LECTURE NOTES, ALGEBRA AND NUMBERS, MATH3172

(BASED ON NOTES OF DUGALD MACPHERSON)

Contents

1. The integers 1
2. Congruences 8
2.1. Three other theorems on congruences 10
2.2. Public Key Encryption 13
3. Equivalence relations 15
4. Rings 18
4.1. Subrings 20
4.2. Polynomial rings 21
5. Ideals 22
5.1. Divisibility in rings 25
6. The rings $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Q}[\sqrt{d}]$ 27

These are essentially the lecture notes I will be lecturing from, though the examples given will be slightly different in lectures. There are one or two small topics covered below that may not be covered in lectures. You should aim to be familiar also with the examples in the problem sheets.

1. The integers

The symbol $\mathbb{Z}$ denotes the set of integers. By *positive integers*, we mean integers x with $0 < x$, and x is a *non-negative* integer if $0 \leqslant x$.

Definition 1.1. Let $n, d \in \mathbb{Z}$. We say that d *divides* n (or that d is a *divisor* of n, or that d is a *factor* of n) if there is some $q \in \mathbb{Z}$ with $n = dq$. We denote this by $d|n$. If d does not divide n, we write $d \not| n$.

Example 1.2. $5|10$, $4 \not| 39$, $-7|21$, $9|0$, $0|0$, $0 \not| 5$.

Convention. Often, by a *divisor*, one means a **positive** divisor. You have to work out what is meant from the context.

Lemma 1.3. *(i)* $d|1 \Rightarrow d = \pm 1$.
(ii) $a|b$ *and* $b|a \Rightarrow a = \pm b$.
(iii) $d|a$ *and* $d|b \Rightarrow d|ax + by$ *for any* $x, y \in \mathbb{Z}$.
(iv) $d|n$ *and* $n|m \Rightarrow d|m$.

Date: Academic year 2011–12.

Proof. (i) Obvious.

(ii) If one of a, b is zero, so are both, and the result holds, so suppose neither is zero. Suppose $aq = b$ and $bq' = a$ (where $q, q' \in \mathbb{Z}$). Then $aqq' = a$, so $qq' = 1$, so by (i), $q = \pm 1$, giving $b = \pm a$.

(iii) As $d|a$ there is $k \in \mathbb{Z}$ with $dk = a$, and as $d|b$ there is $l \in \mathbb{Z}$ with $dl = b$. Now $ax + by = dkx + dly = d(kx + ly)$, and as $kx + ly \in \mathbb{Z}$, $d|ax + by$.

(iv) As $d|n$ there is $k \in \mathbb{Z}$ with $dk = n$ and as $n|m$ there is $l \in \mathbb{Z}$ with $ln = m$. Then $m = d(lk)$, so $d|m$. □

Definition 1.4. (a) An integer u is a *unit* if $u|1$ (so the only units are $1, -1$).
(b) An integer p is *prime* if
 (i) $p \neq 0$, p not a unit,
 (ii) $p|ab \Rightarrow p|a$ or $p|b$,
(c) An integer p is *irreducible* if
 (i) $p \neq 0$, p not a unit,
 (ii) the only divisors of p are $\pm 1, \pm p$.

So the first few positive irreducibles are $2, 3, 5, 7, 11, 13$.

Later, you'll see that in $\mathbb{Z}$, 'irreducible'= 'prime'. This is not true in other number systems (other 'rings'); hence the pedantic-looking definitions.

Question 1.5. Are there infinitely many irreducibles (or primes)?

They seem to keep going, but how could one *prove* that there are infinitely many?

Digression – proof by mathematical induction.

Suppose we want to prove, for some property P of certain numbers, the statement 'for all positive integers n, $P(n)$ is true'. We can't in finite time do infinitely many tasks, i.e. separately prove $P(1)$, $P(2)$, $P(3)$, etc.

Method 1. Prove $P(1)$, and for each n prove 'if $P(n)$, then $P(n+1)$'. If we can do this then it follows that for all integers $n \geqslant 1$, $P(n)$ does hold. Indeed, suppose $P(n)$ does not hold for some n. Let k be the least such n (where it fails). We can't have $k = 1$, since we proved $P(1)$. So $k > 1$, so by the minimality of k, $P(k-1)$ *does hold*. But then since we proved '$P(n) \Rightarrow P(n+1)$' for *every* n, in particular for $n = k-1$, we get $P(k)$, a contradiction.

Example 1.6. (1) Let us prove the statement $1 + 2 + \ldots + n = n(n+1)/2$. Let $P(n)$ be this statement. Clearly $P(1)$ holds. Assume $P(n)$ holds, and aim to show $P(n+1)$. Then $1 + 2 + \ldots + (n+1) = (1 + 2 + \ldots + n) + (n+1) =$ (by inductive hypothesis) $n(n+1)/2 + (n+1) = (n+1)(n+2)/2$. Thus, under the assumption of $P(n)$, $P(n+1)$ holds. It follows by induction that $P(n)$ holds for *all* $n \geqslant 1$.

Sometimes, one handles several initial cases separately, before the inductive step (or one only aims to prove $P(n)$ for n greater than some specified integer).

(2) Let $P(n)$ be the statement 'any product of n odd positive integers is odd.' The statement $P(1)$ is obvious – an odd number is odd! Also, consider the case $n = 2$. Let a_1, a_2 be odd, say $a_1 = 2k+1$, $a_2 = 2l+1$. Then $a_1 a_2 = (2k+1)(2l+1) = 2(2kl + k + l) + 1$, so is odd. So $P(2)$ holds. Assume now $P(n)$ holds, for some $n > 2$, and let $a_1, \ldots, a_{n+1}$ be odd positive integers. Then $b := a_1 \ldots a_{n+1} = (a_1 \ldots a_n).a_{n+1}$.

As $P(n)$ holds, $a_1 \dots a_n$ is odd, so b is the product of two positive integers, so is odd (by the $n = 2$ case).

Method 2. 'Course of values' induction. Sometimes, you replace the step 'if $P(n)$ then $P(n+1)$' by the step 'if $P(k)$ holds for all $k \leqslant n$, then $P(n+1)$.' (Strictly speaking, when arguing in this way one doesn't even need to do a 'base case' – think abou it!). This is also valid: if $P(n)$ is false for some n, then there is a least n such that $P(n)$ fails, and this n violates the inductive step.

We end the digression on induction here – but the proof of the next lemma illustrates 'course of values' induction.

Lemma 1.7. *Let k be an integer with $k \geqslant 2$. Then k can be expressed as a product of positive irreducibles.*

Proof. We use 'course of values' induction on k. It is clearly true for the starting point $k = 2$, which *is* irreducible.

Suppose now that for every integer a with $2 \leqslant a \leqslant k-1$, a can be expressed as a product of positive irreducibles. We prove that k can be so expressed.

Case 1. k is irreducible. Then trivially k is a product of irreducibles.

Case 2. k is reducible. Now k has divisors other than ± 1, and we can assume them positive. So $k = lm$ where $2 \leqslant l < k$, $2 \leqslant m < k$. By induction, each of l, m is a product of positive irreducibles. Hence so is k. □

So, irreducibles are the 'building blocks' for multiplication in $\mathbb{Z}$.

Question 1.8. Is the decomposition into irreducibles *unique*?

Clearly, we have to ignore the order of the irreducibles, as for example $3 \times 3 \times 7 \times 5 = 5 \times 3 \times 7 \times 3$. Apart from this, it *is unique*. This will be shown later, after we've developed some more theory. It is false in some other number systems, and is a source of some famous errors.

Theorem 1.9. *There are infinitely many irreducibles in $\mathbb{Z}$.*

Proof. The argument is by contradiction, 'reduction ad absurdum'. Assume the statement false, do some argument, and get a contradiction. So the statement was true after all!

So suppose that the theorem is false, so in particular there are just finitely many *positive* irreducibles, say $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7, \dots, p_t$. (We cannot specify what t is). Write

$$N := p_1 \dots p_t + 1.$$

By Lemma 1.7, we can write N as a product of positive irreducibles, say $N = q_1 \dots q_s$. Now $q_1 | N$. Also, as the p_i list *all* irreducibles, q_1 is one of the p_i, so $q_1 | N - 1$. Hence, by Lemma 1.3(iii), $q_1 | N - (N-1) = 1$. So $q_1 = 1$, contrary to the definition of 'irreducible'.

So, after all, there are infinitely many positive irreducibles! □

There are many other similar but stronger results, such as the following.

Theorem 1.10. *There are infinitely many irreducibles of the form $4k+3$ with k a positive integer.*

Proof. Suppose there are just finitely many, say $p_1 = 3$, $p_2 = 7$, $p_3 = 11$, $p_4 = 19, \ldots, p_t$. Put $N := 4p_1 \ldots p_t - 1$. Then $N = 4(p_1 \ldots p_t - 1) + 3$, so has form $4k + 3$. Write N as a product of positive irreducibles, $N = q_1 \ldots q_s$. Each of the q_i must have for $4k + 1$ or $4k + 3$, as N is odd. If they were all of the form $4k + 1$, then N would also have this form (Exercises 1, Q2(ii)). But N has form $4k + 3$, so some q_i must have form $4k + 3$, so must be among the p_j. Thus, $q_i | 4p_1 \ldots p_t - N = 1$, which is impossible. This contradiction proves the theorem. □

Definition 1.11. Let a, b be integers. An integer g is called a *greatest common divisor* (g.c.d.) of a, b if
(i) $g|a$ and $g|b$ (so g is a common factor of a, b), and
(ii) for any $c \in \mathbb{Z}$, if $c|a$ and $c|b$, then $c|g$.

This seems a bit odd: part (i) is easy to understand, but (ii) is complicated – why not just say that g is the biggest integer satisfying (i)? The reasons:
– if g is a g.c.d of a, b, so is $-g$ – we do not claim that the g.c.d. is unique.
– (ii) gives more information that just saying g is the biggest of the common divisors.
– in some number systems, there may not be any meaning of 'biggest', but the present definition still works.

Definition 1.12. We say integers a, b are *coprime*, or *relatively prime*, if 1 is a g.c.d. of a, b.

Notation. The positive g.c.d. of a, b is denoted (a, b), and really is the 'biggest' common divisor.

So, for example, $(8, 20) = 4 = (-8, 20)$.
$(15, 26) = 1 = (-15, -26)$.
$(36, 54) = 18$.
$(0, 50) = 50$.
We do not define $(0, 0)$.
To find a g.c.d., e.g to find $(540, 900)$: one way is to express each as a product of powers of distinct irreducibles. So $540 = 2^2 \times 3^3 \times 5$ and $900 = 2^2 \times 3^2 \times 5^2$. Then for each prime divisor, take the minimum of the exponents, and multiply these prime powers together. So $(540, 900) = 2^2 \times 3^2 \times 5 = 180$. But we'll shortly see another way.

Lemma 1.13. *The positive g.c.d. of a, b is unique.*

Proof. Suppose c, c' are both positive g.c.d.'s of a, b. Then as c is a common divisor and c' satisfies Definition 1.11 (ii), $c|c'$. Likewise (reversing c, c') $c'|c$. So by Lemma 1.3(ii), $c = \pm c'$, so as they are both positive, $c = c'$. □

Does every pair of integers have a g.c.d? We'll show the answer is 'Yes'. It follows that $g = (a, b)$ is also the *biggest* common divisor d of a, b; for any such d divides g, so as $g > 0$, $d \leqslant g$.

Theorem 1.14 (The Division Algorithm). *Let $d, n \in \mathbb{Z}$ with $d \neq 0$. Then $n = qd + r$ for some $q, r \in \mathbb{Z}$ with $0 \leqslant r < |d|$.*

The Idea. q stands for 'quotient', and r stands for 'remainder'. We divide d into n as many times as possible (q times) and the remainder r is then less than $|d|$. For example, let $n = 50$, $d = 8$.Then $6 \times 8 = 48$, remainder 2, so $50 = 6 \times 8 + 2$, so $q = 6$, $r = 2$. Or if $n = -80$, $d = -7$, we have $12 \times -7 = -84$, remainder 4, so $-80 = 12 \times (-7) + 4$, so $q = 12$, $r = 4$.

Proof of Theorem 1.14. Let $S := \{n - qd : q \in \mathbb{Z}\}$. Then S contains at least one positive integer: for example, if $n \geqslant 0$ take $q = 0$, or if $n < 0$ take $q = nd$ so $n - qd = n - nd^2 = -n(d^2 - 1) \geqslant 0$.

Let r be the smallest non-negative integer in S. Then $r \geqslant 0$, and $r = n - qd$ for some $q \in \mathbb{Z}$.

Now $r - |d| \in S$: for if $d > 0$, then $r - |d| = r - d = n - qd - d = n - (q+1)d$, and if $d < 0$ then $r - |d| = r + d = n - qd + d = n - (q-1)d$.

Also, $r - |d| < r$. If $r - |d| \geqslant 0$, this contradicts the minimality of r. So $r - |d| < 0$, so $r < |d|$. ☐

Lemma 1.15. *Let $a, b, q, r \in \mathbb{Z}$ with $a = qb + r$, and with a, b not both zero. Then $(a, b) = (b, r)$.*

Proof. Let $g = (a, b)$. Then $g > 0$, so to show $g = (b, r)$ we must show (i) $g|b$ and $g|r$, and (ii) for any $c \in \mathbb{Z}$, if $c|b$ and $c|r$ then $c|g$.

Trivially (by definition of (a, b)), $g|b$, and as $g|a$ and $g|b$, $g|a - qb$ so $g|r$. Thus (i) holds. For (ii), suppose $c|b$ and $c|r$. Then $c|qb + r = a$. Thus, as $c|a$ and $c|b$ and g satisfies part (ii) of the definition of (a, b), $c|g$. Thus (ii) above holds for (b, r), so $g = (b, r)$. ☐

Lemma 1.16. *Given n, d as in the Division Algorithm, q and r are unique.*

Proof. See Exercises 1, Q5. ☐

Theorem 1.17 (Euclid's Algorithm)**.** *Every pair of integers a, b (not both 0) has a positive g.c.d. (a, b). Furthermore, there are integers s, t so that*

$$(a, b) = sa + tb$$

(and we can find *such s, t).*

Proof. Step 1: By the Division Algorithm (Theorem 1.14) there are $m_1, r_1 \in \mathbb{Z}$ so that

$$a = m_1 b + r_1 \qquad \text{and } 0 \leqslant r_1 < |b|.$$

Now (if $r_1 \neq 0$) replace a and b by b and r_1 and repeat.

Step 2. By the Division Algorithm (Theorem 1.14) there are $m_2, r_2 \in \mathbb{Z}$ so that

$$b = m_2 r_1 + r_2 \qquad \text{and } 0 \leqslant r_2 < r_1.$$

Now (if $r_2 \neq 0$) replace b and r_1 by r_1 and r_2 and repeat.

Step 3. By the Division Algorithm (Theorem 1.14) there are $m_3, r_3 \in \mathbb{Z}$ so that

$$r_1 = m_3 r_2 + r_3 \qquad \text{and } 0 \leqslant r_3 < r_2.$$

Continue like this. We have $r_1 > r_2 > r_3 > \ldots$ and $r_i \geqslant 0$ for all i, so eventually (possibly already at Step 1) we get m_l, r_l so that $r_l = 0$, that is,

$$r_{l-2} = m_l r_{l-1} + 0.$$

Claim. $(a, b) = r_{l-1}$ (the last non-zero remainder).

Proof of Claim. Clearly $r_{l-1} = (r_{l-1}, r_{l-2})$, as $r_{l-1} | r_{l-2}$. And by Lemma 1.15,

$$r_{l-1} = (r_{l-1}, r_{l-2}) = \ldots = (r_1, r_2) = (b, r_1) = (a, b).$$

□

To find s, t, go back up the above steps. We have $a = m_1 b + r_1$
$b = m_2 r_1 + r_2$
$r_1 = m_3 r_2 + r_3$
and so on, to $r_{l-4} = m_{l-2} r_{l-3} + r_{l-2}$
$r_{l-3} = m_{l-1} r_{l-2} + r_{l-1}$
$r_{l-2} = m_l r_{l-1}$.
Then $(a, b) = r_{l-1} = r_{l-3} - m_{l-1} r_{l-2}$
$= r_{l-3} - m_{l-1}(r_{l-4} - m_{l-2} r_{l-3})$
$= -m_{l-1} r_{l-4} + (1 + m_{l-1} m_{l-2}) r_{l-3})$
=..., until we get an expression $sa + tb$. □

Example 1.18. (1) To find $(-50, 8)$.
$-50 = (-7) \times 8 + 6$
$8 = 6 \times 1 + 2$
and $6 = 3 \times 2$
so $(-50, 8) = 2$, and $2 = 8 - 6 = 8 - (-50 + 7 \times 8) = (-1) \times (-50) + (-6) \times 8$. So $s = -1$, and $t = -6$.

(2) To find $(6300, 1320)$.
$6300 = 4 \times 1320 + 1020$
$1320 = 1 \times 1020 + 300$
$1020 = 3 \times 300 + 120$
$300 = 2 \times 120 + 60$
$120 = 2 \times 60 + 0$,
so $(6300, 1320) = 60$, the last non-zero remainder. Now
$60 = 300 - 2 \times 120$
$= 300 - 2(1020 - 3 \times 300)$
$= 7 \times 300 - 2 \times 1020$
$= 7(1320 - 1020) - 2 \times 1020$
$= -9 \times 1020 + 7 \times 1320$
$= -9(6300 - 4 \times 1320) + 7 \times 1320$
$= -9 \times 6300 + 43 \times 1320$, so $s = -9$ and $t = 43$.

In general, s and t are not uniquely determined.

Recall, a, b are *coprime* if $(a, b) = 1$. We now have

Lemma 1.19. *Let $a, b \in \mathbb{Z}$. Then a, b are coprime if and only if there are $s, t \in \mathbb{Z}$ with $sa + tb = 1$.*

Proof. $\Rightarrow$ Immediate from Theorem 1.17.

$\Leftarrow$ Suppose $sa + tb = 1$ and that $c|a$ and $c|b$. Then $c | sa + tb$ (by Lemma 1.3(iii)) so $c|1$, and hence $c = \pm 1$ (by Lemma 1.3(i)). □

Next, a long-promised fact.

Theorem 1.20. *Let $p \in \mathbb{Z}$. Then p is prime if and only if p is irreducible.*

Proof. Of course, in the definitions of 'prime' and 'irreducible' in Definition 1.4, clause (i) is the same, so we focus on clause (ii).

$\Rightarrow$. Assume p is prime, and $p = ab$ (for $a, b \in \mathbb{Z}$). We must show that a, b are from $\pm 1, \pm p$. Now $p|p$, so $p|ab$, so $p|a$ or $p|b$ (as p is prime). The situation is symmetrical between a and b, so we may assume $p|a$. Then $a = pc$ for some $c \in \mathbb{Z}$. So $p = ab = pcb$, so $cb = 1$, so $b = \pm 1$, $a = \pm p$.

$\Leftarrow$ Assume p is irreducible, and $p|ab$, say $pc = ab$. If $p|a$ we are done, so assume $p \not| a$. The only divisors of p are $\pm 1, \pm p$, so the only common divisors of a, p are ± 1, so $(p, a) = 1$. Thus, by Euclid's Algorithm (Theorem 1.17), there are $s, t \in \mathbb{Z}$ with

$$sp + ta = 1.$$

Then $spb + tab = b$, so $p(sb + tc) = b$, so $p|b$. So if $p \not| a$ then $p|b$, so p is prime. □

Remark 1.21. The above proof shows that if $n|ab$ and $(n, a) = 1$ then $n|b$. For there are $s, t \in \mathbb{Z}$ with $sn + ta = 1$, so $snb + tab = b$, and $n|snb + tab$.

In $\mathbb{Z}$, we'll now just use the word 'prime'. But remember the two words 'irreducible' and 'prime' with their distinct meanings – in other number systems they differ.

Finally, we give a major theorem which illustrates a key theme of the later parts of the module.

Theorem 1.22 (The Fundamental Theorem of Arithmetic)**.** *Let $a \in \mathbb{Z}$, with $a \neq 0$ and a not a unit. Then*

(1) (Existence) a can be expressed as a product $a = up_1 \dots p_m$ where u is a unit and each p_i is a positive prime;

(2) (Uniqueness) also, if there is another expression $a = vq_1 \dots q_n$ where v is a unit and the q_i are positive primes, then $u = v$, $n = m$, and the p_i, q_j can be paired off so that corresponding pairs are equal.

For example, $-160 = (-1) \times 2 \times 2 \times 2 \times 2 \times 2 \times 5$
$= (-1) \times 2 \times 5 \times 2 \times 2 \times 2 \times 2$, etc.

Proof. (Existence) If $a \geqslant 2$, apply Lemma 1.7. If $a \leqslant -2$, find a decomposition for $-a$, and premultiply by (-1).

(Uniqueness). If a is positive then the unit is 1, and if $a < 0$ then the unit is -1, so $u = v$.

Let us assume that a is positive (otherwise first prove the result for $-a$). Dropping the initial unit, $a = p_1 \dots p_m = q_1 \dots q_n$, where the p_i, q_j are positive primes. Also, $p_1|a$, so $p_1|q_1 \dots q_n$, so as p_1 is prime, $p|q_j$ for some j (see Exercises 1, Q3). As q_j is prime, it is irreducible (here we use Theorem 1.20), so in fact $p_1 = q_j$. So as

$$p_1 \dots p_m = q_1 \dots q_n, \quad \text{we have}$$

$$p_2 \dots p_m = q_1 \dots q_{j-1} q_{j+1} \dots q_n.$$

Continuing this way (or arguing by induction), we get rid of all the p_i, and there can't be any q_i left. So $n = m$, and we've paired off the p_i and q_j. □

2. Congruences

Definition 2.1. Let $a, b, n \in \mathbb{Z}$ with $n > 0$. We say 'a is congruent to b modulo n' and write $a \equiv b \pmod{n}$, if $n|a-b$, that is, $a - b = kn$ for some $k \in \mathbb{Z}$.

Examples. $21 \equiv 3 \pmod{6}$; $18 \equiv -3 \pmod{7}$; $-5 \not\equiv 15 \pmod{3}$. Also, a is even if and only if $a \equiv 0 \pmod{2}$ and a is odd if and only if $a \equiv 1 \pmod{2}$.

In the lemmas below, we always assume n is a positive integer.

Lemma 2.2. *(i) For all $x \in \mathbb{Z}$, $x \equiv x \pmod{n}$;*

(ii) For all $x, y \in \mathbb{Z}$, if $x \equiv y \pmod{n}$ then $y \equiv x \pmod{n}$.

(iii) For all $x, y, z \in \mathbb{Z}$, if $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$ then $x \equiv z \pmod{n}$.

Proof. These are all easy. For (iii), we have $x - y = kn$ and $y - z = ln$, say. Then $x - z = (x - y) + (y - z) = (k + l)n$, so $n|x - z$. □

Lemma 2.3. *(i) Every integer a is congruent modulo n to exactly one integer in the range $0, 1, \ldots, n-1$, namely its remainder on division by n.*

(ii) Integers a, b are conguent modulo n if and only if they have the same remainder on division by n.

Proof. (i) By the Division Algorithm (Theorem 1.14) we can write $a = qn + r$ with $0 \leqslant r < n$. Then $a \equiv r \pmod{n}$.

For the uniqueness of r, note that if also $a \equiv r' \pmod{n}$ with $0 \leqslant r' < n$, then $r' \equiv a \equiv r \pmod{n}$, so $r' \equiv r$ (by Lemma 2.2(iii)), so $n|r' - r$. But $|r' - r| < n$, so actually $r' = r$.

(ii) If a, b have the same remainder r, then $a \equiv r$ and $b \equiv r$, so (using Lemma 2.2) $r \equiv b$, so $a \equiv b$ (all modulo n).

Conversely, if a and b are congruent modulo n, and have remainders r and r' respectively, then $a \equiv r$, $b \equiv r'$, and $a \equiv b$, so $a \equiv r'$, so by (i), $r = r'$. □

Lemma 2.4. *Modulo n we have*

(i) If $a \equiv a'$ and $b \equiv b'$ then $a + b \equiv a' + b'$ and $ab \equiv a'b'$.

(ii) If $a \equiv a'$ then $a^r \equiv (a')^r$ for all $r \geqslant 0$.

(iii) If $a \equiv a'$ then $f(a) \equiv f(a')$ for any polynomial $f(x)$ with integer coefficients.

Proof. (i) We have $n|a - a'$ and $n|b - b'$, so $n|(a - a') + (b - b') = (a + b) - (a' + b')$, so $a + b \equiv a' + b'$.

Also, if $kn = a - a'$ and $ln = b - b'$, say, then

$$ab = (a' + kn)(b' + ln) = a'b' + (kb' + kln + a'l)n,$$

so $n|ab - a'b'$.

(ii),(iii) These follow from (i). □

The last lemma generalises familiar facts about 'even', 'odd'. For example, if $x \equiv 1 \pmod{2}$ and $y \equiv 1 \pmod{2}$ then $xy \equiv 1 \times 1 \equiv 1 \pmod{2}$, so odd × odd=odd. Also if $x \equiv 1 \pmod{2}$ and $y \equiv 0 \pmod{2}$ then $x + y \equiv 1 + 0 = 1 \pmod{2}$, so odd+even=odd.

Lemma 2.5. *Modulo n, we have*

(i) if $a \equiv c$ then $ma \equiv mc$.

(ii) If $ma \equiv mc$ and $(m, n) = 1$ then $a \equiv c$.

Proof. (i) Obvious.

(ii) By Theorem 1.17 there are $s, t \in \mathbb{Z}$ with $sm + tn = 1$. Then as

$$ma \equiv mc \pmod n, \text{ we have}$$

$$sma \equiv smc \pmod n,$$

so $sma + tna \equiv smc + tnc$ (we are just adding multiples of n), so

$$(sm + tn)a \equiv (sm + tn)c, \text{ so}$$

$$a \equiv c \pmod n.$$

□

Note. In (ii) we need the hypothesis $(m, n) = 1$. For example, putting $m = n = 2$, we have $2 \times 1 \cong 2 \times 2 \pmod 2$ but $1 \not\equiv 2 \pmod 2$.

Example 2.6. (1) Suppose we wish to find *all* solutions of $x^3 + 6x \equiv 2 \pmod 5$. First, this is the same as $x^3 + x \equiv 2 \pmod 5$ (so we reduce all coefficients modulo 5). By Lemmas 2.3 and 2.4, it suffices to look for solutions in the *finite* set $\{0, 1, 2, 3, 4\}$ – this is trial and error, but easy. The only such solution is $x = 1$. So the general solution of the congruence is $x = 1 + 5k$ (for any $k \in \mathbb{Z}$). If the number n is small, this is a good general method.

(2) What is the last digit of 3^{81}? This is too large a problem for calculators. We need to find the remainder of 3^{81} modulo 10.

Now, $3^{81} = (3^4)^{20} \times 3$. So working modulo 10, $3^4 = 81 \equiv 1$, so $3^{81} \equiv 1^{20} \times 3 \equiv 3$. So the last digit is 3.

(3) Find the remainder when 12^{12} is divided by 567.

Modulo 567 we have:
$12^2 = 144$, so $12^4 = 14462 = 20736 \equiv 324$, so
$12^8 \equiv 324^2 = 104976 \equiv 81$, so
$12^{12} = 12^8 \times 12^4 \equiv 81 \times 324 = 26244 \equiv 162$, the remainder.

(4) Any positive integer is congruent modulo 9 to the sum of its digits. Indeed, write the number n as $a_r a_{r-1} a_{r-2} ... a_0$ (so a_r is the first digit in base 10, etc.). Then $n = a_r \times 10^r + a_{r-1} \times 10^{r-1} + \ldots a_1 \times 10 + a_0 \equiv a_r \times 1^r + a_{r-1} \times 1^{r-1} + \ldots a_1 \times 1 + a_0 = a_r + a_{r-1} + \ldots + a_0$, the sum of the digits.

(5) Any square is congruent to 0 or 1 modulo 4. For if a is congruent to $0, 1, 2, 3$ modulo 4, then, respectively, a^2 is congruent to $0, 1, 4, 9$, hence to $0, 1, 0, 1$, modulo 4.

Lemma 2.7. *Let $a, n \in \mathbb{Z}$ with $n > 0$ and $(a, n) = 1$. Then the congruence $ax \equiv 1 \pmod n$ has a* unique *solution modulo n.*

Proof. First, existence. As $(a, n) = 1$, by Theorem 1.17 there are $s, t \in \mathbb{Z}$ with $as + tn = 1$. Then as $as \equiv 1 \pmod n$, $x = s$ is a solution.

For uniqueness, suppose that x_0 is a solution, and x is arbitrary. Then x is a solution $\Leftrightarrow ax \equiv ax_0 \pmod n \Leftrightarrow x \equiv x_0 \pmod n$, the last step by Lemma 2.5(ii). □

Example. Solve $15x \equiv 1 \pmod{11}$. Here $(15, 11) = 1$ so there is a unique solution modulo 11. We use Euclid's Algorithm to find s, t as above (or, as the numbers are small, just spot them!)
$15 = 11 \times 1 + 4$
$11 = 4 \times 2 + 3$
$4 = 3 \times 1 = 1$
$3 = 3 \times 1 + 0$,
so $1 = 4 - 3 = 4 - (11 - 4 \times 2) = 3 \times 4 - 11 = 3(15 - 11) - 11 = 3 \times 15 - 4 \times 11$. Thus, the solution is $x \equiv 3 \pmod{11}$.

Remark 2.8. How do we solve (can we solve?) a general congruence $ax \equiv b \pmod{n}$.

(i) Suppose $(a, n) = 1$. Use Euclid's Algorithm as above to find a solution y for $ay \equiv 1 \pmod{n}$. Then $ax \equiv b \Leftrightarrow x \equiv by \pmod{n}$. (Proof. If $ax \equiv b$ then $axy \equiv by$ so $x \equiv x(ay) \equiv by$. Conversely, if $x \equiv by$ then $ax \equiv aby = (ay)b \equiv b$.)

For example, solve $15x \equiv 4 \pmod{11}$. By the example above, take $y = 3$, a solution to $15y \equiv 1 \pmod{11}$. The solutions are $x \equiv 4 \times 3 \equiv 1 \pmod{11}$.

(ii) Suppose $(a, n) = d > 1$.

Now if $d \not| b$ there is no solution; indeed, if x_0 was a solution then $ax_0 = b + kn$, and $d | ax_0 - kn$ but $d \not| b$, a contradiction.

So suppose $d|b$ (with $d > 1$). Now the solutions of $ax \equiv b \pmod{n}$ are exactly the solutions of $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$ (CHECK THIS!). Also, $(\frac{a}{d}, \frac{n}{d}) = 1$ (CHECK THIS TOO!). So $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$ has a unique solution mod $\frac{n}{d}$, by case (i). This gives d solutions modulo n.

Example 2.9. (i) Solve $40x \equiv 12 \pmod{28}$. Now $(40, 28) = 4$, and $4|12$, so the solutions are the same as for $10x \equiv 3 \pmod 7$. As $(10, 7) = 1$, we use Euclid's Algorithm to find s, t with $10s + 7t = 1$, namely $s = -2$, $t = 3$. So the solutions of $10y \equiv 1 \pmod 7$ are $y \equiv -2 \equiv 5 \pmod 7$, so the solutions of $10x \equiv 3 \pmod 7$ are $x \equiv 3 \times 5 \equiv 1 \pmod 7$. (As the numbers are small, you could have just spotted this.) So, the solutions of $40x \equiv 12 \pmod{28}$, written modulo 28, are $x \equiv 1, 8, 15, 22 \pmod{28}$. (Note: you might have been asked to find all solutions in the range $0, 1, \ldots, 27$; these would be $1, 8, 15, 22$.)

(ii) Solve $40x \equiv 14 \pmod{28}$. There are no solutions, as $4 = (40, 28) \not| 14$.

(iii) Solve $3x \equiv 6 \pmod 7$. Here $(3, 7) = 1$, and $(-2) \times 3 + 1 \times 7 = 1$. So the solutions of $3y \equiv 1 \pmod 7$ are $y \equiv -2 \equiv 5 \pmod 7$, and the solutions of $3x \equiv 6 \pmod 7$ are $x \equiv (-2) \times 6 \equiv 2 \pmod 7$.

2.1. **Three other theorems on congruences.** In an equation like $40^{85}x \equiv 20^{102} \pmod{11}$, we can replace 40 by 7, and 20 by 9, but we CANNOT just change the exponents modulo 11. For such equations, we often use Fermat's Little Theorem.

Theorem 2.10 (Fermat's Little Theorem)**.** *Let a be an integer and p a positive prime. Then $a^p \equiv a \pmod{p}$.*

Proof. We first prove it for non-negative a. We use induction on a. For the case $a = 0$ it just says $0^p \equiv 0$, obviously true. So assume it is true for $a = k$, and deduce that it holds for $a = k+1$. Now, by the Binomial Theorem, $(k+1)^p = k^p + \binom{p}{1}k^{p-1} + \binom{p}{2}k^{p-2} + \ldots + 1^p$. This is congruent modulo p to $k^p + 1$, as $p|\binom{p}{i}$ for all i with $1 \leqslant i \leqslant p-1$, by Sheet 1 Q4. By induction hypothesis, $k^p \equiv k \pmod{p}$, so $(k+1)^p \equiv k+1 \pmod{p}$. So we have proved the result for $a = k+1$, so by induction, it holds for all $a \geqslant 0$.

Finally, suppose $a < 0$. If p is odd, then $a^p = -(-a)^p \equiv -(-a) = a$, and if $p = 2$, $a^p = a^2 \equiv -(-a)^2 \equiv -(-a) = a$, in both cases using the last paragraph for the congruence step. □

Corollary 2.11. *Let a be an integer, and p a prime with $(a,p) = 1$. Then $a^{p-1} \equiv 1 \pmod p$.*

Proof. By Theorem 2.10, $a^p \equiv a \pmod p$. Now we may cancel a as $(a,p) = 1$ – we here use Lemma 2.5(ii). □

Example 2.12. We want to solve the congruence mentioned above, $40^{85}x \equiv 20^{102} \pmod{11}$. As noted, this is the same as $7^{85}x \equiv 9^{102} \pmod{11}$. Now by the last Corollary, as 11 is prime and $(11,7) = (11,9) = 1$, $7^{10} \equiv 1$ and $9^{10} \equiv 1 \pmod{11}$. Also, $85 = 8 \times 10 + 5$ and $102 = 10 \times 10 + 2$, so the congruence is

$$(7^{10})^8 \times 7^5 x \equiv (9^{10})^{10} \times 9^2 \pmod{11},$$

which is $1^8 \times 7^5 x \equiv 1^{10} \times 9^2$. Now $7^5 = 7^2 \times 7^2 \times 7 \equiv 5 \times 5 \times 7 \equiv 3 \times 7 \equiv 10 \pmod{11}$, and $9^2 = 81 \equiv 4 \pmod{11}$. So we have $10x \equiv 4 \pmod{11}$. Now $(-1) \times 10 + 1 \times 11 = 1$, so $10y \equiv 1 \pmod{11}$ has solutions $y \equiv -1 \equiv 10 \pmod{11}$, so $10x \equiv 4 \pmod{11}$ has solutions $x \equiv 40 \equiv 7 \pmod{11}$. So the original congruence has solutions $x \equiv 7 \pmod{11}$.

Example 2.13. For $p = 13$, we have $2 \times 7 \equiv 1$, $3 \times 9 \equiv 1$, $4 \times 10 \equiv 1$, $5 \times 8 \equiv 1$, $6 \times 11 \equiv 1$, so $2 \times 7 \times 3 \times 9 \times 4 \times 10 \times 5 \times 8 \times 6 \times 11 \equiv 1^5 = 1$. So $12! \equiv 1^5 \times 12 \equiv -1 \pmod{13}$.

This is the idea of the proof for the following general example:

Lemma 2.14. *Let p be a prime number and a be in the range $2, 3, \dots, p-2$. Then the equation $ax \equiv 1 \pmod p$ has the unique solution in the same range and, moreover, $a \neq x$.*

Proof. Consider the $p-3$ numbers $2, 3, \dots, p-2$. If a is one of these numbers, then, by Lemma 2.7, there is a unique number x in the range $0, 1, \dots, p-1$ with $ax \equiv 1 \pmod p$. Also, x is in the range $2, 3, \dots, p-2$. For if $x = 0$ we get $a \times 0 = 0 \not\equiv 1$; if $x = 1$ we get $ax \equiv a \equiv 1 \bmod p$ so $a = 1$; and if $x = p-1$ then $1 \equiv ax = a(p-1) \equiv -a$, so $a \equiv -1$ so $a = p-1$; in each case the assumptions on a are contradicted.

Also, $x \neq a$, for otherwise $ax = a^2 \equiv$, so $p | a^2 - 1 = (a+1)(a-1)$. Hence, as p is prime, $p | a+1$ or $p | a-1$, so $a \equiv -1$ or $a \equiv 1 \pmod p$. These are impossible as $2 \leqslant a \leqslant p-2$. □

As an immediate corollary we obtain:

Theorem 2.15 (Wilson's Theorem). *Let p be a positive prime. Then $(p-1)! \equiv -1 \pmod p$.*

Proof. If $p = 2$ or $p = 3$ it is clear by calculation, so assume $p > 3$. By Lemma 2.14, we can pair off the numbers in the range $2, \dots, p-2$, so the product of each pair is congruent to 1 modulo p. Thus, $(p-1)! \equiv 1 \times \dots \times 1 \times (p-1) \equiv -1 \pmod p$. □

Theorem 2.16 (Chinese Remainder Theorem). *If $n_1, n_2, \dots, n_k$ are pairwise coprime integers and $a_1, \dots, a_k \in \mathbb{Z}$, then the simultaneous congruences*

$x \equiv a_1 \pmod{n_1}$
$x \equiv a_2 \pmod{n_2}$
.
.

.
$x \equiv a_k \pmod{n_k}$
have a unique solution modulo $n_1 n_2 \dots n_k$.

History. This is the problem of Sun Tsu (AD100?). Suppose we have an unknown number of objects. When counted in threes, 2 are left over. When counted in fives, 3 are left over, and when counted in sevens, 2 are left over. How many objects are there? This is equivalent to solving the simultaneous congruences
$x \equiv 2 \pmod 3$
$x \equiv 3 \pmod 5$
$x \equiv 2 \pmod 7$.

Proof of Theorem 2.16. We use induction on k – the result is clearly true for $k = 1$. Suppose $k = 2$, and consider (with $(n_1, n_2) = 1$) the congruences
$x \equiv a_1 \pmod{n_1}$
$x \equiv a_2 \pmod{n_2} \quad (*)$.

By Euclid's Algorithm, there are $s_1, s_2 \in \mathbb{Z}$ such that $s_1 n_1 + s_2 n_2 = 1$. Define

$$x_0 = a_1 s_2 n_2 + a_2 s_1 n_1.$$

Then x_0 is a solution, for (using $s_2 n_2 = 1 - s_1 n_1$) we find

$$x_0 = a_1 s_2 n_2 + a_2 s_1 n_1 = a_1(1 - s_1 n_1) + a_2 s_1 n_1 \equiv a_1 \pmod{n_1},$$

and, as $s_1 n_1 = 1 - s_2 n_2$,

$$x_0 = a_1 s_2 n_2 + a_2 s_1 n_1 = a_1 s_2 n_2 + a_2(1 - s_2 n_2) \equiv a_2 \pmod{n_2}.$$

Now x is a solution of the simultaneous congruences $(*)$
$\Leftrightarrow x \equiv x_0 \pmod{n_1}$ and $x \equiv x_0 \pmod{n_2}$
$\Leftrightarrow n_1$ and n_2 divide $x - x_0$
$\Leftrightarrow x - x_0 = n_1 q$ for some $q \in \mathbb{Z}$ and $n_2 | n_1 q$
$\Leftrightarrow x - x_0 = n_1 q$ for some $q \in \mathbb{Z}$ and $n_2 | nq$ (by Problem Sheet 2, Q1)
$\Leftrightarrow n_1 n_2 | x - x_0$
$\Leftrightarrow x \equiv x_0 \pmod{n_1 n_2}$.
This gives the result for $k = 2$.

Now assume the result for a system of $k-1$ congruences, where $k-1 \geqslant 2$, and prove it for k congruences (e.g. the system in the statement of the theorem). By the $k = 2$ case, we may replace the first two congruences by the single congruence

$$x \equiv x_0 \pmod{n_1 n_2}.$$

Now $n_1 n_2$ is coprime to $n_3, \dots, n_k$, so the result holds by induction on k.

□

Example 2.17. Find all solutions to the simultaneous congruences
$x \equiv 1 \pmod 3$
$x \equiv 5 \pmod 7$
$x \equiv 2 \pmod{11}$.

To solve the first two, find s_1, s_2 with $3s_1 + 7s_2 = 1$ – e.g. put $s_1 = -2$ and $s_2 = 1$. Then $x_0 = 1 \cdot 7.1 + 5 \cdot 3.(-2) = -23 \equiv 19 \pmod{21}$ is the general solution. (Check it!)

Now solve
$x \equiv 19 \pmod{21}$
$x \equiv 2 \pmod{11}$.
Find t_1, t_2 with $21t_1 + 11t_2 = 1$. Again, rather than use Euclid's Algorithm you can just spot $t_1 = -1$, $t_2 = 2$. So $x_1 = 19 \cdot 11 \cdot 2 + 2 \cdot 21 \cdot (-1) = 418 - 42 = 376$ is a solution. Our general solution is modulo $11 \times 21 = 231$, and $376 \equiv 145 \pmod{231}$, so the general solution is $x \equiv 145 \pmod{231}$. Check this works for all three equations.

2.2. **Public Key Encryption.**

We consider an application of Fermat's Little Theorem (and Euclid's Algorithm) to cryptography.

Traditionally, secret messages are sent as follows. Alice wants to send a message m (the *plaintext*) to Bob. It is assumed that m is already a sequence of digits, obtained by some trivial way of converting letters to numbers. Using some encrypting method, she turns m into a *ciphertext*. This is sent to Bob, who decrypts it. In traditional methods, both Alice and Bob know *both* the encrypting and decrypting methods. This is potentially insecure.

Public key cryptography is based on the idea that certain mathematical operations are computationally feasible, but the inverse operation may be hopelessly unfeasible. In particular, suppose p, q are huge prime numbers (e.g. with 500 digits) and $n = pq$. Recovering this factorisation from n may be unfeasible.

RSA ciphers are named after Rivest, Shamir and Adleman, who published a key paper in 1978. It turned out that the same method had been worked out in GCHQ, Cheltenham several years earlier, but kept secret.

Fix two huge prime numbers p, q (e.g. 500 digits). It may be hard to find such primes (or to determine that they are prime), but it is feasible to produce numbers which are prime with a very high degree of probability. Put $n = pq$, $k = (p-1)(q-1)$. Now choose large d with $(k, d) = 1$ – for example choose d to be a prime greater than p, q.

The person *receiving* messages, namely Bob, knows n, p, q, k, d. He publishes just n, d. Then *anyone* can easily send Bob a message, and can send it publicly, but only Bob can decode it in reasonable time.

The message (the plaintext) will be an integer m with $1 < m < n$. (In practice, it may be a sequence $m_1, m_2, \ldots, m_t$ of such numbers.) To encode m, find c with $0 \leqslant c < n$ with $m^d \equiv c \pmod{n}$. It is not so hard for Alice to find such c – we've done exercises like that.

To decode, Bob uses Euclid's Algorithm to find integers $x > 0$ and $y < 0$ such that $dx + ky = 1$ – these exist as $(d, k) = 1$. This calculation is done by Bob just once – he uses the result for *all* messages he receives. Note that no-one else knows k, so no-one else could find such x, y – as they cannot in reasonable time recover p, q from n, d.

Lemma 2.18. $m \equiv c^x \pmod{n}$.

Thus, to decode the message, Bob finds (reasonably easily) c' with $0 \leqslant c' < n$ such that $c^x \equiv c' \pmod{n}$, and knows that $m = c'$.

Proof of Lemma 2.18. Since $n = pq$ with p, q prime, if $m \equiv c^x \pmod{p}$ and $m \equiv c^x \pmod{q}$ then $p | m - c^x$ and $q | m - c^x$, so $pq | m - c^x$, so $m \equiv c^x \pmod{n}$.

To see $m \equiv c^x \pmod{p}$, note

$$c^x \equiv (m^d)^x = m^{dx} = m^{1-ky} = m.m^{-(ky)} = m.m^{-(p-1)(q-1)y} = m.m^{(p-1)i} \pmod{n},$$

where $i = -(q-1)y > 0$. Hence as $p|n$,

$$c^x \equiv m.m^{(p-1)i} \pmod p.$$

Now either $p|m$, or $(m,p) = 1$, and in the latter case $m^{p-1} \equiv 1 \pmod p$ by the Corollary to Fermat's Little Theorem (Corollary 2.11). Either way, we find $c^x \equiv m \pmod p$.

Likewise $c^x \equiv m \pmod q$, and so $c^x \equiv m \pmod n$.

$\square$

How do we obtain the plaintext from message in English? We could use e.g. a table converting letters, punctuation, and numerals into two digit numbers: for example $A = 00, B = 01, \ldots, M = 12, \ldots, Z = 25$, comma $= 26$, full stop $= 27$, question mark $= 28$, $0 = 29, 1 = 30, \ldots, 9 = 38$, exclamation mark $= 39$, with 99 indicating a space between words. (I've taken this from D.M. Burton *Elementary Number Theory*, p. 148.)

As an example with ridiculously small numbers try $p = 11$, $q = 13$, so $n = 143$ and $k = (p-1)(q-1) = 120$. Choose $d = 7$ – again, far too small, but at least coprime to k. Remember, the recipient Bob knows all these numbers, but only publishes 143, and 7. (Of course, as 143 is so small, it it trivial for everyone to work out p, q, k, but this wouldn't be true for a large such n.) Bob also finds $x, y \in \mathbb{Z}$ with $dx + ky = 1$ and with $y < 0$. We find $120 = 7 \times 17 + 1$, so $(-17) \cdot 7 + 1 \cdot 120 = 1$, but unfortunately $1 > 0$. However, we use the trick from Problem Sheet 1, Q7(b) to get $y < 0$. That is, $(-17)7 + 1 \cdot 120 + 120 \cdot 7 - 120 \cdot 7 = 1$, so $7(120-17) + 120(1-7) = 1$, so $7 \cdot 103 + 120 \cdot (-6) = 1$, that is, $x = 103$ and $y = -6$.

Alice wishes to encode the word SAUSAGE. Using the routine in the last paragraph, this gives plaintext 18002018000604. She encodes this two digits at a time (but she might use longer blocks than two, but each block should as a number be at most n)). The initial 18 is encoded by the remainder of $18^7 \pmod{143}$. Now $18^2 = 324 \equiv 38 \pmod{143}$, so $18^4 \equiv 38^2 = 1444 \equiv 14 \pmod{143}$, so $18^6 = 18^2 \times 18^4 \equiv 38 \times 14 = 532 \equiv 103 \pmod{143}$, and $18^7 = 103 \times 18 = 1854 \equiv 138 \pmod{143}$. So the first three digits of the ciphertext are 138. Since $(00)^7 = 0$, the next three digits are 000. (Note that as 143 has three digits, we should expect our blocks of ciphertext to have three digits.)

However, a better and more instructive approach is to use the Chinese Remainder Theorem and Fermat's Little Theorem as follows. We have $143 = 11 \times 13$. To find 18^7 modulo 143, we first find c_1, c_2 such that
$18^7 \equiv c_1 \pmod{11}$ and
$18^7 \equiv c_2 \pmod{13}$. In fact, the exponent 7 is here so small that Fermat's Little Theorem plays no role (with larger d, as in the lecture notes, it does). Reducing mod 11 and 13, we solve
$7^7 \equiv c_1 \pmod{11}$
$5^7 \equiv c_2 \pmod{13}$.
Now $7^7 = (7^2)^3 \cdot 7 = 49^3 \cdot 7 \equiv 5^3 \cdot 7 = 25 \cdot 35 \equiv 3 \cdot 2 = 6$ and $5^7 = (5^2)^3 \cdot 5 \equiv (-1)^3 \cdot 5 = -5 \equiv 8$. So $c_1 = 6$ and $c_2 = 8$. So our ciphertext will be $x \in \{0, \ldots, 142\}$ with
$x \equiv 6 \pmod{11}$
$x \equiv 8 \pmod{13}$. (For such x will be congruent to 18^7 modulo 11 and modulo 13, and hence modulo $11 \times 13 = 143$.)

We do this by the Chinese Remainder Theorem (Theorem 2.16). First find s, t with $11s + 13t = 1$ – using Euclid's Algorithm or guesswork we get $11 \cdot 6 + 13 \cdot (-5) = 1$. Thus, by the Chinese Remainder Theorem, $x \equiv 6 \cdot (-5) \cdot 13 + 8 \cdot 6.11 = -390 + 528 = 138$ (as before).

The advantage of this way is we can use the same values s, t for other calculations, and it is easier to do by hand. We continue: as before, clearly $00^7 \equiv 0 \pmod{143}$, so plaintext 00 gives ciphertext 000.

We next find $20^7 \pmod{143}$. Now modulo 11, $20^7 \equiv 9^7 = (9^2)^3 \cdot 9 = 4^3 \cdot 9 = 16 \cdot 36 \equiv 5 \cdot 3 = 15 \equiv 4$. And modulo 13, $20^7 \equiv 7^7 = (7^2)^3 \cdot 7 \equiv 5^3 \cdot 7 = 25 \cdot 35 \equiv (-1)(-4) = 4$. So solve
$x \equiv 4 \pmod{11}$
$x \equiv 4 \pmod{13}$
by Chinese Remainder Theoirem as above to get $x \equiv 4 \cdot 13 \cdot (-5) + 4 \cdot 11 \cdot 6 = -260 + 264 = 4$, so plaintext 20 gives ciphertext 004.

Repeating the above 18 gives ciphertext 138, and 00 gives ciphertext 000. We find $(06)^7 \pmod{143}$. Here bare hands quickly gives $6^4 = 1296 \equiv 9 \pmod{143}$, and $6^3 = 216 \equiv 73 \pmod{143}$, so $6^7 \equiv 9 \times 73 = 657 \equiv 85$, so the ciphertext for 06 is 085. Likewise $(04)^7 = 256 \cdot 4^3 \equiv 113 \cdot 4^3 = 452 \cdot 16 \equiv 23 \cdot 16 = 368 \equiv 82$, so plaintext 04 gives ciphertext 082. If I've made no calculation errors (!) the overall message gets encoded as 138000004138000085082.

To decode, Bob will treat the ciphertext as having blocks of length 3, each corresponding to a letter, or number, or punctuation, or space. He first finds the remainder of $(138)^x$, that is $(138)^{103}$, modulo 143. This should be 18. Let's check it, again using Fermat's Little Theorem and the Chinese Remainder Theorem. Modulo 11, $138^{103} \equiv 6^{103} = (6^{10})^{10}.6^3 \equiv 1 \cdot 6^3 = 36 \cdot 6 \equiv 3 \cdot 6 \equiv 7$.
And modulo 13, $138^{103} \equiv 8^{103} = (8^{12})^8 \cdot 8^7 \equiv 1 \cdot 8^7 = 64^3 \cdot 8 \equiv (-1)^3 \cdot 8 = 5$.
Thus, the plaintext x satisfies $x \equiv 7 \pmod{11}$ and $x \equiv 5 \pmod{13}$, so by the Chinese Remainder Theorem, $x \equiv 7 \cdot (-5) \cdot 13 + 5 \cdot 6.11 = -455 + 330 = -125 \equiv 18 \pmod{143}$, as required.

He then finds the remainder of $(000)^{103}$ modulo 143, which of course is 00. Continuing, he recovers the plaintext from the ciphertext.

One further comment: we have used Fermat's Little Theorem and the Chinese Remainder Theorem to save work when finding powers modulo n. Of course, when Alice finds the ciphertext (so when calculating $m^d \pmod{n}$), she can't actually do this, as she doesn't know p and q! Bob, however, can use this method when decoding. Everything is done by computer anyway, and the assumption is that finding powers modulo n doesn't take a computer too long.

3. Equivalence relations

Definition 3.1. Let X be a set. A *relation* R on X is a subset of $X \times X$, so is a set of pairs from X. If $x, y \in X$, write xRy if $(x, y) \in R$.

We say that R is an *equivalence relation* on X if it satisfies the following three properties.
Reflexive: xRx for all $x \in X$;
Symmetric: if xRy then yRx.
Transitive: if xRy and yRz then xRz.

Example 3.2. (1) (1) The relation $\leqslant$ on $\mathbb{Z}$ is reflexive: $x \leqslant x$ holds for all x. It is transitive: if $x \leqslant y$ and $y \leqslant z$ then $x \leqslant z$. It is not symmetric, as $2 \leqslant 3$ but $3 \nleqslant 2$. So it is not an equivalence relation. The relation $<$ on $\mathbb{Z}$ is transitive, but not reflexive or symmetric.

(2) Consider the relation D on $\mathbb{Z}$ defined by xDy if and only if $|x-y| \leqslant 1$. This is reflexive and symmetric, but not transitive: $1D2$ and $2D3$, but not $1D3$. So D is not an equivalence relation.
(3) The relation S on $\mathbb{C}$ defined by putting xSy if $x^4 = y^4$ is an equivalence relation. Here $1S1$ and $1S(-1)$ and $1Si$ and $1S(-i)$.

Lemma 3.3. *Congruence modulo n is an equivalence relation on $\mathbb{Z}$.*

Proof. We did it – see Lemma 2.2. □

Definition 3.4. If R is an equivalence relation on X, and $x \in X$, define

$$\hat{x} := \{y \in X : xRy\}.$$

An *equivalence class* for R is a subset of X of the form $\hat{x}$ for some x. If R is an equivalence relation on X, denote by X/R the set of equivalence classes on X.

For example, for the relation on $\mathbb{Z}$ of congruence modulo 4, we have

$$\widehat{-1} = \{..., -5, -1, 3, 7, 11, ...\}$$
$$\hat{0} = \{..., -4, 0, 4, 8, ...\}$$
$$\hat{1} = \{..., -3, 1, 5, 9, ...\}$$
$$\hat{2} = \{..., -6, -2, 2, 6, ...\}$$
$$\hat{3} = \{..., -5, -1, 3, 7, ...\}$$
$$\hat{4} = \{..., -4, 0, 4, 8, ...\}$$
$$\hat{5} = \{..., -3, 1, 5, 9, ...\}.$$

So there are exactly four different equivalence classes,

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

Definition 3.5. A *partition* of a set X is a collection $\{X_1, X_2, \ldots\}$ of subsets of X such that
(a) each X_i is non-empty,
(b) each element of X lies in exactly one of the X_i (so they are mutually exclusive, and exhaust X).

Theorem 3.6. *Let X be a set.*
(i) If R is an equivalence relation on X, then the set of equivalence classes of R is a partition of X.
(ii) Any partition of X is the set of equivalence classes of some equivalence relation on X.

Examples. For congruence modulo 4, view $\mathbb{Z}$ as split into the four equivalence classes. For the relation S on $\mathbb{C}$ (where xSy means $x^4 = y^4$), view $\mathbb{C}$ as split into equivalence classes $\{0\}$ (a class of size 1) and classes of size 4 of form $\{a, -a, ia, -ia\}$.

Proof of Theorem 3.6. (i) If $x \in X$ then $x \in \hat{x}$ (as R is reflexive), so each equivalence class is non-empty and their union is X. We must now show that distinct equivalence classes have empty intersection. In fact, we note

(1) if xRy then $\hat{x} = \hat{y}$, and

(2) if not xRy, then $\hat{x} \cap \hat{y} = \emptyset$.

Together, these suffice.

To see (1): Suppose xRy: if $z \in \hat{y}$ then yRz, so xRz by transitivity, so $z \in \hat{x}$. Thus, $\hat{y} \subseteq \hat{x}$. By symmetry, also yRx, so the same argument gives $\hat{x} \subseteq \hat{y}$, so together these give $\hat{x} = \hat{y}$.

(2) We prove the 'contrapositive', so suppose $\hat{x} \cap \hat{y} \neq \emptyset$, so there is some $z \in \hat{x} \cap \hat{y}$. Then xRz and yRz, so zRy (symmetry), and so xRy (symmetry).

(ii) Given a partition of X, define a relation R on X, putting xRy if and only if x, y belong to the same set in the partition. Now check R is reflexive, symmetric, and transitive. $\square$

Notation. Let n be a positive integer. We denote by $\mathbb{Z}_n$ the set of equivalence classes in $\mathbb{Z}$ for the equivalence relation of congruence modulo n.

Thus, $x \equiv y \pmod{n}$ if and only if the elements $\hat{x}, \hat{y}$ of $\mathbb{Z}_n$ are equal. As any integer is congruent to exactly one of $0, 1, 2, \dots, n-1$, the set $\mathbb{Z}_n$ has exactly n elements, namely $\mathbb{Z}_n = \{\hat{0}, \hat{1}, \dots, \widehat{n-1}\}$.

For any integer x, $\hat{x}$ is equal to one of $\hat{0}, \hat{1}, \dots, \widehat{n-1}$. For example, for $n = 4$, we have $\mathbb{Z}_4 = \{\hat{0}, \hat{1}, \hat{2}, \hat{3}\}$, and $\hat{4} = \hat{0}$, $\hat{5} = \hat{1} = \widehat{-3}$, etc.

Define operations of addition and multiplication on $\mathbb{Z}_n$ by:

$\hat{x} + \hat{y} = \widehat{x+y}$

$\hat{x} \times \hat{y} = \widehat{xy}$.

e.g., for $n = 4$, $\hat{2} + \hat{3} = \hat{1}$, $\hat{2} \times \hat{3} = 2 \hat{\times} 3 = \hat{6} = \hat{2}$, $\hat{2} \times \hat{2} = \hat{0}$, etc.

Thus, $\mathbb{Z}_n$ is a 'number system'.

Remark 3.7. (i) It is not obvious that $+, \times$ on $\mathbb{Z}_n$ are 'well-defined'. We can write $\hat{x}$ in different ways, e.g. $\hat{4} = \hat{10}$ in $\mathbb{Z}_6$. But the definition of $+$, $\hat{x} + \hat{y} = \widehat{x+y}$, appeared to depend on how we write $\hat{x}$.

For example, in $\mathbb{Z}_6$, $\hat{4} + \hat{5} = \hat{9} = \hat{3}$, and also $\hat{10} + \hat{-1} = \hat{9} = \hat{3}$, so we get the same answer. Likewise, $\hat{4} \times \hat{2} = \hat{8} = \hat{2}$, and $\hat{16} \times \hat{-4} = \widehat{-64} = \hat{2}$.

Does this work out in general? Well, (working in $\mathbb{Z}_n$) if $\hat{x} = \hat{x}'$ and $\hat{y} = \hat{y}'$, then $x \equiv x' \pmod{n}$ and $y \equiv y' \pmod{n}$, so by Lemma 2.4(i), $x + y \equiv x' + y' \pmod{n}$, and $xy \equiv x'y' \pmod{n}$, so $\widehat{x+y} = x' \hat{+} y'$, and $\widehat{xy} = \widehat{x'y'}$, as needed.

(ii) Allenby writes $\oplus$ for $+$, and $\odot$ for $\times$, in $\mathbb{Z}_n$. So he defines $\hat{x} \oplus \hat{y}$, $\hat{x} \odot \hat{y}$. This is useful to remind you that it is not the usual $+, \times$. I will not do this. When you see $+$, think: is this usual addition, or addition in $\mathbb{Z}_n$, or vector or matrix addition, or what?

(iii) The theory of congruences which we have developed can be viewed as a theory about *equations* in $\mathbb{Z}_n$. For example,

(a) to solve $3x \equiv 1 \pmod{5}$ for $x \in \mathbb{Z}$ is *equivalent* to solving $\hat{3}y = \hat{1}$ in $\mathbb{Z}_5$. The solution for the latter is $y = \hat{2}$, so the original congruence had general solution $x \equiv 2 \pmod{5}$.

The equation $40x \equiv 12 \pmod{28}$ (Example 2.9(i)) had general solution $x \equiv 1, 8, 15, 22 \quad$ (m In $\mathbb{Z}_{28}$ this congruence becomes the equation $\hat{12}x = \hat{12}$, which has the four solutions $x = \hat{1}, \hat{8}, \hat{15}, \hat{22}$.

(b) Fermat's Little Theorem (Theorem 2.10) says: if p is prime then $y^p = y$ for all $y \in \mathbb{Z}_p$.

(c) Wilson's Theorem (Theorem 2.15) says that if p is prime then the product of the non-zero elements of $\mathbb{Z}_p$ is $\hat{-1}$.

4. Rings

We have $+, \times$ on $\mathbb{Z}$ (or on $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$) and also (with a new definition), we have these operations on the finite set $\mathbb{Z}_n$. Each gives a rule for obtaining a third element from two elements.

Definition 4.1. Given a set X, a function which associates to each pair of elements of X (equal or distinct) another element of X is called a *binary operation on* X.

For example, if $X = \mathbb{R}$, then $+, \times$ are binary operations, as are $f_1(x, y) = 2x + 3y$ and $f_2(x, y) = x^2 + \mathrm{Cos} y$. But
$f_3(x, y) = (x + y)^{\frac{1}{2}}$ is not a binary operation on $\mathbb{R}$, since $f_3(0, -1) \notin \mathbb{R}$.

We are only interested in binary operations with (unlike f_1, f_2 above) 'nice' properties, i.e. satisfying certain axioms.

Definition 4.2. A *ring* is a set R with two binary operations $+, \times$ satisfying the following axioms.

(A1) (Associativity of $+$): $a + (b + c) = (a + b) + c$ for all $a, b, c \in R$;

(A2) (Commutativity of $+$): $a + b = b + a$ for all $a, b \in R$.

(A3) (Existence of additive identity): There is an element of R, denoted by 0 or 0_R, such that $a + 0 = a$ for all $a \in R$.

(A4) (Existence of additive inverse): For every $a \in R$ there is an element $-a \in R$ with $a + (-a) = 0$.

(So far, the axioms say that $(R, +)$ is an 'abelian group).)

(M1) (Associativity of multiplication): $a \times (b \times c) = (a \times b) \times c$ for all $a, b, c \in R$.

(D) (Distributivity): $a \times (b + c) = a \times b + a \times c$ and $(a + b) \times c = a \times c + b \times c$ for all $a, b, c \in R$.

Note: We write $a.b$ or ab for $a \times b$. We write $a - b$ for $a + (-b)$.

Example 4.3. (1) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all rings (with the usual operations).

(2) $\mathbb{Z}_n$ is a ring. The zero is $0_{\mathbb{Z}_n} = \hat{0}$. The additive inverse of $a = \hat{x}$ is $-a = \widehat{-x}$. The axioms otherwise follow from the corresponding axioms for $\mathbb{Z}$, e.g.

$$\hat{x} + (\hat{y} + \hat{z}) = \hat{x} + \widehat{y + z} = x + \widehat{(y + z)} = \widehat{(x + y) + z} = \widehat{x + y} + \hat{z} = (\hat{x} + \hat{y}) + \hat{z}.$$

(3) The set $M_2(\mathbb{Q})$ of all 2×2 matrices with entries in $\mathbb{Q}$ is a ring under matrix addition and multiplication. In fact, if R is any ring and n is any positive integer, the set $M_n(R)$ of all $n \times n$ matrices with entries in R is a ring.

(4) If R, S are rings, then their Cartesian product $R \times S = \{(r, s) : r \in R, s \in S\}$ is a ring. The ring operations are given by:

$$(r, s) + (r', s') = (r + r', s + s')$$

$$(r, s) \cdot (r', s') = (rr', ss').$$

The zero is $(0, 0)$ and $-(r, s)$ is $(-r, -s)$.

We shall prove some elementary facts from the axioms for rings. Note that this saves us work: because these facts follow from the axioms, they are true for *all* rings, and we don't have to check them separately in each ring.

Also, by seeing properties that are common to different rings, we gain insights about them.

Lemma 4.4. *Let* R *be a ring, and* $a, b, c \in R$.
(i) $(a - b) + b = a$.
(ii) If $a + c = b + c$ *then* $a = b$ *(cancellation law for* $+$*).*
(iii) $a.0 = 0.a = 0$ *for all* a.
(iv) $(-a)b = a(-b) = -(ab)$.
(v) $-(a + b) = (-a) + (-b)$.
(vi) $-(-a) = a$.
(vii) The additive inverse of a *is unique.*

Proof. (i) $(a - b) + b = (a + (-b)) + b =_{A1} a + ((-b) + b) =_{A4} a + 0 =_{A3} a$.
(ii) $a =_{A3} a + 0 =_{A4} a + (c + (-c)) =_{A1} (a + c) + (-c) = (b + c) + (-c) =_{A1} b + (c + (-c)) =_{A4} b + 0 =_{A3} b$.
(iii) $a.0 + a.0 =_D a(0 + 0) =_{A3} a.0 =_{A3} a.0 + 0 =_{A2} 0 + a.0$, so by (ii), $a.0 = 0$.
(iv) $ab + (-a)b =_D (a - a)b = 0b =_{(iii)} 0 = ab - (ab)$ so by cancellation (ii) and (A2), $(-a)b = -(ab)$. Similarly $a(-b) = -(ab)$.
(v) $a + b + (-(a + b)) = 0 = a + b + (-a) + (-b)$. Now cancel.
(vi) $(-a) + (-(-a)) = 0 = (-a) + a$. Now cancel.
(vii) If $a + (-a) = a + b = 0$, then $b = -a$ by cancellation. □

Definition 4.5. (i) A ring R is *commutative* if $ab = ba$ for all $a, b \in R$. So $\mathbb{Z}$ and $\mathbb{Z}_n$ are commutative, but $M_2(\mathbb{R})$ is not (matrix multiplication is not commutative).

(ii) A ring R *has a 1* if there is a *multiplicative identity*, that is, an element in R denoted 1 or 1_R with $a.1 = 1.a = a$ for all $a \in R$, and satisfying $1 \neq 0$.

For example, $\mathbb{Z}$ has a 1, and $\mathbb{Z}_n$ has a 1, namely $\hat{1}$. The ring $2\mathbb{Z}$ of even integers does *not* have a 1. The ring $\hat{2}\mathbb{Z}_{14} = \{\hat{0}, \hat{2}, \hat{4}, \hat{6}, \hat{8}, \hat{10}, \hat{12}\}$ (with operations modulo 14) has a one, namely $\hat{8}$.

Nearly all rings in this module are commutative rings with a 1.

(iii) Let R be a ring with a 1. An element $x \in R$ is *invertible* or is a *unit* if there is an element $x^{-1} \in R$ with $xx^{-1} = x^{-1}x = 1$.

Note: A matrix A in $M_2(\mathbb{R})$ is a unit if and only if $\det A \neq 0$.

(iv) If the ring R is commutative, then an element $a \in R$ is a *zero-divisor* of R if $a \neq 0$ and there is $b \in R$ with $b \neq 0$ such that $ab = 0$.

(v) A ring is an *integral domain* if it is commutative, has a 1, and if it has no zero-divisors.

So, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{Z}_5$ are integral domains, but $\mathbb{Z}_6$ is not, as $\hat{2} \times \hat{3} = \hat{0}$.

(vi) A ring R is a *field* if it is commutative, has a 1, and every non-zero element of R is invertible.

Thus, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields, but $\mathbb{Z}$ is not a field, as 2 is not invertible ($\frac{1}{2} \notin \mathbb{Z}$).

Lemma 4.6. *Every field* F *is an integral domain.*

Proof. We must check that there are no zero-divisors. Suppose $a, b \in F$ with $ab = 0$ and $a \neq 0$. Then a^{-1} exists, so $b = 1.b = (a^{-1}a)b = a^{-1}(ab) = a^{-1}0 = 0$, so $b = 0$. □

Theorem 4.7. *Every* finite *integral domain is a field.*

Proof. Let R be a finite integral domain, say $R = \{r_1, \dots, r_n\}$. Let a be a non-zero element of R, so a is one of the r_i. Consider the elements $ar_1, ar_2, \dots, ar_n$. These are all different, for if $ar_i = ar_j$ then $ar_i - ar_j = 0$, so $a(r_i - r_j) = 0$, so $r_i - r_j = 0$ (as $a \neq 0$), so $r_i = r_j$.

This gives n distinct elements of R, which has size n, so every element of R occurs in this list. In particular, the element 1 occurs in the list, that is, $ar_l = 1$ for some l. Then r_l is the multiplicative inverse of a. So R is a field. $\square$

Theorem 4.8. *Let $n \in \mathbb{Z}$.*
(i) $\hat{a}$ is a unit of $\mathbb{Z}_n$ if and only if $(a, n) = 1$.
(ii) $\mathbb{Z}_n$ is a field if and only if n is a prime (of $\mathbb{Z}$).

Proof. (i) $\hat{a}$ has an inverse if and only if the congruence $ax \equiv 1 \pmod n$ has a solution (for the inverse will then be $\hat{x}$). Thus, the result follows from Remark 2.8. (Explicitly: suppose $\hat{a}$ is a unit, say $\hat{a}\hat{b} = \hat{1}$. Then $ab = 1 + nk$ for some k. If $(a, n) = d$ then $d | ab - nk = 1$, so $d = 1$. Conversely, suppose $(a, n) = 1$. Then by Euclid's Algorithm there are $x, y \in \mathbb{Z}$ with $ax + ny = 1$, and then $\hat{a}\hat{x} = \hat{1}$.)

(ii) Clearly, $\mathbb{Z}_n$ is a commutative ring with a 1. So $\mathbb{Z}_n$ is a field
$\Leftrightarrow$ $\hat{a} = 0$ or $\hat{a}$ is a unit for all $a \in \mathbb{Z}$
$\Leftrightarrow$ $n | a$ or $(a, n) = 1$ for all $a \in \mathbb{Z}$
$\Leftrightarrow$ n is prime. $\square$

4.1. **Subrings.** Recall from linear algebra the notion of *subspace* of $\mathbb{R}^n$ (or of any vector space). This is just a non-empty subset of $\mathbb{R}^n$ which inherits the algebraic structure of $\mathbb{R}^n$ (is a vector space). We look at similar ideas for rings.

Definition 4.9. Let R be a ring, and S a non-empty subset of R. Then S is a *subring* of R if
(i) S is closed under $+, \times$; that is, $a, b \in S \Rightarrow a + b, ab \in S$.
(ii) S is a ring with the same operations.

There is a similar definition of 'subfield' (just replace the word 'ring' by 'field' everywhere).

Remark 4.10. If S is a subring of R, then S and R have the same 0, i.e., $0_S = 0_R$. For
$0_R + a = a + 0_R = a$ for all $a \in R$, and
$0_S + b = b + 0_S = b$ for all $b \in S$. Putting $a = 0_S = b$, we find $0_S + 0_R = 0_S$ and $0_S + 0_S = 0_S$, so $0_S + 0_R = 0_S + 0_S$, so $0_R = 0_S$ by cancellation.

Likewise, if $a \in R$ then $(-a)_S = (-a)_R$ (for if b is an additive inverse of a in S, then $a + b = 0$ and $a + (-a) = 0$, so $b = -a$.

Lemma 4.11 (Subring Test)**.** *If R is a ring, and S is a subset of R, then S is a subring of R if and only if all the following hold.*
(i) if $a, b \in S$ then $a + b \in S$ and $ab \in S$,
(ii) $0 \in S$,
(iii) if $a \in S$ then $-a \in S$.

Note: It is a lot easier to check conditions (i)-(iii) than all the ring axioms. There is a similar way of checking a subset of a vector space is a subspace.

Proof. If (i)–(iii) hold, then by (i), $+, \times$ are binary operations on S. Condition (ii) gives the additive identity, and (iii) gives additive inverses. The other axioms for rings are inherited from R. So S is a subring of R.

Conversely, if S is a subring then (i) holds as $+, \times$ are binary operations on S. Also, S must have an additive identity, and by Remark 4.10, this is 0_R. Every $a \in S$ must have an additive inverse, and by 4.10 again, this is $-a$. $\square$

Note: Any subring (with a 1) of a field is an integral domain.

Example 4.12. (i) $\mathbb{Z}$ is a subring of the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.
(ii) $2\mathbb{Z}$ is a subring of $\mathbb{Z}$ (but has no 1).
(iii) $\mathbb{Z}_3$ is NOT a subring of $\mathbb{Z}$, as its set of elements is not a *subset* of $\mathbb{Z}$.
(iv) The set S of even integers modulo 14 is a subring of $R = \mathbb{Z}_{14}$. But note that $1_R = \hat{1}$, whilst $1_S = \hat{8}$, as $\hat{8} \times \widehat{2a} = \widehat{16a} = \widehat{2a}$. So the multiplicative version of Remark 4.10 can fail.

The next example will be very important later.

Example 4.13. Let $d \in \mathbb{Z}$, d not a square. Define $\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}$. Then $\mathbb{Z}[\sqrt{d}]$ is a subring of the field $\mathbb{C}$, and contains 1, so is an integral domain. Also, $\mathbb{Z}$ is a subring of $\mathbb{Z}[\sqrt{d}]$. In the special case when $d = -1$, $\mathbb{Z}[\sqrt{d}] = \mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$, the ring of *Gaussian integers*.

4.2. **Polynomial rings.** If R is a ring, let $R[X]$ be the collection of all polynomials in X with coefficients in R, i.e., $R[X]$ is the set of all expressions of the form $a_0 + a_1X + a_2X^2 + \dots a_nX^n$ where $a_0, \dots, a_n \in R$. Then $R[X]$ is the *polynomial ring* in variable X with coefficients in R (or 'over R').

So $2 + 3X^2 + 5X^4 \in \mathbb{Z}[X]$, and $2 + 3X^2 + \frac{1}{2}X^4 \in \mathbb{Q}[X]$ but is not in $\mathbb{Z}[X]$.

We use the usual rules for addition and multiplication: for example,

$$(2 + 3X^2 + X^3) + (1 + 2X^2 + X^4) = 3 + 5X^2 + X^3 + X^4, \quad \text{and}$$

$$(2 + 3X^2 + X^3)(1 + 2X^2) = 2 + 7X^2 + X^3 + 6X^4 + 2X^5.$$

Formally, if $f(X) = \Sigma_{i=0}^n a_iX^i$ and $g(X) = \Sigma_{i=0}^n b_iX^i$, then $f(X) + g(X) = \Sigma_{i=0}^n c_iX^i$ where $c_i = a_i + b_i$ for each i, and $f(X) \times g(X) = \Sigma_{i=0}^{2n} d_iX^i$, where, for each i, $d_i = \Sigma_{j=0}^i a_jb_{i-j}$.

In the notation, we often omit terms for which the coefficient is 0. If the coefficient of a power of X is 1 we omit the coefficient. Thus $3 + X^3$ denotes $3 + 0.X + 0.X^2 + 1.X^3$.

Two polynomials are *equal* if the corresponding coefficients are equal.

The *degree* of $f(X)$ is the largest exponent of X with non-zero coefficient – for example, $2 + 3X^2 + \frac{1}{2}X^{17}$ has degree 17, $3 + X$ has degree 1, and 3 has degree 0. We say the corresponding term is the *leading term* – so the leading terms of the above three polynomials are respectively $\frac{1}{2}X^{17}$, X, and 3. The polynomial $f(X)$ is *monic* if the coefficient of the leading term is 1.

Remark 4.14. (i) If R has a 1, then so does $R[X]$ (namely, the constant 1).
(ii) R is a subring of $R[X]$ (identify the element $r \in R$ with the constant polynomial $r.X^0$).
(iii) If S is a subring of R, then $S[X]$ is a subring of $R[X]$.

Lemma 4.15. *(i) If* R *is an integral domain, so is* $R[X]$.
(ii) If F *is a field, then the units of* $F[X]$ *are the non-zero constant polynomials.*

Proof. Suppose that R is an integral domain. We must show that $R[X]$ has no zero-divisors, so let $f(X), g(X)$ be non-zero elements of $R[X]$, say

$$f(X) = a_0 + a_1X + \ldots + a_mX^m,$$

$$g(X) = b_0 + b_1X + \ldots + b_nX^n \text{ with } a_m, b_n \neq 0.$$

Then $f(X)g(X) = \Sigma_{i=0}^{m+n} c_iX^i$, where, in particular, $c_{m+n} = a_mb_n$. As R is an integral domain, $c_{m+n} \neq 0$, so $f(X)g(X) \neq 0$, as required.

(ii) Suppose in (i) that $f(X)g(X) = 1$. Since 1 has degree 0, this forces that $f(X)g(X) = 0$, that is, $m + n = 0$, so $m = n = 0$. Thus $f(X) = a_0$ and $g(X) = 0$, so they are constant polynomials.

Conversely, if $f(X) = a$ is a non-zero constant polynomial, then $a \in F$, and as F is a field there is $b \in F$ with $ab = 1$, and then the *polynomial* b is an inverse of $f(X)$ in $F[X]$. □

5. Ideals

For the rest of the module, all rings will be assumed to be commutative, with a 1.

Definition 5.1. Let I be a non-empty subset of R. Then I is an *ideal* of R if
(i) if $a, b \in I$ then $a + b, -a \in I$, and
(ii) if $a \in I$ and $r \in R$ then $ar \in I$.

Note. (i) Every ideal of R is a subring of R. Apply the Subring Test Lemma 4.11, noting that as $I \neq \emptyset$, there is $a \in I$, so by (ii) of Definition 5.1, $a.0 = 0 \in I$.

(ii) However, not every subring is an ideal. The definition of 'ideal' is stronger, for in (ii) above, we allow r to be any element of R, not just any element of I.

Example 5.2. (1) If R is a ring, then $\{0_R\}$ and R are ideals of R.

(2) Let $d \in \mathbb{Z}$ and $[d] := \{dn : n \in \mathbb{Z}\}$ (the set of all multiples of d). Then $[d]$ is an ideal of $\mathbb{Z}$. For example, $[2]$ is the ideal of all *even* integers.

(3) Let $f \in \mathbb{Z}[X]$ (the ring of polynomials over X). Then $[f] := \{fg : g \in \mathbb{Z}[X]\}$ is an ideal of $\mathbb{Z}[X]$.

More generally,

(4) If R is any commutative ring and $a \in R$, then $[a] := \{ar : r \in R\}$ is an ideal of R. It is called the *principal ideal generated by* a.

Even more generally,

(5) Suppose R is a commutative ring, and $a_1, \ldots, a_m \in R$. Consider ideals containing $a_1, \ldots, a_m$. Any such ideal contains all elements of the form $r_1a_1 + \ldots + r_ma_m$ (where $r_i \in R$). Conversely, $\{r_1a_1 + \ldots + r_ma_m\}$ is an ideal of R. So it is the *smallest* ideal of R containing $a_1, \ldots, a_m$, and is denoted $[a_1, \ldots, a_m]$.

(6) Let $R = \mathbb{Z}[X]$ and

$$I = [2, X] = \{2f_1 + Xf_2 : f_1, f_2 \in \mathbb{Z}[X]\}.$$

Every element of I has even constant term, and conversely, any polynomial with even constant term has the form

$$2a_0 + a_1X + \ldots + a_nX^n = 2.a_0 + X(a_1 + a_2X + \ldots a_nX^{n-1}),$$

so lies in I.

However, I is not a principal ideal. For suppose $[2, X] = [f]$. Then as $2 \in [f]$, f is a constant polynomial, and as $X \in [f]$, X is a multiple of f, so $f = \pm 1$. But $1 \notin [2, X]$.

Theorem 5.3. *Every ideal in $\mathbb{Z}$ is principal.*

Proof. Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$, then $I = [0]$ so is principal. So we may assume $I \neq \{0\}$, so I contains a non-zero integer a, so I contains $-a = (-1) \times a$. Now one of $a, -a$ is positive, so I contains a (strictly) positive integer. Let d be the *smallest* positive integer in I.

Claim. $I = [d]$ (which by definition is $\{dn : n \in \mathbb{Z}\}$).

Proof of Claim. Clearly, $[d] \subseteq I$.

To show $I \subseteq [d]$, suppose $e \in I$. By the Division Algorithm (Theorem 1.14), we have $e = md + r$ for some $m, d \in \mathbb{Z}$ with $0 \leqslant r < d$. □

Now $e \in I$, and $md \in I$, so $r = e - md \in I$. By minimality of d, $r = 0$, so $e = md \in [d]$. □

Is there a similar result for some other rings? Let's try polynomial rings. We know that Theorem 5.3 cannot work for $\mathbb{Z}[X]$, by Example 5.2 (6). What about $\mathbb{Q}[X]$?

We first need a version of the Division Algorithm for polynomials.

Definition 5.4. The *degree* of a polynomial $f(X)$, denoted $\deg f(X)$, is the largest exponent e such that X^e has non-zero coefficient.

For example, $1 + 2X + 5X^2 - X^4$ has degree 4.

Theorem 5.5 (Division Algorithm for Polynomials)**.** *Let F be a field (e.g. $\mathbb{Q}$), and let $f(X), g(X) \in F[X]$, with $g(X) \neq 0$. Then there are $q(X), r(X) \in F[X]$ such that*

$$f(X) = q(X)g(X) + r(X)$$

and either

(i) $r(X) = 0$, *or*

(ii) $r(X) \neq 0$ *and* $\deg r(X) < \deg g(X)$.

Proof. See pp. 49-50 of the book by Allenby. □

In lectures, I will give an example (omitted here).

Theorem 5.6. *Let F be a field. Then every ideal of $F[X]$ is principal.*

Proof. We use the same proof as for $\mathbb{Z}$ (Theorem 5.3), now using the Division Algorithm for polynomials.

So let I be an ideal of $F[X]$, $I \neq \{0\}$. In the proof of 5.3, we chose a smallest positive element of I. This time, we choose a non-zero element of I of smallest degree, say f. (Note: as often, we write f for $f(X)$.)

Claim. $I = [f]$ (which equals $\{fg : g \in F[X]\}$).

Proof of Claim. Clearly, $[f] \subseteq I$, by the definition of 'ideal'.

To show $I \subseteq [f]$, suppose $h \in I$. By Theorem 5.5, there are $q, r \in F[X]$ such that $h = qf + r$, and either $r = 0$, or $r \neq 0$ and $\deg r < \deg f$. □

Now $r = h - qf \in I$, as $h, f \in I$. So by the minimality of $\deg(f)$, $r = 0$, so $h = qf \in [f]$. Thus, $I = [f]$.

□

Definition 5.7. We say that the ring R is a *principal ideal domain* (PID) if it is an integral domain, and every ideal of R is principal.

Note: By Theorems 5.3 and 5.6, $\mathbb{Z}$ and $F[X]$ (for F a field) are PIDs. However, $\mathbb{Z}[X]$ is not (see Example 5.2(6)).

Exercise. If F is a field, then the only ideals of F are $\{0\}$ and F. These have the form $[0]$ and $[1]$ respectively, so F is a PID.

Example 5.8. Consider the ideal $[9, 15] = \{9s + 15t : s, t \in \mathbb{Z}\}$, in $\mathbb{Z}$. We wish to express it as $[d]$, as $\mathbb{Z}$ is a PID.

What should d be? Try $d = (9, 15)$ (the positive g.c.d.).

Then,

(i) $d|9$, $d|15$, so $9, 15 \in [d]$, so $[9, 15] \subset [d]$;

(ii) there are $s, t \in \mathbb{Z}$ with $9s + 15t = d$ (Euclid's Algorithm), so $d \in [9, 15]$, so $[d] \subseteq [9, 15]$.

Thus, by (i) and (ii), $[d] = [9, 15]$. Of course, here $d = 3$. But our argument shows that in general, if $a, b \in \mathbb{Z}$, not both zero, then $[a, b] = [(a, b)]$.

Example 5.9. We do a similar example, but for the ring $\mathbb{Q}[X]$ of polynomials in place of $\mathbb{Z}$. Consider the ideal $I = [X^2 + 4X + 3, X^3 - X^2 - 3X - 1]$ in $\mathbb{Q}[X]$. We wish to express I in the form $[f]$, i.e., (by the proof of Theorem 5.6) to find some nonzero $f \in I$ of smallest possible degree. As in Example 5.8, it will suffice to find some $f \in \mathbb{Q}[X]$ such that

(a) $f|X^2 + 4X + 3$ and $f|X^3 - X^2 - 3X - 1$ in $\mathbb{Q}[X]$, and

(b) there are polynomials $s, t \in \mathbb{Q}[X]$ with $f = s(X^2 + 4X + 3) + t(X^3 - X^2 - 3X - 1)$.

We do this using Euclid's Algorithm for polynomials, just as in $\mathbb{Z}$, but using the Division Algorithm for polynomials (Theorem 5.5).

Easy division of polynomials gives

$$X^3 - X^2 - 3X - 1 = (X^2 + 4X + 3)(X - 5) + (14X + 14)$$

$$X^2 + 4X + 3 = (14X + 14)(\frac{1}{14}X + \frac{3}{14}).$$

The last non-zero remainder is $14X + 14$, so this looks like a g.c.d. Indeed,

(i) $14X + 14$ divides $X^2 + 4X + 3$ and $X^3 - X^2 - 3X - 1$ in $\mathbb{Q}[X]$, and

(ii) $14X + 14 = (X^3 - X^2 - 3X - 1) - (X - 5)(X^2 + 4X + 3)$.

Thus, $I = [14X + 14] = [X + 1]$. Note here that if $q \in \mathbb{Q}$ is nonzero then $[f] = [qf]$. So we usually choose the generator of I to be *monic*, i.e. so that the coefficient of the highest power of X is 1.

(I chose this example so the calculation is short – usually it would be messier.)

The last example suggests that notions like g.c.d. work well in rings other than $\mathbb{Z}$, such as $\mathbb{Q}[X]$. We now explore this.

5.1. Divisibility in rings.

Definition 5.10. Let R be a commutative ring with a 1.
(i) If $a, b \in R$, we say a *divides* b, written $a|b$, if there is $c \in R$ with $ac = b$.
(ii) Given $a, b \in R$, not both zero, we say $d \in R$ is a *greatest common divisor* (g.c.d.) of a, b if
  (a) $d|a$ and $d|b$, and
  (b) for any $c \in R$, if $c|a$ and $c|b$ then $c|d$.
(iii) $u \in R$ is a *unit* if $u|1$.
(iv) If $a, b \in R$ then a is an *associate* of b if there is unit $u \in R$ with $a = ub$.
(v) $r \in R$ is *irreducible* if
  (a) $r \neq 0$ and r is not a unit, and
  (b) if $r = ab$ where $a, b \in R$, then a or b is a unit.
(vi) $r \in R$ is *prime* if
  (a) $r \neq 0$ and r is not a unit, and
  (b) if $r|ab$, where $a, b \in R$, then $r|a$ or $r|b$.
(vii) $a, b \in R$ are *coprime* if $1 = \text{g.c.d.}(a, b)$.

Example 5.11. (i) In $\mathbb{Q}[X]$, $X^2 + \frac{1}{3}X + \frac{1}{2} | X^3 + \frac{5}{6}X^2 + \frac{2}{3}X + \frac{1}{4}$, as

$$X^3 + \frac{5}{6}X^2 + \frac{2}{3}X + \frac{1}{4} = (X^2 + \frac{1}{3}X + \frac{1}{2})(X + \frac{1}{2}).$$

We have $2X + 4 | X + 2$ in $\mathbb{Q}[X]$ but not in $\mathbb{Z}[X]$.
In $\mathbb{Z}[i]$, $2 + i | 1 + 13i$, as $(2 + i)(3 + 5i) = 1 + 13i$.
(ii) In Example 5.9, we showed that in $\mathbb{Q}[X]$, the polynomials X^2+4X+3 and X^3-X^2-3X-1 have g.c.d. $X + 1$.
(iii) In $\mathbb{Q}[X]$ the units are the constant polynomials except for 0, so the associates of X^2+3 are the polynomials of the form $c(X^2 + 3)$ where $c \in \mathbb{Q}$ with $c \neq 0$.
In $\mathbb{Z}[X]$, the units are $1, -1$.
In $\mathbb{Z}[i]$, $1, -1, i, -i$ are units. Are there others? We will investigate in the next section.
(iv) In $\mathbb{Q}[X]$, 'irreducible' has the usual meaning for 'irreducible polynomial'. Are irreducibles in $\mathbb{Q}[X]$ the same as primes? Can we factorise every polynomial uniquely into irreducibles, as in Theorem 1.20?
(v) If d is a g.c.d. of a, b, then the other g.c.d.'s are exactly the *associates* of d. Why is this?
(vi) For which rings is there something like the Division Algorithm?

Lemma 5.12. *Let R be a commutative ring with a 1. Then the relation 'a is an associate of b' is an equivalence relation on R. The equivalence class of a is $\{au : u \text{ is a unit of } R\}$.*

Proof. (i) (Reflexivity) a is an associate of a as $a.1 = a$.
(ii) (Symmetry) Suppose a is an associate of b, so $a = ub$ for some unit u of R. There is $v \in R$ with $vu = uv = 1$, and v is also a unit. Then $va = vub = 1.b = b$, so b is an associate of a.
(iii) (Transitivity) Suppose a is an associate of b, and b is an associate of c, say $a = ub$ and $b = vc$ with u, v both units. Now uv is a unit, for if $uu' = 1$ and $vv' = 1$ then $(uv)(u'v') = 1$. Thus, as $a = (uv)c$, a is an associate of c.
The second assertion is direct from the definitions. $\square$

We now discuss divisibility in polynomial rings over fields.

Exercise 5.13. Show that if F is a field, then every non-zero element of $F[X]$ is an associate of a unique monic polynomial over F.

Theorem 5.14 (Factor Theorem)**.** *Let F be a field, $f(X) \in F[X]$, and $a \in F$. Then $X - a$ divides $f(X)$ if and only if $f(a) = 0$ (calculated in F).*

Proof. $\Rightarrow$ Suppose $(X - a)g(X) = f(X)$. Then substituting a for X, we find $(a - a)g(a) = f(a)$, so $f(a) = 0$.

$\Leftarrow$ Suppose $f(a) = 0$. We use the Division Algorithm for polynomials (Theorem 5.5) to write $f(X) = (X - a)g(X) + r(X)$, where $\deg r(X) < \deg(X - a) = 1$. So $0 = f(a) = (a - a)g(a) + r(a)$, so $r(a) = 0$. As $r(X)$ is constant, this forces $r(X) = 0$, so $X - a$ divides $f(X)$, $\square$

Corollary 5.15. *If F is a field, then every polynomial $f(X) \in F[X]$ of degree n has at most n roots. It has exactly n roots in F (counted with multiplicity) if, when you write $f(X)$ as a product of irreducible factors, all the irreducible factors have degree 1.*

Proof. For the first part, suppose the distinct roots of $f(X)$ are $a_1, \ldots, a_t$. Then $f(X) = (X - a_1)f_1(X)$ (for some $f_1(X) \in F[X]$), and $X - a_2$ divides $f(X)$, so $X - a_2$ divides $f_1(X)$ (why? Compare Sheet 2 Q1). So $f(X) = (X - a_1)(X - a_2)f_2(X)$, for some polynomial $f_2(X)$. Now $X - a_3$ divides $f(X)$ and so divides $f_2(X)$, and so on. So we find $f(X) = (X - a_1)(X - a_2)...(X - a_t)h(X)$ for some polynomial $h(X)$, and so f has degree at least t, so $t \leqslant n$.

The second assertion is an exercise. $\square$

Theorem 5.16 (Fundamental Theorem of Algebra)**.** *Every polynomial over $\mathbb{C}$ of degree at least 1 has a root in $\mathbb{C}$.*

Proof. Omitted – this really belongs to analysis, not algebra! $\square$

It follows from the last theorem that every polynomial over $\mathbb{C}$ can be written as a product of linear factors over $\mathbb{C}$. More precisely, we have

Corollary 5.17. *(i) The irreducible polynomials in $\mathbb{C}[X]$ are the linear polynomials $aX + b$,*

(ii) The irreducible polynomials in $\mathbb{R}[X]$ are the linear polynomials $aX + b$ and the quadratics $aX^2 + bX + c$ with no real root.

Proof. (i) Clearly any linear polynomial is irreducible. Conversely, if $f(X)$ is irreducible, then by the Fundamental Theorem of Algebra (5.16), it has a root $\alpha \in \mathbb{C}$. Then by the Factor Theorem (5.14), $f(X)$ has a linear factor $X - \alpha$. As $F(X)$ is irreducible, we have $f(X) = c(X - \alpha)$ for some $c \in \mathbb{C}$, so $f(X)$ is linear. (ii) Clearly the polynomials of the described types are irreducible. Conversely, suppose $f(X) \in \mathbb{R}[X]$ is irreducible of degree > 1. Then $f(X)$ has no root in $\mathbb{R}$. However, $f(X)$ has a root $\alpha = +iv \in \mathbb{C}$. Since $f(X)$ has real coefficients, $f(\bar{\alpha}) = \overline{f(\alpha)} = 0$, where $\bar{\alpha}$ denotes the complex conjugate of α (we use here that for complex numbers z, w, $\overline{z+} = \bar{z} + \bar{w}$ and $\overline{zw} = \bar{z}\bar{w}$). So $\bar{\alpha}$ is also a root of $f(X)$. Thus, $f(X)$ is divisible in $\mathbb{C}[X]$ by

$$(X - \alpha)(X - \bar{\alpha}) = (X - u - iv)(X - u + iv) = (X - u)^2 + v^2 = aX^2 + bX + c,$$

where $a = 1$, $b = -2u$ and $c = u^2 + v^2$ (all real). Thus, $f(X) = (aX^2 + bX + c)g(X)$ with $g(X) \in \mathbb{C}[X]$. Calculating $g(X)$ by polynomial division, we see that as $f(X)$, $aX^2 + bX + c \in$

$\mathbb{R}[X]$, also $g(X) \in \mathbb{R}[X]$. Thus, as $f(X)$ is irreducible in $\mathbb{R}[X]$, we find $f(X) = aX^2 + X + c$, so is quadratic. □

6. The rings $\mathbb{Z}[\sqrt{d}]$ and $\mathbb{Q}[\sqrt{d}]$

We consider $d \in \mathbb{Z}$ with $d \neq 0, 1$, and d *square-free*, that is, for any prime $p \in \mathbb{Z}$, $p^2 \nmid d$.

Lemma 6.1. *If $a, b \in \mathbb{Q}$ are not both zero then $a^2 - db^2 \neq 0$.*

Proof. Suppose $a^2 = db^2$. Multiplying out denominators, we may suppose $a, b \in \mathbb{Z}$. We may also suppose $(a, b) = 1$, by factoring out any common factors. Let $p|d$. Then $p|a^2$, so $p^2|a^2$. (Note here that if $a = p_1^{a_1} \dots p_r^{a_r}$, expressed as a product of distinct prime powers p_i,then $a^2 = p_1^{2a_1} \dots p_r^{2a_r}$so for each i, $p_i^2|a^2$.) Thus, $p^2|db^2$, so as $p^2 \nmid d$, $p|b^2$. Hence $p|b$, which contradicts the assumption that $(a, b) = 1$. □

Definition 6.2. Define $\mathbb{Q}[\sqrt{d}] := \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}$ and $\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}$.

It is really $\mathbb{Z}[\sqrt{d}]$ that we are interested in. But first note

Proposition 6.3. *$\mathbb{Q}[\sqrt{d}]$ is a field.*

Proof. We just need to check the existence of multiplicative inverses. But

$$(a + b\sqrt{d})^{-1} = \frac{1}{a + b\sqrt{d}} = \frac{a - b\sqrt{d}}{(a + b\sqrt{d})(a - b\sqrt{d})} = \frac{a}{a^2 - db^2} - \frac{b}{a^2 - db^2}\sqrt{d}.$$

Note here that $a^2 - db^2 \neq 0$, by Lemma 6.1. □

Definition 6.4. In $\mathbb{Z}[\sqrt{d}]$ (or in $\mathbb{Q}[\sqrt{d}]$) define the *norm* $N(\alpha)$ as follows. Let $\alpha = a+b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$. Then

$$N(\alpha) := |a^2 - db^2| = |(a + b\sqrt{d})(a - b\sqrt{d})|.$$

Example 6.5. If $d = -1$, then $\mathbb{Z}[\sqrt{d}] = \mathbb{Z}[i]$ and if $\alpha = a + bi$, then $N(\alpha) = |a^2 + b^2|$, the square of the modulus (in the sense of complex numbers).

Recall that modulus has nice properties, such as $|zw| = |z|.|w|$. Fortunately, this generalises.

Lemma 6.6. *Let $\alpha, \beta \in \mathbb{Z}[\sqrt{d}]$. Then*
(i) $N(\alpha\beta) = N(\alpha)N(\beta)$,
(ii) $N(\alpha) = 0 \Leftrightarrow \alpha = 0$,
(iii) $N(\alpha)$ is a non-negative integer.

Proof. (i) Let $\alpha = a + b\sqrt{d}$ and $\beta = a' + b'\sqrt{d}$. Then

$$\alpha\beta = (aa' + dbb') + (ab' + ba')\sqrt{d}, \text{ so}$$
$$N(\alpha\beta) = |(aa' + dbb')^2 - (ab' + ba')^2 d|$$
$$= |(a^2 - db^2)(a'^2 - db'^2)|$$
$$|a^2 - db^2|.|a'^2 - db'^2| = N(\alpha)N(\beta).$$

(ii) See Lemma 6.1.
(iii) This is immediate from the definition of 'norm'. □

The idea of norm is that, because of (i) in the last lemma, it reduces multiplicative questions in $\mathbb{Z}[\sqrt{d}]$ to similar but easier questions in $\mathbb{Z}$. Also, N behaves a bit like absolute value in $\mathbb{Z}$ or the degree of a polynomial in $\mathbb{Q}[X]$, so perhaps there is a version of the Division Algorithm for rings $\mathbb{Z}[\sqrt{d}]$.

We first give a nice consequence of Lemma 6.6.

Example 6.7. If $n, m \in \mathbb{Z}$ can be written in the form $a^2 + 5b^2$, so can nm.

Indeed, let $n = a^2 + 5b^2$, $m = e^2 + 5f^2$. Then, working in the ring $\mathbb{Z}[\sqrt{-5}]$, $n = N(a + b\sqrt{-5})$ and $m = N(e + f\sqrt{-5})$, so

$$nm = N((a + b\sqrt{-5})(e + f\sqrt{-5}))$$

$$= N((ae - 5bf) + (af + be\sqrt{-5}))$$

$$= (ae - 5bf)^2 + 5(af + be)^2.$$

Example 6.8. Write $10824 = 88 \times 123$ in the form $a^2 + 2b^2$, for $a, b \in \mathbb{Z}$.

We have $88 = 4^2 + 2 \times 6^2$ and $123 = 5^2 + 2 \times 7^2$. So, working in $\mathbb{Z}[\sqrt{-2}]$,

$$10824 = 88 \times 123 = N(4 + 6\sqrt{-2})N(5 + 7\sqrt{-2})$$

$$= N((4 + 6\sqrt{-2})(5 + 7\sqrt{-2}))$$

$$= N(-64 + 58\sqrt{-2})$$

$$= 64^2 + 2 \times 58^2.$$

We can find other expressions for 10824 in form $a^2 + 2b^2$. For also

$$10824 = N(4 - 6\sqrt{-2})N(5 + 7\sqrt{-2})$$

$$N((4 - 6\sqrt{-2})(5 + 7\sqrt{-2})$$

$$= N(104 - 2\sqrt{-2})$$

$$= 104^2 + 2 \times 2^2.$$

We now investigate units, a possible division algorithm, primes, irreducibles, and uniqueness of factorisation in the rings $\mathbb{Z}[\sqrt{d}]$. It turns out that the results depend crucially on the choice of d.

Lemma 6.9. *$\alpha \in \mathbb{Z}[\sqrt{d}]$ is a unit if and only if $N(\alpha) = 1$.*

Proof. If α is a unit then there is $\beta \in \mathbb{Z}[\sqrt{d}]$ with $\alpha\beta = 1$. Then $N(\alpha\beta) = N(\alpha)N(\beta) = N(1) = 1$. As $N(\alpha)$, $N(\beta)$ are non-negative integers, this forces $N(\alpha) = 1$.

Conversely, if $\alpha = a + b\sqrt{d}$ then, working in $\mathbb{Q}[\sqrt{d}]$, $\alpha^{-1} = \frac{a}{a^2 - db^2} - \frac{b}{a^2 - db^2}\sqrt{d}$ (see the proof of Proposition 6.3). By our assumption, $a^2 - db^2 = \pm 1$, so $\alpha^{-1} \in \mathbb{Z}[\sqrt{d}]$. □

Theorem 6.10. *In the rings $\mathbb{Z}[\sqrt{d}]$, the units are:*

(i) $1, -1, i, -i$ *if* $d = -1$,

(ii) $1, -1$ *if* $d < -1$,

(ii) $1, -1$ *and infinitely many others, if* $d > 1$.

Proof. (i) We must solve $N(a+b\sqrt{-1}) = a^2+b^2 = 1$ in $\mathbb{Z}$. The only solutions are $(a,b) = (1,0),(-1,0),(0,1),(0,-1)$, so $\alpha = 1,-1,i,-i$, respectively.

(ii) We must here solve (in $\mathbb{Z}$) $a^2 - db^2 = 1$, where $d < -1$. This forces $a = 1$ and $b = 0$.

(iii) I do not give a general proof, but suppose for example $d = 3$. Now $2^2 - 3 \times 1^2 = 1$, so $(2+\sqrt{3})(2-\sqrt{3}) = 1$. Hence $2+\sqrt{3}$ and $2-\sqrt{3}$ are units. But also, for every integer $n > 1$, $(2+\sqrt{3})^n$ is a unit, as

$$(2+\sqrt{3})^n(2-\sqrt{3})^n = [(2+\sqrt{3})(2-\sqrt{3})]^n = 1^n = 1.$$

Clearly the real numbers $(2+\sqrt{3})^n$ are all distinct, as n varies. □

We now show that in *all* the rings $\mathbb{Z}[\sqrt{d}]$, all primes are irreducible. In the other direction, all irreducibles are primes only for special values of d.

Recall that an *integral domain* is a commutative ring with a 1 and with no zero-divisors: that is, if $ab = 0$ then $a = 0$ or $b = 0$ (unlike say the elements $\hat{2}, \hat{3}$ in $\mathbb{Z}_6$.) The proof below is basically the same as the proof of the corresponding direction of Theorem 1.20.

Theorem 6.11. *Let D be an integral domain. Then every prime of D is irreducible.*

Proof. Suppose that p is prime and $p = ab$, where $a,b \in D$. We must show that one of a,b is a unit. Now $p|ab$ (as $p.1 = ab$), so by the definition of 'prime', $p|a$ or $p|b$. We suppose $p|a$ (the other case is similar). Then $a = pc$ for some $c \in D$. So

$$p = ab = pcb = p.1,$$

so $p(cb-1) = 0$. As D is an integral domain it has no zero divisors, so as $p \neq 0$, this forces $cb - 1 = 0$. Hence $cb = 1$, so b is a unit. □

The (partial) converse, which we now give, requires a stronger assumption.

Theorem 6.12. *Let R be a principal ideal domain. Then every irreducible of R is prime.*

Proof. Suppose that r is irreducible, and $r|ab$, where $a,b \in R$. We must show $r|a$ or $r|b$, so suppose $r \not| a$. As R is a PID, the ideal $[r,a]$ of R is a principal ideal, so has the form $[d]$ for some $d \in R$. Now $d|r$, so there is $m \in R$ with $dm = r$, and as r is irreducible, one of d,m is a unit.

Claim. d is a unit.

Proof of Claim. If m is a unit, then m^{-1} exists, so $d = rm^{-1}$. Then as $d|a$ we get $r|a$, a contradiction. Hence d is a unit. □

Now choose e with $de = 1$. As $[r,a] = [d]$, there are $s,t \in R$ with $rs + at = d$. Then

$$rse + ate = de = 1, \text{ so}$$

$$rseb + ateb = b.$$

Hence, as $r|ab$ we have $r|rseb + ateb$, so $r|b$. □

Theorem 6.13 (Division Algorithm). *Let $\alpha, \beta \in \mathbb{Z}[\sqrt{-2}]$. Then there are $m, r \in \mathbb{Z}[\sqrt{-2}]$ with $\alpha = m\beta + r$ and $0 \leqslant N(r) < N(\beta)$.*

Remark. The proof below also works for $d = 2, 3, -1$.

Corollary 6.14. *(i)* $\mathbb{Z}[\sqrt{d}]$ *is a PID for* $d = -2, -1, 2, 3$.
(ii) For $d = -2, -1, 2, 3$, *'prime'='irreducible'.*
(iii) In any ring $\mathbb{Z}[\sqrt{d}]$, *every prime is irreducible.*

Proof. (i) Apply the method of proof of Theorems 5.3 and 5.6.
(ii) Use (i) and Theorem 6.12.
(iii) This is direct from Theorem 6.11. □

Proof of Theorem 6.13. Let $\alpha = a + b\sqrt{-2}$ and $\beta = c + d\sqrt{-2}$. Then

$$\frac{\alpha}{\beta} = \frac{a + b\sqrt{-2}}{c + d\sqrt{-2}} = \frac{(a + b\sqrt{-2})(c - d\sqrt{-2})}{(c + d\sqrt{-2})(c - d\sqrt{-2})} = \frac{x}{c^2 + 2d^2} \text{ for some } x,$$
$$= u + v\sqrt{-2} \text{ where } u, v \in \mathbb{Q}.$$

Let U be the nearest integer to u, so $|u - U| \leqslant \frac{1}{2}$, and let V be the nearest integer to v, so $|V - v| \leqslant \frac{1}{2}$. Then

$$\frac{\alpha}{\beta} = u + v\sqrt{-2} = (U + V\sqrt{-2}) + [(u - U) + (v - V)\sqrt{-2}].$$

Multiplying out by β, we get

$$\alpha = (U + V\sqrt{-2})\beta + [(u - U) + (v - V)\sqrt{-2}]\beta,$$
$$= m\beta + r \text{ where } m = U + V\sqrt{-2} \text{ and } r = [(u - U) + (v - V)\sqrt{-2}]\beta.$$

As $r = \alpha - m\beta$ and $m \in \mathbb{Z}[\sqrt{-2}]$, we get $r \in \mathbb{Z}[\sqrt{-2}]$. Now

$$N(r) = N(\beta)N[(u - U) + (v - V)\sqrt{-2}]$$
$$= N(\beta)[(u - U)^2 + 2(v - V)^2]$$
$$\leqslant N(\beta)[(\frac{1}{2})^2 + 2(\frac{1}{2})^2] = \frac{3}{4}N(\beta) < N(\beta).$$

□

The above proof wouldn't work for $\mathbb{Z}[\sqrt{-3}]$, as $(\frac{1}{2})^2 + 3(\frac{1}{2})^2 = 1$. The proof also works for $d = \sqrt{2}, \sqrt{3}, \sqrt{-1}$.

Definition 6.15. A *unique factorisation domain* (UFD) is an integral domain R such that
(i) every element not equal to 0 or a unit is a product of irreducibles, and
(ii) if $x = p_1 \dots p_n = q_1 \dots q_m$ where the p_i and q_j are irreducible, then $n = m$ and the p_i and q_j can be paired so that corresponding pairs are associates of each other.

Remark. $\mathbb{Z}$ is a UFD, by Theorem 1.22. Infact

Theorem 6.16. *Every PID is a UFD.*

Proof. (i) *Existence of factorisation into irreducibles.* Let D be a principal ideal domain, and $a \in D$ be a non-zero non-unit. Suppose for a contradiction that a is not a product of irreducibles. Then a is not irreducible, so $a = a_1b_1$ say, and a_1, b_1 are non-units. By assumption, one of a_1, b_1, say a_1, is not a product of irreducibles, so there are non-units a_2, b_2 such that $a_1 = a_2b_2$. One of a_2, b_2 is not a product of irreducibles (otherwise a_1 is, a contradiction), so there are non-units a_3, b_3 such that $a_2 = a_3b_3$. We continue in this way forever. We always find $a_{i+1}|a_i$, so we obtain a sequence of ideals

$$[a] \subseteq [a_1] \subseteq [a_2] \subseteq [a_3] \subseteq \dots.$$

The union of this sequence of ideals of D is again an ideal of D (check this!), denoted I, say, and since D is a principal ideal domain, $I = [d]$ for some $d \in D$. Now $d \in I$, so $d \in [a_j]$ for some j. Thus, $[d] \subseteq [a_j] \subseteq [a_{j+1}] \subseteq [d]$, so $[a_j] = [a_{j+1}]$. Hence $a_{j+1} \in [a_j]$, so there is $c \in D$ such that $a_{j+1} = a_j c$. But $a_j = a_{j+1}b_{j+1}$, so $a_{j+1} = a_{j+1}b_{j+1}c$. Since D is an integral domain, it follows that b_{j+1} is a unit, a contradiction.

(ii) *Uniqueness of factorisation into irreducibles.* This is almost exactly as in the proof of Theorem 1.22 (using that primes are the same as irreducibles, which holds by Theorems 6.11 and 6.12).

□

Example 6.17. We work in $\mathbb{Z}[\sqrt{-7}]$. We have

$$8 = 2 \times 2 \times 2 = (1+\sqrt{-7})(1-\sqrt{-7}).$$

Claim. Each of $2, 1+\sqrt{-7}, 1-\sqrt{-7}$ is irreducible in $\mathbb{Z}[\sqrt{-7}]$.

Proof of Claim. By Theorem 6.10(ii), they are not units. We'll show $1+\sqrt{-7}$ is irreducible using the norm, which reduces questions about $\mathbb{Z}[\sqrt{-7}]$ to questions about $\mathbb{Z}$. (The proofs for 2 and $1-\sqrt{-7}$ are similar.) So suppose $\alpha = 1+\sqrt{-7} = \beta\gamma$, where $\beta = a + b\sqrt{-7}$ and $\gamma = e + f\sqrt{-7}$. Then

$$N(\alpha) = 8 = N(\beta)N(\gamma) = (a^2+7b^2)(e^2+7f^2), \text{ an equation in } \mathbb{Z}.$$

The only such factorisations are $8 = 8\times 1, 4\times 2, 2\times 4, 1\times 8$. We cannot solve $a^2+7b^2 = 2$ in $\mathbb{Z}$, so we must have $a^2+7b^2 = 1$ (so $a = \pm 1, b = 0$ and β is a unit) or $e^2+7f^2 = 1$ (so γ is a unit).

However, 2 is not prime in $\mathbb{Z}[\sqrt{-7}]$. For $2|8 = (1+\sqrt{-7})(1-\sqrt{-7})$, but $2 \not| 1+\sqrt{-7}$ and $2 \not| 1-\sqrt{-7}$. Indeed, suppose that $2(p+q\sqrt{-7}) = 1+\sqrt{-7}$. Then we have $2p = 1$, so $p \notin \mathbb{Z}$. □

Similar arguments show that $1+\sqrt{-7}$ and $1-\sqrt{-7}$ are not prime. So in the ring $\mathbb{Z}[\sqrt{-7}]$, every prime is irreducible (by Theorem 6.11), but some irreducibles are not prime. Also, $\mathbb{Z}[\sqrt{-7}]$ is not a UFD, since we have two essentially different factorisations of 8 into irreducibles. Hence, by Theorem 6.16, $\mathbb{Z}[\sqrt{-7}]$ is not a PID.

Example 6.18. We work in $\mathbb{Z}[i]$ so $d = -1$. The units are $\pm 1, \pm i$, by Theorem 6.10(i). As noted after the proof of Theorem 5.5, $\mathbb{Z}[i]$ has a Division Algorithm, so it is also a UFD.

(i) Find the g.c.d.in $\mathbb{Z}[i]$ of $\alpha = 5+8i$ and $\beta = 3+5i$. We use the Division Algorithm (which holds in $\mathbb{Z}[i]$ by Theorem 6.13) and Euclid's Algorithm. (Actually, some of the equations below could be written down directly, without recourse to Euclid's Algorithm, but we aim to illustrate the general technique.)

First,

$$\frac{\alpha}{\beta} = \frac{5+8i}{3+5i} = \frac{(5+8i)(3-5i)}{(3+5i)(3-5i)} = \frac{55-i}{34} = 1 + \frac{21-i}{34}. \text{ So}$$

$\alpha = \beta + r_1$ where $r_1 = \frac{(21-i)\beta}{34} = 2+3i$ (note that r *must* lie in $\mathbb{Z}[i]$). Next, divide r_1 into β. We find

$$\frac{\beta}{r_1} = \frac{3+5i}{2+3i} = \frac{(3+5i)(2-3i)}{(2+3i)(2-3i)} = \frac{21+i}{13} = 1 + \frac{8+i}{13}. \text{ Multiplying out,}$$

$$(3+5i) = (2+3i) + \frac{(8+i)(2+3i)}{13} = (2+3i) + r_2 \text{ where } r_2 = 1+2i.$$

Now

$$\frac{r_1}{r_2} = \frac{2+3i}{1+2i} = \frac{8-i}{5} = 1 + \frac{3-i}{5}, \text{ so multiplying out}$$

$$(2+3i) = (1+2i) + \frac{(3-i)(1+2i)}{5} = (1+2i) + r_3 \text{ where } r_3 = 1+i.$$

Similarly, $(1+2i) = (1+i) + i$ (so $r_4 = i$), and $(1+i) = i(1-i)$, so $r_5 = 0$. Thus, the last non-zero remainder is i, the g.c.d..

In Euclid's Algorithm above, we obtained the equations

$$(5+8i) = (3+5i) + (2+3i)$$
$$(3+5i) = (2+3i) + (1+2i)$$
$$(2+3i) = (1+2i) + (1+i)$$
$$(1+2i) = (1+i) + i.$$

Going back up these equations, we find

$$i = (1+2i) - (1+i) = (1+2i) - [(2+3i) - (1+2i)] = -(2+3i) + 2(1+2i)$$
$$= -(2+3i) + 2[(3+5i) - (2+3i)] = 2(3+5i) - 3(2+3i) = 2(3+5i) - 3[(5+8i) - (3+5i)]$$
$$= 5(3+5i) - 3(5+8i).$$

Thus, we have expressed the g.c.d. i in the form $i = s(3+5i) + t(5+8i)$ where $s = 5$ and $t = -3$. Of course, the associates of i, namely $1, -1, -i$, are also g.c.d.'s of $5+8i$, $3+5i$. In particular, $5+8i$ and $3+5i$ are coprime.

We have $1 = -5i(3+5i) + 3i(5+8i)$.

(ii) Factorise 20 into primes of $\mathbb{Z}[i]$ (remember that in $\mathbb{Z}[i]$, primes are the same as irreducibles).

We have $20 = 2 \times 2 \times 5$, and

$$2 = (1+i)(1-i)$$
$$5 = (2+i)(2-i).$$
$$\text{So, } 20 = (1+i)^2(1-i)^2(2+i)(2-i).$$

Are these irreducible? Well, suppose $1+i = \alpha\beta$, a factorisation in $\mathbb{Z}[i]$. Taking norms, $2 = N(\alpha)N(\beta)$, so either $N(\alpha) = 1$ (when α is a unit, by 6.9), or $N(\beta) = 1$, and β is a unit. Thus, $1+i$, and similarly $1-i$, are irreducible.

Likewise, $2+i$ is irreducible, as $N(2+i) = 5$ is an irreducible of $\mathbb{Z}$.

Another factorisation of 20 into irreducibles is

$$20 = (1+i)^3(-1+i)(-1+2i)(2-i).$$

These factorisations are not really different, as they can be matched into associate pairs. For $(-1+i) = i(1+i)$, $(1+i) = i(1-i)$, and $(-1+2i) = i(2+i)$. So though we appear to have two different factorisations of 20 in $\mathbb{Z}[i]$, this does not contradict the fact that $\mathbb{Z}[i]$ is a UFD.