symmetric are called *order relations*. They can be thought of as generalisations of $\leqslant$.

Equivalence relations are very important in mathematics — principally because they enable us to consider some elements of a set as being somehow the 'same.' The idea here is that if two elements of a set are related under an equivalence relation then we view these elements as sharing some property and then we collect together all elements sharing this property. (In some sense, equivalence relations are then generalisations of equality.)

Example 6.9 typifies this idea: elements that are related share the power of 2 that they are divisible by and we split the set X into 4 subclasses.

We formalise this idea in the following definition.

Definition 6.15 Let R be an equivalence relation on the set X and let $x \in X$. The *equivalence class* of x is the following subset of X:

$$[x] = \{\, y \in X \mid xRy \,\},$$

the set of all elements which are related to x.

Example 6.16 The equivalence classes for the equivalence relation R in Example 6.9 are:

$$[1] = \{1, 3, 5, 7, 9, 11\}$$
$$[2] = \{2, 6, 10\}$$
$$[4] = \{4, 12\}$$
$$[8] = \{8\}$$

Theorem 6.17 *Let R be an equivalence relation on a set X.*

(i) *$x \in [x]$ for all $x \in X$;*

(ii) *$\bigcup_{x \in X} [x] = X$;*

(iii) *if $x, y \in X$, then either $[x] = [y]$ or $[x] \cap [y] = \varnothing$.*

PROOF: (i) R is reflexive, so xRx. Hence $x \in [x]$.

(ii) follows directly from (i).

(iii) Let $x, y \in X$ and suppose that $[x] \cap [y] \neq \varnothing$. This means that there is at least one element, say z, belonging to both $[x]$ and $[y]$. This means

$$xRz \qquad \text{and} \qquad yRz.$$

Since R is symmetric, the latter implies zRy. Then as R is transitive, xRz and zRy together imply that

$$xRy.$$

We shall now use this observation to prove that $[x] = [y]$. We do this by showing that every element of $[x]$ is also an element in $[y]$ and *vice versa*.

First let $u \in [x]$. This means xRu. Now applying the symmetry of R to the fact that xRy, we deduce yRx. Then the transitivity of R means that yRx and xRu imply yRu. Thus $u \in [y]$.

Now let $v \in [y]$. Then we have yRv. Taking this together with xRy leads us to xRv (by transitivity), so $v \in [x]$.

Hence $[x]$ and $[y]$ contain exactly the same elements, so are equal. □

Our conclusion is that the set X is the union of the equivalence classes and any two different equivalence classes are disjoint.

Corollary 6.18 *If R is an equivalence relation on a set X, then X is the disjoint union of the equivalence classes.*

We refer to this situation by saying that the equivalence classes of R *partition* X.

Example 6.19 Define a relation R on $\mathbb{Q}$ by

$$xRy \qquad \text{if and only if} \qquad x - y \in \mathbb{Z}.$$

R: $x - x = 0$, so $x - x \in \mathbb{Z}$ and hence xRx for all $x \in \mathbb{Q}$.

S: Suppose xRy, so that $x - y \in \mathbb{Z}$. Then $y - x = -(x - y) \in \mathbb{Z}$, so yRx.

T: Suppose xRy and yRz. Then $x - y \in \mathbb{Z}$ and $y - z \in \mathbb{Z}$, so $x - z = (x - y) + (y - z) \in \mathbb{Z}$. Hence xRz.

Thus R is an equivalence relation on $\mathbb{Q}$. Note that

$$xR0 \qquad \text{means} \qquad x \in \mathbb{Z},$$

so

$$[0] = \mathbb{Z}.$$

Similarly

$$[\tfrac{1}{2}] = \{\dots, -\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2}, \dots\}$$

and, in general, for any $q \in \mathbb{Q}$,

$$[q] = \{\, q + a \mid a \in \mathbb{Z} \,\}.$$

We shall finish this section by considering congruence modulo m in detail.

Let m be an integer with $m > 1$. Theorem 5.5 tells us that congruence modulo m ($\equiv \pmod m$) is an equivalence relation on the set $\mathbb{Z}$. We desire to describe the equivalence classes. To do this we need an alternative way to describe the congruence condition.

Lemma 6.20 *$a \equiv b \pmod m$ if and only if a and b have the same remainder when divided by m.*

PROOF: Suppose a and b have the same remainder upon dividing by m:

$$a = mq_1 + r \qquad \text{and} \qquad b = mq_2 + r.$$

Then

$$a - b = (mq_1 + r) - (mq_2 + r) = m(q_1 - q_2),$$

so $a \equiv b \pmod m$.

Conversely assume $a \equiv b \pmod m$. Divide a and b by m:

$$a = mq_1 + r_1 \qquad \text{and} \qquad b = mq_2 + r_2$$

where $0 \leqslant r_1, r_2 < m$. Then

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

Now $m \mid (a - b)$ by assumption, so we deduce $m \mid (r_1 - r_2)$. But $0 \leqslant r_1, r_2 < m$, so $-m < r_1 - r_2 < m$. Since $r_1 - r_2$ is divisible by m, we deduce $r_1 - r_2 = 0$; that is, $r_1 = r_2$ as required. □

This has the following consequence:

Theorem 6.21 *Let $m > 1$. The equivalence relation of being congruent modulo m has precisely m equivalence classes, namely*

$$[r] = \{\, km + r \mid k \in \mathbb{Z} \,\}$$

for $r = 0, 1, \ldots, m-1$.

These equivalence classes are sometimes also referred to as *congruence classes*.

PROOF: The first thing to do is to apply Lemma 6.20: for $0 \leqslant r < m$ we see

$$\begin{array}{lll} a \in [r] & \iff & r \equiv a \pmod{m} \\ & \iff & r \text{ and } a \text{ have the same remainder upon dividing by } m \\ & \iff & a = km + r \quad \text{for some } k. \end{array}$$

So

$$[r] = \{\, km + r \mid k \in \mathbb{Z} \,\}.$$

This shows that these equivalence classes do have the form we claimed.

We still need to show that this list gives us all the equivalence classes and that they are distinct.

If $a \in \mathbb{Z}$, then a has remainder r upon dividing by m where $0 \leqslant r < m$, and so $a \equiv r \pmod{m}$. This shows that $a \in [r]$. Hence every element of $\mathbb{Z}$ lies in one of the equivalence classes

$$[0],\ [1], \ldots, [m-1].$$

It follows that this list gives all the equivalence classes.

If $[r] = [s]$ where $0 \leqslant r, s < m$, then $r \equiv s \pmod{m}$, so they have the same remainder upon dividing by m, so $r = s$.

Hence there are *precisely* m equivalence classes:

$$[0],\ [1], \ldots, [m-1].$$

□

To finish this section, we shall interpret Theorem 5.6 in the context of these equivalence classes. Part (i) of that theorem stated:

- If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then

$$\begin{aligned} a + c &\equiv b + d \pmod{m} \\ ac &\equiv bd \pmod{m}. \end{aligned}$$

The upshot of these conditions is that we can think of addition and multiplication as operating on the congruence classes as well as the integers. For example, the first can be interpreted as the statement:

$$[a] = [b] \text{ and } [c] = [d] \quad \text{imply} \quad [a + c] = [b + d].$$

This allows us to define operations on the equivalence classes by:

$$[a] + [b] = [a + b] \qquad \text{and} \qquad [a] \cdot [b] = [ab]$$

and here we can replace $a + b$ and ab by the integers in the range $0 \leqslant r < m$ congruent to $a + b$ and ab modulo m.

Example 6.22 Take $m = 6$. For simplification, let us write 0, 1, ..., 5 for the congruence classes modulo 6 (instead of [0], [1], ...). We then have the following multiplication tables:

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

This is *modular arithmetic*, which plays an important role in abstract algebra and appears in the MT2002 course.

Section 7

Higher Degree Diophantine Equations

In Section 4 we established great control over the solutions of the linear Diophantine equation $ax + by = c$. We shall finish the our discussion of number theory by considering solutions to some Diophantine equations which are not linear.

Pythagorean triples

Consider the Diophantine equation

$$x^2 + y^2 = z^2. \tag{7.1}$$

By Pythagoras' Theorem, a solution to this Diophantine equation corresponds to a right-angled triangle with sides of integer length.

Suppose x, y, z are solutions with

$$\gcd(x, y, z) = d > 1.$$

Write $x = dx_1$, $y = dy_1$ and $z = dz_1$. Substituting into Equation (7.1) and dividing by d^2 yields:

$$x_1^2 + y_1^2 = z_1^2$$

and here $\gcd(x_1, y_1, z_1) = 1$.

Definition 7.1 A solution (x, y, z) of Equation (7.1) is called a *Pythagorean triple*. If $x, y, z > 0$ and $\gcd(x, y, z) = 1$, then the Pythagorean triple (x, y, z) is called *primitive*.

Our earlier discussion tells us that it is enough to find all primitive Pythagorean triples. We can then find all Pythagorean triples by multiplying by an integer.

Our goal is then to find all primitive Pythagorean triples. We begin with three auxiliary results.

Lemma 7.2 *Let $a, b, c, n \in \mathbb{Z}$ with $n \geqslant 1$. If $ab = c^n$ and $\gcd(a, b) = 1$, then both a and b are nth powers.*

PROOF: Factorise c as a product of prime powers:

$$c = p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}.$$

Then

$$ab = c^n = p_1^{nk_1} p_2^{nk_2} \dots p_m^{nk_m}.$$

Now each p_i divides only one of a and b (as $\gcd(a, b) = 1$), so it follows that a and b are both products of nth powers of prime powers, so a and b are nth powers. □

Lemma 7.3 *The sum of two odd squares is not a square.*

PROOF: First recall Theorem 1.4: The square of an integer is either divisible by 4 or it gives remainder 1 when divided by 8. Thus if a and b are odd integers,

$$a^2 \equiv 1 \pmod 8 \qquad \text{and} \qquad b^2 \equiv 1 \pmod 8.$$

Thus

$$a^2 + b^2 \equiv 2 \pmod 8,$$

so

$$a^2 + b^2 \equiv 2 \pmod 4.$$

Hence $a^2 + b^2$ is not a square (it is not divisible by 4). □

Corollary 7.4 *If (x, y, z) is a primitive Pythagorean triple then one of x and y is even, while the other is odd (and consequently z is odd).*

PROOF: If both x and y are even, then so would be z, contradicting $\gcd(x, y, z) = 1$. The previous lemma shows that they cannot both be odd. Hence one is even and the other is odd. □

Now let (x, y, z) be a primitive Pythagorean triple. Without loss of generality assume that x is even and that y and z is odd. Then $z - y$ and $z + y$ are even. Write

$$z - y = 2u, \qquad z + y = 2v$$

where $u, v \in \mathbb{Z}$. Then $2z = 2(u + v)$, so $z = u + v$. Similarly $y = v - u$.

Claim: $\gcd(u, v) = 1$.

Let $d = \gcd(u, v)$. If $d > 1$, let p be a prime dividing d. Then $p \mid u$ and $p \mid v$, so

$$p \mid (u + v) = z \qquad \text{and} \qquad p \mid (v - u) = y.$$

Then $p^2 \mid (z^2 - y^2) = x^2$ and it follows that $p \mid x$ also. This is a contradiction since $\gcd(x, y, z) = 1$.

Thus $\gcd(u, v) = 1$, as claimed.

Remember x is even and observe

$$(x/2)^2 = (z^2 - y^2)/4 = \left(\frac{z-y}{2}\right)\left(\frac{z+y}{2}\right) = uv$$

and then, by Lemma 7.2, both u and v are squares. Write

$$u = t^2, \qquad v = s^2$$

where $s, t \in \mathbb{Z}$. Then

$$\begin{aligned} z &= u + v = s^2 + t^2 \\ y &= v - u = s^2 - t^2 \\ x &= \sqrt{4uv} = 2st. \end{aligned}$$

Moreover $\gcd(s, t) = 1$ because $\gcd(u, v) = 1$. Also s and t are not both odd, since otherwise this would cause x, y, z all to be even.

Conversely suppose x, y, z are given by the above formulae. Then

$$\begin{aligned} x^2 + y^2 &= 4s^2t^2 + (s^2 - t^2)^2 \\ &= s^4 + 2s^2t^2 + t^4 \\ &= (s^2 + t^2)^2 = z^2, \end{aligned}$$

so (x, y, z) is a Pythagorean triple. If $\gcd(s, t) = 1$ and s and t are not both odd, then y and z are odd. If p is a prime dividing both y and z, then firstly it is odd and secondly

$$p \mid (y + z) = 2s^2 \qquad \text{and} \qquad p \mid (z - y) = 2t^2,$$

so $p \mid s$ and $p \mid t$, contrary to $\gcd(s, t) = 1$. Hence $\gcd(y, z) = 1$ and we deduce (x, y, z) is a *primitive* Pythagorean triple.

We summarise this discussion as follows:

Theorem 7.5 (Pythagorean Triples) *All the solutions of the equation*

$$x^2 + y^2 = z^2$$

satisfying

$$x, y, z > 0, \qquad \gcd(x, y, z) = 1, \qquad 2 \mid x$$

are given by

$$x = 2st, \qquad y = s^2 - t^2, \qquad z = s^2 + t^2$$

for integers $s > t > 0$ such that $\gcd(s, t) = 1$ and s and t are not both odd.

All other solutions of the equation can be obtained from these by multiplying by an integer, interchanging x and y, and changing the sign of some of x, y, z.

The solutions for small s and t $(0 < t < s \leqslant 7)$ are as follows:

s	t	x	y	z
2	1	4	3	5
3	2	12	5	13
4	1	8	15	17
4	3	24	7	25
5	2	20	21	29
5	4	40	9	41
6	1	12	35	37
6	5	60	11	61
7	2	28	45	53
7	4	56	33	65
7	6	84	13	85

Example 7.6 Let (x, y, z) be a primitive Pythagorean triple. We may suppose (without loss of generality) that x is even. So

$$x = 2st, \qquad y = s^2 - t^2, \qquad z = s^2 + t^2$$

for some integers s and t with

$$s > t > 0, \qquad \gcd(s, t) = 1, \qquad s, t \text{ not both odd.}$$

Suppose that s and t are not divisible by 3. (Note that if $3 \mid s$ or $3 \mid t$, then $3 \mid x$.)

Then $s \equiv \pm 1 \pmod 3$ and $t \equiv \pm 1 \pmod 3$, so

$$s^2 \equiv 1 \pmod 3 \qquad \text{and} \qquad t^2 \equiv 1 \pmod 3.$$

Hence

$$y = s^2 - t^2 \equiv 1 - 1 = 0 \pmod 3.$$

So if $3 \nmid s$ and $3 \nmid t$, then $3 \mid y$.

Conclusion: *If (x, y, z) is a primitive Pythagorean triple, then exactly one of x and y is divisible by 3 (and hence z is not divisible by 3).*

We finish this section by considering two theorems about higher Diophantine equations. The following was first discovered by Fermat.

Theorem 7.7 *The equation*

$$x^4 + y^4 = z^2$$

has no positive integer solutions.

I shall omit the proof for the moment. If time permits in the last week of this lecture course I shall return and prove this. The method is to observe if (x_0, y_0, z_0) is a solution (all positive) with z_0 as small as possible, then (x_0^2, y_0^2, z_0) is a primitive Pythagorean triple. We then apply Theorem 7.5 and eventually reduce to a solution to the original equation with smaller value of z. This will be a contradiction and would prove the theorem.

An immediate corollary is:

Theorem 7.8 *The equation*

$$x^4 + y^4 = z^4$$

has no positive integer solutions.

PROOF: If (x_0, y_0, z_0) were a solution, then (x_0, y_0, z_0^2) would be a solution to $x^4 + y^4 = z^2$ and the previous theorem says this is impossible. □

In the 17th century, Fermat conjectured that the equation

$$x^n + y^n = z^n$$

has no positive integer solutions when $n > 2$. (Actually, he claimed to have a proof, but modern mathematicians suspect he was wrong — as have many who have attempted to solve it.) In the end, establishing the truth of the conjecture resisted attempts until very recently when Wiles finally showed it was true. It is very interesting to note how the solution of such an innocuous appearing question finally depended on deep developments in new and exciting areas of pure mathematics.

Section 8

Graphs

The theory of graphs started with a paper of Euler who was interested in a problem known as "The Seven Bridges of Königsberg. We shall meet this problem later, but we shall begin by motivating a study by recalling our earlier discussion of relations.

Recall that a relation R on a set V is a collection of ordered pairs and that we wrote aRb to denote that (a, b) belonged to this collection. We previously thought of this as indicating some link from a to b and denoted it by a diagram. [DRAW!]

We considered these sort of diagrams and noted, for example, that equivalence relations were relations satisfying particular properties. In this section, we shall study these diagrams and give particular names to them. Principally we shall see that such diagrams can be used to represent other situations as well as relations.

Example 8.1 Consider a collection of villages $\{a, b, c, d, e, f\}$ joined by a number of roads as follows:

PICTURE HERE – CONVERT TO TIKZ

This will be an example of a graph (without having direction attached to it). We begin by defining what we mean by a *directed* graph; i.e., when all these roads are one-way! Such objects are useful for describing many situations, for example, road networks, communication networks, etc.

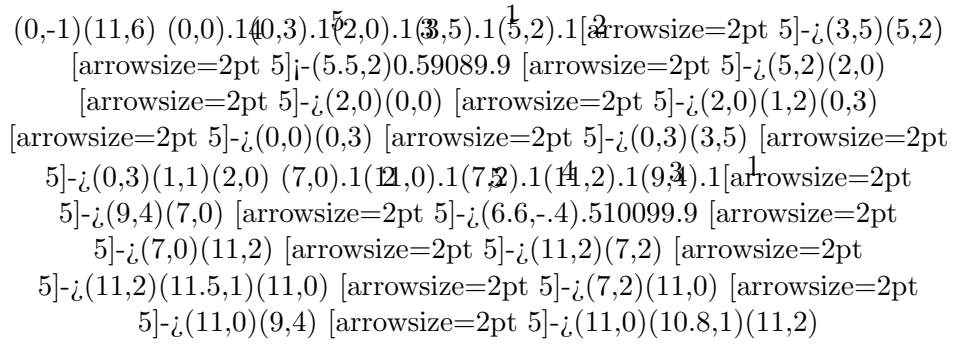
Directed graphs

Definition 8.2 A *directed graph* (or *digraph*) consists of a set V of points, called the *vertices*, and a set E of ordered pairs from V, called the *edges*. We write $\Gamma = (V, E)$ to denote such a graph.

Example 8.3 Let $V = \{1, 2, 3, 4, 5\}$ and

$$E = \{(1, 2), (2, 2), (2, 3), (3, 4), (3, 5), (4, 5), (5, 1), (5, 3)\}.$$

We represent the graph $\Gamma = (V, E)$ by the following diagrams:



We usually think of the term "graph" as referring to such a pictorial representation. It is for this reason we refer to the "vertices" and the "edges" of the graph. It does raise one question:

Question: When are two graphs the 'same'?

The above example illustrates the same description can be drawn in two different ways — however, we would like to think of them as somehow the same. We can formalise this sort of idea in the following definitions.

Definition 8.4 Let $\Gamma = (V, E)$ be a directed graph. We say that a vertex u is *adjacent* to a vertex v if (u, v) is an edge in Γ.

We will use this to describe what it means for graphs to be 'the same.'

Definition 8.5 Let $\Gamma = (V, E)$ and $\Gamma' = (V', E')$ be directed graphs. We say that Γ and Γ' are *isomorphic* (written $\Gamma \cong \Gamma'$) if there is a bijection $f\colon V \to V'$ such that if v_1, v_2 are vertices of Γ then

$$v_1 \text{ and } v_2 \text{ are adjacent in } \Gamma \iff f(v_1) \text{ and } f(v_2) \text{ are adjacent in } \Gamma'.$$

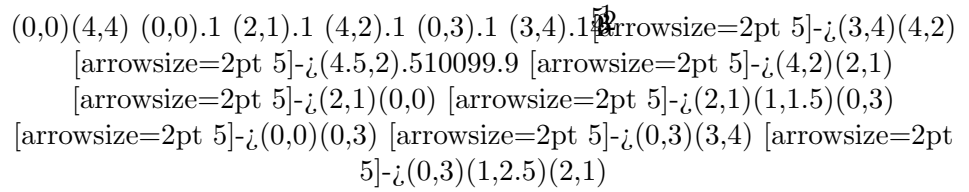
Thus these graphs are isomorphic if they have essentially the same vertices and the same edges. In view of this we usually do not wish to distinguish between isomorphic graphs.

(This term comes from Greek: *isos* means 'equal', while *morphe* means 'shape'.)

Deciding when two graphs are isomorphic can be quite difficult though and it is useful to consider alternative ways of describing graphs to help us answer this problem.

Definition 8.6 Let $\Gamma = (V, E)$ be a directed graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$. The *adjacency matrix* of Γ is the matrix $A(\Gamma)$ whose (i, j)th entry is 1 if (v_i, v_j) is an edge in Γ and whose (i, j)th entry is 0 if v_i is not adjacent to v_j.

Example 8.7 Consider the following graph:



The adjacency matrix for this directed graphs is:

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

The adjacency matrix entirely encodes the edges present in the graph. Consequently the following observation is immediate:

Lemma 8.8 *Two directed graphs are isomorphic if and only if they have identical adjacency matrices following some relabelling of the vertices.*

Walks in directed graphs

Definition 8.9 Let $\Gamma = (V, E)$ be a directed graph. A *walk* in Γ is a sequence of vertices and edges where each edge is directed from the vertex preceding it to the vertex following it.

Thus a walk has the form

$$v_0, e_1, v_1, e_2, v_2, \ldots, v_{n-1}, e_n, v_n$$

where edge e_i has the form $e_i = (v_{i-1}, v_i)$. (We often omit the reference to the vertices in this sequence for this reason.) We define the *length* of the walk to be the number of edges occurring.

Thus, in Example 8.7 the following is a walk of length 6 from 1 to 4:

$$(1,2),\ (2,2),\ (2,3),\ (3,5),\ (5,3),\ (3,4).$$

Related to this definition we have:

Definition 8.10 (i) A *path* in a directed graph Γ is a walk in which no vertex occurs more than once.

(ii) A *circuit* in a directed graph Γ is a *closed* walk (that is, a walk where the first and last vertex *are* the same).

So in Example 8.7

$$(2,3),\ (3,4),\ (4,5)$$

is a path, while

$$(5,3),\ (3,4),\ (4,5)$$

is a circuit.

Theorem 8.11 *Let $\Gamma = (V, E)$ be a directed graph with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and adjacency matrix A. Then the (i,j)th entry of A^m is the number of walks of length m from v_i to v_j in Γ.*

PROOF: We proceed by induction on m. The case $m = 1$ is immediate: there is a walk from v_i to v_j of length 1 precisely when there is an edge (v_i, v_j) in Γ, which is precisely when there is a 1 (and not a 0) in the (i,j)th entry of the adjacency matrix A.

Let $A = (a_{ij})$. Suppose that $B = A^m = (b_{ij})$ and that b_{ij} is the number of walks of length m from v_i to v_j. According to the rules for matrix multiplication, the (i,j)th entry of $A^{m+1} = BA$ is

$$\sum_{k=1}^{n} b_{ik}a_{kj}.$$

Now b_{ik} is the number of walks of length m from v_i to v_k. We can extend such a walk to one of length $m+1$ to v_j precisely when there is an edge from v_k to v_j (i.e, when $a_{kj} = 1$ rather than 0). Hence $b_{ik}a_{kj}$ is equal to the number of walks of length $m+1$ from v_i to v_j stopping at v_k at the mth step and therefore

$$\sum_{k=1}^{n} b_{ik}a_{kj}$$

is the *total* number of walks from v_i to v_j of length $m+1$. This completes the inductive step and proves the theorem. □

Other forms of graph

We have met the definition of a directed graph. Two important variations are the following:

Definition 8.12 If $\Gamma = (V, E)$ is a directed graph, an edge of the form (v, v) (with $v \in V$) is called a *loop*.

A directed graph $\Gamma = (V, E)$ is said to be *loop-free* if it has no loops.

We said that a directed graph was useful for describing a relation on a set V. Under this, a loop-free directed graph corresponds to a relation R that is *irreflexive*; i.e., for which xRx does not hold for any x.

Definition 8.13 A directed graph $\Gamma = (V, E)$ with the property that whenever (v_i, v_j) is an edge then also (v_j, v_i) is an edge is called a *graph* (or *undirected graph*).

Thus an (undirected) graph corresponds to a relation R which is symmetric. In this situation we can replace to the two directed edges between a pair of adjacent vertices by a single undirected edge. For example [draw any graph] represents an undirected graph.

Sometimes we permit graphs to have more than one edge between a pair of vertices. We then say that our graph has *multiple edges* and call it a *multigraph*. The following definition restricts us away from this situation:

Definition 8.14 A *simple graph* is an (undirected) graph that possesses no multiple edges and no loops.

Definition 8.15 A graph Γ is *connected* if there is a path between any two distinct vertices of Γ.

We usually employ this definition only for undirected graphs — the reason being that in directed graphs things get rather complicated:

This directed graph is disconnected according to the definition: it is impossible to travel from the left-hand vertex to the right-hand one along a path.

Degrees of vertices

Definition 8.16 Let $\Gamma = (V, E)$ be an (undirected) graph. The *degree* of a vertex v is the number of edges incident to that vertex. We denote this by $\rho(v)$.

If Γ is a graph in which every vertex has the same degree, then we say Γ is *regular*.

Example 8.17 The following graphs are regular:

unit=9.7mm (0,0)(13,3) (0,0).1 (1.5,3).1 (3,0).1 (0,0)(1.5,3) (1.5,3)(3,0) (0,0)(3,0) (5,0).1 (6.5,1).1 (6.5,3).1 (8,0).1 (5,0)(8,0) (5,0)(6.5,3) (5,0)(6.5,1) (6.5,1)(6.5,3) (6.5,1)(8,0) (6.5,3)(8,0) (10,0).1 (10,3).1 (13,0).1 (13,3).1 (10,0)(10,3) (10,0)(13,0) (10,3)(13,3) (13,0)(13,3)

Examples of graphs

Example 8.18 (i) The *complete graph* K_n is the (simple) graph with n vertices in which every pair of distinct vertices is adjacent.

(0,-1)(10,2) (0,0).1 (0,-.5)K_1 (2,0).1 (2,2).1 (2,0)(2,2) (2,-.5)K_2
(4,0).1 (6,0).1 (5,1.732).1 (4,0)(6,0) (6,0)(5,1.732) (5,1.732)(4,0)
(5,-.5)K_3 (8,0).1 (8,2).1 (10,2).1 (10,0).1 (8,0)(8,2) (8,0)(10,2)
(8,0)(10,0) (8,2)(10,2) (8,2)(10,0) (10,2)(10,0) (9,-.5)K_4

The graph K_n has $n(n-1)/2$ edges.

(ii) The *null graph* N_n is the graph with n vertices and no edges.

(0,-1)(4,.4) (0,0).1 (1,0).1 (2,0).1 (3,0).1 (4,0).1 (2,-.5)N_5

(iii) The *cycle of length* n is the graph C_n with vertices $V = \{v_1, v_2, \ldots, v_n\}$ and edges

$$E = \{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n), (v_n, v_1)\}.$$

(-2,-3)(2,2) (0,2).1 (1.564,1.247).1 (-1.564,1.247).1 (1.95,-.445).1
(-1.95,-.445).1 (.868,-1.802).1 (-.868,-1.802).1 (0,2)(1.564,1.247)
(1.564,1.247)(1.95,-.445) (1.95,-.445)(.868,-1.802)
(.868,-1.802)(-.868,-1.802) (-.868,-1.802)(-1.95,-.445)
(-1.95,-.445)(-1.564,1.247) (-1.564,1.247)(0,2) (0,-2.2)C_7

Trees

Definition 8.19 A simple graph Γ is called a *tree* if it is connected and contains no circuits.

Definition 8.20 Given any graph Γ_1, a *spanning tree* Γ of Γ_1 is a subgraph which contains all the vertices of Γ_1 and is a tree.

We think of a spanning tree as providing minimal connectivity for the graph and as a minimal skeletal framework holding the vertices together.

Theorem 8.21 *If a and b are vertices in a tree T, then there is a unique path that connects these vertices.*

PROOF: Since T is connected, there is at least one path from a to b. If there were more than one, from two such paths we could construct a circuit using some of the edges. This contradicts T having no circuits. □

Theorem 8.22 *Let Γ be an undirected graph. Then Γ is connected if and only if it has a spanning tree.*

Proof is omitted. Illustrate with diagram.

Example 8.23 There are three non-isomorphic trees that exist on five vertices.

unit=.5cm (0,0)(10,8) (0,0).2 (0,2).2 (0,4).2 (0,6).2 (0,8).2 (0,0)(0,8) (2,0).2 (2,2).2 (2,4).2 (2,6).2 (4,2).2 (2,0)(2,6) (2,2)(4,2) (6,2).2 (8,0).2 (8,2).2 (8,4).2 (10,2).2 (6,2)(10,2) (8,0)(8,4)

Note that all three have the same number of edges: 4. This is a general feature of trees.

Theorem 8.24 *Let $T = (V, E)$ be a tree. Then $|E| = |V| - 1$.*

PROOF: Let $n = |V|$, the number of vertices in T. We proceed by induction on n. If $n = 1$, then T has a single vertex and must have no edges. Hence $|E| = n - 1$ and the result holds in this case.

Suppose that the result holds for all trees with fewer than n vertices. Pick an edge $\{u, v\}$ in T. If we delete the edge $\{u, v\}$, then we obtain a new graph Γ, which is now disconnected. (There can be no path from u to v in the new graph since T is a tree, so $\{u, v\}$ can be the only path from u to v.) Since T is connected, it must be the case that deleting the edge $\{u, v\}$ breaks it into two connected subgraphs, T_1 and T_2. These must both be trees since deleting an edge cannot suddenly introduce loops or circuits. Let n_1 and n_2 be the number of vertices in T_1 and T_2, respectively. By induction T_1 has $n_1 - 1$ edges and T_2 has $n_2 - 1$ edges. Hence T has

$$(n_1 - 1) + (n_2 - 1) + 1 = n_1 + n_2 - 1 = n - 1$$

edges. This completes the induction. □

This completes our discussion of basic properties and types of graphs. In the next section we consider particular properties these graphs may have.

Section 9

Eulerian and Hamiltonian Graphs

Graph theory began with Euler's study of a particular problem: the Seven Bridges of Königsberg. During the eighteenth century the city of Königsberg (in East Prussia) was divided into four sections (including the island of Kneiphof) by the Pregel river. Seven bridges connected these regions and it was said that residents spent their Sunday walks trying to find a starting point so that they could walk about the city, cross each bridge exactly once, and return to their starting point.

To apply graph theory to this, let us represent each of the four sections of the city by a node in a graph and represent each bridge by an edge:

(0,-.5)(2,2) (0,0).1(0,1).1(0,2).1(2,1).1(0,0)(2,1) (0,1)(2,1) (0,2)(2,1)
(-.5,.5).707-4545 (-.5,1.5).707-4545 (.5,.5).707135225 (.5,1.5).707135225

(This is an undirected multigraph.)

Note that the degrees of the vertices are the following values:

$$\rho(a) = \rho(c) = \rho(d) = 3, \qquad \rho(b) = 5.$$

The question we seek to ask is the following:

> Is there a circuit (a closed walk) that traverses every edge in the graph once?

Euler established that the answer to this question depended upon the number of vertices of odd degree in the graph.

We make the following definition:

Definition 9.1 A graph $\Gamma = (V, E)$ is called *Eulerian* if there is a circuit in Γ that passes through every vertex $v \in V$ and that traverses every edge of Γ exactly once.

A weakening is the following:

Definition 9.2 A graph $\Gamma = (V, E)$ is called *semi-Eulerian* if there is a walk in Γ that passes through every vertex $v \in V$ and that traverses every edge of Γ exactly once.

(In a semi-Eulerian graph, we do not require that we end up back where we started!)

Theorem 9.3 *Let $\Gamma = (V, E)$ be a connected graph. Then Γ is Eulerian if and only if every vertex has even degree.*

Corollary 9.4 *A connected graph $\Gamma = (V, E)$ is semi-Eulerian if and only if Γ has at most two vertices of odd degree.*

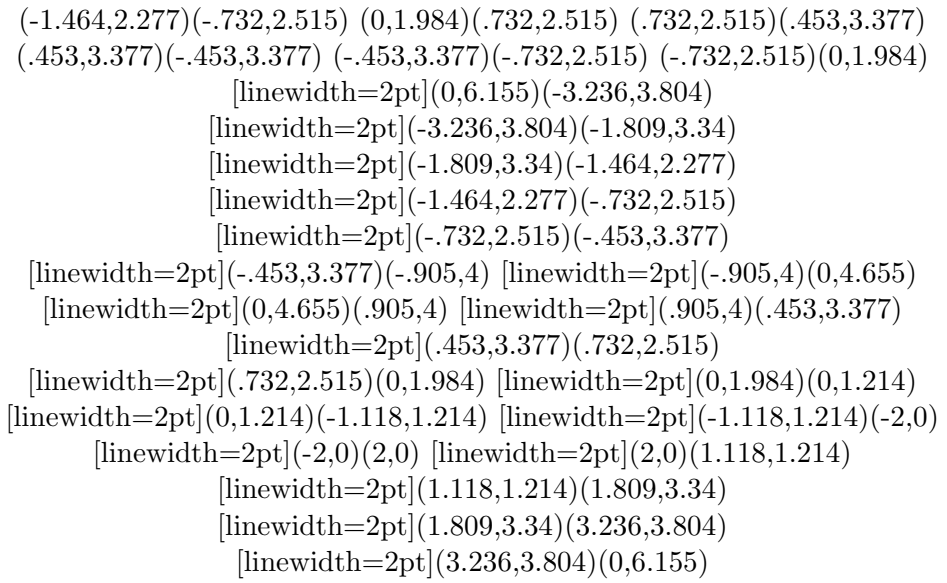
The graph for the Seven Bridges of Königsberg has four vertices of odd degree. Consequently it is neither Eulerian nor semi-Eulerian: the people of Königsberg were wasting their time on Sunday afternoons trying to find such a route!

Hamiltonian paths and cycles

In 1859, the Irish mathematician Sir William Rowan Hamilton developed a game that he sold to a Dublin toy manufacturer. The game consisted of a wood regular dodecahedron with the twenty corner points (vertex) labelled with the names of prominent cities. The object of the game was to find a circuit along the edges of the solid so that each city on the circuit exactly once.

We represent the solid by a graph: the vertices of the graph correspond to the vertices of the solid and the edges similarly correspond:

(-4,-1)(4,7) (2,0).1 (-2,0).1 (3.236,3.804).1 (-3.236,3.804).1 (0,6.155).1
(-2,0)(2,0) (2,0)(3.236,3.804) (3.236,3.804)(0,6.155) (0,6.155)(-3.236,3.804)
(-3.236,3.804)(-2,0) (-1.118,1.214).1 (1.118,1.214).1 (1.809,3.34).1
(-1.809,3.34).1 (0,4.655).1 (-2,0)(-1.118,1.214) (2,0)(1.118,1.214)
(3.236,3.804)(1.809,3.34) (-3.236,3.804)(-1.809,3.34) (0,6.155)(0,4.655)
(-1.118,1.214)(1.118,1.214) (1.118,1.214)(1.809,3.34) (1.809,3.34)(0,4.655)
(0,4.655)(-1.809,3.34) (-1.809,3.34)(-1.118,1.214) (0,1.214).1 (1.464,2.277).1
(.905,4).1 (-.905,4).1 (-1.464,2.277).1 (0,1.984).1 (.732,2.515).1
(.453,3.377).1 (-.453,3.377).1 (-.732,2.515).1 (0,1.214)(0,1.984)
(1.464,2.277)(.732,2.515) (.905,4)(.453,3.377) (-.905,4)(-.453,3.377)



Definition 9.5 Let $\Gamma = (V, E)$ be a graph. A *Hamiltonian circuit* is a circuit which passes through every vertex exactly once (with only the first and last vertex being a repeat).

A graph is called *Hamiltonian* if it possesses a Hamiltonian circuit.

Unsolved Problem: What is a necessary and sufficient condition for a graph to be Hamiltonian?

This question appears to be extremely difficult to solve. The following gives a sufficient condition:

Theorem 9.6 (Dirac 1952) *Let $\Gamma = (V, E)$ be a simple graph with n vertices and suppose $\rho(v) \geqslant n/2$ for every vertex v. Then Γ is Hamiltonian.*

(We can see easily that this is not a necessary condition. The dodecahedron graph corresponding to Hamilton's original game has $n = 20$, $\rho(v) = 3$ for every vertex v, yet the graph *is* Hamiltonian.)

Proof: Suppose Γ is not Hamiltonian. If we were to add more edges to Γ, then eventually we would have to create a graph which is Hamiltonian. Therefore we may add a number of edges to Γ and create a simple graph Γ' which is not Hamiltonian, but for which the addition of a single further edge gives a Hamiltonian graph. Note that $\rho'(v) \geqslant \rho(v) \geqslant n/2$ where $\rho'(v)$ denotes the degree of the vertex v in the new graph Γ'.

Let $\{v_1, v_2\}$ be the edge which when added creates a Hamiltonian circuit. This circuit necessarily has the form

$$v_1 \to v_2 \to v_3 \to \cdots \to v_{i-1} \to v_i \to \cdots \to v_n \to v_1.$$

Now if for some i (with $3 \leqslant i \leqslant n$) there is an edge $\{v_2, v_i\}$ and an edge $\{v_1, v_{i-1}\}$ in Γ', then we can create a new Hamiltonian circuit

$$v_2 \to \underbrace{v_i \to v_{i+1} \to \cdots \to v_n} \to v_1 \to \underbrace{v_{i-1} \to v_{i-2} \to \cdots \to v_3 \to v_2}.$$

This circuit does not involve the edge $\{v_1, v_2\}$ and so exists in the graph Γ'. This contradicts the assumption that Γ' is not Hamiltonian.

Hence for each of $i = 3, \ldots, n$, it is not the case that there is both an edge $\{v_1, v_{i-1}\}$ and an edge $\{v_2, v_i\}$. Let $A' = (a'_{kl})$ be the adjacency matrix of Γ'. This assertion is that

$$a'_{1,i-1} + a'_{2i} \leqslant 1 \qquad \text{for } 3 \leqslant i \leqslant n.$$

Let us sum over all i:

$$\sum_{j=2}^{n-1} a'_{1j} + \sum_{i=3}^{n} a'_{2j} \leqslant n - 2$$

so

$$\sum_{j=1}^{n} a'_{1j} + \sum_{i=1}^{n} a'_{2j} \leqslant n - 1.$$

(Note $a'_{11}, a'_{12}, a'_{21}, a'_{22} = 0$, as Γ' is simple and does not have $\{v_1, v_2\}$ as an edge.) Hence

$$\rho'(v_1) + \rho'(v_2) \leqslant n - 1.$$

Yet

$$\rho'(v_1) + \rho'(v_2) \geqslant n/2 + n/2 = n,$$

and we have a contradiction. Hence Γ is Hamiltonian. □

Example 9.7 (The Knight's Tour) The Knight's Tour Problem is concerned with the use of a chessboard and the piece known as the knight. Can a knight vist each square of a chessboard by a sequence of knight's moves and finish on the same square that it began on?

(0,0)(8,8) (0,0)(1,0)9(0,0)(0,8) (0,0)(0,1)9(0,0)(8,0) (2.5,1.5)4pt
(3.5,3.5)4pt -¿(2.5,1.5)(3.4,3.3)

The solution is to find a link between this problem and the finding of Hamiltonian circuits in a graph. To simplify the explanation, we consider the situation of a 4×4 chessboard (rather than the usual 8×8 one). We represent each square on this chessboard by a vertex of a graph and we join two vertices by an edge if a knight could make a move from between the corresponding squares. The following illustrates some of the edges (more need to be added):

(0,0)(3,3) (0,0).1 (1,0).1 (2,0).1 (3,0).1 (0,1).1 (1,1).1 (2,1).1 (3,1).1 (0,2).1
(1,2).1 (2,2).1 (3,2).1 (0,3).1 (1,3).1 (2,3).1 (3,3).1 (0,0)(2,1) (0,0)(1,2)
(2,1)(3,3) (1,2)(3,3) (3,0)(1,1) (3,0)(2,2) (0,3)(1,1) (0,3)(2,2)

This graph is not Hamiltonian. The corner vertices all have degree precisely two and consequently the eight edges shown would have to be included in any Hamiltonian circuit. This shows that no such Hamiltonian circuit can exist since each of the centre four vertices must be visited at least twice.

It can be shown that there is no Knight's Tour for a chessboard with an odd number of squares (e.g., a 5×5 board). However for some other boards there are solutions (e.g., for 6×6). The solution for the standard 8×8 board was given by Euler in 1759.

Section 10

Planar Graphs

Definition 10.1 We say that a graph Γ is *planar* if it *can be* drawn in the plane with its edges only intersecting at vertices of Γ.

Sometimes we use the term '*plane*' to refer to a graph that actually *is* drawn in the plane (as opposed to one that has the potential to be drawn as such).

Example 10.2 The following are planar graphs:

(0,0)(8,2.7) (0,0).1 (.6,.4).1 (1,1.2).1 (1,2).1 (1.4,.4).1 (2,0).1 (0,0)(2,0)
(0,0)(.6,.4) (0,0)(1,2) (.6,.4)(1.4,.4) (.6,.4)(1,1.2) (1,1.2)(1,2) (1,1.2)(1.4,.4)
(1,2)(2,0) (1.4,.4)(2,0) (3,0).1 (3,2).1 (5,0).1 (5,2).1 (3,0)(5,0) (3,0)(3,2)
(3,0)(5,2) (3,2)(5,2) (3,2)(5,0) (5,0)(5,2) (6,0).1 (6,2).1 (8,0).1 (8,2).1
(6,0)(8,0) (6,0)(6,2) (6,0)(8,2) (6,2)(8,2) (8,0)(8,2) (6,2)(7.5,3)(9,1.5)(8,0)

Theorem 10.3 *The complete graph K_n is planar for $n = 1, 2, 3, 4$.*

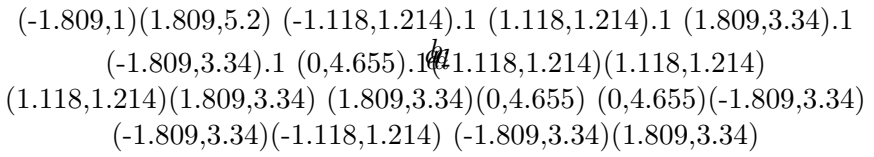
Proof: Draw them! □

Theorem 10.4 *The complete graph K_5 is non-planar.*

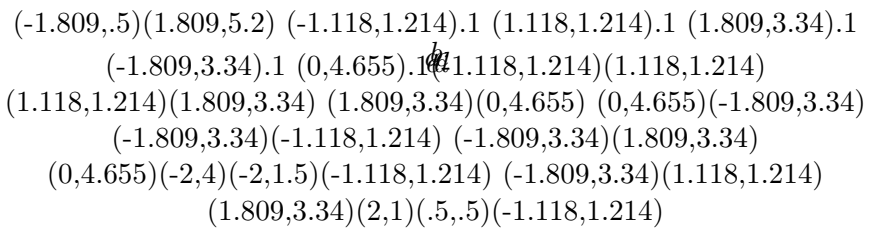
Proof: We attempt to draw K_5 in the plane. We first start with a pentagon:

(-1.809,1)(1.809,5) (-1.118,1.214).1 (1.118,1.214).1 (1.809,3.34).1
(-1.809,3.34).1 (0,4.655).1 (1.118,1.214)(1.118,1.214)
(1.118,1.214)(1.809,3.34) (1.809,3.34)(0,4.655) (0,4.655)(-1.809,3.34)
(-1.809,3.34)(-1.118,1.214)

A complete graph contains an edge between every pair of vertices, so there is an edge between a and c. This may as well be inside the pentagon (as if it is outside then we just adjust the following argument appropriately):



Now we add the edge between b and e (this must be outside the pentagon as it cannot cross $\{a, c\}$), the edge between a and d (inside so as to not cross $\{b, e\}$), and then between c and e (outside so as to not cross $\{a, d\}$):

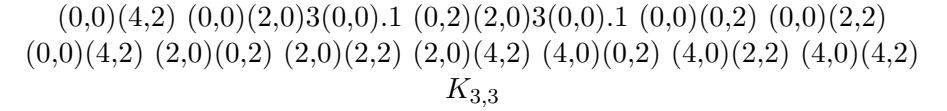


All these edges were forced into position and we have no choice. It remains to add an edge between b and d. We cannot add it inside (since it would cross $\{a, c\}$) nor can we add it outside (since it would cross $\{c, e\}$).

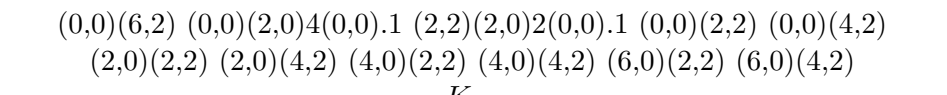
Consequently K_5 is non-planar. □

Definition 10.5 A graph $\Gamma = (V, E)$ is called *bipartite* if $V = V_1 \cup V_2$ with $V_1 \cap V_2 = \varnothing$ and every edge of Γ is of the form $\{a, b\}$ with one of the vertices a and b in V_1 and the other in V_2.

If every vertex in V_1 is joined to every vertex in V_2 we obtain a *complete bipartite graph*. We write $K_{m,n}$ for the complete bipartite graph with $|V_1| = m$ and $|V_2| = n$. Here $|E| = mn$.



$K_{3,3}$

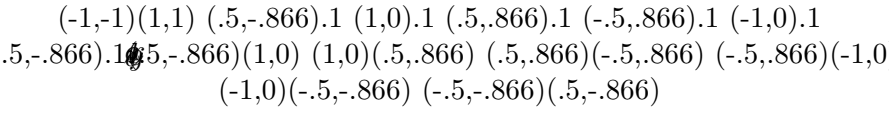


$K_{4,2}$

Theorem 10.6 *The complete bipartite graph $K_{3,3}$ is non-planar.*

PROOF: Let $V_1 = \{a, b, c\}$ and $V_2 = \{x, y, z\}$ and draw a hexagonal circuit:

$$a \to x \to b \to y \to c \to z \to a$$



The proof is completed by observing that two of the edges $\{a, y\}$, $\{b, z\}$ or $\{c, x\}$ must both lie inside or both outside the hexagon and hence must cross. □

The significance of our observation that K_5 and $K_{3,3}$ are non-planar is in the next theorem (which we shall only state). We need a definition first.

Definition 10.7 Two graphs $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ are *homeomorphic* if Γ_2 can be obtained from Γ_1 by the insertion or deletion of a number of vertices of degree 2.

Example 10.8 The following three graphs are homeomorphic:

(0,0)(10,2) (0,0).1 (0,2).1 (1,1).1 (1.5,1.5).1 (2,0).1 (2,2).1 (0,0)(0,2) (0,0)(2,2) (0,2)(2,0) (0,2)(2,2) (2,0)(2,2) (4,0).1 (4,2).1 (5,1).1 (6,0).1 (6,2).1 (4,0)(4,2) (4,0)(6,2) (4,2)(6,0) (4,2)(6,2) (6,0)(6,2) (8,0).1 (8,1).1 (8,2).1 (8.667,2).1 (9,1).1 (9.333,2).1 (9.5,1.5).1 (10,0).1 (10,.667).1 (10,1.333).1 (10,2).1 (8,0)(8,2) (8,0)(10,2) (8,2)(10,0) (8,2)(10,2) (10,0)(10,2)

One can think of homeomorphic graphs as being as the same shape. Adding or deleting a vertex of degree two does not change the shape of the edges but simply replaces a single edge by a pair of edges taking the same shape (or vice versa).

Theorem 10.9 (Kuratowski 1930) *A graph is non-planar if and only if it contains a subgraph that is homeomorphic to either K_5 or $K_{3,3}$.*

The proof is omitted — one direction is very hard to prove. It is clear that if a graph Γ contains either K_5 or $K_{3,3}$ then it cannot be planar (since then K_5 or $K_{3,3}$ would be planar). Similarly the same happens for homeomorphic subgraphs, since homeomorphism does not change the planarity of a graph.

If Γ is a planar graph, then it divides the plane into a number of regions, each of which is bounded by edges (and these meet at the vertices, by planarity). These regions are called the *faces*. This includes one region of infinite area, called the *infinite face*. (Consider the planar graph K_4 for example.)

Theorem 10.10 (Euler) *Let $\Gamma = (V, E)$ be a connected planar graph (not necessarily simple). Let v, e and f be the number of vertices, edges and faces of Γ. Then*

$$v - e + f = 2.$$

PROOF: We proceed by induction on the number e of edges of Γ. If $e = 0$, then $v = 1$ (since Γ is connected) and so $f = 1$ (the infinite face). Hence the result is true when $e = 0$.

Now suppose that the theorem is true for all connected planar graphs with fewer than e edges. Select some edge m in Γ and delete it. This produces a graph Γ' with v' vertices, e' edges and f' faces that satisfies $v' - e' + f' = 2$ by induction. Of course, $e' = e - 1$.

The edge m can be selected so that one of the following conditions hold:

(i) m is a loop. Adding this loop back introduces a new face, so $f' = f - 1$, but does not change the number of vertices so $v' = v$. Hence $v - e + f = v' - (e' + 1) + (f' + 1) = v' - e' + f' = 2$.

(ii) m joins two distinct vertices of Γ. Adding the edge back splits one of the faces in Γ' into two, so $f = f' + 1$. The number of vertices are unchanged, so $v' = v$. Hence $v - e + f = v' - (e' + 1) + (f' + 1) = v' - e' + f' = 2$.

(iii) m is incident to only one vertex in Γ'. This means that to delete the edge m from Γ we needed to remove a vertex of Γ. When we reinstate these, we find that $v = v' + 1$ but that $f = f'$. Hence $v - e + f = (v' + 1) - (e' + 1) + f' = v' - e' + f' = 2$.

These are the only possibilities and so the induction is complete. □

Corollary 10.11 *Let Γ be a connected simple planar graph with v vertices, e edges ($e \geqslant 2$) and f faces. Then*

$$3f \leqslant 2e$$
$$e \leqslant 3v - 6.$$

PROOF: We assume that Γ is drawn in the plane. We have assumed that the graph has no multiple edges and no loops. This has the consequence that the only way that a face could be bounded by less than three edges is if this face is the *infinite face* and Γ is the following graph:

(0,0)(2,0) (0,0)(1,0)3(0,0).1 (0,0)(2,0)

In this case, we verify immediately that $3f = 1 \leqslant 2e = 4$ and $e = 2 \leqslant 3v - 6 = 3$.

Thus, from now one, we may assume that each face is bounded by at least three edges. Add up the number of edges around each face:

$$\sum_{\text{faces } F} (\text{no. of edges bounding } F) \geqslant \sum_{\text{faces } F} 3 = 3f.$$

Each edge lies either side of at most two faces, so we deduce

$$2e \geqslant 3f.$$

Now by Euler's Theorem, $v - e + f = 2$. Hence $e = v + f - 2$, so

$$\begin{aligned} 3e &= 3v + 3f - 6 \\ &\leqslant 3v + 2e - 6, \end{aligned}$$

so

$$e \leqslant 3v - 6.$$

□

An example of an application of this result appears on the second graph theory problem sheet. We can apply this result to prove two results we already know:

Example 10.12 K_5 and $K_{3,3}$ are non-planar.

PROOF: K_5 has 5 vertices and 10 edges. But $3v - 6 = 9 < e$ which is impossible for a planar graph.

$K_{3,3}$ has 6 vertices and 9 edges. If this graph were planar, each face would have to be bounded by at least four edges. The argument used above then shows

$$4f \leqslant 2e = 18.$$

However, Euler's Theorem gives $f = e - v + 2 = 5$ and we have a contradication. Thus $K_{3,3}$ is not planar. □

Section 11

Permutations

Definition 11.1 Let X be a non-empty set. A bijective function $f\colon X \to X$ will be called a *permutation* of X.

Consider the case when X is the finite set with n elements:

$$X = \{1, 2, \ldots, n\}.$$

The collection of all permutations of this set X will be called the *symmetric group* on n symbols and is denoted by S_n.

(We shall meet the definition of the term *group* in the next section.)

Observation: S_n contains $n!$ permutations.

This holds since we have n choices for the image of 1, then $n-1$ choices for the image of 2, etc. We conclude that

$$|S_n| = n(n-1)(n-2)\ldots 2 \cdot 1 = n!.$$

If f is a permutation of the set X, we shall write xf for the image of the element $x \in X$ under f (rather than $f(x)$). The principal reason for doing this is that it makes composition of permutations much easier: fg will mean apply f first and then apply g rather than the other way around.

If $f \in S_n$, then we denote it as follows:

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1f & 2f & 3f & \ldots & nf \end{pmatrix}$$

In this two-row notation, we write the image of an element k in the second row below the occurrence of k in the first row. Thus

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

denotes the permutation of $\{1, 2, 3, 4\}$ which maps 1 to 2, 2 to 4, 3 to 1 and finally 4 to 3.

Note that as f is a bijective function, all n of the elements in $X = \{1, 2, \ldots, n\}$ must occur in the second row. It is for this reason that such functions are termed "permutations": one can think of them as simply re-ordering the elements in X.

The composite of two permutations f and g is the function obtained by applying f first and then applying g. Since we are writing maps on the right, we denote this by fg. It is easy to calculate the permutation obtained by composing two permutations written in the above two-row notation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} = ?$$

Here $1 \mapsto 2$ by the first permutation and then $2 \mapsto 2$ by the second. Thus the composite does the first then the second, so $1 \mapsto 2$. Equally the composite has the following effects:

$$2 \mapsto 4 \mapsto 1, \quad 3 \mapsto 1 \mapsto 3, \quad 4 \mapsto 3 \mapsto 4$$

Hence

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

Similarly

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

(Since $1 \mapsto 3 \mapsto 1$, $2 \mapsto 2 \mapsto 4$, $3 \mapsto 4 \mapsto 3$ and $4 \mapsto 1 \mapsto 2$.) Note this already illustrates one phenomenon: in general,

$$fg \neq gf$$

for two permutations f and g.

Cycle Notation

The above two-row notation is quite inefficient and also difficult to understand the permutations in great detail. For example, the permutation

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

fixes both 1 and 3, while swaps round 2 and 4. It would be nice to have a more efficient way to describe this element (ideally one which misses out 1 and 3 since they are not moved by the permutation).

Definition 11.2 Let x_1, x_2, ..., x_r be r distinct elements of $\{1, 2, \ldots, n\}$ (so $1 \leqslant r \leqslant n$). The *$r$-cycle* $(x_1\, x_2 \ldots x_r)$ is the permutation in S_n which maps

$$x_1 \mapsto x_2, \quad x_2 \mapsto x_3, \quad \ldots, \quad x_{r-1} \mapsto x_r, \quad x_r \mapsto x_1$$

and fixes all other points in $\{1, 2, \ldots, n\}$.

Such a cycle may be described by drawing the points x_i in a circular picture. Thus the cycle could also be written as

$$(x_2\, x_3 \ldots x_r\, x_1), \quad \text{or} \quad (x_3\, x_4 \ldots x_r\, x_1\, x_2), \quad \text{etc.}$$

For example, the above permutation

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

could be written more simply as

$$(2\,4) \qquad \text{or} \qquad (4\,2).$$

This tells us that this permutation fixes both 1 and 3.

What about the identity permutation? This is the permutation

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & 2 & 3 & \ldots & n \end{pmatrix}.$$

This is often written as the cycle (1). (Such a cycle fixes all elements except 1 and moves 1 to 1: so it really is the identity.) Of course, it could also be written (x) for any $x \in \{1, 2, \ldots, n\}$, but to avoid confusion its probably best to stick to $x = 1$.

Definition 11.3 Two cycles $(x_1\, x_2 \ldots x_r)$ and $(y_1\, y_2 \ldots y_s)$ in S_n are *disjoint* if no element in $\{1, 2, \ldots, n\}$ is moved by both cycles.

If these cycles are non-identity (i.e., if $r \geqslant 2$ and $s \geqslant 2$) then this condition can be expressed as

$$\{x_1, x_2, \ldots, x_r\} \cap \{y_1, y_2, \ldots, y_s\} = \varnothing.$$

The crucial observation that will enable us to make use of cycles is the following:

Theorem 11.4 *Every permutation (of n points) can be written as a product of disjoint cycles.*

The proof is omitted. A proof is not too hard, but it is more helpful to give an example to illustrate and this should be fairly convincing of the truth of the theorem.

Example 11.5

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 4 & 3 & 5 & 1 & 6 & 8 & 7 \end{pmatrix} = (1\,2\,4\,5)(3)(6)(7\,8)$$
$$= (1\,2\,4\,5)(7\,8)$$

To calculate this one starts with 1, follow the images round until we get back to 1. Then we start again with the next symbol not accounted for.

It should be reasonably clear that we can follow this process with any permutation and consequently the truth of the above theorem is assured (if not proved in careful detail).

Example 11.6

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 5 & 7 & 4 & 2 & 8 & 1 & 6 \end{pmatrix} = (1\,3\,7)(2\,5)(6\,8)$$

We can use the same method used to calculate the composite of two permutations when these permutations are expressed as products of cycles:

Example 11.7 $(4\,5\,3) \circ (1\,2\,3\,4\,5) = (1\,2\,3\,5\,4)$
[Can be done by following images.]

Definition 11.8 We say two permutations f and g *commute* if $fg = gf$.

We have noticed that we cannot expect two permutations commute. Of course, a permutation f always commutes with itself. The following is easy to establish:

Lemma 11.9 *Disjoint cycles commute.*

(This may simply be described: the effect of a product of disjoint cycles is the same no matter which was around it is calculated.)

This has the consequence that

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 5 & 7 & 4 & 2 & 8 & 1 & 6 \end{pmatrix} = (1\,3\,7)(2\,5)(6\,8)$$
$$= (2\,5)(1\,3\,7)(6\,8)$$
$$= (2\,5)(3\,7\,1)(6\,8)$$

etc. Note that $(2\,5)$ commutes with this permutation:

$$(2\,5)f = (2\,5)(1\,3\,7)(2\,5)(6\,8)$$
$$= (1\,3\,7)(2\,5)(2\,5)(6\,8)$$
$$= (1\,3\,7)(2\,5)(6\,8)(2\,5) = f(2\,5)$$

since disjoint cycles commute. Similar calculations can be done for other examples.

Definition 11.10 The *order* of a permutation f is the smallest positive integer m such that f^m is the identity.

The idea here is that the order of f is the number of permutations we can produce by taking powers of f. Once we have reached the identity, any further powers just produce ones we already have calculated.

How do we calculate the order of f? One method is just to calculate powers (f, f^2, f^3, ...) and wait until we hit the identity. The problem is that this can be laborious. Instead we can exploit the way we can write permutations as products of disjoint cycles.

First consider a cycle $f = (x_1\, x_2 \ldots x_r)$. Note that powers of f first map x_1 to x_2, then to x_3, then to x_4, and so on. Hence to produce the identity, we need to use f^r. (So the order of an r-cycle is r.)

Now consider any permutation f and write it as a product of disjoint cycles:

$$f = f_1 f_2 \ldots f_s.$$

Since disjoint cycles commute, we find

$$f^m = f_1^m f_2^m \ldots f_s^m.$$

To obtain the identity we therefore need to take the m which makes the power of each cycle the identity. We thus have:

Theorem 11.11 *The order of a permutation is equal to the lowest common multiple of the lengths of the cycles occurring in its decomposition into disjoint cycles.*

Example 11.12 Take

$$\begin{aligned} f &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 4 & 7 & 9 & 5 & 10 & 8 & 1 & 2 & 6 \end{pmatrix} \\ &= (1\ 3\ 7\ 8)(2\ 4\ 9)(6\ 10). \end{aligned}$$

We need fourth powers to make the first cycle the identity, cubes to make the second cycle the identity, and squares for the final cycle. Thus the order of σ is $4 \times 3 = 12$.

We give a special name to the following very short cycles.

Definition 11.13 A 2-cycle (that is, a cycle of length 2) is also called a *transposition*.

Thus a transposition is a permutation $(x\, y)$ which simply swaps round the two elements x and y. Transpositions are useful for the following reason:

Theorem 11.14 *Every permutation can be expressed as a product of transpositions.*

PROOF: We can express every permutation as a product of disjoint cycles. The next step is to express any cycle as a product of transpositions. For example,

$$(1\,2)(1\,3)(1\,4)(1\,5) = (1\,2\,3\,4\,5)$$

does what we want for a 5-cycle. Analogous calculations establish the same for other lengths. □

The final thing we can do with permutations is *invert* them:

Definition 11.15 If f is a permutation of the set X, then the *inverse* f^{-1} of f is the permutation that undoes the effect of applying f; i.e., if $f\colon x \mapsto y$, then $f^{-1}\colon y \mapsto x$.

Calculating Inverses, Method 1: If f is written in two-row notation, then interchanging the rows produces its inverse:

e.g., if

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 4 & 3 & 5 & 1 & 6 & 8 & 7 \end{pmatrix}$$

then

$$\begin{aligned} f^{-1} &= \begin{pmatrix} 2 & 4 & 3 & 5 & 1 & 6 & 8 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 1 & 3 & 2 & 4 & 6 & 8 & 7 \end{pmatrix}. \end{aligned}$$

Calculating Inverses, Method 2: If g is a cycle, say $g = (x_1\,x_2 \ldots x_r)$, so

$$g\colon x_1 \mapsto x_2, \quad x_2 \mapsto x_3, \quad \ldots, \quad x_{r-1} \mapsto x_r, \quad x_r \mapsto x_1,$$

then

$$g^{-1}\colon x_r \mapsto x_{r-1}, \quad x_{r-1} \mapsto x_{r-2}, \quad \ldots, \quad x_2 \mapsto x_1, \quad x_1 \mapsto x_r.$$

That is,

$$g^{-1} = (x_r\,x_{r-1} \ldots x_2\,x_1),$$

i.e., we write the cycle for g backwards.

Hence for $f = (1\,2\,4\,5)(7\,8)$, we have

$$\begin{aligned} f^{-1} &= (5\,4\,2\,1)(8\,7) \\ &= (1\,5\,4\,2)(7\,8). \end{aligned}$$

(Note this agrees with the answer obtained by Method 1!)

Section 12

Groups

Definition 12.1 A *binary operation* on a set X is a function $X \times X \to X$. We shall think of it as a method for combining two elements of X to give another and write, for example, $x * y$ for the image of the pair (x, y) under the function.

Definition 12.2 Let G be a set and let $*$ denote a binary operation on G. (Sometimes we use multiplication or addition to denote the binary operation.) So given $a, b \in G$, we have an element $a * b$ in G.

We say $(G, *)$ is a *group* if the following axioms are satisfied:

(i) $*$ is *associative*:

$$a * (b * c) = (a * b) * c \qquad \text{for all } a, b, c \in G;$$

(ii) there is an *identity element* $e \in G$ such that

$$a * e = e * a = a \qquad \text{for all } a \in G;$$

(iii) for each $a \in G$ there is an element $a^{-1} \in G$ such that

$$a * a^{-1} = a^{-1} * a = e.$$

(This is called the *inverse* of a.)

We shall see later that the identity element in a group is unique and that the inverse of an element a is uniquely determined by a. Note that many authors denote the identity element by 1: I would do such in a more advanced course, but shall avoid doing so in this more elementary one.

Some people will list "closure" as a necessary axiom for a group. This, however, is built into our definition of a binary operation: the definition of $*$ being a binary operation on G is that $a * b \in G$ for all $a, b \in G$.

Note that we have not assumed that $a * b$ and $b * a$ are always the same element of our group. In general they are different, but we give a special name to groups where they are equal.

Definition 12.3 Let $(G, *)$ be a group. If $a * b = b * a$ for all elements $a, b \in G$, then we say that G is *abelian* (or *commutative*).

Most groups are *non-abelian*: the ones where every pair of elements commute turn out to be rather special.

Definition 12.4 The *order* of a group $(G, *)$ is the number of elements in the set G. We denote this by $|G|$.

Examples of groups

Example 12.5 $(\mathbb{Z}, +)$, the set of all integers forms an abelian group under addition. It is well known that $+$ is an associative binary operation on $\mathbb{Z}$. The identity element is 0 and $-a$ is the inverse for a:

$$a + (-a) = (-a) + a = 0.$$

Similarly $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$ are groups.

Example 12.6 Let us write $\mathbb{Q}^+$ for $\{x \in \mathbb{Q} \mid x > 0\}$, the set of positive rational numbers. We claim that $(\mathbb{Q}^+, \cdot)$ is a group.

Note that the product of two positive rational numbers is a positive rational number, so multiplication is a binary operation on $\mathbb{Q}^+$. Multiplication is an associative binary operation on $\mathbb{Q}^+$. The identity element is 1. If $x = p/q \in \mathbb{Q}^+$, then $1/x = q/p \in \mathbb{Q}^+$ and this is the inverse we seek for x:

$$x \cdot (1/x) = (1/x) \cdot x = 1.$$

Similarly $(\mathbb{R}^+, \cdot)$ is a group.

However, the set $\mathbb{Z}^+$ of positive integers does not form a group under multiplication. We have 1 as the identity element, but no element (other than 1) has an inverse.

We also have groups $(\mathbb{R} \setminus \{0\}, \cdot)$ and $(\mathbb{Q} \setminus \{0\}, \cdot)$.

The set $\mathbb{R}^-$ of *negative* real numbers does not form a group under multiplication: multiplication is not a binary operation on $\mathbb{R}^-$ (it is not "closed").

Example 12.7 Consider the set S of n-tuples $(a_1, a_2, \ldots, a_n)$ where each a_i is a real number. We can define a binary operation on S by addition in each component:

$$(a_1, a_2, \ldots, a_n) * (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

This binary operation is associative because addition is an associative binary operation on $\mathbb{R}$. The identity element is $(0, 0, \ldots, 0)$. The inverse of $(a_1, a_2, \ldots, a_n) = (-a_1, -a_2, \ldots, -a_n)$.

Example 12.8 Recall that S_n, the *symmetric group* on n points, consists of all permutations of $X = \{1, 2, \ldots, n\}$. We know how to compose two permutations f and g to produce another permutation $f \circ g$. We claim that $(S_n, \circ)$ forms a group.

Associativity: Let $f, g, h \in S_n$. If $x \in X$, then $xf(gh)$ means apply f to x and then apply gh. The latter means apply g and then h. Thus $xf(gh) = ((xf)g)h$. Similarly $x(fg)h$ means first apply fg and then apply h. However to apply fg means apply f and then g. Thus $x(fg)h = ((xf)g)h$. Hence $f(gh) = (fg)h$.

The identity permutation e moves nothing. Thus $f \circ e = e \circ f = f$ for all permutations f.

Finally if f is a permutation, we have seen how to construct the inverse f^{-1}. If $f\colon x \mapsto y$, then $f^{-1}\colon y \mapsto x$. Thus $f \circ f^{-1} = f^{-1} \circ f = e$, so f^{-1} is also the inverse in the sense of Definition 12.2. Hence we do indeed have a group.

Multiplication Tables

One way one can present the information encoded in the binary operation on a group is in a multiplication table (or *Cayley table*). Suppose G is a (fairly small) finite set and we have some binary operation $*$ defined on it. We label the rows and the columns of the table with the elements of G and place the element $a*b$ in the entry with row labelled a and column labelled b. For example:

Group of order 2:

	e	a
e	e	a
a	a	e

Group of order 3:

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Such tables are very good for checking for "closure", finding the identity and inverses, but very bad at checking associativity and not terribly useful for obtaining detailed information as the order of the group grows larger.

Example 12.9 The following is the multiplication table for a group of or-

der 6: $G = \{1, a, b, c, d, e\}$.

	1	a	b	c	d	e
1	1	a	b	c	d	e
a	a	b	1	d	e	c
b	b	1	a	e	c	d
c	c	e	d	1	b	a
d	d	c	e	a	1	b
e	e	d	c	b	a	1

We can see that 1 is the identity element (as the entries in the row and column labelled 1 are appropriate). We can also check for inverses: there is precisely one entry in each row and column equal to 1. Associativity holds but is extremely difficult to check for! The group is *not* abelian:

$$cd = b, \quad dc = a$$

so $cd \neq dc$.

More examples

First recall that we defined

$$a \equiv b \pmod{m}$$

to mean $m \mid (a - b)$. We observed that this was an equivalence relation on the set of integers with m equivalence classes:

$$[0], [1], \ldots, [m-1].$$

Let use $\mathbb{Z}/m$ to denote the set of these m equivalence classes (some authors use $\mathbb{Z}_m$). We have also observed that addition can be used to define an addition on the equivalence classes. Let us drop the brackets, so write (for example)

$$\mathbb{Z}/4 = \{0, 1, 2, 3\}.$$

We then have $2 + 3 = 1$, $3 + 1 = 0$, etc. This gives the following table:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Example 12.10 $(\mathbb{Z}/m, +)$ is a group.

Addition is associative on $\mathbb{Z}/m$, since it is an associative operation on all integers and this is inherited by the equivalence classes.

0 is the identity element, and the inverse of a is $(m-a)$ (as $a+(m-a) \equiv 0 \pmod{m}$).

This examples shows that there is at least one group with any given (finite) order.

What about modular multiplication?

We cannot hope to get a group if we include 0, since $0 \cdot a = 0$ for all a. (This is the same as what happened with multiplication on $\mathbb{Q}$, or $\mathbb{R}$, etc.)

Theorem 12.11 *If p is a prime number, $(\mathbb{Z}/p) \setminus \{0\}$ is a group under multiplication modulo p.*

(If m is not prime, then $(\mathbb{Z}/m) \setminus \{0\}$ is not closed under multiplication: for example, $2 \cdot 3 = 0$ in $\mathbb{Z}/6$.)

PROOF: The operation is closed: if $a, b \in (\mathbb{Z}/p) \setminus \{0\}$, then $p \nmid a$ and $p \nmid b$ (as integers), so $p \nmid ab$. Hence $ab \in (\mathbb{Z}/p) \setminus \{0\}$.

Normal multiplication on integers is associative, consequently the modular multiplication is associative. Clearly 1 is the identity element for the multiplication.

Finally let $a \in (\mathbb{Z}/p) \setminus \{0\}$. Then $p \nmid a$, so p and a are coprime. Hence, by the Euclidean Algorithm, there exist $u, v \in Z$ such that $ua + vp = 1$. Hence

$$ua \equiv 1 \pmod{p}.$$

This means that u is our inverse for a in $(\mathbb{Z}/p) \setminus \{0\}$. □

As an example, the multiplication table of $(\mathbb{Z}/5) \setminus \{0\}$ is:

$\cdot$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Example 12.12 Consider the set of 2×2 matrices with real entries. We might ask whether we can turn this (or some related structure) into a group under matrix multiplication. Matrix multiplication is indeed associative. The only obvious candidate for an identity element is the identity matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

It is easy to check that $IA = AI = A$ for all matrices A. We seek to be able to invert matrices. Accordingly we cannot work with *all* 2×2 matrices, since, for example,

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

Instead, we work with the set of all *invertible* matrices, that is, those with non-zero determinant:

$$G = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; ad - bc \neq 0 \right\}$$

This set does form a group under matrix multiplication: the inverse of

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

is

$$\begin{aligned} A^{-1} &= \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} \end{aligned}$$

Example 12.13 Given a (geometric) shape (for the moment restricted to lying in 2-dimensional space), the collection of invertible transformations that move the shape back to itself form a group. This is called the *symmetry group* of the original shape.

For example, for the square [DRAW!], there are eight symmetries: the identity transformation, three rotations through angles of 90°, 180° and 270°, respectively, and four reflections in axes either through opposite vertices or through the mid-points of opposite edges.

In general, the symmetry group of the regular n-gon is called the *dihedral group* D_{2n} of order $2n$ (though some authors use D_n). The square is the case $n = 4$ and we have observed that its symmetry group does indeed contain 8 transformations.

If we label the vertices of the square as 1, 2, 3, 4, then each symmetry corresponds to a permutation of these numbers. For example, the anticlockwise rotation through 90° corresponds to the cycle (1 2 3 4).

Section 13

Some group theory

We finish the course by considering some examples of the development of the theory of groups.

Lemma 13.1 *Let $(G, *)$ be a group.*

(i) *The identity element e of G is unique.*

(ii) *Each element in G has a unique inverse.*

PROOF: (i) If e and f are identities, then $e * f = f$ since e is an identity, while $e * f = e$ as f is an identity. Thus $e = f$.

(ii) Suppose $a \in G$ and that x and y are inverses for a. Then

$$x * a = a * x = e \qquad \text{and} \qquad y * a = a * y = e.$$

Then using associativity:

$$y = e * y = (x * a) * y = x * (a * y) = x * e = x.$$

□

The upshot of (ii) is that we may safely write a^{-1} for the (unique) inverse of a without worrying that we may be referring to more than one element of our group.

Definition 13.2 If $(G, *)$ is a group and $a \in G$, we define powers of a as follows:

$$a^n = \underbrace{a * a * \cdots * a}_{n \text{ times}}$$

and

$$a^{-n} = (a^n)^{-1}$$

for all $n \in \mathbb{N}$. Also $a^0 = e$, the (unique) identity element.

This, together with the fact that the group operation is associative, has the consequence that standard power laws hold:

$$a^m * a^n = a^{m+n} \qquad (a^m)^n = a^{mn}.$$

Definition 13.3 A group $(G, *)$ is called *cyclic* if there exists some element $a \in G$ such that every element in G has the form a^n. This element a is called the *generator*.

Cyclic groups are very special. For example:

Lemma 13.4 *Cyclic groups are abelian.*

PROOF: Let a be a generator for our cyclic group $(G, *)$. If x and y are elements in G, then $x = a^m$ and $y = a^n$ for some $m, n \in \mathbb{Z}$. Then

$$x * y = a^m * a^n = a^{m+n} = a^n * a^m = y * x.$$

Hence every pair of elements commute, so our group is abelian. □

We already defined the concept of order for a permutation (an element in the symmetric group S_n). We make the same definition for groups.

Definition 13.5 Let $(G, *)$ be a group and a be an element in the group. The *order* of a is the least positive integer m such that $a^m = e$ (the identity element).

If there is no such m, then a is said to have *infinite order*.

Lemma 13.6 *If $(G, *)$ is a cyclic group with generator a, then the order of a equals $|G|$.*

So far we have considered mainly the structure of groups just referring to its elements. To achieve more we need to know about groups that contain groups within them.

Definition 13.7 Let $(G, *)$ be a group. A non-empty subset H of G is called a *subgroup* if $x * y \in H$ and $x^{-1} \in H$ for all $x, y \in H$.

So a subset is a subgroup if it is closed under the group operation and under taking inverses.

The idea here is that if H is a subgroup, then from the binary operation $G \times G \to G$ (given by $(x, y) \mapsto x * y$) we can construct a binary operation on H:

$$\begin{aligned} H \times H &\to H \\ (x, y) &\mapsto x * y. \end{aligned}$$

(The first condition to be a subgroup ensures this always lies in H if we start with elements in H.) Since

$$x * (y * z) = (x * y) * z$$

holds for all $x, y, z \in G$, it certainly holds when we only consider elements which belong to H.

Now our subgroup H is non-empty, so it contains some element x. By assumption, $x^{-1} \in H$. Therefore

$$e = x * x^{-1} \in H.$$

Hence H contains the identity element from G. This behaves like an identity with respect to the elements of H, so H has an identity for its binary operation.

Finally if $x \in H$, then x^{-1} (the inverse as an element of G) also belongs to H, and this behaves like an inverse for x with respect to the binary operation on H.

We therefore have:

Lemma 13.8 *Let $(G, *)$ be a group. A subset H of G is a subgroup of $(G, *)$ if and only if H forms a group under a binary operation induced from $*$.*

Subgroups are useful since they enable us to break our group in particular ways. This will be discussed further in later courses (e.g., MT2002 and later).

We finish with an extended example.

Example 13.9 Let $G = \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ and define a binary operation on G by

$$(a, b) * (c, d) = (ad + c, bd).$$

Then $(G, *)$ is a group. Then $*$ is associative (check!).

We calculate:

$$(a, b) * (0, 1) = (a1 + 0, b1) = (a, b)$$

and

$$(0, 1) * (a, b) = (0b + a, 1b) = (a, b)$$

so $(0, 1)$ is the identity. Also

$$(a, b) * (-a/b, 1/b) = (a(1/b) - a/b, b(1/b)) = (0, 1)$$

and

$$(-a/b, 1/b) * (a, b) = (-(a/b)b + a, (1/b)b) = (0, 1)$$

so $(-a/b, 1/b)$ is the inverse of (a, b).

Let

$$H = \{\, (a,1) \mid a \in \mathbb{R} \,\}$$
$$K = \{\, (a,a) \mid a \in \mathbb{R} \setminus \{0\} \,\}.$$

If $(a,1),(b,1) \in H$, then

$$(a,1) * (b,1) = (a1 + b, 1) = (a+b, 1) \in H$$

and

$$(a,1)^{-1} = (-a/1, 1/1) = (-a, 1) \in H.$$

Hence H is a subgroup of G.

On the other hand, $(1,1),(2,2) \in K$ but

$$(1,1) * (2,2) = (1 \cdot 2 + 2, 1 \cdot 2) = (4,2) \notin K.$$

Hence K is not a subgroup of G.