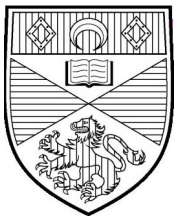
MT4517 Rings & Fields

J. D. Mitchell & V. Maltcev

with some changes by M. Neunhöffer

2012



School of Mathematics and Statistics
MT4517 Rings & Fields
Part 1 - Introduction

A brief introduction

A *ring* is a set together with two binary operations $+$ and $\cdot$ satisfying various natural axioms.

The 'prototype' example is the set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ with the usual arithmetic. The fact that this example on its own yields the whole of 'Number Theory' shows what a rich structure rings can have.

In fact, many of the 'usual' examples where one can 'add' or 'multiply' give us rings. For example: the integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$, real valued functions and so on...

However, starting with the axioms and looking for examples of things that satisfied them is not the way rings first came into mathematics.

Some historical background

In about 1630 Fermat was reading a recently published translation of Arithmetica by Diophantus of Alexandria. He was making notes in the margin and at one point he entered:

> *To divide a cube into two other cubes, a fourth power or in general any power whatever into two powers of the same denomination above the second is impossible, and I have assuredly found an admirable proof of this, but the margin is too narrow to contain it.*

That is, if $n > 2$ there are no integer solutions x, y and z of the equation $x^n + y^n = z^n$. This is the famous Fermat's Last Theorem which resisted all attempts to prove it until recently. Investigations of this result led to much interesting mathematics, including some of the first systematic investigations of ring theory.

Fermat was able to prove the case $n = 4$ (by something called the 'method of descent') but all other cases proved much harder.

In his 1770 book Algebra Euler published a proof for the n = 3 case. The following is a sketch of this proof. (We will return to this later in the semester!)

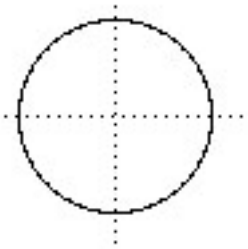
Assume that $x^3 + y^3 = z^3$ and that x and y are both odd and coprime. Then put $x = p + q$, $y = p - q$ and $z = 2r$ and then get $(p + q)^3 + (p - q)^3 = 8r^3$ or $p(p^2 + 3q^2) = 4r^3$. Since p, q are coprime it is possible to deduce that p is divisible by 4 and $p^2 + 3q^2$ must be a perfect cube. Then Euler factorised $p^2 + 3q^2$ into $(p + q\sqrt{-3})(p - q\sqrt{-3})$ and observed that in the ring (he

didn't use that term!) of complex numbers of the form $\{ a+b\sqrt{-3} : a, b \in \mathbb{Z} \}$ these two factors were coprime and so each factor is a perfect cube. That leads to a contradiction.

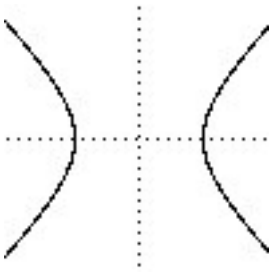
Unfortunately Euler's proof relies on the fact that in this ring we can factor elements into a unique product of primes just as one can in $\mathbb{Z}$. But in the ring Euler used we have $4 = 2 \times 2 = (1+\sqrt{-3})(1-\sqrt{-3})$ and these two factorisations are distinct. Hence this proof is incorrect, though nobody noticed the problem at the time.

It only became apparent much later in 1847 when Lamé claimed he had proved Fermat's Last Theorem by factoring $x^n + y^n = (x-\eta)(x-\eta^2)...(x-\eta^{n-1})$ where η is an n^{th} complex root of 1. It was swiftly realised that the rings in which this factorisation was done did not have unique factorisation and it was left to Kummer to introduce the idea of an 'ideal number' to restore unique factorisation and allow Fermat's Theorem to be proved for some values of n. This invention of Kummer led to the development of the idea of an ideal of a ring which we will meet later.

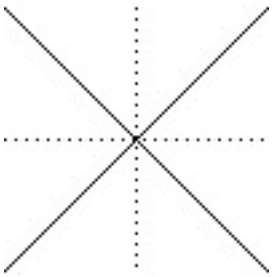
Another area of mathematics which was important in the development of ring theory is geometry. Many interesting curves and surfaces have equations which involve polynomials.



A circle $x^2 + y^2 - 1 = 0$



A hyperbola $x^2 - y^2 - 1 = 0$



Two lines $x + y = 0$

A single point $x^2 + y^2 = 0$

A paraboloid $x^2 + y^2 - z = 0$

A hyperboloid $x^2 + y^2 - z^2 - 1 = 0$

It turns out that geometric objects like this are associated with particular rings of polynomials and the algebra of these rings gives insight into the geometric properties. This area of mathematics is called *algebraic geometry*.

In the 19th Century it was realised that the complex numbers parametrise the plane $\mathbb{R}^2$ in a very useful way. Mathematicians realised that it would be nice to find a similar way to parametrise the space $\mathbb{R}^3$. The mathematician William Rowan Hamilton worked on this for a long time with no success. Each day at breakfast his daughter would ask:

> *Well, Papa can you multiply triplets?*

but he had to admit that he could still only add and subtract them.

His breakthrough came in two stages. First he realised that one had to move from 3 to 4 dimensions and so (by analogy with the $\mathbb{C}$ case) one had numbers of the form $a + ib + jc + kd$ for a, b, c, d and $i^2 = j^2 = k^2 = -1$. Then one could put $ij = k$ but to avoid a contradiction one had to make the multiplication non-commutative with $ji = -k$.

> *And here there dawned on me the notion that we must admit, in some sense, a fourth dimension of space for the purpose of calculating with triples... An electric circuit seemed to close, and a spark flashed forth.*

This revelation came to Hamilton when he was walking with his wife by a canal in Dublin in 1843 and he was so taken with it that he stopped and carved two rules for multiplication on the Brougham bridge.

This system is called the *quaternions* or *hamiltonians*. These proved to be very important in the development of mechanics and other areas of applied mathematics in the 19th century. In fact, the quaternions contain what we now call the scalar and vector product of 3-dimensional vectors and it is these products which are now used.

This was the first example of a non-commutative ring and when Cayley and Sylvester developed the ideas of matrices later in the century this too gave examples of such structures.

1. Numbers

Before beginning our study of abstract rings, we recall some well-known properties of the integers $\mathbb{Z}$.

If $a, b \in \mathbb{Z}$ such that $a = qb$ for some $q \in \mathbb{Z}$, then we say that b *divides* a and we write $b \mid a$. Note that $1 \mid a$, $-1 \mid a$, $a \mid 0$ for all $a \in \mathbb{Z}$, and $0 \mid a$ if and only if $a = 0$.

Example 1.1. The numbers

$$1, 2, 3, 5, 6, 10, 15, 30$$

divide 30 and

$$1, 2, 19, 38$$

divide 38.

1.1. The Division Algorithm

Theorem 1.2 (Division Algorithm) *If $a, b \in \mathbb{Z}$ with $b \neq 0$, then there exist unique quotient $q \in \mathbb{Z}$ and remainder $r \in \mathbb{Z}$ such that*

$$a = qb + r$$

where $q \in \mathbb{Z}$ and $0 \leqslant r < b$.

Proof. To prove that r is unique assume that

$$a = q_1 b + r_1 \text{ and } a = q_2 b + r_2.$$

We must prove that $r_1 = r_2$. Seeking a contradiction assume that $r_1 < r_2$. It follows that $(q_1 - q_2)b = r_2 - r_1 > 0$. But then b divides $r_2 - r_1 \leq r_2 < b$, a contradiction.

To prove that r exists, let $M = \{\, a - qb \;:\; q \in \mathbb{Z} \,\}$ and let r be the smallest number in M such that $r \geq 0$ (exercise: why does r exist?). Then $r = a - qb$ for some $q \in \mathbb{Z}$. If $r \geq b$, then $r > r - b \geq 0$ and $r - b = a - qb - b = a - (q+1)b \in M$. Thus we have a contradiction since r is the smallest number in M with $r \geq 0$. Since r is unique also $q = (a - r)/b$ is unique. ■

Example 1.3. Let $a = 13281$ and $b = 17$. Then $a = 781b + 4$. So, the quotient is 781 and the remainder is 4.

1.2. The Euclidean Algorithm

Definition 1.4. Let $a, b \in \mathbb{Z}$. Then $d \in \mathbb{N}$ is the *greatest common divisor* of a and b if $d \mid a$ and $d \mid b$ and $d' \leqslant d$ whenever $d' \mid a$ and $d' \mid b$. The greatest common divisor d is denoted by $\gcd(a, b)$.

The next lemma will help us find a method for determining the $\gcd$ of any two integers.

Lemma 1.5. *Let $a, b \in \mathbb{Z}$. Then*

(i) $\gcd(a, 0) = |a|$;

(ii) $\gcd(a, b) = \gcd(a - qb, b)$ *for all* $q \in \mathbb{Z}$.

Proof. **(i).** As noted above, all integers divide 0 and the largest number dividing a is a. Hence $\gcd(a, 0) = a$.

(ii). Let $d_1 = \gcd(a, b)$ and $d_2 = \gcd(a - qb, b)$. Then $d_1 \mid a$ and $d_1 \mid b$. Hence d_1 divides $a - qb$ and so $d_1 \leqslant d_2$. Conversely, if $n \mid a - qb$ and $n \mid b$ for some $n \in \mathbb{N}$, then $n \mid a$. In particular, $d_2 \mid a$ and $d_2 \mid b$ and so $d_2 \leqslant d_1$. ■

Theorem 1.6. *Let $a, b \in \mathbb{Z}$ and $d = \gcd(a, b)$. Then there exist $x, y \in \mathbb{Z}$ such that $d = ax + by$.*

Before proving Theorem 1.6 let us consider an example.

Example 1.7. Let $a = 76$ and $b = 32$. Then the divisors of a are

$$1, 2, 4, 19, 38, 76$$

and the divisors of b are

$$1, 2, 4, 8, 16, 32.$$

So, the $\gcd(a, b) = 4$. How do we find x and y? We use the extended Euclidean algorithm:

$$\begin{array}{rrl|l} & a = 76 & 32 = b & \\ & 2b = 64 & 24 = 2a - 4b & \\ \hline r_1 = & a - 2b = 12 & 8 = -2a + 5b & = r_2 \\ & -2a + 5b = 8 & 8 = 6a - 14b & \\ \hline r_3 = & 3a - 7b = 4 & 0 = -8a + 19b & \end{array}$$

Proof of Theorem 1.6. [Omitted from lectures.] If $a, b \in \mathbb{N}$, then the extended Euclidean algorithm described above can be used, as shown below, to find x and y. Assume without loss of generality that $a > b$. Set $a = r_0$ and $b = r_1$. Applying the Division Algorithm, there exists $q_2 \in \mathbb{Z}$ such that $r_0 = q_2 r_1 + r_2$ where $0 \leqslant r_2 < r_1$. It follows from Lemma 1.5(ii) that $\gcd(a, b) = \gcd(r_0 - q_2 r_1) = \gcd(r_2, r_1)$. If $r_2 = 0$, then from Lemma 1.5(ii) that $\gcd(a, b) = \gcd(0, b) = b$ and $a - (q_2 - 1)b = b$, and we can stop.

If $r_2 \neq 0$, then apply the Division Algorithm to find $q_3 \in \mathbb{Z}$ such that $r_1 = q_3 r_2 + r_3$ where $0 \leqslant r_3 < r_2$. Again, $\gcd(a, b) = \gcd(r_2, r_1) = \gcd(r_2, r_1 - q_3 r_2) = \gcd(r_2, r_3)$. So, if $r_3 = 0$, then $\gcd(a, b) = r_2$ and $a - q_2 b = r_2$, and we can stop. If $r_3 \neq 0$, then we repeat this procedure by applying the Division Algorithm to r_2 and r_3 and so on.

Now, $b > r_0 > r_1 > \cdots \geq 0$ and so $r_{i+1} = 0$ for some i. Hence we have found $\gcd(a, b) = r_i$ and the equations

$$r_0 = q_2 r_1 + r_2, r_1 = q_3 r_2 + r_3, \ldots, r_{i-2} = q_i r_{i-1} + r_i, r_{i-1} = q_i r_i.$$

The integers x and y can be obtained as a combination of q_js by substituting (from the second last to first) in these equations.

If a or b (or both) are in $\mathbb{Z} \setminus \mathbb{N}$, then $-a$ and/or $-b$ are in $\mathbb{N}$. Hence changing the signs if necessary, there exist $x, y \in \mathbb{Z}$ such that $x(-a) + y(-b) = d$. Hence $(-x)a + (-y)b = d$, as required. ■

Definition 1.8. Let $a, b \in \mathbb{Z}$. Then a and b are called *coprime* if $\gcd(a, b) = 1$.

Corollary 1.9. *Let $a, b, c \in \mathbb{Z}$ such that a and b are coprime. Then*

(i) *if $a \mid bc$, then $a \mid c$;*

(ii) *if $a \mid c$ and $b \mid c$, then $ab \mid c$;*

(iii) *if a and c are coprime, then a and bc are coprime.*

Proof. As an exercise. ■

Example 1.10. Let $a = 15$, $b = 16$, and $c = 30$. Then $\gcd(a, b) = 1$ and $a \mid bc$ and so $15 = a \mid c = 30$ by Corollary 1.9(i).

Let $c = 240$. Then $a \mid c$ and $b \mid c$, and so $ab = 240 \mid 240 = c$ by Corollary 1.9(ii).

Let $c = 17$. Then $a = 15$ and $bc = 272$ are coprime by Corollary 1.9(iii).

1.3. Modular arithmetic

In this section we will recall the notions involved in modular arithmetic. Let $m \in \mathbb{N}$ and let $\mathbb{Z}/(m)$ denote the set

$$\{ n \in \mathbb{N} : 0 \leqslant n < m \} = \{0, 1, 2, \ldots, m - 1\}.$$

(The reason for choosing this notation should become apparent later in the course!) If $a, b \in \mathbb{Z}/(m)$, then $a + b, a \cdot b \in \mathbb{N}$ and so by the Division Algorithm 1.2 we can write

$$a + b = q_1 m + r_1 \text{ and } a \cdot b = q_2 m + r_2$$

where r_1, r_2 are the unique remainders such that $0 \leqslant r_1, r_2 < m$. Note that $r_1, r_2 \in \mathbb{Z}/(m)$. This defines two operations *addition modulo m* and *multiplication modulo m* on $\mathbb{Z}/(m)$.

We will write

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

if x and y leave the same remainder on division by m. Note that this is equivalent to the fact that $x - y$ is divisible by m. So for modular arithmetic we would write

$$a + b \equiv r_1 \pmod{m} \text{ and } a \cdot b \equiv r_2 \pmod{m}.$$

If $n \in \mathbb{N}$ and $n \equiv r \pmod{m}$ with $0 \leq r < m$, then we may also say that r is n *reduced modulo* m. Sometimes we denote the modulo m reduction of a number by putting a bar over the number (if it is clear from the context modulo which number we reduce).

Example 1.11.

$$\begin{aligned} 1+1 &\equiv 2 \pmod{7} \\ 4+6 &\equiv 3 \pmod{7} \\ 3+4 &\equiv 0 \pmod{7} \\ 2 \cdot 5 &\equiv 3 \pmod{7} \\ \overline{17} &= 3 \qquad \text{(if it is clear that we are reducing modulo 7).} \end{aligned}$$

You can (and should!) verify as an exercise that (modulo m)

$$\overline{x \cdot y} = \overline{\overline{x} \cdot \overline{y}}$$

and

$$\overline{x + y} = \overline{\overline{x} + \overline{y}}.$$

From this follows that addition and multiplication modulo m satisfy the following properties:

$$\begin{gathered} \overline{\overline{(x+y)} + z} = \overline{x + \overline{(y+z)}} \\ \overline{\overline{(xy)}z} = \overline{x\overline{(yz)}} \\ \overline{x+0} = \overline{0+x} = x \\ \overline{1 \cdot x} = \overline{x \cdot 1} = x \\ \overline{x + (m-x)} = \overline{(m-x) + x} = 0 \\ \overline{x+y} = \overline{y+x} \\ \overline{xy} = \overline{yx} \\ \overline{x \cdot 0} = \overline{0 \cdot x} = 0 \end{gathered}$$

for all $x, y, z \in \mathbb{Z}/(m)$.

Example 1.12. The following is the table of multiplication for $\mathbb{Z}/(7)$ modulo 7.

$\cdot$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

1.4. The Chinese Remainder Theorem

Let us now consider the following problem. Player 1 thinks of any natural number x less than 30 and calculates the remainder of x modulo 2, modulo 3, and modulo 5. Player 1 then tells Player 2 the three remainders Player 2 wins if he can tell Player 1 what the original x is, and Player 1 wins if Player 2 cannot. For example, if $x = 18$, then $x \equiv 0 \pmod 2$, $x \equiv 0 \pmod 3$, and $x \equiv 3 \pmod 5$.

Is there a way for Player 2 to always win this game? The answer is yes! To see this let us reformulate the game. If $q_1, q_2, \ldots, q_t \in \mathbb{N} \setminus \{0\}$ and $N = q_1 q_2 \cdots q_t \neq 0$, then define

$$\rho : \mathbb{Z}/(N) \longrightarrow \mathbb{Z}/(q_1) \times \mathbb{Z}/(q_2) \times \cdots \times \mathbb{Z}/(q_t)$$

by

$$\rho(x) = (\overline{x} \pmod{q_1}, \overline{x} \pmod{q_2}, \ldots, \overline{x} \pmod{q_t}). \tag{4.1}$$

(Note that the bars here indicate reductions modulo the different numbers q_1 up to q_t!)

Lemma 1.13. *Let $q_1, q_2, \ldots, q_t \in \mathbb{N} \setminus \{0\}$ such that $q_1, q_2, \ldots, q_t$ are pairwise coprime and let $N = q_1 q_2 \cdots q_t$. Then $\rho : \mathbb{Z}/(N) \longrightarrow \mathbb{Z}/(q_1) \times \mathbb{Z}/(q_2) \times \cdots \times \mathbb{Z}/(q_t)$ defined in* (4.1) *is a bijection.*

Proof. Let $x, y \in \mathbb{Z}/(N)$ such that $\rho(x) = \rho(y)$. Then $x \equiv y \pmod{q_i}$ for all i. It follows that $x - y \equiv 0 \pmod{q_i}$ for all i and so $q_1 \mid x - y$, $q_2 \mid x - y$, ..., $q_t \mid x - y$. Now, $q_1, q_2, \ldots, q_t$ are coprime and so by repeatedly applying Corollary 1.9(ii) and (iii) we obtain $N = q_1 \cdots q_t \mid x - y$. But $x, y \in \mathbb{Z}/(N)$ and so $0 \leqslant x, y < N$. Hence $x = y$ and ρ is injective. Moreover, ρ must be surjective as $\mathbb{Z}/(N)$ and $\mathbb{Z}/(q_1) \times \mathbb{Z}/(q_2) \times \cdots \times \mathbb{Z}/(q_t)$ have equal size. ■

So, in our example, we known that the number 18 is in 1-1 correspondence with the triple $(0, 0, 3)$. To recover 18 from the triple $(0, 0, 3)$ we only have to apply ρ^{-1}.

Theorem 1.14 (Chinese Remainder Theorem) *Let $q_1, q_2, \ldots, q_t \in \mathbb{N} \setminus \{0\}$ such that $q_1, q_2, \ldots, q_t$ are pairwise coprime, let $N = q_1 q_2 \cdots q_t$, and let $a_1, a_2, \ldots a_t \in \mathbb{Z}$ be arbitrary. Then the system of equations:*

$$\begin{aligned} x &\equiv a_1 \pmod{q_1} \\ x &\equiv a_2 \pmod{q_2} \\ &\vdots \\ x &\equiv a_t \pmod{q_t} \end{aligned}$$

has a solution $x \in \mathbb{Z}$. Moreover, if x is a solution, then $y \in \mathbb{Z}$ is one if and only if $x \equiv y \pmod N$.

Proof. We start by finding the solution x to the equations given in the theorem. By repeatedly applying Corollary 1.9(iii) we deduce that q_i and N/q_i are coprime for all i. Hence from the extended Euclidean Algorithm 1.6 there exist integers x_i and y_i such that

$$\begin{aligned} x_1 q_1 + y_1(N/q_1) &= 1 \\ x_2 q_2 + y_2(N/q_2) &= 1 \\ &\vdots \\ x_t q_t + y_t(N/q_t) &= 1. \end{aligned}$$

Set $b_i = y_i(N/q_i)$. Then

$$b_i = 1 - x_i q_i \equiv 1 \pmod{q_i}$$

and since $q_j \mid (N/q_i)$ for all $i \neq j$

$$b_i \equiv 0 \pmod{q_j}.$$

So, $x = a_1b_1 + a_2b_2 + \cdots + a_tb_t \in \mathbb{Z}$ is a solution to the equations in the theorem.

Now let $x \in \mathbb{Z}$ be a solution. An integer y is a solution if and only if x and y leave the same remainder on division by all the q_i, that is, the difference is divisible by q_i for $1 \leq i \leq t$. However, the proof of Lemma 1.13 shows that this is the case if and only if $x - y$ is divisible by N. ■

The proof of Theorem 1.14 can be used to find the solutions to specific examples of systems of equations like those given in Theorem 1.14.

Example 1.15. Continuing with the example given above. Let $N = 30$ and $(0, 0, 3) \in \mathbb{Z}/(2) \times \mathbb{Z}/(3) \times \mathbb{Z}/(5)$. Then we want to find a solution x to the equations

$$\begin{aligned} x &\equiv 0 \pmod{2} \\ x &\equiv 0 \pmod{3} \\ x &\equiv 3 \pmod{5}. \end{aligned}$$

So, applying the extended Euclidean Algorithm to the pairs $(2, 15)$, $(3, 10)$, and $(5, 6)$ we obtain:

$N/q_1 = 15$	$2 = q_1$	$N/q_2 = 10$	$3 = q_2$	$N/q_3 = 6$	$5 = q_3$
$7q_1 = 14$		$3q_2 = 9$		$q_3 = 5$	
$(N/q_1) - 7q_1 = 1$		$(N/q_2) - 3q_2 = 1$		$(N/q_3) - q_3 = 1$	

So, $b_1 = N/q_1 = 15$, $b_2 = N/q_2 = 10$, and $b_3 = N/q_3 = 6$. Thus the solution we are looking for is

$$x = a_1b_1 + a_2b_2 + a_3b_3 = 0 \cdot 15 + 0 \cdot 10 + 3 \cdot 6 = 18.$$

1.5. Prime Numbers

Recall that a prime number is a natural number $p > 1$ that is only divisible by itself and 1. The first few primes are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots.$$

They can be found as follows using a method called the *sieve of Eratosthenes*. Write out all the natural numbers

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25$$

and then cross out the numbers that are multiples of 2 but not 2 itself

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, 9, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, 15, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, 21, \cancel{22}, 23, \cancel{24}, 25.$$

The numbers not crossed out are not multiples of 2. Continue by crossing out the multiples of 3

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, \cancel{15}, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, \cancel{21}, \cancel{22}, 23, \cancel{24}, 25.$$

then 5, then 7, then 11 and so on

$$2, 3, \cancel{4}, 5, \cancel{6}, 7, \cancel{8}, \cancel{9}, \cancel{10}, 11, \cancel{12}, 13, \cancel{14}, \cancel{15}, \cancel{16}, 17, \cancel{18}, 19, \cancel{20}, \cancel{21}, \cancel{22}, 23, \cancel{24}, \cancel{25}.$$

Repeating this procedure over and over again you can eventually tell if any number is prime or not. The largest known prime (as of 2008!) is

$$2^{43112609} - 1$$

and it has 12978189 digits!

Theorem 1.16. *Every non-zero natural number is a product of primes.*

Proof. We will prove that every natural number $n \geq 1$ is a product of primes by induction on n.

The number 1 is the empty product of primes. Assume that for all $m < n$ we know that m is a product of primes. If n is a prime number, then it is a product of primes (of length 1). If n is not a prime, then there exist $a, b \in \mathbb{N}$ such that $a, b > 1$ and $n = ab$. Then by induction a and b are products of primes and hence so is n. ■

Theorem 1.17. *There are infinitely many prime numbers.*

Proof. Seeking a contradiction assume the contrary. Then the primes can be listed as

$$p_1, p_2, \ldots, p_m$$

for some natural number $m \geq 1$. Now, the natural number $n = p_1 p_2 \cdots p_m + 1$ is a product of primes and hence divisible by some p_i for $1 \leqslant i \leqslant m$. But then $p_i \mid n$ and $p_i \mid p_1 p_2 \cdots p_m$ and so $p_i \mid (n - p_1 p_2 \cdots p_m) = 1$, a contradiction as 1 is not a prime. ■

A crucial property of the primes is given in the following lemma. We will revisit this lemma again in later sections.

Lemma 1.18. *If $p \in \mathbb{N}$ is a prime and $p \mid ab$, then $p \mid a$ or $p \mid b$.*

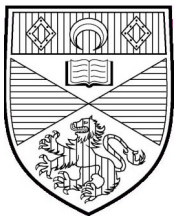
Proof. If $p \nmid a$, then $\gcd(p, a) = 1$. Hence $p \mid b$ by Corollary 1.9. On the other hand, if $p \mid a$, then we are finished. ■

Theorem 1.19. *Let $m \in \mathbb{N}$. Then there exists a unique (up to changing the order of the factors) factorization of m into a product of prime numbers $m = p_1 p_2 \cdots p_n$ for some $n \in \mathbb{N}$.*

Proof. We may assume that $m > 1$. Let $p_1 p_2 \cdots p_n$ and $q_1 q_2 \cdots q_r$ be two factorizations of m into products of primes. Then

$$m = p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_r.$$

If $p_i = q_j$ for some i and j, then we can cancel these factors. Hence we can also assume that $p_i \neq q_j$ for all i, j. There must also be at least 1 prime on the left hand side (say) and at least 2 on the right hand side. So, we can assume that $n \geq 1$ and $r > 1$. Now, $p_1 \mid m$ and so by repeatedly applying Lemma 1.18 we find that $p_1 \mid q_j$ for some j. But q_j is a prime and so $p_1 = q_j$, a contradiction to the assumption that $p_i \neq q_j$ for all i, j. ■



School of Mathematics and Statistics

MT4517 Rings & Fields

Ring Theory – the basics

2. Definitions and first examples

The example of the integers in the previous chapter serves as a prototype for the more abstract study we start in this chapter.

Definition 2.1. A *ring* is a set R together with a pair of binary operations $+$ and $*$ satisfying the axioms:

A1. $r + (s + t) = (r + s) + t$ for all $r, s, t \in R$;

A2. there exists $0 \in R$ such that $r + 0 = 0 + r = r$ for all $r \in R$;

A3. for all $r \in R$ there exists $-r \in R$ such that $r + (-r) = 0 = (-r) + r$;

A4. $r + s = s + r$ for all $r, s \in R$;

M1. $r * (s * t) = (r * s) * t$ for all $r, s, t \in R$;

D. $(r + s) * t = r * t + s * t$ and $r * (s + t) = r * s + r * t$ for all $r, s, t \in R$.

A is for Addition, **M** is for multiplication, and **D** is for Distributive. **A1, A2, A3, A4, M1, D** are called the *ring axioms*. The final Axiom **D** is called the *distributive law*. Note that we will write $r - s$ to mean $r + (-s)$.

[Aside: $(R, +, *)$ is a ring if $(R, +)$ is an abelian group, $(R, *)$ is a semigroup, and it satisfies the distributive law.]

Example 2.2. The set of integers $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ with the usual addition and multiplication are all rings.

The element 0 from Axiom **A2** is called the *zero* of the ring R. The use of 0 here is symbolic and it does not always mean the integer 0. The *order* of a ring R is the number of elements it contains, that is, $|R|$. In the previous example, all the rings had infinite order.

Example 2.3. The integers $\mathbb{Z}/(n) = \{0, 1, \ldots, n - 1\}$ with addition and multiplication modulo n satisfy the ring axioms. Hence $\mathbb{Z}/(n)$ is a ring for all $n \in \mathbb{N}$. The order of $\mathbb{Z}/(n)$ is n.

Example 2.4. Let R be a finite set. Then it is possible to specify operations $+$ and $*$ on R using addition and multiplication tables.

For example, if $R = \{0, a, b, c\}$, then such tables might look like

$+$	0	a	b	c
0	0	a	b	c
a	a	0	c	b
b	b	c	0	a
c	c	b	a	0

$.$	0	a	b	c
0	0	0	0	0
a	0	0	a	a
b	0	0	b	b
c	0	0	c	c

.

It is almost impossible to check by hand that these tables satisfy the ring axioms. But they do and R is a ring.

Example 2.5. Let $M_2(\mathbb{R})$ denote the set of all 2×2 real matrices. Then $M_2(\mathbb{R})$ forms a ring under the usual matrix addition $+$ and multiplication $*$. For instance,

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 2 & 4 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 5 & 8 \end{pmatrix}.$$

In fact, the set of $n \times n$ matrices with entries in any ring R forms a ring denoted $M_n(R)$ (see Exercise 2.14).

Example 2.6. The quaternions are defined to be the set $\mathbb{H} = \{\, a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \;:\; a, b, c, d \in \mathbb{R} \,\}$ with addition

$$(a_1 + a_2\mathbf{i} + a_3\mathbf{j} + a_4\mathbf{k}) + (b_1 + b_2\mathbf{i} + b_3\mathbf{j} + b_4\mathbf{k}) = (a_1 + b_1) + (a_2 + b_2)\mathbf{i} + (a_3 + b_3)\mathbf{j} + (a_4 + b_4)\mathbf{k}$$

and multiplication

$$\begin{aligned}(a_1+a_2\mathbf{i}+a_3\mathbf{j}+a_4\mathbf{k})*(b_1+b_2\mathbf{i}+b_3\mathbf{j}+b_4\mathbf{k}) &= (a_1b_1-a_2b_2-a_3b_3-a_4b_4)+(a_1b_2+a_2b_1+a_3b_4-a_4b_3)\mathbf{i} \\ &\quad + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)\mathbf{j} + (a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1)\mathbf{k}.\end{aligned}$$

Sometimes the quaternions are denoted Q_8 or $\mathbf{H}$.

The above multiplication is rather cumbersome. It is often more useful to multiply elements of $\mathbb{H}$ together as if they are polynomials (with real coefficients and indeterminants **i, j,** and **k**) and then apply the rules given in the following table:

*	1	**i**	**j**	**k**
1	1	**i**	**j**	**k**
i	**i**	-1	**k**	**-j**
j	**j**	**-k**	-1	**i**
k	**k**	**j**	**-i**	-1

By rather laborious calculations it is possible to show that $\mathbb{H}$ together with the addition and multiplication defined above satisfy the ring axioms. Thus $\mathbb{H}$ is a ring.

3. Further axioms for rings

Definition 3.1. Let R be a ring satisfying the axiom

M2. there exists $1 \in R$ such that $1 * r = r * 1 = r$ for all $r \in R$.

Then R is called a *ring with identity*. The element $1 \in R$ is referred to as the *multiplicative identity* or *one* of R.

Again note that the 1 in Axiom **M2** is symbolic and not necessarily the integer 1. Some mathematicians include the existence of a multiplicative identity as an axiom in the definition of a ring. In this course we will not assume, unless otherwise stated, that our rings have a multiplicative identity.

Example 3.2. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/(n)$ all have identity $1 \in \mathbb{Z}$. The ring R in Example 2.4 has no multiplicative identity. The one of the ring $M_2(\mathbb{R})$ is the *identity matrix*

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

The multiplicative identity of the quaternions $\mathbb{H}$ is $1 + 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k} = 1$.

Definition 3.3. If R is a ring with identity 1, then the *multiplicative inverse* of an element $x \in R$ is an element $x^{-1} \in R$ such that

$$x * x^{-1} = x^{-1} * x = 1.$$

If $x \in R$ has a multiplicative inverse, then x is called a *unit*.

Even if a ring R has a multiplicative identity, it may not be possible to find a multiplicative inverse for every element in R. In particular, if $|R| > 1$, then the element 0 will never have an inverse.

Example 3.4. Let $x \in \mathbb{Z}$ be arbitrary. Then x is a unit if there exists $y \in \mathbb{Z}$ such that $x * y = y * x = 1$. It follows that the only units in $\mathbb{Z}$ are 1 and -1. Let $x \in \mathbb{Z}/(n)$. Then x is a unit if and only if x is coprime to n. A matrix $A \in M_2(\mathbb{R})$ is invertible (i.e. is a unit) if and only if $\det(A) \neq 0$. It can be shown the multiplicative inverse of a non-zero element $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ in the quaternions $\mathbb{H}$ is

$$(a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}) \cdot \frac{1}{a^2 + b^2 + c^2 + d^2}.$$

Definition 3.5. Let R be a ring with identity satisfying the axiom

M3. every $r \in R \setminus \{0\}$ is a unit.

Then R is called a *division ring* (or sometimes a *skew field*.)

Example 3.6. Every non-zero element of $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ is a unit. Hence the rings $\mathbb{R}, \mathbb{C}, \mathbb{Q}$ are division rings. The integers $\mathbb{Z}$ is not a division ring. The ring $\mathbb{Z}/(n)$ is a division ring if and only if n is prime. The ring $M_2(\mathbb{R})$ is not a division ring. The quaternions form a division ring.

Definition 3.7. If R is a ring satisfying the axiom

M4. $r * s = s * r$ for all $r, s \in R$.

Then R is called a *commutative ring*.

Example 3.8. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and $\mathbb{Z}/(n)$ are commutative (for all n). In $M_2(\mathbb{R})$

$$\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 7 & 11 \\ 3 & 5 \end{pmatrix} \neq \begin{pmatrix} 4 & 6 \\ 5 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} * \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

Hence is $M_2(\mathbb{R})$ is a non-commutative ring.

Since $\mathbf{i} * \mathbf{j} = \mathbf{k} \neq -\mathbf{k} = \mathbf{j} * \mathbf{i}$, it follows that $\mathbb{H}$ is non-commutative.

Definition 3.9. A commutative division ring is called a *field*. That is, a field is a set F together with a pair of binary operations $+$ and $*$ satisfying the axioms: **A1, A2, A3, A4, M1, M2, M3, M4**, and **D1**.

Example 3.10. So, $\mathbb{Z}$ is not a field but $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. The ring $\mathbb{Z}/(n)$ is a field if and only if n is a prime. The ring $M_2(\mathbb{R})$ fails to satisfy **M4** and so $M_2(\mathbb{R})$ is not a field. Likewise, the quaternions $\mathbb{H}$ fail **M4** and so $\mathbb{H}$ is not a field.

If p is a prime, we will denote $\mathbb{Z}/(p)$ by $\mathbb{F}_p$ and we will refer to it as the *Galois field of order* p.

If R is a ring and $a, b \in R$ with $a, b \neq 0$ and $a * b = 0$, then a and b are called *zero divisors*.

Example 3.11. In the commutative ring with identity $\mathbb{Z}/(6)$ we have $2 * 3 = 0$ and so 2 and 3 are zero divisors.

More generally, $x \in \mathbb{Z}/(n)$ is a zero divisor if and only if $\gcd(x, n) \neq 1$.

Definition 3.12. If R is a commutative ring with identity ($1 \neq 0$) where no element is a zero divisor, then R is called an *integral domain*.

So, if R is an integral domain and $a * b = 0$, then $a = 0$ or $b = 0$.

Example 3.13. The commutative ring $\mathbb{Z}$ is an integral domain, as are the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$. As seen in Example 3.11, $\mathbb{Z}/(n)$ is an integral domain if and only if n is a prime. In $M_2(\mathbb{R})$

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

and so $M_2(\mathbb{R})$ is not an integral domain. The quaternions $\mathbb{H}$ is not an integral domain as $*$ is not commutative. However, $\mathbb{H}$ has no zero divisors (see Exercise 2.10).

Lemma 3.14. *Every field F is an integral domain.*

Proof. Let $a, b \in F$ with $a * b = 0$. If $a \neq 0$, then $0 = a^{-1} * 0 = a^{-1} * a * b = 1 * b = b$. Thus F has no zero divisors. ■

The following table shows the different axioms satisfied by the different types of rings defined in this section. **Z** denotes that the ring has no zero divisors.

	A1	**A2**	**A3**	**A4**	**M1**	**D**	**M2**	**M3**	**M4**	**Z**	
1	✔	✔	✔	✔	✔	✔	–	–	–	–	ring
2	✔	✔	✔	✔	✔	✔	–	–	✔	–	commutative ring
3	✔	✔	✔	✔	✔	✔	✔	–	–	–	ring with identity
4	✔	✔	✔	✔	✔	✔	–	–	–	✔	ring with no zero divisors
5	✔	✔	✔	✔	✔	✔	✔	–	✔	–	comm. ring with one
6	✔	✔	✔	✔	✔	✔	–	–	✔	✔	comm. ring with no zero divisors
7	✔	✔	✔	✔	✔	✔	✔	–	–	✔	ring with one and no zero divisors
8	✔	✔	✔	✔	✔	✔	✔	–	✔	✔	integral domain
9	✔	✔	✔	✔	✔	✔	✔	✔	–	✔	division ring
10	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	field

The following table shows which of the properties given in the table above are satisfied by the examples we have followed throughout this section.

	1	2	3	4	5	6	7	8	9	10	
$\mathbb{Z}$	✔	✔	✔	✔	✔	✔	✔	✔	–	–	
$\mathbb{Q}, \mathbb{R}, \mathbb{C}$	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	
$\mathbb{Z}/(n)$	✔	✔	✔	–	✔	–	–	–	–	–	n composite
$\mathbb{Z}/(p)$	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	p prime
$M_2(\mathbb{R})$	✔	–	✔	–	–	–	–	–	–	–	
$\mathbb{H}$	✔	–	✔	✔	–	–	✔	–	✔	–	

We have seen that the following are equivalent

(i) n is a prime;

(ii) $\mathbb{Z}/(n)$ is a division ring;

(iii) $\mathbb{Z}/(n)$ is a field;

(iv) $\mathbb{Z}/(n)$ is an integral domain.

In fact, if $\mathbb{Z}/(n)$ is replaced with any finite ring R in (ii), (iii), and (iv), then they are still equivalent! Clearly (iii) implies (ii) and (iv). The following shows that (iv) implies (iii) but proving that (ii) implies (iii) is very hard!

Theorem 3.15. *Every finite integral domain I is a field.*

Proof. The only thing we need to show is that an arbitrary non-zero element $a \in I$ has a multiplicative inverse. The sequence $a, a^2, a^3, \ldots$ can only contain finitely many elements of I, since there are only finitely many elements in I. Therefore $a^m = a^n$ for some $m < n$ (say). Then $0 = a^m - a^n = a^m(1 - a^{n-m})$. Since there are no zero-divisors and $a \neq 0$ it follows that $a^m \neq 0$. Hence $1 - a^{n-m} = 0$ and so $1 = a * a^{n-m-1}$. It follows that $a^{n-m-1} = a^{-1}$ is a multiplicative inverse for a. ■

3.1. Further Examples

Example 3.16. Let $R = \{ a + b\sqrt{2} : a, b \in \mathbb{Z} \} \subseteq \mathbb{R}$ with the usual $+$ and $*$ on real numbers. It is straightforward to check that R satisfies Axioms **A1, A2, A3, A4, M1**, and **D** (and you should check it!). Hence R is a ring. Let us consider which of the other axioms **M2, M3, M4** and **Z** are satisfied by R.

The integer $1 = 1 + 0\sqrt{2}$ is an element of R and $1 * (a + b\sqrt{2}) = (a + b\sqrt{2}) * 1 = a + b\sqrt{2}$. Hence R satisfies **M2** and is a ring with identity. An element $a + b\sqrt{2} \in R$ is a unit if and only if $a^2 - 2b^2 = \pm 1$ (see Exercise 2.6). Let $a + b\sqrt{2}, c + d\sqrt{2} \in R$. Then

$$(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (bc + ad)\sqrt{2} = (c + d\sqrt{2})(a + b\sqrt{2})$$

and so R is commutative. Note that we only use that fact that the ring $\mathbb{Z}$ is commutative to prove that R is commutative. Finally, suppose that $(a + b\sqrt{2}).(c + d\sqrt{2}) = 0 = 0 + 0\sqrt{2}$. Since $a + b\sqrt{2}$ and $c + d\sqrt{2}$ belong to the field $\mathbb{R}$, it follows that either $a + b\sqrt{2} = 0$ or $c + d\sqrt{2} = 0$. So, R is an integral domain.

Example 3.17. Let $\mathbb{Z}[i]$ denote the subset $\{\ a+bi\ :\ a,b\in\mathbb{Z}\ \}$ of $\mathbb{C}$ where $i=\sqrt{-1}$. It is straightforward to check that $(a+bi)*(c+di)\in\mathbb{Z}[i]$ for all $a+bi, c+di\in\mathbb{Z}[i]$ and that $\mathbb{Z}[i]$ satisfies Axioms **A1, A2, A3, A4, M1**, and **D**. Hence $\mathbb{Z}[i]$ is a ring called the *Gaussian integers*.

The integer $1=1+0i$ is an element of $\mathbb{Z}[i]$ and $1*(a+bi)=(a+bi)*1=a+bi$. Hence $\mathbb{Z}[i]$ has a multiplicative identity. We will determine the units of $\mathbb{Z}[i]$ in a later section. Again $\mathbb{Z}[i]$ is commutative since the integral domain $\mathbb{Z}$ is commutative. As in the previous example, since $\mathbb{Z}[i]$ is contained in the field $\mathbb{C}$, it follows that $\mathbb{Z}[i]$ has no zero divisors. Hence $\mathbb{Z}[i]$ is an integral domain.

Example 3.18. Let M denote the set of real 2×2 matrices of the form

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

If

$$A=\begin{pmatrix} a_1 & b_1 \\ 0 & d_1 \end{pmatrix}, B=\begin{pmatrix} a_2 & b_2 \\ 0 & d_2 \end{pmatrix}\in M,$$

then

$$A+B=\begin{pmatrix} a_1+a_2 & b_1+b_1 \\ 0 & d_1+d_2 \end{pmatrix}\in M \text{ and } A*B=\begin{pmatrix} a_1a_2 & a_1b_2+b_1c_2 \\ 0 & d_1d_2 \end{pmatrix}\in M.$$

So, it makes sense to talk about the usual matrix operations of $+$ and $*$ on M. It follows that M satisfies Axioms **A1, A2, A3, A4, M1**, and **D** as $M_2(\mathbb{R})$ does. The identity of M is the identity matrix

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

and $A\in M$ has an inverse if and only if $a\neq 0$ and $b\neq 0$. To prove that M is not commutative it suffices to see that

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}*\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix}\neq\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}*\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

The matrices

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

are zero divisors in M. Hence M is a non-commutative ring with one and zero divisors that is not a division ring.

3.2. Polynomial Rings

Polynomials are the source of some of the most important examples of rings.

Let R be a ring. Then a *polynomial over* R is an expression of the form

$$f=\sum_{i=0}^{n} a_i x^i = a_0+a_1x+a_2x^2+\cdots+a_nx^n$$

where n is a non-negative integer, the *coefficients* $a_0, a_1, \ldots, a_n$ are elements of R and x is a symbol not in R called the *indeterminant*. To reiterate, the indeterminate x is not a member of R, and neither are x^2, x^3, ... They are simply markers that indicate how to add and multiply.

The *degree* of a polynomial f is the largest n such that $a_n\neq 0$ and is denoted $\deg(f)=n$. By convention, $\deg(0)=-\infty$. Polynomials of degree 0 are called *constant polynomials*.

Addition and multiplication of polynomials is done in the usual way. That is, if $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$ (taking coefficients to be 0 if necessary to ensure that the degrees are equal), then

$$f + g = \sum_{i=0}^{n} (a_i + b_i) x^i$$

and

$$f * g = \sum_{k=0}^{2n} c_k x^k \text{ where } c_k = \sum_{i+j=k,\, 0 \leqslant i,j \leqslant n} a_i b_j.$$

Note that

$$\deg(f+g) \leqslant \max(\deg(f), \deg(g)) \text{ and } \deg(f*g) \leqslant \deg(f) + \deg(g).$$

If R is an integral domain, then

$$\deg(f*g) = \deg(f) + \deg(g).$$

Two polynomials are equal if and only if all of their coefficients are equal.

Example 3.19. If $f = 1 + 13x + 3x^2 + x^3$ and $g = x + 3x^3$ are polynomials over the ring $\mathbb{Z}/(14)$, then

$$f + g = 1 + 3x^2 + 4x^3 \text{ and } f * g = x + 13x^2 + 6x^3 + 12x^4 + 9x^5 + 3x^6.$$

We denote by $R[x]$ the set of polynomials over R with the operations $+$ and $*$ given above.

Theorem 3.20. *Let R be a ring. Then $R[x]$ is a ring called the* ring of polynomials over *R, and its zero element is the* zero polynomial *all of whose coefficients are zero.*

Proof. As an exercise. ■

Example 3.21. Let us consider the polynomial ring $\mathbb{Z}[x]$. The one of $\mathbb{Z}[x]$ is just $1 \in \mathbb{Z}[x]$ (the polynomial of degree 0) and $\mathbb{Z}[x]$ is commutative as $\mathbb{Z}$ is commutative.

What are the units of $\mathbb{Z}[x]$? If $f = \sum_{i=0}^{n} a_i x^i, g = \sum_{i=0}^{n} b_i x^i \in \mathbb{Z}[x]$ and $f * g = 1$, then the only non-zero coefficients must be a_0 and b_0. Hence either $f = 1$ and $g = 1$ or $f = -1$ and $g = -1$. So, $\mathbb{Z}[x]$ is not a field.

Let $f, g \in \mathbb{Z}[x]$ be non-zero polynomials with $\deg(f) = m$ and $\deg(g) = n$. Then $\deg(f*g) = m + n \in \mathbb{N}$. In particular, $f * g \neq 0$ and so we have shown that $\mathbb{Z}[x]$ is an integral domain.

You can prove using an analogous argument that $\mathbb{R}[x]$ is a commutative ring with identity, that the units of $\mathbb{R}[x]$ are the constant polynomials, and that $\mathbb{R}[x]$ is an integral domain.

Lemma 3.22. *$R[x]$ is commutative if and only if R is commutative.*

4. Subrings

Definition 4.1. A *subring* S of a ring R is a subset of R which is a ring under the same operations as R. We will write $S \leqslant R$ to denote that S is a subring of R.

Rather than rechecking all the ring axioms for S we can simply apply the following lemma.

Lemma 4.2. *Let S be a non-empty subset of a ring R such that $a - b, a * b \in S$ for all $a, b \in S$. (That is, S is closed under subtraction and multiplication.) Then S is a subring of R.*

Proof. Associativity and commutativity of addition **A1** and **A4**, associativity of multiplication **M1**, and the distributive laws **D** in S follow from the respective properties of R. Since $a-b \in S$ for all $a, b \in S$, in particular, $0 = a-a \in S$ and so S satisfies **A2**. Also since $0 \in S$, $-a = 0-a \in S$ for all $a \in S$ and so S satisfies **A3**. ■

Example 4.3. Let R denote the set $\{ a+b\sqrt{5} \, : \, a, b \in \mathbb{Z} \}$. If $+$ and $*$ denote the usual operations on $\mathbb{R}$, then

$$(a + b\sqrt{5}) - (c + d\sqrt{5}) = (a - c) + (b - d)\sqrt{5} \in R$$

and

$$(a + b\sqrt{5})(c + d\sqrt{5}) = (ac + 5bd) + (ad + bc)\sqrt{5} \in R.$$

Hence R is a subring of $\mathbb{R}$. An analogous argument shows that $\{ x + y\sqrt{5} \, : \, x, y \in \mathbb{Q} \}$ with $+$ and $*$ is also a subring of $\mathbb{R}$.

We showed in Example 3.16 that $\{ a + b\sqrt{2} \, : \, a, b \in \mathbb{Z} \}$ is a ring, and hence a subring of $\mathbb{R}$.

Example 4.4. Let (2) denote the even integers. Then

$$2i - 2j = 2(i - j) \in (2)$$

and

$$2i * 2j = 2 * 2ij \in (2).$$

Hence, by Lemma 4.2, (2) is a subring of $\mathbb{Z}$.

More generally, if n is any integer, then the same reasoning shows that the set (n) of all multiples of n is a subring (n) of $\mathbb{Z}$. On the other hand, the odd integers do not form a subring of $\mathbb{Z}$ since $5 - 3 = 2$ is not odd.

Example 4.5. The subsets $\{0, 2, 4\}$ and $\{0, 3\}$ are subrings of $\mathbb{Z}/(6)$.

Lemma 4.6. *Let R be a ring and let S, T be subrings of R. Then $S \cap T$ is a subring of R.*

Proof. Since $0 \in S \cap T$, it follows that $S \cap T$ is non-empty. If $s, t \in S \cap T$, then $s - t \in S \cap T$ and $s * t \in S \cap T$. ■

5. Ideals

Definition 5.1. A subring I of R is an *ideal* if for all $a \in I$ and for all $r \in R$, then $r * i, i * r \in I$.

If R is any ring, then $\{0\}$ and R are ideals in R. An ideal of a ring R is called *proper* if it is not equal to R.

Example 5.2. In Example 4.4, we saw that the even integers (2) are a subring of $\mathbb{Z}$. If $i \in (2)$ and $r \in \mathbb{Z}$, then $i = 2 * j$, for some j, and so $i * r = r * i = 2 * jr \in (2)$. Hence (2) is an ideal in $\mathbb{Z}$.

In Example 4.5, we showed that $I = \{0, 2, 4\}$ and $J = \{0, 3\}$ are subrings of $\mathbb{Z}/(6)$. If $i \in I$ and $r \in \mathbb{Z}/(6)$, then from the table below, $i * r = r * i \in I$

*	0	1	2	3	4	5
0	0	0	0	0	0	0
2	0	2	4	0	2	4
4	0	4	2	0	4	2

Likewise, J is an ideal of $\mathbb{Z}/(6)$.

Example 5.3. If $a + bi \in \mathbb{Z}[i]$ from Example 3.17, and $r = c/d$ where $d \neq 0$ and $d \nmid ac$ or $d \nmid bc$. Then $(c/d) * (a + bi) \notin \mathbb{Z}[i]$. Hence $\mathbb{Z}[i]$ is not an ideal of $\mathbb{C}$ even though it is a subring.

Likewise $\mathbb{Z}[\sqrt{5}]$ is not an ideal in $\mathbb{R}$ or $\mathbb{C}$, although it is a subring of both.

Example 5.4. Let I be all the polynomials over any commutative ring R with 0 constant coefficient. That is,

$$I = \{\ f = \sum_{i=0}^{n} a_i x^i \in R[x]\ :\ a_0 = 0\ \}.$$

Thus $f \in I$ if and only if $f = x * g$ for some $g \in R[x]$. Hence if $f_1, f_2 \in I$ and $h \in R[x]$, then

$$f_1 = x * g_1,\ f_2 = x * g_2$$

for some $g_1, g_2 \in R[x]$. Hence

$$f_1 - f_2 = x * g_1 - x * g_2 = x * (g_1 - g_2) \in I$$

and

$$f_1 * h = x * g_1 * h = x * (g_1 * h) \in I \text{ and } h * f_1 = h * x * g_1 = x * (h * g_1) \in I$$

It follows that I is an ideal in $R[x]$.

Example 5.5. The set of all polynomials in $\mathbb{Z}[x]$ with even coefficients is an ideal. So is the set of those polynomials with even constant coefficient.

Lemma 5.6. *Let I be an ideal of a ring R with identity. Then*

(i) *if $1 \in I$, then $I = R$;*

(ii) *if R is a division ring, then $I = R$ or $I = \{0\}$.*

Proof. **(i).** If $r \in R$, then $r = r * 1 \in I$. Hence $I = R$.

(ii). Assume that $I \neq \{0\}$. Then there exists $a \in I$ such that $a \neq 0$. Hence $1 = a^{-1} * a \in I$. It follows that from part (i) that $I = R$. ■

Corollary 5.7. *Let R be a commutative ring. Then $R[x]$ is not a field.*

Proof. From Example 5.4 we know that the polynomials with 0 constant coefficient form a proper ideal in $R[x]$. Hence by Lemma 5.6(ii), $R[x]$ is not a division ring and hence not a field. ■

Corollary 5.8. *Let R be a ring with identity. Then $(1) = \{\ r * 1\ :\ r \in R\ \} = R$.*

Example 5.9. Let R denote the ring of all 2×2 matrices with real entries and let I and J denote the sets of matrices of the form

$$\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}, \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

respectively. Now,

$$\begin{pmatrix} x & y \\ z & t \end{pmatrix} * \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} 0 & xb + yd \\ 0 & zb + td \end{pmatrix} \in I \text{ and } \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} - \begin{pmatrix} 0 & y \\ 0 & t \end{pmatrix} = \begin{pmatrix} 0 & b - y \\ 0 & d - t \end{pmatrix} \in I.$$

Thus I is a subring and $r * i \in I$ for all $i \in I$ and for all $r \in R$. But

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \notin I.$$

Hence it is not true that $i * r \notin I$ for all $i \in I$ and for all $r \in R$. It follows that I is not an ideal of R. It can also be shown that J is not an ideal of R. In fact, R has no proper ideals except $\{0\}$.

Definition 5.10. Let R be a commutative ring with identity and let $r_1, r_2, \ldots, r_n \in R$. Then define

$$(r_1, r_2, \ldots, r_n) = \{ \lambda_1 r_1 + \lambda_2 r_2 + \cdots + \lambda_n r_n \ : \ \lambda_1, \lambda_2, \ldots, \lambda_n \in R \}.$$

If $i = \lambda_1 r_1 + \lambda_2 r_2 + \cdots + \lambda_n r_n, j = \mu_1 r_1 + \mu_2 r_2 + \cdots + \mu_n r_n \in (r_1, r_2, \ldots, r_n)$, then

$$i - j = (\lambda_1 - \mu_1) r_1 + (\lambda_2 - \mu_2) r_2 + \cdots + (\lambda_n - \mu_n) r_n \in (r_1, r_2, \ldots, r_n).$$

If $x \in R$ is arbitrary, then

$$xi = ix = (x\lambda_1) r_1 + (x\lambda_2) r_2 + \cdots + (x\lambda_n) r_n \in (r_1, r_2, \ldots, r_n).$$

It follows that $(r_1, r_2, \ldots, r_n)$ is an ideal in R. We say that $(r_1, r_2, \ldots, r_n)$ is the *ideal generated by* $r_1, r_2, \ldots, r_n$.

If I and J are ideals of a ring R, then define $I + J = \{ i + j \ : \ i \in I, \ j \in J \}$.

If R did not have an identity, then $r_1, r_2, \ldots, r_n$ would not necessarily contain the elements of $(r_1, r_2, \ldots, r_n)$;

Example 5.11. We have already seen that the even integers (2) form an ideal in $\mathbb{Z}$. Moreover,

$$(2) = \{ 2n \ : \ n \in \mathbb{Z} \}$$

and so (2) is the ideal generated by 2.

The ideal (r) generated by a single element $r \in R$ is called the *principal ideal generated by* r. So, with this nomenclature the even integers are the principal ideal generated by 2.

Example 5.12. In Example 5.4 we showed that the set I of all polynomials in $R[x]$ with 0 constant coefficient is an ideal. We also saw that $f \in I$ if and only if $f = x * g$ for some $g \in R[x]$. It follows that

$$I = (x) = \{ x * g \ : \ g \in R[x] \}.$$

Lemma 5.13. *Let R be a commutative ring with identity, let I be an ideal of R, and let $r_1, r_2, \ldots, r_n \in I$. Then $(r_1, r_2, \ldots, r_n) \subseteq I$.*

Proof. As an exercise. ■

The previous lemma states that $(r_1, r_2, \ldots, r_n)$ is the least ideal containing $r_1, r_2, \ldots, r_n$.

Definition 5.14. A *principal ideal domain (PID)* is an integral domain where every ideal is prinicipal.

Lemma 5.15. *The ring of integers $\mathbb{Z}$ is a principal ideal domain.*

Proof. Let $a, b \in \mathbb{Z}$ be arbitrary. Then

$$(a, b) = \{ ax + by \ : \ x, y \in \mathbb{Z} \}.$$

It follows from the euclidean algorithm that $d = \gcd(a, b) \in (a, b)$. Hence $(d) \subseteq (a, b)$. But

$$(d) = \{ dc \ : \ c \in \mathbb{Z} \}$$

and so $a, b \in (d)$. It follows that $(a, b) = (d)$ and so $\mathbb{Z}$ is a principal ideal domain. ■

6. Quotients of rings

Definition 6.1. Let R be a commutative ring with identity, I be an ideal of R, and $a \in R$. Then the *coset* of I with *representative* x is the subset

$$a + I = \{\, a + s \;:\; s \in I \,\}$$

of R.

Theorem 6.2. *Let R be a commutative ring with identity and let I be an ideal of R. Then*

(i) *if $a, b \in R$, then $a + I = b + I$ if and only if $a - b \in I$;*

(ii) *any two cosets of I are either equal or disjoint: if $a, b \in R$, then either $a + I = b + I$ or $(a + I) \cap (b + I) = \emptyset$;*

(iii) *R is a disjoint union of the cosets of I;*

(iv) *if $a \in R$, then the map $r \mapsto a + r$ is a bijection from I to the coset $a + I$.*

Proof. For a proof see MT4003. ■

From Theorem 6.2(ii) and (iii) we can talk about the set of cosets an ideal I in a commutative ring with identity R; we will denote by R/I the set

$$\{\, a + I \;:\; a \in R \,\}$$

of cosets of the ideal I. The ring R/I is called a *factor ring* (or sometimes a *quotient ring* or *residue class ring*); we may also say that R/I is the quotient of R by I. We may refer to the factor ring R/I as R *modulo* I.

Theorem 6.3. *Let R be a commutative ring with identity, I be an ideal of R. Then R/I is a ring under the operations defined by*

$$(a + I) + (b + I) = (a + b) + I \text{ and } (a + I) * (b + I) = (a * b) + I.$$

Proof. We must prove that the operations $+$ and $*$ are well-defined and that $(R/I, +, *)$ satisfies the ring axioms. To show that $+$ and $*$ are well-defined we must prove that if $a_1 + I = a_2 + I$ and $b_1 + I = b_2 + I$, then

$$(a_1 + b_1) + I = (a_2 + b_2) + I$$

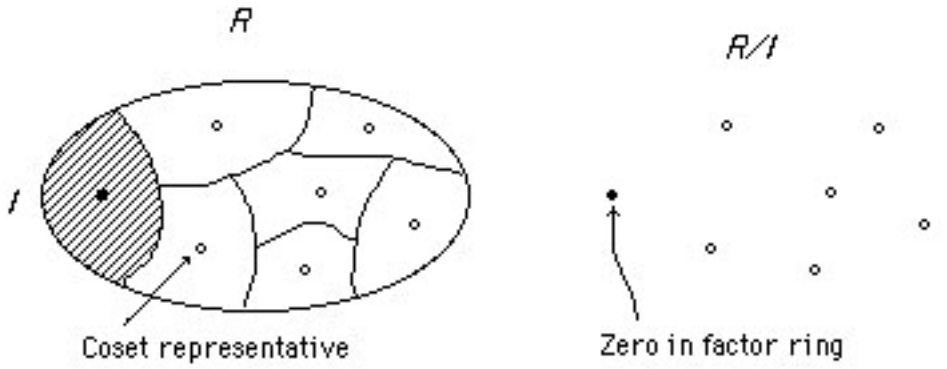
and

$$(a_1 * b_1) + I = (a_2 * b_2) + I.$$

From the direct implication of Theorem 6.2(i), $a_1 - a_2 \in I$ and $b_1 - b_2 \in I$. Hence $(a_1 + b_1) - (a_2 + b_2) \in I$ and from the converse implication of Theorem 6.2(i) it follows that $(a_1 + b_1) + I = (a_2 + b_2) + I$. Likewise $a_1 b_1 - a_2 b_2 = (a_1 - a_2)b_1 + a_2(b_1 - b_2) \in I$ since $a_1 - a_2, b_1 - b_2 \in I$ and so $(a_1 * b_1) + I = (a_2 * b_2) + I$.

It is a straightforward exercise to prove that R/I satisfies the ring axioms. ■

The way you should think of the factor ring R/I is that it is the ring R where all the elements in the ideal I have been 'made' into zero. Here are two picture:



Example 6.4. Let $n \in \mathbb{Z}$ and let (n) be the principal ideal generated by n. Then we can form the quotient $\mathbb{Z}/(n)$ of $\mathbb{Z}$ by (n). Two elements $x, y \in \mathbb{Z}$ are representatives of the same coset in $\mathbb{Z}/(n)$ if and only if $x - y \in (n)$ if and only if $n \mid (x - y)$ if and only if $x \equiv y \pmod{n}$. It follows that the cosets in $\mathbb{Z}/(n)$ are

$$0 + (n), 1 + (n), \ldots, n - 1 + (n)$$

and we add and multiply the representatives of cosets modulo n. For the sake of brevity, we omit the (n) when referring to elements of $\mathbb{Z}/(n)$ and we get back to the elements $\{0, 1, 2, \ldots, n-1\}$ modulo n!

The moral of the story: you've been working with factor rings since the start of the course!

We will now consider another very important class of factor rings. Recall that if p is a prime, we may write $\mathbb{F}_p$ to mean $\mathbb{Z}/(p)$.

Example 6.5. Let $f = x$. Then the elements of $\mathbb{F}_2[x]/(f)$ are the cosets

$$g + (f)$$

where $g \in \mathbb{F}_2[x]$. The zero of $\mathbb{F}_2[x]/(f)$ is the coset

$$0 + (f) = x + (f).$$

So, if $g \in \mathbb{F}_2[x]$, then

$$g + (f) \in \{0 + (f), 1 + (f)\},$$

as whenever an x appears in g it can be replaced by 0. For example,

$$x^4 + x^2 + 1 + (f) = 0 + 0 + 1 + (f) = 1 + (f).$$

Example 6.6. Let $f = x^2 + x + 1$. Then the elements of $\mathbb{F}_2[x]/(f)$ are

$$0 + (f), 1 + (f), x + (f), x + 1 + (f)$$

as again whenever x^2 appears it can be replaced by $x + 1$. For example,

$$[1 + x + (f)] * [1 + x + (f)] = 1 + 2x + x^2 = x.$$

The multiplicative inverses of $1, x, 1 + x$ are $1, 1 + x, x$, respectively. So, $\mathbb{F}_2[x]/(f)$ is a field with 4 elements.

We will return to the study of polynomial factor rings later in the course.

7. Homomorphisms

Just as in many branches of mathematics, functions that preserve the structure of a ring are very important.

Definition 7.1. A map $f : R \to S$ between rings is called a *ring homomorphism* if

$$f(x+y) = f(x) + f(y) \text{ and } f(x*y) = f(x) * f(y) \text{ for all } x, y \in R.$$

Note that $+$ in $f(x+y)$ is the operation in R and $+$ in $f(x)+f(y)$ is the operation in S. Likewise, $*$ is the operation in R in $f(x*y)$ and in S in $f(x)*f(y)$.

Let R and S be rings and let $f : R \to S$ be a homomorphism. If f is a bijection, then it is calle a *ring isomorphism*. We say that the rings R and S *isomorphic* and write $R \cong S$. Isomorphic rings have identical ring-theoretic properties, that is, commutativity, the existence of an identity, being a field etc. Any pair of isomorphic rings can be regarded as the same.

Lemma 7.2. *Let $f : R \to S$ be a homomorphism from the ring R to the ring S. Then*

(i) *if 0_R and 0_S are the zeros in R and S, respectively, then $f(0_R) = 0_S$;*

(ii) *if 1_R and 1_S are the identities of R and S, respectively, then it is not always true that $f(1_R) = 1_S$;*

(iii) *if T is a ring and $g : S \to T$ is a ring homomorphism, then so is $g \circ f : R \to T$.*

Proof. As an exercise. ■

Example 7.3. Let $f : \mathbb{Z} \to \mathbb{Z}/(n)$ be defined by $f : x \mapsto x \pmod n$. Then

$$f(i+j) = i + j \pmod n \equiv i \pmod n + j \pmod n = f(i) + f(j)$$

and

$$f(i*j) = i * j \pmod n = i \pmod n * j \pmod n = f(i) * f(j)$$

for all $i, j \in \mathbb{Z}$. It is not (of course) a ring isomorphism since it is not injective.

Example 7.4. Let $f : \mathbb{Z} \to (2)$ be defined by $f(x) = 2x$. Then

$$f(i+j) = 2(i+j) = 2i + 2j = f(i) + f(j)$$

for all $i, j \in \mathbb{Z}$. However,

$$f(2*2) = 8 \neq 16 = f(2) * f(2)$$

so f is not a homomorphism.

Example 7.5. Let $f : \mathbb{Z} \to M_2(\mathbb{R})$ be defined by

$$f(x) = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$f(x+y) = \begin{pmatrix} x+y & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} = f(x) + f(y)$$

and

$$f(x*y) = \begin{pmatrix} x*y & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} * \begin{pmatrix} y & 0 \\ 0 & 0 \end{pmatrix} = f(x) * f(y).$$

Note that

$$f(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

But the identity of $\mathbb{Z}$ is 0 and the $M_2(\mathbb{R})$ is

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

Example 7.6. The set $C[0,1]$ of real-valued continuous functions on the interval $[0,1]$ is a ring under the operations $+$ and $*$ defined by

$$(f+g)(x) = f(x) + g(x)$$

and

$$(f*g)(x) = f(x) * g(x).$$

The 'evaluation at $1/2$' map $\phi : C[0,1] \to \mathbb{R}$ defined by $\phi(f) = f(1/2)$ is a ring homomorphism since

$$\phi(f+g) = (f+g)(1/2) = f(1/2) + g(1/2) = \phi(f) + \phi(g)$$

and

$$\phi(f*g) = (f*g)(1/2) = f(1/2) * g(1/2) = \phi(f) * \phi(g).$$

Example 7.7. Let R denote the set of all real-valued matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

The map f from $\mathbb{C}$ to R defined by

$$a + bi \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

is a ring isomorphism. The mapping is clearly a bijection. Let $a+bi, c+di \in \mathbb{C}$. Then

$$f(a+bi) + f(c+di) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ -b-d & a+c \end{pmatrix} = f((a+bi)+(c+di))$$

and

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

Example 7.8. Let $\phi : \mathbb{Q}[x] \to \mathbb{Q}$ be defined by

$$\phi(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0 + a_1 + \cdots + a_n.$$

Then ϕ is a ring homomorphism.

Example 7.9. Let $\phi : \mathbb{Z}[x] \to \mathbb{R}$ given by

$$\phi(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0 + a_1\sqrt{2} + \cdots + a_n\sqrt{2}^n.$$

Then this is a ring homomorphism.

Definition 7.10. Let R and S be rings and let $f : R \to S$ be a homomorphism. Then the *kernel* of f is $\ker(f) = f^{-1}(0) = \{ r \in R \ : \ f(r) = 0_S \}$.

Lemma 7.11. *Let R and S be $f : R \to S$ be a ring homomorphism. Then $\ker(f)$ is an ideal of R.*
Proof. As an exercise. ■

Example 7.12. It was shown in Example 7.3 that the map from $\mathbb{Z}$ to $\mathbb{Z}/(n)$ defined by $f : x \mapsto x \mod n$ is a homomorphism. The kernel of f is the set (n) of all multiples of n.

Example 7.13. In Example 7.5 we showed that the mapping $f : \mathbb{Z} \to M_2(\mathbb{R})$ defined by

$$f(x) = \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}$$

is a homomorphism. The kernel of f is $\{0\}$.

Example 7.14. In Example 7.6 we proved that the function $\phi : C[0,1] \to \mathbb{R}$ defined by $\phi(f) = f(1/2)$ is a homomorphism. The kernel of ϕ is

$$\{ f \in C[0,1] \ : \ f(1/2) = 0 \}.$$

Example 7.15. In Example 7.7 we saw that the function $f : \mathbb{C} \to M_2(\mathbb{R})$ defined by

$$a + bi \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

is a homomorphism. If

$$f(a + bi) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

then $a = b = 0$. Hence $\ker(f) = \{0\}$.

Example 7.16. Let $\phi : \mathbb{R}[x] \to \mathbb{R}$ be defined by

$$\phi(a_0 + a_1 x + \cdots + a_n x^n) = a_0.$$

Then the kernel of ϕ is the set of polynomials with zero constant term. We proved that $\ker(\phi)$ is the principal ideal (x) in Example 5.12.

Example 7.17. Let $\phi : \mathbb{Z}[x] \to \mathbb{Z}_2$ be defined by

$$\phi(a_0 + a_1 x + \cdots + a_n x^n) = a_0 \pmod 2.$$

Then the kernel of ϕ is the set of polynomials with even constant terms.

Lemma 7.18. *Let R and S be rings and let $f : R \to S$ be a homomorphism. Then $\ker(f) = \{0\}$ if and only if f is injective.*
Proof. ($\Rightarrow$) We have to prove that f is injective. Let $f(r) = f(s)$ for some $r, s \in R$. Then $f(r - s) = f(r) - f(s) = f(r) - f(r) = 0$ and so $r - s \in \ker(f)$ and so $r - s = 0$. It follows that $r = s$ and so f is injective.
($\Leftarrow$) If $a \in \ker(f)$ and $a \neq 0$, then $f(a) = f(0) = 0$ and so f is not injective. ■

Let R and S be rings and let $f : R \to S$ be a homomorphism. Then the image of f is the set

$$\mathrm{im}(f) = \{ s \in S \ : \ s = f(r) \text{ for some } r \in R \}.$$

Lemma 7.19. *Let R and S be rings and let $f : R \to S$ be a homomorphism. Then $\mathrm{im}(f)$ is a subring of S.*
Proof. As an exercise. ■

8. Three Isomorphism Theorems

Theorem 8.1 (First Isomorphism Theorem) *Let R and S be rings and let $f : R \to S$ be a ring homomorphism. Then*

$$\phi : R/\ker(f) \cong \mathrm{im}(f), a + \ker(f) \mapsto f(a)$$

is an isomorphism of rings.

Proof. First we prove that ϕ is well-defined: If $a+\ker(f) = b+\ker(f)$, then $a-b \in \ker(f)$ and thus $0 = f(a-b) = f(a) - f(b)$ so $f(a) = f(b)$.

Next we will prove that ϕ is an isomorphism.

Homomorphism. Let $a+\ker(f), b+\ker(f) \in R/\ker(f)$ and remember that f is a homomorphism. Then

$$\begin{aligned}\phi(a+\ker(f)) + \phi(b+\ker(f)) = f(a) + f(b) = f(a+b) = \phi((a+b)+\ker(f)) \\ = \phi((a+\ker(f)) + (b+\ker(f))\end{aligned}$$

and

$$\phi(a+\ker(f))*\phi(b+\ker(f)) = f(a)*f(b) = f(a*b) = \phi((a*b)+\ker(f)) = \phi((a+\ker(f))*(b+\ker(f)).$$

Surjective. If $y \in \mathrm{im}(f)$, then there exists $x \in R$ such that $f(x) = y$. It follows that $y = f(x) = \phi(x + \ker(f))$ and so ϕ is surjective.

Injective. It is straightforward to prove this directly. Alternatively, $\phi(x+\ker(f)) = f(x) = 0$ if and only if $x \in \ker(f)$. Hence $\ker(\phi) = \{0 + \ker(f)\}$. But $0 + \ker(f)$ is the zero of $R/\ker(f)$ and so, by Lemma 7.18, ϕ is injective. ■

The preceding theorem and section indicate that there is a close relationship between ideals and kernels of homomorphisms.

Let R be a ring and I be an ideal of R. Then define $f : R \to R/I$ by $f(r) = r + I$. It is straightforward to verify that f is a homomorphism. It is called the "natural homomorphism" from R to R/I. Moreover, $\ker(f) = \{ r \in R : f(r) = I \} = \{ r \in R : r \in I \} = I$. It follows that every ideal I of a ring is the kernel of a homomorphism.

Furthermore, it is straightforward to verify that for every subring S of a ring R, the map $\iota : S \to R, s \mapsto s$ (the so-called "inclusion map") is a ring homomorphism.

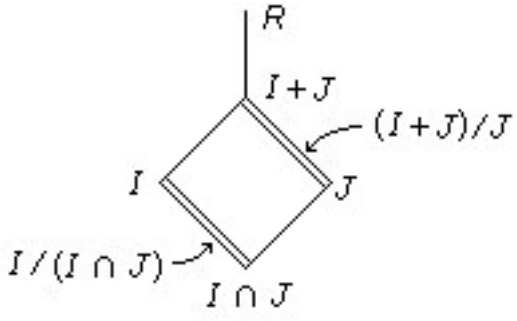
The name of the First Isomorphism Theorem indicates that there may be other isomorphism theorems. Here's one.

Theorem 8.2 (Second Isomorphism Theorem) *Let I and J be ideals of a ring R. Then*

$$(I + J)/J \cong I/(I \cap J).$$

Proof. We first note that the set $I + J := \{i + j \mid i \in I, j \in J\}$ is a subring of R and that $I \cap J$ is an ideal in I. To show these use the subring criterion Lemma 4.2 and the facts that an intersection of ideals is an ideal and that every ideal of R is an ideal in every subring in which it is contained.

Here is a picture showing the inclusions. The double lines represent the two factor rings.



To prove the result we will use Theorem 8.1. Define

$$f : I \to (I + J)/J, i \mapsto i + J.$$

The map f is the composition of the inclusion homomorphism $I \to I + J$ with the natural map $I + J \to (I + J)/J$ and as such a ring homomorphism.

All we need to show is that f is surjective and that its kernel is $I \cap J$. To prove that f is surjective, consider an arbitrary element $i + j \in I + J$, that is, $i \in I$ and $j \in J$. Since $(i + j) + J = i + J$ it follows, that $f(i) = i + J = (i + j) + J$. An element $i \in I$ is mapped to $0 + J$ by f if and only if it lies in J and therefore in $I \cap J$, so $\ker(f) = I \cap J$.

It follows by Theorem 8.1 that $I/\ker(f) \cong \mathrm{im}(f) = (I + J)/J$. ■

Example 8.3. Let $m, n \in \mathbb{Z}$ and consider the principal ideals (m) and (n). Then

$$(m) + (n) = (\gcd(m, n))$$

and

$$(m) \cap (n) = (\mathrm{lcm}(m, n)).$$

From Theorem 8.2 we deduce that

$$(\gcd(m, n))/(n) \cong (m)/(\mathrm{lcm}(m, n))$$

and so

$$n/\gcd(m, n) = \mathrm{lcm}(m, n)/m.$$

Hence

$$\gcd(m, n)\mathrm{lcm}(m, n) = mn.$$

Theorem 8.4 (Third Isomorphism Theorem) *Let I and J be ideals of a ring R with $I \subseteq J$. Then J/I is an ideal of R/I and*

$$(R/I)/(J/I) \cong R/J.$$

Proof. Let $j + I \in J/I$ and $r + I \in R/I$. Then $j * r, r * j \in J$ since J is an ideal. It follows that

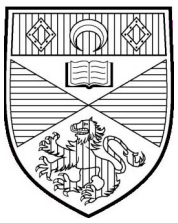
$$(j + I) * (r + I) = (j * r) + I \in J/I \text{ and } (r + I) * (j + I) = (r * j) + I.$$

Thus J/I is an ideal of R/I (it clearly is a subring!).

We will now prove that $(R/I)/(J/I) \cong R/J$. Define $\phi : R/I \to R/J$ by $a + I \to a + J$ for any coset $a + I$ of I in R. Since $a \in R$ is arbitrary it follows that ϕ is surjective. By the definition of $+$ and $*$ in R/I and R/J we have that ϕ is a homomorphism. An element $a + I$ is in the kernel of ϕ if and only if $a \in J$. Thus from the definition of J/I, $\ker(\phi) = J/I$ and

$$(R/I)/(J/I) = (R/I)/\ker(\phi) \cong \mathrm{im}(\phi) = R/J$$

by the First Isomorphism Theorem 8.1. ■



School of Mathematics and Statistics
MT4517 Rings & Fields
Factorization

We will also write ab instead of $a * b$ whenever there is no ambiguity! We will also use R^* to denote the set of all units in R. Throughout this section we will assume that all rings encountered are commutative with identity unless specified otherwise.

9. Prime Ideals

The purpose of this and the following section is to answer the questions: when is a factor ring R/I an integral domain? When is R/I a field?

Definition 9.1. An ideal I of a ring R is said to be *prime* if $I \neq R$ and if $a, b \in R$ with $ab \in I$, then either $a \in I$ or $b \in I$.

Theorem 9.2. *Let I be an ideal of a ring R. Then I is a prime ideal if and only if R/I is an integral domain.*

Proof. ($\Leftarrow$) Since R/I is an integral domain, $R/I \neq \{0\}$ and so $R \neq I$. If $a, b \in R$ with $ab \in I$, then $ab + I = (a + I)(b + I) = 0 + I$. Hence since R/I is an integral domain either $a + I = 0 + I$ or $b + I = 0 + I$. Thus either $a \in I$ or $b \in I$ by Theorem 6.2(i) and so I is a prime ideal.

($\Rightarrow$) As an exercise. ■

10. Maximal Ideals

Definition 10.1. An ideal I of a ring R is said to be *maximal* if whenever $I \subsetneq J$ for some ideal J of R, then $J = R$.

Theorem 10.2. *Let I be an ideal of a ring R. Then I is a maximal ideal if and only if R/I is a field.*

Proof. ($\Leftarrow$) Since R/I is a field, $R/I \neq \{0\}$ and for all $a + I \in R/I$ there exists $b + I \in R/I$ such that $(a + I)(b + I) = ab + I = 1 + I$. Hence for all $a \notin I$ there exists $b \in R$ such that $ab - 1 \in I$.

Let J be an ideal of R where $I \subseteq J \subseteq R$. If $a \in J \setminus I$, then there exists $b \notin I$ such that $ab - 1 \in I \subseteq J$. But $ab \in J$ since $a \in J$ and J is an ideal. Hence $1 = ab - (ab - 1) \in J$ and so $J = R$. We have shown that I is a maximal ideal of R.

($\Rightarrow$) Let $a + I \in R/I$ such that $a + I \neq 0 + I$. Then $a \notin I$. Set

$$J = I + Ra = \{\, i + ra \,:\, i \in I,\ r \in R \,\}.$$

It is straightforward to verify that J is an ideal of R and that $I \subsetneq J$. Since I is maximal, it follows that $J = R$. Hence $1 \in J$ and so $1 = i + ra$ for some $i \in I$ and $r \in R$. Now,

$$1 + I = [i + ra] + I = ra + I = (r + I)(a + I)$$

It follows that $a + I$ is a unit in R/I and, since a was arbitrary, R is a field. ■

Let I be a maximal ideal. Then R/I is a field and hence an integral domain. Hence I is a prime ideal.

Example 10.3. The ring of integers $\mathbb{Z}$ is a principal ideal domain. Hence every ideal of $\mathbb{Z}$ is of the form (n) for some $n \in \mathbb{Z}$. If $p \in \mathbb{Z}$ is a prime number and $n \in \mathbb{Z}$ such that $(p) \subseteq (n)$, then $p \in (n)$ and so $n \mid p$. Thus $n = \pm 1$ or $n = \pm p$. It follows that $(n) = \mathbb{Z}$ or $(n) = (p)$, and so (p) is a maximal ideal.

On the other hand, if $a, b \in \mathbb{Z}$ with $a, b \neq \pm 1$, then $ab \in (ab)$ but $a \notin (ab)$ and $b \notin (ab)$. Hence (ab) is not a prime ideal and hence not a maximal ideal.

Corollary 10.4. *Let $n \in \mathbb{Z}$. Then the following are equivalent:*

(i) *n or $-n$ is a prime number;*

(ii) *(n) is a prime ideal;*

(iii) *(n) is a maximal ideal.*

Proof. The proof that (i) and (iii) are equivalent is contained in Example 10.3.

By Theorem 9.2 and Example 3.11, (n) is a prime ideal if and only if $\mathbb{Z}/(n)$ is an integral domain if and only if n is prime. ■

11. Divisors

Lemma 11.1. *Let R be an integral domain and let $a, b, c \in R$ with $c \neq 0$ and $ac = bc$. Then $a = b$.*

Proof. As $ac = bc$, we have that $(a - b)c = ac - bc = 0$. But R is an integral domain and so $a - b = 0$. Hence $a = b$, as required. ■

Throughout the remainder of this chapter we will assume, unless otherwise specified, that R is an integral domain.

Definition 11.2. An element b of R is said to be a *divisor* of $a \in R$ if there exists $c \in R$ such that $a = bc$.

If b is a divisor of a, we write $b \mid a$. Note that $a \mid a$ and if $a \mid b$ and $b \mid c$, then $a \mid c$. Also note that if $0 \mid a$, then $a = 0$, and that $1 \mid a$ for all $a \in R$.

Definition 11.3. Elements $a, b \in R$ are said to be *associates* if $a \mid b$ and $b \mid a$. Throughout we will denote this by $a \sim b$.

The only associate of 0 is 0. Let $u \in R$ be arbitrary. Then u is a unit if and only if there exists $v \in R$ such that $uv = 1$ if and only if $u \mid 1$.

Lemma 11.4. *Let $a, b \in R$. Then $a \sim b$ if and only if $a = bu$ for some unit $u \in R$.*

Proof. ($\Rightarrow$) If $a = 0$, then $b = 0$ and $a = b1$. Likewise, if $b = 0$, then $a = 0$ and $a = b1$.

If $a, b \neq 0$, then as $a \sim b$, $a \mid b$ and $b \mid a$. Hence there exist $c, d \in R$ such that $b = ac$ and $a = bd$. Thus $a = acd$ and so $cd = 1$. Therefore both c and d are units.

($\Leftarrow$) If $a = 0$ or $b = 0$, then $b = 0$ or $a = 0$, respectively, and so $a \sim b$. If $a, b \neq 0$, then $a = bu$ and so $b \mid a$. It follows that $b = au^{-1}$ and so $a \mid b$. ■

Lemma 11.5. *If $a, b \in R$, then*

(i) *$b \mid a$ if and only if $(a) \leqslant (b)$;*

(ii) *a and b are associates if and only if $(a) = (b)$.*

Proof. **(i).** $b \mid a$ if and only if there exists $x \in R$ such that $a = bx$ if and only if $a \in (b)$ if and only if $(a) \leqslant (b)$.

(ii). a and b are associates if and only if $a \mid b$ and $b \mid a$ if and only if $(a) \leqslant (b)$ and $(b) \leqslant (a)$ if and only if $(a) = (b)$. ■

Example 11.6. The units in $\mathbb{Z}$ are $\mathbb{Z}^* = \{-1, 1\}$. The associates of $a \in \mathbb{Z}$ are a and $-a$.

Definition 11.7. The element d of an integral domain R is said to be the *greatest common divisor* of $a, b \in R$ if

(i) $d \mid a$ and $d \mid b$; and

(ii) if $c \mid a$ and $c \mid b$, then $c \mid d$.

If R is a principal ideal domain and $a, b \in R$, then there exists $d \in R$ such that $(a, b) = \{\, ax + by \;:\; x, y \in R \,\} = (d)$. Hence $(a) \leqslant (d)$ and $(b) \leqslant (d)$, and so, by Lemma 11.5, $d \mid a$ and $d \mid b$. On the other hand, if $c \in R$ divides a and b, then $(a) \leqslant (c)$ and $(b) \leqslant (c)$. Thus $(d) = (a, b) \leqslant (c)$ and so $c \mid d$. It follows that d is a greatest common divisor of a and b.

12. Irreducibles and primes

Definition 12.1. An element $a \in R \setminus R^*$ is *irreducible* if $a = bc$ for $b, c \in R$ implies that b or c is a unit.

Example 12.2. The irreducibles elements of $\mathbb{Z}$ are the primes and their negatives.

Definition 12.3. An element $p \in R \setminus R^*$ is said to be *prime* if whenever $p \mid ab$ for some $a, b \in R$, we have that $p \mid a$ or $p \mid b$.

If q is an associative of a prime p and $q \mid ab$, then $p \mid ab$ and so $p \mid a$ or $p \mid b$. It follows that $q \mid a$ or $q \mid b$. Hence the associates of primes are also primes.

Lemma 12.4. *Every prime element of an integral domain R is irreducible.*

Proof. Let $p \in R \setminus R^*$ be a prime element and assume that $p = ab$ for some $a, b \in R$. Then $p \mid ab$ and so $p \mid a$ or $p \mid b$. If $p \mid a$, then $a = px$ for some $x \in R$. But then $a = abx$ and so $bx = 1$ and b is a unit. Likewise, if $p \mid b$, then a is a unit. Hence p is irreducible. ■

We require the following notion so that we can give some specific examples of irreducible elements. A function $N : R \longrightarrow \mathbb{Z}$ is called a *norm* if

$$N(ab) = N(a)N(b)$$

for all $a, b \in R$.

Example 12.5. Let n be a non-square in $\mathbb{Z}$ (i.e. $n \neq m^2$ for all $m \in \mathbb{Z}$) and let $\mathbb{Z}[\sqrt{n}] = \{ a+b\sqrt{n} : a, b \in \mathbb{Z} \}$. Then we define $N : \mathbb{Z}[\sqrt{n}] \longrightarrow \mathbb{Z}$ by

$$N(a + b\sqrt{n}) = a^2 - nb^2.$$

It is straightforward to verify that

$$N(a + b\sqrt{n})N(c + d\sqrt{n}) = N((a + b\sqrt{n})(c + d\sqrt{n}))$$

and so N is a norm on $\mathbb{Z}[\sqrt{n}]$.

Lemma. Let n be a non-square in $\mathbb{Z}$. Then $\mathbb{Z}[\sqrt{n}]$ is an integral domain.

Proof. As an exercise. ■

What goes wrong if n in the previous lemma and example is not a non-square?

Example 12.6. We will show that the element $2 + \sqrt{-5}$ is irreducible in

$$\mathbb{Z}[\sqrt{-5}] = \{ a + b\sqrt{-5} : a, b \in \mathbb{Z} \}.$$

Let us start by determining the units of $\mathbb{Z}[\sqrt{-5}]$. If $\alpha \in \mathbb{Z}[\sqrt{-5}]$ is a unit, then there exists $\beta \in \mathbb{Z}[\sqrt{-5}]$ such that $\alpha\beta = 1$. Hence

$$N(\alpha)N(\beta) = N(\alpha\beta) = N(1) = 1.$$

Hence $N(\alpha)|1$ and so $N(\alpha) = \pm 1$. Hence $\alpha = \pm 1$.

If $a+b\sqrt{-5} \mid 2+\sqrt{-5}$, then $N(a+b\sqrt{-5}) \mid N(2+\sqrt{-5}) = 9$. Hence $N(a+b\sqrt{-5}) = 1, 3$, or 9. If $N(a+b\sqrt{-5}) = 1$, then $a = \pm 1$ and $b = 0$. That is, $a+b\sqrt{-5}$ is a unit. If $N(a+b\sqrt{-5}) = 3$, then there exist $a, b \in \mathbb{Z}$ such that $a^2 + 5b^2 = 3$. But there are no integer solutions to this equation. If $N(a + b\sqrt{-5}) = 9$, then $(a + b\sqrt{-5})(c + d\sqrt{-5}) = 9$ for some $c + d\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$. Hence $N(c + d\sqrt{-5}) = 1$ and so, as above, $c + d\sqrt{5}$ is a unit. It follows that $2 + \sqrt{-5}$ is irreducible.

Note that the element $2 + \sqrt{-5}$ is not prime in $\mathbb{Z}[\sqrt{-5}]$ since

$$2 + \sqrt{-5} \mid 9 = 3 \cdot 3$$

but $N(2 + \sqrt{-5}) = 9 = N(3)$ and $3 \neq \pm 1 \cdot (2 + \sqrt{-5})$ and so $2 + \sqrt{-5}$ does not divide 3.

Example 12.7. Let $\mathbb{Z}[\sqrt{10}] = \{ a + b\sqrt{10} : a, b \in \mathbb{Z} \}$. Then the norm on $\mathbb{Z}[\sqrt{10}]$ is defined by

$$N(a + b\sqrt{10}) = a^2 - 10b^2.$$

The norms of 2, 3, $(4 + \sqrt{10})$, $(4 - \sqrt{10}) \in \mathbb{Z}[\sqrt{10}]$ are

$$N(2) = 4,\ N(3) = 9, N(4 + \sqrt{10}) = N(4 - \sqrt{10}) = 6.$$

If $a + b\sqrt{10} \in \mathbb{Z}[\sqrt{10}]$, then

$$N(a + b\sqrt{10}) \pmod{10} = a^2 \pmod{10}.$$

But $x^2 \pmod{10} \in \{0, 1, 4, 5, 6, 9\}$ for all $x \in \mathbb{Z}$ and so $N(a + b\sqrt{10}) \pmod{10} \in \{0, 1, 4, 5, 6, 9\}$. Thus $N(a + b\sqrt{10}) \neq \pm 2$ or ± 3. It follows that 2, 3, $(4 + \sqrt{10})$, and $(4 - \sqrt{10})$ are irreducible in $\mathbb{Z}[\sqrt{10}]$.

On the other hand,

$$6 = 2 \cdot 3 = (4 + \sqrt{10})(4 - \sqrt{10})$$

and $2 \nmid (4 + \sqrt{10})$ and $2 \nmid (4 - \sqrt{10})$ as $N(2) = 4 \nmid 6 = N(4 + \sqrt{10}) = N(4 - \sqrt{10})$. Hence 2 is not prime and by similar arguments neither are 3, $4 + \sqrt{10}$, and $4 - \sqrt{10}$.

Example 12.8. Let $\mathbb{Z}[i]$ denote the Gaussian integers. Then the norm on $\mathbb{Z}[i] = \mathbb{Z}[\sqrt{-1}]$ is defined by by $N(a+bi) = a^2 + b^2$. Let p be a prime in $\mathbb{N}$ such that there exist $a, b, c \in \mathbb{Z}$ with $\gcd(p, c) = 1$ and

$$a^2 + b^2 = cp.$$

(For example, $a = 3$, $b = 5$, $c = 2$ and $p = 17$.) Then we will prove that p is not prime in $\mathbb{Z}[i]$.

Seeking a contradiction assume that p is prime in $\mathbb{Z}[i]$. Now,

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

and $p \mid cp$. Therefore $p \mid (a+bi)$ or $p \mid (a-bi)$. If $p \mid (a+bi)$, then $a + bi = (u+vi)p$ for some $u, v \in \mathbb{Z}$ and so $a = pu$ and $b = pv$. Thus $p \mid (a-bi)$. Hence $p^2 \mid (a+bi)(a-bi) = cp$ and so $p \mid c$. But $\gcd(p, c) = 1$, a contradiction.

13. Factorization domains

Definition 13.1. Let $x \in R \setminus R^*$ be an arbitrary non-zero element. Then x is said to have a *factorization into irreducibles* if there exists irreducible elements $p_1, p_2, \ldots, p_m \in R$ such that $x = p_1 p_2 \cdots p_m$.

Definition 13.2. An integral domain is said to be a *factorization domain* if every non-zero element in $R \setminus R^*$ has a factorization into irreducibles.

Example 13.3. The integers $\mathbb{Z}$ are a factorization domain, as every element can be given as a product of primes and primes are irreducible. If n is a non-square integer, then the integral domain $\mathbb{Z}[\sqrt{n}]$ is a factorization domain (Prove it as an exercise!).

Let $x \in R \setminus R^*$. Then x is said to have a *unique factorization* into irreducibles if for any two factorizations of x into irreducibles $p_1 p_2 \cdots p_m$ and $q_1 q_2 \cdots q_n$ we have that for all i there exists j such that $p_i \mid q_j$. That is, for all i there exists j such that $q_j = p_i u_i$ for some unit u_i. In particular, $m = n$.

Example 13.4. In the integers $\mathbb{Z}$, the following factorizations of 6 has a unique factorization into irreducibles even though

$$6 = 3 \cdot 2 = 2 \cdot 3 = (-2) \cdot (-3).$$

Definition 13.5. An integral domain R is a *unique factorization domain* if every non-zero element of $R \setminus R^*$ has unique factorization into irreducibles.

Theorem 13.6. *Let R be a factorization domain. Then R is a unique factorization domain if and only if every irreducible element in R is a prime.*

Proof. ($\Rightarrow$) Let $a \in R$ be an irreducible and assume that $a \mid bc$ for some $b, c \in R$. Then we must prove that $a \mid b$ or $a \mid c$.

Since $a \mid bc$, there exists $x \in R$ such that $ax = bc$. If $x \in R^*$, then, by Exercise 5.3, ax is irreducible. Hence ax is not a unit and so bc is not a unit either. If either b or c is a unit, then $b^{-1}ax = c$ or $axc^{-1} = b$ and so $a \mid c$ and $a \mid b$, respectively.

If neither b nor c is a unit, then there exist factorizations into irreducibles for both b and c. But a is an irreducible and so, by unique factorization, a is a divisor of one of the irreducibles dividing b or c. Hence a divides b or c, as required.

($\Leftarrow$) Let $x \in R \setminus R^*$ be an arbitrary non-zero element and let

$$x = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n$$

be factorizations of x into irreducibles. Since every irreducible element is also a prime element, $p_1, p_2, \ldots, p_m$ are prime elements of R. Hence for all i as $p_i \mid q_1 q_2 \cdots q_n$ it follows that $p_i \mid q_j$ for some j, as required. ■

Example 13.7. The integers $\mathbb{Z}$ are a unique factorization domain because the irreducible and prime elements coincide.

Example 13.8. We saw in Example 12.6 that in $\mathbb{Z}[\sqrt{-5}]$ the element $2 + \sqrt{-5}$ is irreducible but not prime. It follows that $\mathbb{Z}[\sqrt{-5}]$ is not a unique factorization domain.

Likewise in Example 12.7 we saw that in $\mathbb{Z}[\sqrt{10}]$ the element 2 is irreducible but not prime. It follows that $\mathbb{Z}[\sqrt{10}]$ is not a unique factorization domain.

What about the Gaussian integers? We will see later on that they are a unique factorization domain.

Lemma 13.9. *Let R be a principal ideal domain. Then R is a factorization domain.*

Proof. To prove this lemma we require the following. If $a_1, a_2, \ldots \in R$ are such that $(a_1) \subseteq (a_2) \subseteq \cdots$, then by Exercise 3.12 the union

$$I = \bigcup_{i=1}^{\infty} (a_i)$$

is an ideal of R. It follows that $I = (d)$ for some $d \in R$. From the definition of I, there exists $N \in \mathbb{N}$ such that $d \in (a_N)$. Thus $I = (d) \subseteq (a_n)$ for all $n \geq N$. That is, $(a_N) = (a_{N+1}) = (a_{N+2}) = \cdots = I$.

Let $r_0 \in R \setminus R^*$ be a non-zero element. We must prove that r_0 has a factorization into irreducibles. Assume the contrary. Then r_0 is not an irreducible. So there exist $r_1, s_1 \in R \setminus R^*$ such that $r_0 = r_1 s_1$. Thus $r_1 \mid r_0$ and $s_1 \mid r_0$ and so $(r_0) \subseteq (r_1)$ and $(r_0) \subseteq (s_1)$. If $(r_0) = (r_1)$, say, then $r_0 \sim r_1$ and so there exists a unit $u \in R^*$ such that $r_1 = r_0 u$. But then $r_0 = r_1 s_1 = r_0 u s_1$ and so $us_1 = 1$, or in other words $s_1 \in R^*$, a contradiction. Hence $(r_0) \neq (r_1)$ and $(r_0) \neq (s_1)$.

If both r_1 and s_1 had factorizations into irreducibles, then so would r_0. Hence, say, r_1 has no factorization into irreducibles. Then repeating the above argument replacing r_0 with r_1 we find $r_2, s_2 \in R \setminus R^*$ such that $r_1 = r_2 s_2$, $(r_1) \subsetneq (r_2)$, and $(r_1) \subsetneq (s_2)$. Again not both r_2 and s_2 can have factorizations into irreducibles, and so we can assume that r_2 has no such factorization.

Continuing in this way we obtain

$$(r_0) \subseteq (r_1) \subseteq \cdots$$

But $(r_i) \neq (r_{i+1})$ for all $i \in \mathbb{N}$, a contradiction to the first part of the proof. ■

Theorem 13.10. *Let R be a principal ideal domain that is not a field and let $r \in R$. Then the ideal (r) is maximal if and only if r is an irreducible element of R.*

Proof. ($\Leftarrow$) Let (s) be any ideal of R with $(r) \leqslant (s)$. Then $r = sx$ for some $x \in R$. But r is irreducible and so either s or x is a unit. If s is a unit, then $(s) = R$. If x is a unit, then $r \sim s$ and so $(r) = (s)$. Hence (r) is a maximal ideal.

($\Rightarrow$) Assume that $r = ab$ for some $a, b \in R$. We must prove that a or b is a unit. Assume the contrary that neither a nor b is a unit. Then since $a \mid r$ we have that $(r) \leqslant (a)$. If $(r) = (a)$, then $r \mid a$ and so there exists $x \in R$ such that $a = rx$. Hence $a = rx = abx$ and so $bx = 1$ and b is a unit, a contradiction. Hence $(r) \neq (a)$. On the other hand, a is not a unit and so $(a) \neq R$. Hence $(r) \lneq (a) \lneq R$, a contradiction to the assumption that (r) is maximal. ■

Corollary 13.11. *Let R be a principal ideal domain that is not a field. Then an ideal in R is maximal if and only if it is prime.*

Proof. We saw earlier that every maximal ideal is prime. Hence it suffices to prove that every prime ideal is maximal.

Let (r) be a prime ideal of R with $(r) \neq \{0\}$. Recall from the definition of a prime ideal that $(r) \neq R$. If $r \in R^*$, then $1 \in (r)$ and so $(r) = R$. Hence $r \notin R^*$. Since R is a factorization domain, it follows that there exist irreducibles $x_1, x_2, \ldots, x_n \in R$ such that

$$r = x_1 x_2 \cdots x_n.$$

But (r) is prime and so $x_i \in (r)$ for some i. Now, by Theorem 13.10, (x_i) is maximal in R. But certainly $(x_i) \subseteq (r) \neq R$ and so $(x_i) = (r)$. Hence (r) is a maximal ideal, as required. ■

Theorem 13.12. *Let R be a principal ideal domain. Then R is a unique factorization domain.*

Proof. We proved in Lemma 13.9 that R is a factorization domain. By Theorem 13.6 it suffices to prove that every irreducible element in R is prime. Let $x \in R$ be an irreducible element such that $x \mid ab$ for some $a, b \in R$. Then $(ab) \leqslant (x)$ and in particular, $ab \in (x)$. But (x) is a maximal ideal by Theorem 13.10 and hence a prime ideal by Corollary 13.11. It follows that either $a \in (x)$ or $b \in (x)$ and so $(a) \leqslant (x)$ or $(b) \leqslant (x)$, respectively. Therefore $x \mid a$ or $x \mid b$ and so x is prime. ■

The converse of Theorem 13.12 is not true.

Example 13.13. We saw earlier that the rings $\mathbb{Z}[\sqrt{-5}]$ and $\mathbb{Z}[\sqrt{10}]$ are not unique factorization domains. Hence they are not principal ideal domains either.

14. Euclidean rings

Let R be a unique factorization domain and let $a, b \in R \setminus R^*$. Then a and b can be given as a product of irreducible (and prime) elements $x_1, x_2, \ldots, x_k \in R$ as follows

$$a = x_1^{m_1} x_2^{m_2} \cdots x_k^{m_k}$$

$$b = x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$$

where $m_i, n_i \geq 0$. Then the greatest common divisor of a and b is

$$c = x_1^{p_1} x_2^{p_2} \cdots x_k^{p_k}$$

where $p_i = \min(m_i, n_i)$ for all i. However, except for very small examples, even in the integers it exceeds human patience and soon after the ability of a computer to use this method to find $\gcd(a, b)$. In the integers we use the division algorithm and the Euclidean algorithm instead.

Definition 14.1. Let R be an integral domain. Then a function $N : R \setminus \{0\} \longrightarrow \mathbb{N}$ satisfying the following is called a *Euclidean function*: for all $a \in R$ and $b \in R \setminus \{0\}$ there exist $q, r \in R$ such that

$$a = qb + r$$

and either $r = 0$ or $N(r) < N(b)$.

If R is an integral domain such that there exists a Euclidean function from $R \setminus \{0\}$ to $\mathbb{N}$, then R is called a *Euclidean ring*.

For example, by the Division Algorithm (Theorem 1.2), the integers $\mathbb{Z}$ have a Euclidean function $N : \mathbb{Z} \setminus \{0\} \longrightarrow \mathbb{N}$ defined by $N(x) = |x|$. Hence the integers are a Euclidean ring.

Theorem 14.2. *Let R be a Euclidean ring. Then R is a principal ideal domain.*

Proof. Let $N : R \setminus \{0\} \longrightarrow \mathbb{N}$ denote the Euclidean function. Let $I \neq \{0\}$ be an ideal in R and let $x \in I$ be any non-zero element where $N(x)$ is minimal in I. We will prove that $I = (x)$.

If $y \in I$ is any element, then there exists $q, r \in R$ such that $y = qx + r$ where $r = 0$ or $N(r) < N(x)$. But $N(x)$ is minimal and so $r = 0$. In other words, $y = qx \in (x)$, as required. ■

The converse of Theorem 14.2 does not hold. The ring $\mathbb{Z}[(1 + \sqrt{-19})/2]$ is an example of a principal ideal domain that is not a Euclidean ring. Proving that a ring R is Euclidean only requires us to find a Euclidean function for R. To prove that R is not Euclidean we must show that none of the functions from $R \setminus \{0\}$ to $\mathbb{N}$ is a Euclidean function.

14.1. Gaussian integers – the division algorithm

A norm is defined on the Gaussian integers $\mathbb{Z}[i]$ by

$$N(x + yi) = (x + yi)(x - yi) = |x + yi|^2 = x^2 + y^2.$$

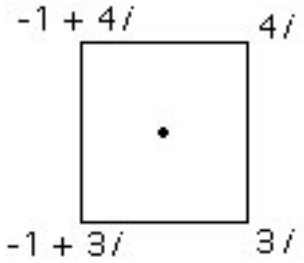
Recall from Exercise 5.4 that $N(\alpha)N(\beta) = N(\alpha\beta)$ for all $\alpha, \beta \in \mathbb{Z}[i]$.

Theorem 14.3. *$\mathbb{Z}[i]$ is a Euclidean ring.*

Proof. Let $N : \mathbb{Z}[i] \longrightarrow \mathbb{N}$ be defined by $N(x + yi) = x^2 + y^2$. We must prove that for all $\alpha \in \mathbb{Z}[i]$ and $\beta \in \mathbb{Z}[i] \setminus \{0\}$ there exist $q, r \in \mathbb{Z}[i]$ such that

$$\alpha = q\beta + r$$

and either $r = 0$ or $N(r) < N(\beta)$. So, α/β is an element of $\mathbb{Q}[i] = \{\, c + di \;:\; c, d \in \mathbb{Q} \,\} \subseteq \mathbb{C}$. We can then plot α/β in the complex plane, for example:



From this picture we deduce that we may find $q = c + di \in \mathbb{Z}[i]$ such that

$$N(\alpha/\beta - q) = N(\alpha\beta^{-1} - q) \leqslant \left(\frac{\sqrt{2}}{2}\right)^2 < 1$$

where N denotes the norm in $\mathbb{C}$, that is, $|a + bi| = a^2 + b^2$, $a, b \in \mathbb{R}$. Multiplying both sides of the inequality by $N(\beta)$ and using $N(\alpha)N(\beta) = N(\alpha\beta)$ we get

$$N(\alpha - q\beta) < N(\beta).$$

Hence N is a Euclidean function and $\mathbb{Z}[i]$ is a Euclidean ring. ■

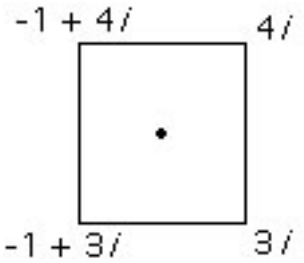
Corollary 14.4. *$\mathbb{Z}[i]$ is a principal ideal domain, and $\mathbb{Z}[i]$ is a unique factorization domain.*

Proof. This follows immediately from Theorems 14.2 and 13.12. ■

Example 14.5. We will find $q, r \in \mathbb{Z}[i]$ such that $3+4i = (1-i)q + r$ where $r = 0$ or $N(r) < N(1-i) = 2$. So, dividing $3+4i$ by $1-i$ in $\mathbb{C}$ we obtain

$$\frac{3+4i}{1-i} = \frac{3+4i}{1-i}\frac{1+i}{1+i} = \frac{-1}{2} + \frac{7i}{2} \in \mathbb{Q}[i].$$

Now, $-1 < -1/2 < 0$ and $3 < 7/2 < 4$ and so plotting $(-1+7i)/2$ in the plane gives:



And so the four possible quotients are $1-3i$, $1-4i$, $3i$, and $4i$. Now,

$$N\left(\frac{-1+7i}{2} - (1-3i)\right) = N\left(\frac{1+i}{2}\right) = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} < 1$$

$$N\left(\frac{-1+7i}{2} - (1-4i)\right) = N\left(\frac{1-i}{2}\right) = \frac{1}{2} < 1$$

$$N\left(\frac{-1+7i}{2} - 3i\right) = N\left(\frac{-1+i}{2}\right) = \frac{1}{2} < 1$$

$$N\left(\frac{-1+7i}{2} - 4i\right) = N\left(\frac{-1-i}{2}\right) = \frac{1}{2} < 1.$$

So, from the proof of Theorem 14.3, any of $1-3i$, $1-4i$, $3i$, and $4i$ can be used as the quotient q. If $q = -1+3i$, then

$$r = 3+4i-(-1+3i)(1-i) = 3+4i-(2+4i) = 1$$

and $N(r) = 1 < 2 = N(1-i)$, as required.

If $q = -1+4i$, then $r = -i$. If $q = 3i$, then $r = i$, and if $q = 4i$, then $r = -1$.

Note that if $\alpha, \beta \in \mathbb{Z}[i]$, then there are at most 4 quotients q and remainders r such that $\alpha = q\beta + r$ and $r = 0$ or $N(r) < N(\beta)$.

14.2. Polynomial rings – the division algorithm

Let us begin by recalling some definitions relating to rings. The degree of a polynomial $f = a_0 + a_1x + \cdots + a_nx^n \in R[x]$ is the largest n such that $a_n \neq 0$, denoted by $\deg(f)$, and where $\deg(0) = -\infty$. If R is an integral domain and $f, g \in R[x] \setminus \{0\}$, then

$$\deg(fg) = \deg(f) + \deg(g).$$

Lemma 14.6. *Let R be an integral domain. Then $f \in R[x]$ is a unit if and only if $f \in R$ is a unit.*

Proof. Note that if $f \in R[x]$, then $\deg(f) = 0$ if and only if $f \in R$. If $f \in R \subsetneq R[x]$ is a unit, then there exists $g \in R \subsetneq R[x]$ such that $fg = 1$. Hence f is a unit in $R[x]$.

Conversely, if $f \in R[x]$ is a unit, then there exists $g \in R[x]$ such that $fg = 1$. Hence $0 = \deg(1) = \deg(fg) = \deg(f) + \deg(g)$ and so $\deg(f) = \deg(g) = 0$. Thus $f \in R$ is a unit in R.

■improve the previous proof.

Example 14.7. Let $f = 3x^4 + 2x^3 + x^2 - 4x + 1, g = x^2 + x + 1 \in \mathbb{Z}[x]$. Then

$$\begin{array}{r|l} & 3x^2 - x - 1. \\ \hline x^2 + x + 1 & 3x^4 + 2x^3 + x^2 - 4x + 1 \\ & -3x^4 - 3x^3 - 3x^2 \\ \hline & -x^3 - 2x^2 - 4x \\ & x^3 + x^2 + x \\ \hline & -x^2 - 3x + 1 \\ & x^2 + x + 1 \\ \hline & -2x + 2 \end{array}$$

Here $q = 3x^2 - x - 1$ and $r = -2x + 2$.

Example 14.8. Let $f = 10x^5 - 2x^2 + 4x - 3, g = 3x^2 + 3x + 3 \in \mathbb{Q}[x]$. Then

$$\begin{array}{r|l} & \frac{10}{3}x^3 - \frac{10}{3}x^2 + \frac{8}{3}. \\ \hline 3x^2 + 3x + 3 & 10x^5 - 2x^2 + 4x - 3 \\ & -10x^5 - 10x^4 - 10x^3 \\ \hline & -10x^4 - 10x^3 - 2x^2 \\ & 10x^4 + 10x^3 + 10x^2 \\ \hline & 8x^2 + 4x - 3 \\ & -8x^2 - 8x - 8 \\ \hline & -4x - 11 \end{array}$$

In this case, $q = (10/3)x^3 - (10/3)x^2 + (8/3)$ and $r = -4x - 11$. ■

Theorem 14.9. *Let R be an integral domain and $f, g \in R[x]$ be arbitrary non-zero polynomials. Then there exist $q, r \in R[x]$ such that*

$$f = qg + r$$

and either $r = 0$ or none of the terms in r is divisible by the leading term of g.

Proof. Omitted. ■

Corollary 14.10. *Let F be a field and $f, g \in F[x]$ be arbitrary be arbitary non-zero polynomials. Then there exist unique $q, r \in F[x]$ such that*

$$f = qg + r$$

and either $r = 0$ or $\deg(r) < \deg(g)$.

Proof. By Theorem 14.9 there exist $q, r \in F[x]$ such that $f = qg + r$ and either $r = 0$ or none of the terms in r is divisible by the leading term of g. In the latter case, assume that $\deg(r) \geq \deg(g)$. Let $a_n x^n$ be the leading term of g. Then a_n is a unit (we're in a field!) and so divides every element of F. Since $\deg(r) \geq \deg(g)$, there exists $m \geq n$ such that $b_m x^m$ is a term in r and $b_m \neq 0$. Hence $a_n x^n$ divides $b_m x^m$, a contradiction. Hence $\deg(r) < \deg(g)$.

Assume that $f = q_1 g + r_1 = q_2 g + r_2$ for some $q_1, q_2, r_1, r_2 \in F[x]$ and $\deg(r_1), \deg(r_2) < \deg(g)$. Then $(q_1 - q_2)g = r_2 - r_1$. If $r_2 - r_1 \neq 0$, then $\deg(q_1 - q_2) + \deg(g) = \deg((q_1 - q_2)g) = \deg(r_2 - r_1) \leqslant \max\{\deg(r_2), \deg(r_1)\}$ and so $\deg(r_2 - r_1) \leqslant \deg(r_1)$ or $\deg(r_2 - r_1) \leqslant \deg(r_2)$, say the former. Hence $\deg(g) \leqslant \deg(q_1 - q_2) + \deg(g) \leqslant \deg(r_1)$, a contradiction. Thus $r_1 = r_2$ and so $q_1 = q_2$, as required. ■

The proof of Theorem 14.9 only relies on the fact that the leading coefficient of g is not a zero divisor. Hence it is possible to prove Theorem 14.9 under the weaker assumption that R is a commutative ring with one and the leading coefficient of g is not a zero divisor. Likewise, Corollary 14.10 holds when F is replaced with a commutative ring with one and the leading coefficient of g is invertible.

If R is an arbitrary ring, then it can be difficult to prove that $R[x]$ is a unique factorization domain, a Euclidean ring, or a principal ideal domain. Likewise it can be hard to determine what the units are or what the irreducible elements are.

Theorem 14.11 (Gauss) *Let R be a unique factorization domain. Then $R[x]$ is a unique factorization domain.*

Proof. Omitted. ■

Theorem 14.12. *Let F be a field. Then $F[x]$ is a Euclidean ring, and hence a principal ideal domain and a unique factorization domain.*

Proof. It follows from Corollary 14.10 that $F[x]$ the function $N : F[x] \setminus \{0\} \longrightarrow \mathbb{N}$ defined by $N(f) = \deg(f)$ is a Euclidean function. Hence $F[x]$ is a Euclidean ring and hence a principal ideal domain and a unique factorization domain. ■

15. Greatest common divisors again

Recall that an element d of an integral domain R is a *greatest common divisor* of $a, b \in R$ if

(i) $d \mid a$ and $d \mid b$; and

(ii) if $c \mid a$ and $c \mid b$, then $c \mid d$.

Let R be a principal ideal domain. Then we also proved that d is a $\gcd$ of $a, b \in R$ if and only if $(d) = (a, b)$. In particular, if R is a Euclidean ring, then $\gcd(a, b)$ exists for all $a, b \in R$.

The next example shows that the $\gcd$ of two elements of an integral domain does not always exist.

Example 15.1. $\mathbb{Z}[\sqrt{-5}]$ is an integral domain but not a unique factorization domain since

$$9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5})$$

and the elements $3, 2 + \sqrt{-5}$, and $2 - \sqrt{-5}$ are irreducible but not prime.

Let $a + b\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$ such that $a + b\sqrt{-5} \mid 9$. Then $N(a + b\sqrt{-5}) \mid N(9) = 81$ and so $N(a+b\sqrt{-5}) \in \{1, 3, 9, 27, 81\}$. If $N(a+b\sqrt{-5}) = 1$, then $a+b\sqrt{-5}$ is a unit and so $a+b\sqrt{-5} = \pm 1$. If $N(a + b\sqrt{-5}) = 3$, then $a^2 + 5b^2 = 3$, which is not possible. If $N(a + b\sqrt{-5}) = 9$, then $a^2 + 5b^2 = 9$ and so $a = \pm 3$ and $b = 0$ or $a = \pm 2$ and $b = \pm 1$, that is, $a + b\sqrt{-5} = \pm 3$ or $\pm(2 + \sqrt{-5})$. If $N(a + b\sqrt{-5}) = 81$, then $a + b\sqrt{-5} = \pm 9$. Hence the only divisors of 9 are

$$\pm 1, \pm 3, \pm(2 + \sqrt{-5}), \text{ and } \pm 9.$$

By a similar argument the only divisors of $3 \cdot (2 + \sqrt{-5})$ are

$$\pm 1, \pm 3, \pm(2 + \sqrt{-5}), \text{ and } \pm [3 \cdot (2 + \sqrt{-5})].$$

It follows that if d is a greatest common divisor of 9 and $3 \cdot (2 + \sqrt{-5})$, then $d = \pm 1$, $d = \pm 3$ or $d = \pm(2 + \sqrt{-5})$. Now, $d \neq \pm 1$, as $3 \mid 9$ and $3 \mid 3 \cdot (2 + \sqrt{-5})$ but $3 \nmid 1$.

If $d = \pm 3$, then as $2 + \sqrt{-5} \mid 9$ and $2 + \sqrt{-5} \mid 3 \cdot (2 + \sqrt{-5})$, we deduce that $(2 + \sqrt{-5}) \mid \pm 3$. But $(2 + \sqrt{-5}) \mid \pm 3$ implies that $\pm 3 = \alpha \cdot (2 + \sqrt{-5})$ where α is a unit, a contradiction. Likewise, if $d = \pm(2 + \sqrt{-5})$, then $3 \mid 2 + \sqrt{-5}$, a contradiction.

Hence the greatest common divisor of 9 and $3 \cdot (2 + \sqrt{-5})$ does not exist.

Theorem 15.2. *Let R be a factorization domain. Then R is a unique factorization domain if and only if* $\gcd(a,b)$ *exists for all $a, b \in R$.*

Proof. Omitted. ■

Let R be a Euclidean ring. Then the following algorithm can be used to find $\gcd(a,b)$ for all $a, b \in R$. Note that $d = \gcd(a,b)$ if and only if $(a,b) = (d)$. Moreover, if $(d') = (d) = (a,b)$, then $d' \sim d$.

Let $N : R \setminus \{0\} \longrightarrow \mathbb{N}$ be a Euclidean function and let $a, b \in R$ be arbitrary. If $a = 0$, then $\gcd(a,b) = b$ and if $b = 0$, then $\gcd(a,b) = a$. In either case, we have found the $\gcd$ and we can stop.

Assume without loss of generality that $a, b \neq 0$ and $N(a) \geq N(b)$. Set $a_0 = a$ and $a_1 = b$. Then there exists $q_0, a_2 \in R$ such that $a_0 = q_0 a_1 + a_2$ and $a_2 = 0$ or $N(a_2) < N(a_1)$. Hence $a_0 \in (a_1, a_2)$ and $a_2 = a_0 - q_0 a_1 \in (a_0, a_1)$ and so $(a_0, a_1) = (a_1, a_2)$. If $a_2 = 0$, then $\gcd(a,b) = \gcd(a_0, a_1) = \gcd(a_1, a_2) = a_1 = b$. Otherwise we repeatedly apply this procedure on all subsequent a_i and a_{i+1}, $i > 0$ until $a_j = 0$ for some j. This is guaranteed to happen as $N(a_0) > N(a_1) > N(a_2) > \cdots > 0$. It follows that $\gcd(a,b) = a_{j-1}$.

Example 15.3. Let $a_0 = 7 + 8i$ and $a_1 = 4 + 5i$. Then

$$\frac{a_0}{a_1} = \frac{7+8i}{4+5i}\frac{4-5i}{4-5i} = \frac{68-3i}{41}.$$

Plotting a_0/a_1 in the complex plane we note that the nearest elements of $\mathbb{Z}[i]$ are $1, 2, 1-i, 2-i$. Since

$$N\left(\frac{68-3i}{41} - 1\right) = \left(\frac{27}{41}\right)^2 + \left(\frac{3}{41}\right)^2 < 1,$$

we have that $7 + 8i = (4 + 5i) + (3 + 3i)$ and $N(3 + 3i) < N(4 + 5i)$.

Set $a_2 = 3 + 3i$ and repeat the previous steps on a_1 and a_2. Then

$$4 + 5i = (3 + 3i) + (1 + 2i)$$

and $N(1 + 2i) < N(3 + 3i)$.

Set $a_3 = 1 + 2i$ and repeat the previous steps on a_2 and a_3. Then

$$3 + 3i = (1 + 2i)2 + (1 - i)$$

and $N(1 - i) < N(1 + 2i)$.

Set $a_4 = 1 - i$ and repeat the previous steps on a_3 and a_4. Then

$$1 + 2i = (1 - i)i + i$$

and $N(i) < N(1 - i)$.

Set $a_5 = i$ and repeat the previous steps on a_4 and a_5. Then

$$1 - i = i(-i - 1) + 0.$$

Hence $\gcd(7 + 8i, 4 + 5i) = i$.

If d' is any other $\gcd$ of a_0 and a_1, then $d' \sim d$ and so there exists a unit u such that $d = d'u$. In $\mathbb{Z}[i]$ the units are ± 1 and $\pm i$ and so ± 1, $\pm i$ are the only $\gcd$s of a_0 and a_1.

Example 15.4. Let $a_0 = x^5 + x + 1, a_1 = x^4 + x^3 + x + 1 \in \mathbb{F}_2[x]$. Then using the division algorithm for polynomials we get:

$$a_0 = (x + 1)a_1 + (x^3 + x^2 + x).$$

Set $a_2 = x^3 + x^2 + x$ and divide again:

$$a_1 = x \cdot a_2 + (x^2 + x + 1).$$

Set $a_3 = x^2 + x + 1$ and divide again:

$$a_2 = x \cdot a_3 + 0.$$

Hence $\gcd(x^5 + x + 1, x^4 + x^3 + x + 1) = x^2 + x + 1$.

Note that if you calculate a gcd of (say) 2 and 4 as polynomials you can get the answer 1, while as integers the result is always 2.

16. Some nasty examples

In this section we present a few examples showing odd behaviour with respect to the theoretical things we did so far.

Example 16.1. Let $R := \mathbb{F}_2 \times \mathbb{F}_2$ with componentwise addition and multiplication. Then of course the zero is $(0, 0)$ and the identity is $(1, 1)$. The units are $\{(1, 1)\}$.

However, R does not have irreducible elements at all: $(1, 0) = (1, 0) \cdot (1, 0)$, where both factors are no units.

In particular, R is not a factorization domain, since no element is a product of irreducibles.

Example 16.2. Let $R := \mathbb{Z} \times \mathbb{Z}$ with componentwise addition and multiplication. Then of course the zero is $(0, 0)$ and the identity is $(1, 1)$. The units are $\{(1, 1), (-1, 1), (1, -1), (-1, -1)\}$.

The ring R does have irreducible elements, namely (up to units) the set $\{(p, 1) : p \text{ a prime}\} \cup \{(1, p) : p \text{ a prime}\}$. Any other pair (m, n) with $m, n \neq 0$ that is not a unit can be written as a product those.

However, R is no factorization domain, since again $(1, 0) = (1, 0) \cdot (1, 0)$, where both factors are no units, and $(1, 0)$ (and indeed all $(m, 0)$ and $(0, n)$ for $m, n \neq 0$) cannot be written as a product of irreducible elements.

The previous two examples failed to be factorization domains because of a lack of irreducible elements. Now we are looking for infinite chains of principal ideals.

Example 16.3. Let $R := \mathbb{Z}^{\mathbb{Z}}$, by which we mean the set of all functions from $\mathbb{Z}$ to $\mathbb{Z}$. We denote elements of R by two-sided sequences $(a_i)_{i \in \mathbb{Z}}$ where all $a_i \in \mathbb{Z}$. We add and multiply componentwise:

$$[(a_i)_{i \in \mathbb{Z}}] + [(b_i)_{i \in \mathbb{Z}}] := (a_i + b_i)_{i \in \mathbb{Z}} \text{ and } [(a_i)_{i \in \mathbb{Z}}] * [(b_i)_{i \in \mathbb{Z}}] := (a_i \cdot b_i)_{i \in \mathbb{Z}}$$

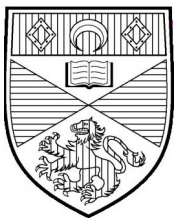
Similarly to the previous example, the units are all elements in which all components are in $\{\pm 1\}$. All elements (up to units) with a prime in one position and ± 1 in all others are irreducible and these are all irreducibles.

As soon as at least one component of $a = (a_i)_{i \in \mathbb{Z}}$ is equal to 0, the element a is not a product of irreducibles as before.

Now take the elements $b^{(n)} = (b_i^{(n)})_{i\in\mathbb{Z}}$ where $b_i^{(n)} = 0$ for $i >= n$ and $b_i^{(n)} = 1$ for $i < n$. The principal ideal generated by $b^{(n)}$ is:

$$(b^{(n)}) = \{a = (a_i)_{i\in\mathbb{Z}} \in R : a_i = 0 \text{ for } i \geq n\}.$$

Thus, we have for $m < n$ that $(b^{(m)}) \subsetneq (b^{(n)})$ and so there is an infinite ascending chain of principal ideals.



School of Mathematics and Statistics
MT4517 Rings & Fields
Part 4 - Finite Fields

16. The classification of finite fields

Finite fields are one of the few examples of algebraic structures that are completely classified. That is, the finite fields of any given order are classified. No such classification is known for finite rings in general. No such classification is known for finite groups or finite semigroups. The classification of finite simple groups is arguably one of the more important mathematical results of the late twentieth century.

In this part of the course we will only outline the very rich topic of finite fields. Many of the proofs in this section are omitted. If you are interested in the details, take MT5826 Finite Fields next year!

Let p be a prime. Then recall that we defined the *Galois field* $\mathbb{F}_p$ *of order* p to be the field over the set $\{0, 1, \ldots, p-1\}$ with $+$ and $*$ the usual arithmetic of the integers modulo p.

16.1. Characteristic

Let R be a ring such that there exists $n > 0$ with

$$nr = \underbrace{r + r + \cdots + r}_{n \text{ times}} = 0$$

for all $r \in R$. Then the minimum such n is called the *characteristic* of R.

If no such n exists, then R is said to have *characteristic* 0.

Example 16.1. The rings $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{n}]$ have characteristic 0.

The fields $\mathbb{F}_2 = \mathbb{Z}/(2)$ or $\mathbb{F}_5 = \mathbb{Z}/(5)$ have characteristic 2 and 5, respectively.

Theorem 16.2. *Let $R \neq \{0\}$ be a ring with identity 1, positive characteristic, and no zero divisors. Then R has prime characteristic.*

Proof. Since R contains non-zero elements, R has characteristic $n \geq 2$. If n were not prime, we could write $n = km$ with $k, m \in \mathbb{Z}$, $1 < k, m < n$. Then $0 = n \cdot 1 = (km)1 = (k \cdot 1)(m \cdot 1)$, so either $k \cdot 1 = 0$ or $m \cdot 1 = 0$, since R has no zero divisors. Hence either $kr = (k \cdot 1)r = 0$ for all $r \in R$ or $mr = (m \cdot 1)r = 0$ for all $r \in R$, contradicting the definition of characteristic. ■

Corollary 16.3. *$\mathbb{F}_p$ has characteristic p.*

Proof. From Theorem 16.2, we only have to show that a finite field F has positive characteristic. Consider the sums $1 \cdot 1, 2 \cdot 1, 3 \cdot 1, \ldots$ of the identity. Since F has finitely many elements, there must exist integers k and m with $1 \leqslant k < m$ such that $k \cdot 1 = m \cdot 1$, i.e. $(k-m)1 = 0$, and thus $(k-m)f = (k-m)1f = 0f = 0$ for all $f \in F$ so F has positive characteristic. ■

16.2. Polynomials again

Let R be an integral domain. Then an element $a \in R \setminus R^*$ is *irreducible* if $a = bc$ for $b, c \in R$ implies that b or c is a unit. If $R = F[x]$ for some field F, then $R^* = F \setminus \{0\}$. Hence $f \in R \setminus R^*$ is *irreducible* if and only if whenever $f = gh$ either $g \in F \setminus \{0\}$ or $h \in F \setminus \{0\}$. In other words, $f \in R \setminus R^*$ is *irreducible* if and only if whenever $f = gh$ either $\deg(g) = 0$ or $\deg(h) = 0$.

In $F[x]$, an element is *reducible* if it is not irreducible.

Example 16.4. In $\mathbb{R}[x]$, $x^2 + 1$ is irreducible but in $\mathbb{C}[x]$

$$x^2 + 1 = (x + i)(x - i)$$

is reducible.

Example 16.5. The polynomial

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

is reducible in $\mathbb{Z}[x]$. The polynomial $x^2 + 1$ is irreducible in $\mathbb{F}_3[x]$ but

$$x^2 + 1 = (x - 2)(x - 3) \in \mathbb{F}_5[x].$$

Theorem 16.6. *Let F be a field and let $f \in F[x]$ be arbitrary. Then $F[x]/(f)$ is a field if and only if f is an irreducible element of $F[x]$.*

Proof. From Theorem 14.12, $F[x]$ is a principal ideal domain and it is not a field since the units are $F \setminus \{0\}$. Hence by Theorem 13.10, (f) is maximal if and only if f is an irreducible in $F[x]$. Moreover, by Theorem 10.2, (f) is maximal if and only if $F[x]/(f)$ is a field. ■

The following is a summary of the steps required to describe the field $F[x]/(f)$:

- the elements of $F[x]/(f)$ are cosets $g + (f)$ with $g \in F[x]$;
- $g + (f) = h + (f)$ if and only if $g - h \in (f)$ if and only if $f \mid g - h$;
- if $g = qf + r$, then $g + (f) = r + (f)$, $\deg(r) < \deg(f)$, and r is unique (by Theorem 14.10);
- the cosets in $F[x]/(f)$ are precisely $g + (f)$ where g runs through all the polynomials in $F[x]$ with $\deg(g) < \deg(f)$;
- if $F = \mathbb{F}_p$ and $\deg(f) = n$, then $\mathbb{F}_p[x]/(f)$ has p^n elements.

A *root* of a polynomial $f \in F[x]$ is an element $a \in F$ such that $f(a) = 0$.

Example 16.7. (i) The elements $2, 3 \in \mathbb{Q}$ are roots of $x^2 - 5x + 6 \in \mathbb{Q}[x]$.

(ii) The polynomial $x^2 + 1 \in \mathbb{Q}[x]$ has no roots in $\mathbb{Q}$, but two roots $\pm i \in \mathbb{C}$.

Theorem 16.8. *Let $f \in F[x]$ and $a \in F$. Then a is a root of $f \in F[x]$ if and only if $x - a \mid f$.*

Proof. ($\Leftarrow$) Dividing f by $x-a$ we obtain

$$f = q \cdot (x-a) + c$$

with $q \in F[x]$ and $c \in F$. Substituting $x = a$, we get $f(a) = c$. Hence $f = q \cdot (x-a) + f(a) = q \cdot (x-a)$ and so $x - a \mid f$, as required.

($\Rightarrow$) $x - a \mid f$ implies that $f = g \cdot (x-a)$ for some $g \in F[x]$. Hence $f(a) = g(a) \cdot (a-a) = g(a) \cdot 0 = 0$, and so a is a root of f. ■

Corollary 16.9. *Let $f \in F[x]$. Then*

(i) *if $\deg(f) = 1$, then f is irreducible;*

(ii) *if f is irreducible and $\deg(f) > 1$, then f has no roots;*

(iii) *if $\deg(f) = 2$ or 3, then f is irreducible if and only if it has no roots.*

Example 16.10. (See Example 6.5.) Let $f = x$. Then $\deg(f) = 1$ and so f is irreducible. Hence $\mathbb{F}_2[x]/(f)$ is a field and from the comments above the elements are:

$$0 + (f), 1 + (f).$$

Example 16.11. (See Example 6.6.) Let $f = x^2 + x + 1$. Then

$$f(0) = f(1) = f(2) = 1$$

and so f is irreducible. Hence $\mathbb{F}_2[x]/(f)$ is a field and its elements correspond to the elements of $\mathbb{F}_2[x]$ with degree at most 1. That is,

$$0 + (f), 1 + (f), x + (f), x + 1 + (f)$$

16.3. Field extensions

If F is a field and K a subfield of F, then F is called an *extension field* of K. If $K \neq F$, then K is called a *proper* subfield. For example, $\mathbb{Q}$ is a proper subfield of $\mathbb{C}$ and $\mathbb{R}$ is an extension field of $\mathbb{Q}$.

The *prime subfield* of F is the intersection of all subfields of a field F, and contains no proper subfields.

Theorem 16.12. *Let K be the prime subfield of a field F. Then*

(i) *if F has characteristic 0, then $K \cong \mathbb{Q}$;*

(ii) *if F has characteristic p, then $K \cong \mathbb{F}_p$.*

Proof. Let F be a field of characteristic 0 and let P be the prime subfield of F. If $m, n \in \mathbb{Z}$ such that $m \neq n$ and $m \cdot 1 = n \cdot 1$, then $(m-n) \cdot 1 = 0$ and so $(m-n)x = 0$ for all $x \in F$. This implies that the characteristic of F is at most $m - n$, a contradiction. Hence the elements $n \cdot 1$ ($n \in \mathbb{Z}$) are all distinct. It is straightforward to prove that they form a subring of F isomorphic to $\mathbb{Z}$. The set

$$Q = \{m \cdot 1 \cdot (n \cdot 1)^{-1} : m, n \in \mathbb{Z}, n \neq 0\}$$

is a subfield of F isomorphic to $\mathbb{Q}$. Any subfield of F must contain 1 and 0 and so must contain Q, so $Q(F) \subseteq P$. Since Q is itself a subfield of F, we also have $P \subseteq Q$ and so in fact $Q = P$.

If F has characteristic p, a similar argument proves that

$$P = \{0 \cdot 1, 1 \cdot 1, 2 \cdot 1, \ldots, (p-1) \cdot 1\},$$

and that $P \cong \mathbb{F}_p$. ■

A polynomial $a_0 + a_1 x + \cdots + a_n x^n$ is *monic* if the largest non-zero coefficient a_n equals 1.

Lemma 16.13. *Let K be a subfield of F and $\alpha \in F$ be the root of a non-zero polynommial in $K[x]$. Then there exists a unique monic polynomial in $F[x]$ of minimal degree with α as a root.*

Proof. As an exercise you can prove that

$$\{\, f \in K[x] \,:\, f(\alpha) = 0 \,\}$$

is an ideal of $F[x]$.

Since $F[x]$ is a principal ideal domain, $I = \{\, f \in K[x] \,:\, f(\alpha) = 0 \,\} = (g)$ for some $g \in K[x]$. Using the Division Algorithm in $F[x]$, we can assume without loss of generality that $\deg(g) = \min\{\deg(f) \,:\, f \in I\}$. Furthermore, if $\deg(g) = n$ and $g = a_0 + a_1x + \cdots + a_nx^n \in F[x]$, then $h = a_n^{-1}a_0 + a_n^{-1}a_1x + \cdots + a_n^{-1}a_{n-1}x^{n-1} + x^n \in (g)$ and $g \in (h)$. Hence $(g) = (h)$ and h is a monic polynomial in $F[x]$ of minimal degree with α as a root.

Let h' be any other monic polynomial in $F[x]$ of minimal degree with α as a root. Then $h' \mid h$ and $h \mid h'$. Hence there exists a unit $u \in F$ such that $h'u = h$. But h' and h are both monic and so $h' = h$. ■

The unique monic polynomial of minimal degree from Lemma 16.13 is called the *minimal polynomial* of α over K.

Theorem 16.14. *Let $\alpha \in F$ be the root of a non-zero polynomial in $K[x]$, let g be the minimal polynomial of α, and let $f \in K[x]$. Then*

(i) *g is irreducible in $K[x]$;*

(ii) *$f(\alpha) = 0$ if and only if g divides f.*

Proof. (i). Since g has the root α, it has positive degree. Suppose $g = h_1h_2$ in $K[x]$ with $1 \leqslant \deg(h_i) < \deg(g)$ $(i = 1, 2)$. This implies $0 = g(\alpha) = h_1(\alpha)h_2(\alpha)$, and so one of h_1 or h_2 must lie in the ideal

$$\{\, f \in K[x] \,:\, f(\alpha) = 0 \,\} = (g)$$

from Lemma 16.13. Hence either h_1 or h_2 is divisible by g, a contradiction.

(ii). This follows immediately from the definition of g. ■

Example 16.15.

- $\sqrt[3]{3} \in \mathbb{R}$ is a root of $x^3 - 3 \in \mathbb{Q}[x]$. Since $x^3 - 3$ is irreducible over $\mathbb{Q}$, it is the minimal polynomial of $\sqrt[3]{3}$ over $\mathbb{Q}$.
- $i = \sqrt{-1} \in \mathbb{C}$ is a root of the polynomial $x^2 + 1 \in \mathbb{R}[x]$. Since $x^2 + 1$ is irreducible over $\mathbb{R}$, it is the minimal polynomial of i over $\mathbb{R}$.

Let L be an extension field of K. An important observation is that L may be viewed as a vector space over K. The elements of L are the "vectors" and the elements of K are the "scalars". We omit the definition of a vector space and the proof that L is a vector space over K (it follows almost immediately from the definition of a field!).

A *basis* of a vector space V over F can be defined as a subset $\{v_1, \ldots, v_n\}$ of vectors in V such that every $v \in V$ can be *uniquely* written as

$$v = a_1v_1 + \cdots + a_nv_n$$

where $a_1, \ldots, a_n \in F$.

Vector spaces can have many different bases, but there are always the same number of basis vectors; called the *dimension* of V over F.

Let L be an extension field of K. If L is finite-dimensional as a vector space over K, then L is said to be a *finite extension* of K. The dimension of the vector space L over K is written $[L : K]$.

Example 16.16. Let $L = \mathbb{C}$ and K be the subfield $\mathbb{R}$. Then we can easily check that $\mathbb{C}$ is a vector space over $\mathbb{R}$. Since $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$, it is clear that $\{1, i\}$ is a basis and so $[\mathbb{C} : \mathbb{R}] = 2$.

Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$ and let K be a subfield of F. Then the least, with respect to containment, subfield of F containing $\alpha_1, \alpha_2, \ldots, \alpha_n$ and K is denoted by $K(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

Theorem 16.17. *Let F be an extension field of K and $\alpha \in F$ be the root of a non-zero polynomial in $K[x]$, and let g be the minimal polynomial of α over K. Then*

(i) *$K(\alpha)$ is isomorphic to $K[x]/(g)$;*

(ii) *$[K(\alpha) : K] = \deg(g)$ and $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is a basis of $K(\alpha)$ over K;*

(iii) *if $\beta \in K(\alpha)$ is a root of a non-zero polynomial in $K[x]$, then the degree of the minimal polynomial of β is a divisor of $\deg(g)$.*

Proof. Omitted. ■

This theorem tells us that the elements of the simple extension $K(\alpha)$ of K are polynomial expressions in α, and any $\beta \in K(\alpha)$ can be uniquely expressed in the form $\beta = a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$ for some $a_i \in K$.

Example 16.18. Consider the simple extension $\mathbb{R}(i)$ of $\mathbb{R}$. We saw earlier that i has minimal polynomial $x^2 + 1$ over $\mathbb{R}$.

So $\mathbb{R}(i) \cong \mathbb{R}[x]/(x^2 + 1)$, and $\{1, i\}$ is a basis for $\mathbb{R}(i)$ over $\mathbb{R}$. So

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

Example 16.19. Consider the simple extension $\mathbb{Q}(\sqrt[3]{3})$ of $\mathbb{Q}$. We saw earlier that $\sqrt[3]{3}$ has minimal polynomial $x^3 - 3$ over $\mathbb{Q}$.

So $\mathbb{Q}(\sqrt[3]{3}) \cong \mathbb{Q}[x]/(x^3 - 3)$, and $\{1, \sqrt[3]{3}, (\sqrt[3]{3})^2\}$ is a basis for $\mathbb{Q}(\sqrt[3]{3})$ over $\mathbb{Q}$. So

$$\mathbb{Q}(\sqrt[3]{3}) = \{a + b\sqrt[3]{3} + c(\sqrt[3]{3})^2 : a, b, c \in \mathbb{Q}\}.$$

16.4. The Main Theorems

In the remainder of this section we will outline the proofs of the following theorems.

Theorem 16.20. *Let F be a finite field. Then F has p^n elements where the prime p is the characteristic of F and n is the dimension of F over its prime subfield.*

Proof. Since F is finite, it has characteristic p for some prime p by Corollary 16.3. Hence by Theorem 16.12 the prime subfield of F is isomorphic to $\mathbb{F}_p$. Thus F is an extension field of $\mathbb{F}_p$. In particular, F is a vector space over $\mathbb{F}_p$ of dimension n, for some $n \in \mathbb{N}$. Hence $|F| = p^n$, as required. ■

Theorem 16.21. *Let p be a prime and n be a positive integer. Then, up to isomorphism, there exists a unique finite field with p^n elements.*

Proof. Omitted. ■

Example 16.22. In Exercise 8.1 we proved that the polynomials $x^2 + 1$ and $y^2 + 2y_2$ are irreducible in $\mathbb{F}_3[x]$. Hence it follows that $F_1 = \mathbb{F}_3[x]/(x^2 + 1)$ and $F_2 = \mathbb{F}_3[y]/(y^2 + 2y + 2)$ are fields with $3^2 = 9$ elements. It follows from Theorem 16.20 that F_1 and F_2 are isomorphic.

Theorem 16.23 (Subfield Criterion) *Let $\mathbb{F}_q$ be a finite field where $q = p^n$ for some prime p. Then every subfield of $\mathbb{F}_q$ has p^m elements where m is some divisor of n.*

If m is a divisor of n, then there exists a unique subfield of $\mathbb{F}_q$ with p^m elements

Proof. Let K be a subfield of $\mathbb{F}_q$ and let P be the prime subfield of F. Since P is the intersection of all subfields of F, it follows that $P \subseteq K$. Hence by Theorem 16.20, $|K| = p^m$ for some $m \in \mathbb{N}$. Since F is a vector space over K, it follows that $q = p^n$ is divisible by p^m and so m divides n.

The proof of the second part of the theorem is omitted. ■

Theorem 16.24. *Let $\mathbb{F}_q$ be a finite field. Then $\mathbb{F}_q \setminus \{0\} = \{x, x^2, \ldots, x^{q-1} = 1\}$ for some $x \in \mathbb{F}_q$, that is $\mathbb{F}_q \setminus \{0\}$ is a cyclic group.*

Proof. Omitted. ■