Rings and Modules
Old Syllabus for O4

T. W. Körner

October 5, 2004

Small print The syllabus for the course is defined by the Faculty Board Schedules (which are minimal for lecturing and maximal for examining). *Please note that, throughout, ring means commutative ring with one.* I should **very much** appreciate being told of any corrections or possible improvements and might even part with a small reward to the first finder of particular errors. This document is written in LaTeX2e and stored in the file labelled `~twk/1B/Rings.tex` on emu in (I hope) read permitted form. My e-mail address is `twk@dpmms`.

Contents

1 Rings

The same ideas and proofs occur in the study of the integers (number theory), polynomials (leading to algebraic geometry), parts of the theory of matrices and in the theory of Abelian groups. They may be unified by using the theory of commutative rings and modules following a programme laid out by Emmy Noether and others. We start by looking at commutative rings with one.

Definition 1 *We say that* $(R,+,.)$ *is a* commutative ring with a one *if*
(i) $(R,+)$ *is an Abelian group.*
(ii) $a(bc)=(ab)c$ *for all* $a,b,c\in R$. *[Associative law of multiplication.]*
(iii) $a(b+c)=ab+ac$, $(b+c)a=ba+ca$ *for all* $a,b,c\in R$. *[Distributive law.]*
(iv) There exists a $1\in R$ *such that* $1a=a1=a$ *for all* $a\in R$. *[Existence of a multiplicative identity.]*
(v) $ab=ba$ *for all* $a,b\in R$. *[Commutative law of multiplication.]*

Rules (iii) and (iv) could be shortened using rule (v). We usually write 0 for the identity of the group $(R,+)$ and call 0 the zero of R.

Rule (iv) is made easier to use by the following simple remark.

Lemma 2 (Uniqueness of multiplicative identities) *If* $(M,.)$ *is an object with multiplication and* $1,1'\in M$ *are identities in the sense that*

$$1a=a1=a \text{ and } 1'a=a1'=a \text{ for all } a\in M,$$

then $1=1'$.

Thus R has a unique multiplicative identity 1. (We shall usually refer to 1 as ‘one’. It is sometimes called ‘the unit element of R’ but the word ‘unit’ means something different in the context of this course, see Definition 42.)

There are important examples of *non-commutative* rings (that is systems obeying all the rules in Definition 1 except (v) the commutative law of multiplication) such as the set of $n \times n$ matrices with the usual addition and multiplication $[n \geq 2]$. However, there are many beautiful results which are only true for commutative rings. Rule (iv) (the existence of a one) is less important. It gives some of our theorems and proofs a more elegant form but commutative rings without one are not much harder to deal with.

In this course we shall only deal with commutative rings with 1 *and 'ring' will mean 'commutative ring with* 1*'.*

Rings have many of the properties of the 'ordinary number systems' with which we are familiar from school. The integers $\mathbb{Z}$ with the usual operations form one of the most important examples. Note that the equation $2m = 1$ has no solution in $\mathbb{Z}$ (in other words 2 has no multiplicative inverse). The system $(\mathbb{Z}_n, +, \times)$ of the integers modulo n is another example. (The reader is certainly familiar with this system but definition freaks will find a neat definition using ring theory in Definition 16.) Note that in $\mathbb{Z}_{12}$

$$3 \neq 0 \text{ and } 4 \neq 0 \text{ yet } 3 \times 4 = 0$$

(we call 3 and 4 divisors of zero) and

$$2 \neq 6 \text{ yet } 2 \times 3 = 6 \times 3$$

(thus we can not use cancellation to get from $a \times b = a \times c$ to $b = c$). In $\mathbb{Z}_{81}$ we have $3 \neq 0$, $3^2 \neq 0$, $3^3 \neq 0$ yet $3^4 = 0$ (we say that 3 is nilpotent). These examples suggest that when dealing with rings we should first try methods and ideas which work for 'ordinary number systems' but be prepared to modify or, if the worst comes to the worst, abandon those parts which depend on division or cancellation.

However we have access to another fertile source of inspiration. We have already met two examples of abstract algebraic systems:- groups and vector spaces. Techniques and ideas which were useful for these are likely to be useful for rings.

Here are a few definitions and results along familiar lines.

Definition 3 *Let $(R, +, .)$ be a ring. If S is a subset of R such that*

(i) $a - b \in S$ and $ab \in S$ whenever $a, b \in S$,

(ii) $1 \in S$,

then we call S a subring of R.

(Condition (ii) excludes the possibility $S = \{0\}$.)

Lemma 4 *Let $(R, +, .)$ be a ring and S subring of R. Then S equipped with the addition and multiplication inherited from R is itself a ring.*

Lemma 5 *Let $(A, +_A, \times_A)$ and $(B, +_B, \times_B)$ be rings. If we define addition and multiplication on $A \times B$ by*

$$\begin{aligned}(a_1, b_1) + (a_2, b_2) =& (a_1 +_A a_2, b_1 +_B b_2)\\ (a_1, b_1) \times (a_2, b_2)a =& (a_1 \times_A a_2, b_1 \times_B b_2)\end{aligned}$$

then $(A \times B, +, \times)$ is a ring.

We often write $A \oplus B$ for the ring just defined and call it the external direct sum.

Definition 6 *Let R and S be rings with multiplicative identities 1_R and 1_S. We say that a map $\alpha : R \to S$ is a homomorphism (more precisely a ring homomorphism) if*
(i) $\alpha(r_1 + r_2) = \alpha(r_1) + \alpha(r_2)$, $\alpha(r_1 r_2) = \alpha(r_1)\alpha(r_2)$ for all $r_1, r_2 \in R$
(ii) $\alpha(1_R) = 1_S$.

(Condition (ii) excludes the possibility $\alpha(r) = 0$ for all $r \in R$.)

Lemma 7 *Let R and S be rings and $\alpha : R \to S$ a homomorphism. Then $\alpha(R)$ is a subring of S.*

We often write $\operatorname{im} \alpha = \alpha(R)$ and call it the image of α.

Definition 8 *Let R and S be rings and $\alpha : R \to S$ a homomorphism. If α is a bijection we say that α is an isomorphism (more exactly a ring isomorphism) and that R and S are isomorphic. We write $R \overset{\alpha}{\cong} S$ (R is isomorphic to S by the map α) and $R \cong S$ (R is isomorphic to S).*

Lemma 9 *Isomorphism is an equivalence relation. That is*
(i) $R \cong R$.
(ii) If $R \cong S$, $S \cong T$ then $R \cong T$.
(iii) If $R \cong S$ then $S \cong R$.

2 Ideals, quotients and the isomorphism theorem

In many ways subrings are less important for ring theory than ideals.

Definition 10 *Let $(R, +, .)$ be a ring. If I is a non-empty subset of R such that*
(i) $a - b \in I$ whenever $a, b \in I$,
(ii) $ab \in I$ whenever $a \in R$ and $b \in I$,
then we call I an ideal of R.

We observe that I is a subgroup of $(R,+)$ the ring R considered as an Abelian group under addition. We take over from group theory the idea of a coset

$$r+I=\{r+s:s\in I\}$$

and observe that the first part of the proof of Lagrange's theorem shows that the cosets form a disjoint cover of R.

Lemma 11 *Let I be an ideal of a ring R. Then*
(i) $\bigcup_{r\in R}(r+I)=R$.
(ii) If $r,s\in R$ then either $(r+I)\cap(s+I)=\emptyset$ or $r+I=s+I$.

The remarkable thing is that we can define addition and multiplication of cosets in a natural way.

Lemma 12 *If I is an ideal of a ring R and*

$$r_1+I=r_2+I,\ s_1+I=s_2+I$$

then

$$(r_1+s_1)+I=(r_2+s_2)+I,\ r_1s_1+I=r_2s_2+I.$$

Definition 13 *If I is an ideal of a ring R we write R/I for the set of cosets of I and define addition and multiplication on R/I by*

$$(r+I)+(s+I)=(r+s)+I,\ (r+I)(s+I)=rs+I.$$

Lemma 14 *If I is an ideal of a ring R then R/I with addition and multiplication as in the previous definition is a ring.*

We call R/I a quotient ring.

The idea of a quotient ring gives a clean definition of arithmetic modulo m.

Lemma 15 *If $m\in\mathbb{Z}$ then*

$$m\mathbb{Z}=\{mr:r\in\mathbb{Z}\}$$

is an ideal of $\mathbb{Z}$.

Definition 16 *If $m\geq 2$ we write*

$$\mathbb{Z}_m=\mathbb{Z}/m\mathbb{Z}.$$

The reader will readily identify $\mathbb{Z}/m\mathbb{Z}$ for all $m\in\mathbb{Z}$.

The next example warns us to stick to Definition 13.

Example 17 *The set $I = \{0, 2\}$ is an ideal of the ring $\mathbb{Z}_4$. We have*

$$(0 + I)(0 + I) = 0 + I$$

but $\{rs : r, s \in I\} = \{0\} \neq I$.

Quotient rings as closely linked with homomorphisms.

Definition 18 *If R and S are rings and $\phi : R \to S$ is a homomorphism we write*

$$\ker \phi = \phi^{-1}(0) = \{r \in R : \phi(r) = 0\}$$

and call $\ker \phi$ the kernel of ϕ.

Lemma 19 *Suppose that R and S are rings and $\phi : R \to S$ is a homomorphism. Then*

(i) $\ker \phi$ is an ideal of R.

(ii) $\phi(r) = s$ has a solution $r \in R$ if and only if $s \in \operatorname{im} \phi$.

(iii) If $\phi(r) = s$ then $\phi(r') = s$ if and only if $r' \in r + \ker \phi$.

We have just shown that every kernel of a homomorphism is an ideal. The next remark shows that every ideal is the kernel of a homomorphism.

Lemma 20 *Let I be an ideal of the ring R. Then the map $\pi : R \to R/I$ given by*

$$\pi(r) = r + I$$

is a homomorphism with kernel I.

The machinery is now in place to state and prove our first key theorem.

Theorem 21 (The isomorphism theorem) *Suppose that R and S are rings and $\phi : R \to S$ is a homomorphism. Then*

$$R/\ker \phi \cong \operatorname{im} \phi.$$

3 Integral domains, fields and fractions

The fact that we can not necessarily cancel or divide in rings means that they are too general for many purposes.

Definition 22 *A ring $(D, +, .)$ is called an integral domain if, whenever $ab = 0$, we can deduce that $a = 0$ or $b = 0$.*

Definition 23 *A ring* $(\mathbb{F},+,.)$ *is called a field if* $(\mathbb{F}\setminus\{0\},.)$ *is an Abelian group.*

Thus a ring $(\mathbb{F},+,.)$ is a field if, whenever $a\in\mathbb{F}$ and $a\neq 0$ we can find a^{-1} with $aa^{-1}=1$. The element a^{-1} (unique by a simple argument from elementary group theory) is called the multiplicative inverse of a.

Lemma 24 *(i) If* $(D,+,.)$ *is an integral domain and* $ab=ac$ *with* $a\neq 0$ *then* $b=c$.
(ii) Every field is an integral domain.
(iii) Every subring of an integral domain is an integral domain.

Lemma 26 below is not in the printed syllabus but this has not deterred examiners from setting it in the past. We need definitions which, important though they are in a more general context, are only included here in order to allow us to state the lemma.

Definition 25 *(i) We say that an ideal* I *of a ring* R *is* maximal *if* $I\neq R$ *but if* J *is an ideal with* $J\supseteq I$ *and* $J\neq I$ *then* $J=R$.
(ii) We say that an ideal P *in a ring* R *is* prime *if* $ab\in P$ *implies* $a\in P$ *or* $b\in P$.

Lemma 26 *Suppose that* I *is an ideal in a ring* R.
(i) I *is maximal if and only if* R/I *is a field.*
(ii) I *is prime if and only if* R/I *is an integral domain.*

We already know quite a lot of fields including $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Q}$. We also know some finite fields.

Lemma 27 *(i) If* p *is a prime then* $\mathbb{Z}_p$ *is a field.*
(ii) If m *is not a prime then* $\mathbb{Z}_m$ *is not an integral domain.* $[m\geq 2]$

We digress briefly to discuss *characteristics*. If R is a ring, n a strictly positive integer and a an element of R let us write

$$na=\underbrace{a+a+\cdots+a}_{n},$$

$(-n)a=-na$ and $0a=a$.

Lemma 28 *Let* R *be a ring with multiplicative identity* 1_R.
(i) The map $\theta:\mathbb{Z}\to R$ *given by* $\theta(m)=m1_R$ *is a homomorphism.*
(ii) The set $\operatorname{im}\theta$ *of all elements of the form* $m1_R$ *is isomorphic to* $\mathbb{Z}$ *or* $\mathbb{Z}_n$ *for some* $n\geq 2$.

Definition 29 *With the notation of Lemma 28, if* $\operatorname{im}\theta$ *is isomorphic to* $\mathbb{Z}_n$ *we say that* R *has* characteristic n. *If* $\operatorname{im}\theta$ *is isomorphic to* $\mathbb{Z}$ *we say that* R *has* characteristic ∞ *(or, in some texts* characteristic 0*).*

There is another way of viewing this idea.

Lemma 30 *If* A *is a subset of a ring* R *then there is a smallest subring* B *containing* A. *(In other words there exists a subring* B *of* R *such that* $B \supseteq A$ *and if* C *is any subring of* R *with* $C \supseteq A$ *then* $C \supseteq B$.*)*

We call B the ring generated by A. If $A = \emptyset$ so that B is the smallest ring in R we call B the *prime subring* of R. (Here prime is used as in 'primal scream', the first or underlying scream.)

Lemma 31 *With the notation of Lemma 28,* $\operatorname{im}\theta$ *is the smallest subring of* R. *Thus the primal ring of* R *is isomorphic to* $\mathbb{Z}$ *or* $\mathbb{Z}_n$ *for some* $n \geq 2$.

It is natural to identify the prime subring with $\mathbb{Z}$ or $\mathbb{Z}_n$ and write $m = \theta(m) = m1_R$.

The notion of characteristic is most useful when applied to integral domains.

Lemma 32 *(i) The characteristic of an integral domain is either a prime or* ∞.

(ii) The prime subring of an integral domain may be identified with $\mathbb{Z}$ *or* $\mathbb{Z}_p$ *where* p *is a prime.*

(iii) If $(R, +, .)$ *is an integral domain then every non-zero element of the additive group* $(R, +)$ *has order the characteristic of the integral domain.*

Later on we shall see that polynomials provide important examples of integral domains which are not fields. For the moment the only obviously interesting example we know of an integral domain which is not a field is $\mathbb{Z}$. However this is such an important example that it justifies by itself all the work we shall do in the remainder of this section.

From the point of view of late nineteenth century mathematics we shall be showing that the rationals can be constructed from the integers. 'God created the integers, all the rest is the work of man.' As a bonus we find that the same proof gives the more modern sounding result that 'every integral domain can be embedded in a field'. (From the point of view of the 'plain man' we are just describing fractions with a great deal of caution.)

Lemma 33 *If* $(D, +, .)$ *is an integral domain write* $D^* = D \setminus \{0\}$. *The relation* $\sim$ *defined on* $D \times D^*$ *by*

$$(r_1, s_1) \sim (r_2, s_2) \text{ if } r_1 s_2 = r_2 s_1$$

is an equivalence relation.

If $(r_1,s_1)\sim(r_2,s_2)$ *and* $(u_1,v_1)\sim(u_2,v_2)$ *then*

$$(r_1v_1+s_1u_1,s_1v_1)\sim(r_2v_2+s_2u_2,s_2v_2)\text{ and }(r_1u_1,s_1v_1)\sim(r_2u_2,s_2v_2).$$

Lemma 34 *Continuing with the assumptions and notation of Lemma 33 let us write* $\mathbf{k}$ *for the set* $D/\sim$ *of equivalence classes*

$$\frac{r}{s}=\{(r',s')\in D\times D^*:(r',s')\sim(r,s)\}.$$

Then we may define addition and multiplication on $\mathbf{k}$ *by*

$$\frac{r}{s}+\frac{u}{v}=\frac{rv+su}{sv}\text{ and }\frac{r}{s}\frac{u}{v}=\frac{ru}{sv}.$$

With this addition and multiplication, $(\mathbf{k},+,.)$ *is a field.*

If we define $\theta:D\to\mathbf{k}$ *by*

$$\theta(r)=\frac{r}{1}$$

then θ *is an injective homomorphism and so* $\tilde{D}=\operatorname{im}\theta$ *is a subring of* $\mathbf{k}$ *isomorphic to* D.

It is natural to identify $\tilde{D}$ with D by writing

$$r=\frac{r}{1}$$

for each $r\in D$. We call $\mathbf{k}$ the field of fractions of D.

We have thus characterised integral domains.

Lemma 35 *A ring* D *is an integral domain if and only if it is isomorphic to a subring of a field.*

If we use the natural identification of $\tilde{D}$ with D we can restate Lemma 35 in a more striking manner.

Lemma 36 *A ring* D *is an integral domain if and only if it embeds in a field.*

The naturalness of our construction is emphasised by the Lemma 38 below. We need a preliminary remark.

Lemma 37 *If* A *is a subset of a field* $\mathbb{F}$ *then there is a smallest subfield* B *containing* A. *(In other words there exists a subfield* B *of* $\mathbb{F}$ *such that* $B\supseteq A$ *and if* C *is any subfield of* $\mathbb{F}$ *with* $C\supseteq A$ *then* $C\supseteq B$.*)*

We call B the field generated by A. If $A = \emptyset$ so that B is the smallest field in R we call B the *prime subfield* of R.

Lemma 38 *Suppose that* $(\mathbb{F}, +, .)$ *is a field and* D *a subring of* $\mathbb{F}$. *Let* Q *be the smallest subfield of* $\mathbb{F}$ *containing* D. *Then there is an isomorphism* $\phi : Q \to \mathbf{k}$ *such that* $\phi(r) = \dfrac{r}{1}$ *for all* $r \in D$.

Lemma 32 (ii) tells us that the prime subring of a field may be identified either with $\mathbb{Z}$ or $\mathbb{Z}_p$ where p is a prime. If the prime subring is $\mathbb{Z}_p$ then it is also a field and so the prime subfield of $\mathbb{F}$. If the prime subring is $\mathbb{Z}$ we may use Lemma 38 to identify the prime subfield.

Lemma 39 *The prime subfield of a field may be identified in a natural manner with* $\mathbb{Q}$ *or* $\mathbb{Z}_p$ *where* p *is a prime.*

So far as the syllabus is concerned this concludes the section. What follows is easy but not on the syllabus.

If we start with $D = \mathbb{Z}$ the construction above yields $\mathbf{k} = \mathbb{Q}$ as a field. But mathematicians are also interested in order. Recall that there is a relation $>$ on $\mathbb{Z}$. We say that $a > b$ if $b - a > 0$. The properties of $>$ follow from the following rules

(A) If $a \in \mathbb{Z}$ then exactly one of the following is true: $a = 0$ or $a > 0$ or $-a > 0$.

(B) If $a, b \in \mathbb{Z}$, $a > 0$ and $b > 0$ then $a + b > 0$ and $ab > 0$.

Lemma 40 *Let* $D = \mathbb{Z}$ *in Lemma 33. If* $(r_1, s_1) \sim (r_2, s_2)$ *and* $r_1 s_1 > 0$ *then* $r_2 s_2 > 0$.

Lemma 41 *Let* $D = \mathbb{Z}$ *in Lemma 34. Then we may define a relation* $>$ *on* $\mathbf{k} = \mathbb{Q}$ *by the conditions*

$$\frac{r}{s} > \frac{u}{v} \quad \text{if} \quad \frac{r}{s} - \frac{u}{v} > 0$$

and

$$\frac{r}{s} > 0 \quad \text{if} \quad rs > 0.$$

The following results hold

(A) If $a \in \mathbb{Q}$ *then exactly one of the following is true:* $a = 0$ *or* $a > 0$ *or* $-a > 0$.

(B) If $a, b \in \mathbb{Q}$, $a > 0$ *and* $b > 0$ *then* $a + b > 0$ *and* $ab > 0$.

In the language of the analysis course C9, $\mathbb{Q}$ is an ordered field.

4 Unique factorisation, Euclidean and principal ideal domains

In this section I shall give a rather cold blooded and abstract treatment of factorisation in rings. Historically the subject was an exciting and confusing one. There are several theorems in number theory and elsewhere, in particular the Wiles-Taylor theorem (formerly Fermat's last theorem), which looked easy to prove provided 'the obvious factorisation theorem holds' and very distinguished mathematicians fell into the trap of assuming that which is obvious is true. On the other hand when unique factorisation did indeed hold, it provided a very powerful tool. We give a simple example by proving an elegant theorem of Fermat (Theorem 57) via unique factorisation at the end of this section.

There are two immediate problems, the first obvious and easily overcome, the second less so. The easy problem is illustrated when we try to extend the unique factorisation theorem from $\mathbb{N}$ (which is, of course, not a ring) to the ring $\mathbb{Z}$. We observe that

$$-15 = (-3) \times 5 = 3 \times (-5)$$

and that

$$15 = (-3) \times (-5) = 3 \times 5$$

so some restatement of the theorem is necessary. We set up the machinery to deal with this in the next definition and the lemma that follows.

Definition 42 *Let R be a ring. We say that $u \in R$ is a* unit *if there exists an $v \in R$ such that $uv = 1$. (Thus u is a unit if it has a multiplicative inverse). We say that r and s are* associates *if there exists a unit u with $r = su$.*

We extend a standard notation of elementary number theory to any ring R. If $a, b, c \in R$ and $a = bc$ we say that 'b divides a' and write $b|a$.

Lemma 43 *(i) Consider a ring R. The relation r is an associate of s is an equivalence relation on R.*

(ii) Consider an integral domain D. Two elements $a, b \in D$ are associates if and only if $a|b$ and $b|a$.

As examples we note that all non-zero elements in a field are units and so all pairs of non-zero elements are associates. In $\mathbb{Z}$ the units are 1 and -1 and the only associate of n is $-n$.

The second problem is clearly marked by the two definitions that follow together with Example 47

Definition 44 *Let R be a ring. We say that $q \in R$ is* irreducible *if it is not a unit and whenever $a|q$ then a is either a unit or an associate of q.*

Definition 45 *Let R be a ring. We say that $p \in R$ is* prime *if it is neither* 0 *nor a unit and whenever $p|ab$ $[a, b \in R]$ then $p|a$ or $p|b$.*

Lemma 46 *Any prime is irreducible.*

Unfortunately there exist rings in which not all irreducible elements are prime.

Example 47 *Let*

$$r = \{n + m\surd(-5) : n, m \in \mathbb{Z}\}$$

and let $N : R \to \mathbb{Z}^+$ be given by

$$N(n + m\surd(-5)) = |n + m\surd(-5)|^2 = n^2 + 5m^2.$$

(i) R is a subring of $\mathbb{C}$ so an integral domain.
(ii) $N(ab) = N(a)N(b)$ for all $a, b \in R$.
(iii) The units of R are 1 and -1.
(iv) $6 = 2 \times 3 = (1 + \surd(-5)) \times (1 - \surd(-5))$.
(v) The elements 2, 3, $(1 + \surd(-5))$ and $(1 - \surd(-5))$ are irreducible.

In the development of the theory of factorisation for $\mathbb{Z}$ (strictly speaking for $\mathbb{N}$, which is not a ring) carried out in Course C3 we showed that every irreducible element is prime by using Bezout's theorem. Fortunately there exist a large class of integral domains for which something rather close to Bezout's theorem holds — the so called *principal ideal domains.*

Definition 48 *If R is a ring we say that an ideal I of R is* principal *if it is generated by a single element a, in other words*

$$I = aR = \{ar : r \in R\}.$$

We also write $I = (a)$.

Definition 49 *An integral domain D is said to be a* principal ideal domain *if every ideal I of D is principal.*

Lemma 50 *In a principal ideal domain every irreducible element is prime.*

Lemma 51 *In a principal ideal domain every element which is neither a unit nor* 0 *is the product of a finite number of irreducible elements.*

Once we have Lemmas 50 and 51 the same easy, if slightly tedious, arguments that we used to prove unique factorisation for the integers in Course C3 give us a unique factorisation theorem for principal ideal domains.

Theorem 52 *Let D be a principal ideal domains.*

(i) If $r \in D$ is non-zero we can find a unit u and irreducible elements a_1, a_2, $\dots a_n$ such that

$$r = ua_1a_2 \dots a_n.$$

(ii) Suppose that u and v are units and a_1, a_2, $\dots a_n$, b_1, b_2, $\dots b_m$ are irreducible with

$$ua_1a_2 \dots a_n = vb_1b_2 \dots b_m.$$

Then $m = n$ and by renumbering we can ensure that a_j and b_j are associates for all $1 \leq j \leq n$.

I said that principal ideal domains are common but I have given no technique for proving that a domain is a principal ideal domain. Not surprisingly, one way is to seek an analogue of Euclid's algorithm from Course C3.

Definition 53 *We say that an integral domain D is a* Euclidean domain *if we can find a function $\phi : D \setminus \{0\} \rightarrow \mathbb{Z}^+$ (called a Euclidean function) such that*

(i) if $a|b$ then $\phi(a) \leq \phi(b)$,

(ii) given $a \in R$ and $b \in R$ with $b \neq 0$ we can find q and r such that $a = qb + r$ and either $r = 0$ or $\phi(r) < \phi(b)$.

Lemma 54 *If D is a Euclidean domain with Euclidean function ϕ then u is a unit of D if and only if $u \neq 0$ and $\phi(u) = \phi(1)$.*

Theorem 55 *Every Euclidean domain is a principal ideal domain.*

We are now in position to give the reader a genuinely novel example of a domain with unique factorisation.

Example 56 (The Gaussian integers) *Consider*

$$R = \{n + mi : n, m \in \mathbb{Z}\}$$

and let $\phi : R \rightarrow \mathbb{Z}^+$ be given by

$$\phi(n + mi) = |n + mi|^2 = n^2 + m^2.$$

(i) R is a subring of $\mathbb{C}$ so an integral domain (called the Gaussian integers).

(ii) ϕ is a Euclidean function, so R is a Euclidean domain.

It is quite hard to give examples of a principal ideal domains which are not Euclidean (presumably, not because they are uncommon but because it is hard to show that no Euclidean function could possibly exist). However, they exist and are given, or at least referenced, in the heavier algebra texts.

The remainder of this section is not on the syllabus. In it we use factorisation in the Gaussian integers to prove a theorem of Fermat.

Theorem 57 (Fermat) *We work in $\mathbb{N}$. An odd prime p can expressed as the sum of the squares of two integers*

$$p = n^2 + m^2$$

if and only if p is of the form $4N + 1$ for some integer N.

The only if part is easy, but to prove the if part we need the following lemma on Gaussian integers.

Lemma 58 *We work in $\mathbb{N}$ except in part (i). Suppose that p is a prime such that we can find integers x and y and an integer c coprime to p such that $x^2 + y^2 = cp$. Then*

(i) p is not a prime for the Gaussian integers,

(ii) there exist integers n and m such that $p = n^2 + m^2$.

Combining this with the following simple consequence of Wilson's theorem (Course C3) we obtain Fermat's theorem (Theorem 57).

Lemma 59 *Suppose that p is of the form $4N + 1$ for some integer N.*

(i) We can solve the congruence $x^2 \equiv -1 \mod p$.

(ii) We can find an integer x with $1 \leq x \leq p/2$ such that $x^2 + 1^2 \equiv 0 \mod p$.

5 Polynomials over rings

The definition of polynomials over rings is complicated by the phenomenon illustrated in the next example.

Example 60 *Let $f : \mathbb{Z}_2 \to \mathbb{Z}_2$ be defined by $f(x) = x^2 + x$. Then $f(x) = 0$ for all x.*

We must thus decide whether to define a polynomial by its values (which is what an analyst would do) or by its coefficients. As algebraists we decide to define it by its coefficients and enshrine our choice in the following definition.

Definition 61 *The polynomial ring $R[X]$ over R is the collection of sequences*

$$\mathbf{r} = (r_0, r_1, r_2, \dots)$$

where each $r_j \in R$ and only finitely many of the r_j are non-zero. We define

$$\mathbf{r} + \mathbf{s} = (r_0 + s_0, r_1 + s_1, r_2 + s_2, \dots)$$

and

$$\mathbf{rs} = \mathbf{t}$$

where $t_j = \sum_{k=0}^{j} r_j s_{k-j}$.

Neither the next lemma nor its proof present any surprises.

Lemma 62 *The polynomial ring $R[X]$ over R is a ring.*

Finally we remove the mask of the mysterious stranger and write

$$\mathbf{r} = \sum_{j=0}^{\infty} r_j X^j = \sum_{j=0}^{N} r_j X^j$$

where N is any integer sufficiently large that $r_j = 0$ for all $j \geq N$. Of course, the X^j are simple place holders (we call X an 'indeterminate'). If we want to talk about the value of a polynomial we need a simple homomorphism (the pont evaluation map).

Lemma 63 (Point evaluation) *If $x \in R$ the map $\delta_x : R[X] \to R$ given by*

$$\delta_x \left(\sum_{j=0}^{N} r_j X^j \right) = \sum_{j=0}^{N} r_j x^j$$

is a homomorphism.

As might be expected, we write $\delta_x p = p(x)$. The degree of a polynomial is defined in the obvious manner.

Definition 64 *If*

$$p(X) = \sum_{j=0}^{N} r_j X^j$$

and $r_N \neq 0$ then we say that p has degree N and write $\partial p = N$. If $p = 0$ we write $\partial p = -\infty$.

In this course we confine ourselves to polynomials over integral domains.

Lemma 65 *If D is an integral domain then so is the polynomial ring $D[X]$.*

Lemma 66 *If p and q are polynomials over an integral domain then*
(i) $\partial(p+q) \leq \max(\partial p, \partial q)$,
(ii) $\partial(pq) = \partial p + \partial q$.

If we restrict ourselves still further to fields we can use a very powerful result.

Theorem 67 (Euclidean division) *If a and b are polynomials over a field $\mathbb{F}$ and $b \neq 0$ then we can find polynomials q and r such that $a = qb + r$ and $\partial r < \partial a$.*

As an immediate corollary we have a key result.

Lemma 68 *The polynomial ring over a field is a Euclidean domain and so a principal ideal domain.*

Notice that Lemma 68 does not extend even to such a well behaved integral domain as $\mathbb{Z}$.

Example 69 *The ideal generated by 2 and X is not principal in $\mathbb{Z}$.*

In the next section we shall see how this problem can be partially overcome by embedding the integral domain in its quotient field. A rather trivial example of this technique is used to derive Lemma 71 from Lemma 70 (iii) below.

Lemma 70 *Let us work in the ring of polynomials over a field $\mathbb{F}$.*
(i) If p is a polynomial and $p(a) = 0$ for some $a \in \mathbb{F}$ then we can find a polynomial q such that $p(X) = (X - a)q(X)$.
(ii) If p is a polynomial and $p(a_1) = p(a_2) = \cdots = p(a_m) = 0$ for some distinct $a_1, a_2, \ldots, a_m \in \mathbb{F}$ then we can find a polynomial q such that

$$p(X) = (X - a_1)(X - a_2)\ldots(X - a_m)q(X).$$

(iii) A polynomial of degree n has at most n zeros in $\mathbb{F}$.

Lemma 71 *Suppose that D is an integral domain and p is a polynomial in $D[X]$ of degree $n \geq 0$. Then there are at most n distinct solutions of $p(x) = 0$ with $x \in D$.*

So far as the syllabus is concerned this concludes the section. What follows is easy but not on the syllabus.

Suppose we consider the particular field $\mathbb{R}$. We know that the polynomials on $\mathbb{R}$ form an integral domain but in this special case we can also define an order. If $p(X) = \sum_{j=0}^{n} a_j X^j$ with $a_n \neq 0$ we say that $p > 0$ if $a_n > 0$. If p is the zero polynomial we say that $p \not> 0$. The following two rules are easy to check.

(A) If $p \in \mathbb{R}[X]$ then exactly one of the following is true: $p = 0$ or $p > 0$ or $-p > 0$.

(B) If $p, q \in \mathbb{R}[X]$, $p > 0$ and $q > 0$ then $p + q > 0$ and $pq > 0$.
If $p, q \in \mathbb{R}[X]$ we write $p > q$ if $p - q > 0$.

Exactly as Lemmas 40 and 41 we can extend this order to the field of quotients.

Lemma 72 *Let $D = \mathbb{R}[X]$ in Lemma 33. If $(r_1, s_1) \sim (r_2, s_2)$ and $r_1 s_1 > 0$ then $r_2 s_2 > 0$.*

Lemma 73 *Let $D = \mathbb{R}[X]$ in Lemma 34. Then we may define a relation $>$ on $\mathbf{k} = \mathbb{K}$ by the conditions*

$$\frac{r}{s} > \frac{u}{v} \quad \textit{if} \quad \frac{r}{s} - \frac{u}{v} > 0$$

and

$$\frac{r}{s} > 0 \quad \textit{if} \quad rs > 0.$$

The following results hold

(A) If $a \in \mathbb{K}$ then exactly one of the following is true: $a = 0$ or $a > 0$ or $-a > 0$.

(B) If $a, b \in \mathbb{K}$, $a > 0$ and $b > 0$ then $a + b > 0$ and $ab > 0$.

In the language of the analysis course C9, $\mathbb{K}$ is an ordered field but of a type rather different from $\mathbb{Q}$ and $\mathbb{R}$.

Remember that $\mathbb{Q}$ and $\mathbb{R}$ obeyed the axiom of Archimedes. 'If $a, b > 0$ then we can find an $n \in \mathbb{Z}^+$ such that

$$na = \underbrace{a + a + \cdots + a}_{n} > b.\text{'}$$

However, in $\mathbb{K}$, we have $X, 1 > 0$ yet

$$n = n1 = \underbrace{1 + 1 + \cdots + 1}_{n} \not> X$$

for all n. In a more striking, but equivalent, formulation

$$\frac{1}{n} > \frac{1}{X}$$

for all $n \in \mathbb{Z}$ with $n \geq 1$. Thus we have an ordered field containing $\mathbb{Z}$ for which $1/n \nrightarrow 0$. Ordered fields like $\mathbb{K}$ which do not obey the axiom of Archimedes are called non-Archimedean.

6 Unique factorisation for polynomials

Once we have a definition for the ring $R[X]$ of polynomials over a ring R it is easy to define the ring $R[X_1, X_2, \dots, X_n]$ of polynomials in n indeterminates X_1, X_2, …, X_n by using the inductive definition

$$R[X_1, X_2, \dots, X_{k+1}] = R[X_1, X_2, \dots, X_k][X_{k+1}].$$

It is not hard to see that this abstract definition corresponds to our intuitive picture of polynomials in several variables provided that we define the polynomial by its coefficients rather than its values. The typical element of $R[X_1, X_2, \dots, X_n]$ can be written

$$P(X_1, X_2, \dots, X_n) = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_n=1}^{N} a_{i_1,i_2,\dots,i_n} X^{i_1} X^{i_2} \dots X^{i_n}$$

and addition, multiplication and point evaluation

$$P(x_1, x_2, \dots, x_n) = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_n=1}^{N} a_{i_1,i_2,\dots,i_n} x^{i_1} x^{i_2} \dots x^{i_n}$$

for $x_1, x_2, \dots, x_n \in R$. The details which echo the previous section are just as trivial here as they were there and I shall omit them.

Simple induction using Lemma 62 and Lemma 65 gives the appropriate version of those lemmas.

Lemma 74 *If R is ring then so is $R[X_1, X_2, \dots, X_n]$.*

Lemma 75 *If D is an integral domain then so is $D[X_1, X_2, \dots, X_n]$.*

Unfortunately, although Lemma 68 tells us that the polynomial ring $\mathbb{F}[X]$ over a field $\mathbb{F}$ is a Euclidean domain and so a principal ideal domain this result does not extend to polynomials in several indeterminates.

Example 76 *If $\mathbb{F}$ is a field and we work in the ring $\mathbb{F}[X_1, X_2]$ then the ideal generated by X_1 and X_2 is not principal.*

In spite of this, it turns out that unique factorisation still holds for $\mathbb{F}[X_1, X_2, \dots, X_n]$. If we reflect on how we might prove this, it seems natural to use induction on n. In order to set out the induction it is natural to make the following definition based on the statement of Theorem 52

Definition 77 *Let D be an integral domain. We say that D is a* unique factorisation domain *if the following two statements hold.*

(i) If $r \in D$ is non-zero we can find a unit u and irreducible elements a_1, a_2, $\dots a_n$ such that

$$r = ua_1a_2 \dots a_n.$$

(ii) Suppose that u and v are units and a_1, a_2, $\dots a_n$, b_1, b_2, $\dots b_m$ are irreducible with

$$ua_1a_2 \dots a_n = vb_1b_2 \dots b_m.$$

Then $m = n$ and by renumbering we can ensure that a_j and b_j are associates for all $1 \le j \le n$.

The following point should be noted.

Lemma 78 *In a unique factorisation domain every irreducible is a prime (so the two terms are synonymous).*

Our aim would be achieved if we could prove the following theorem.

Theorem 79 *If D is a unique factorisation domain then so is $D[X]$.*

Simple induction gives the next result.

Theorem 80 *If D is a unique factorisation domain then so is $D[X_1, X_2, \dots, X_n]$.*

By Theorem 52 every principal ideal domain is a unique factorisation domain and we have a very strong result.

Theorem 81 *If D is a principal ideal domain then $D[X_1, X_2, \dots, X_n]$ is a unique factorisation domain.*

How might we prove Theorem 79? Consider the special case when $D = \mathbb{Z}$. We know nothing about $\mathbb{Z}[X]$ but we do know that $\mathbb{Z}$ embeds naturally in its field of fractions $\mathbb{Q}$ and that unique factorisation holds for $\mathbb{Q}[X]$ (by Theorem 52). Since $\mathbb{Z}[X]$ embeds naturally in $\mathbb{Q}[X]$ we can proceed as follows. Suppose we have a polynomial $6X^3 + 24X^2 + 24X + 6$ in $\mathbb{Z}[X]$. We may not be able to factorise it in $\mathbb{Z}[X]$ but we can certainly factorise it in $\mathbb{Q}[X]$. Take one such factorisation

$$6X^3 + 24X^2 + 24X + 6 = \frac{42}{25}\left(\frac{5}{2}X + \frac{5}{2}\right)\left(\frac{10}{7}X^2 + \frac{30}{7}X + \frac{10}{7}\right).$$

By clearing fractions and cancelling ($\mathbb{Z}$ is, after all, the quintessential integer domain) we arrive at

$$6X^3 + 24X^2 + 24X + 6 = 2.3(X + 1)(X^2 + 3X + 1)$$

and a little thought shows that if $(\frac{5}{2}X + \frac{5}{2})$ and $(\frac{10}{7}X^2 + \frac{30}{7}X + \frac{10}{7})$ were irreducible in $\mathbb{Q}[X]$ then $(X + 1)$ and $(X^2 + 3X + 1)$ are irreducible in $\mathbb{Z}[X]$.

Although the proof of Theorem 79 given below is quite complicated it is my belief that any one seeking to develop the idea just given into a cast iron proof of the uniqueness of factorisation for $\mathbb{Z}[X]$ would be lead almost inevitably to something like it. One the proof is written down it is a simple matter to replace $\mathbb{Z}$ by a general unique factorisation domain.

Definition 82 *Let A be a subset of a ring R, such that A contains a non-zero element. We say that a is a highest common factor of A if*

(i) $a|x$ for all $x \in A$,

(ii) if $a'|x$ for all $x \in A$ then $a'|a$.

Lemma 83 *Any finite subset A of a unique factorisation domain such that A contains a non-zero element has a highest common factor.*

In fact the following is true though we shall not use it.

Lemma 84 *Any subset A of a unique factorisation domain such that A contains a non-zero element has a highest common factor.*

In what follows we work under the following standing hypothesis.

Standing hypothesis *We have a unique factorisation domain D embedded in its field of fractions F. We use the natural embeddings of D in F, F in $F[X]$, $D[X]$ in $F[X]$ and D in $D[X]$.*

We say that a polynomial $p(X) = \sum_{j=1}^{n} a_j X^j$ in $D[X]$ is *primitive* if 1 is a highest common factor of $\{a_j : 0 \leq j \leq n\}$. We observe that any $q \in D[X]$ can be written as $q = \gamma p$ with $\gamma \in D$ and p primitive.

Lemma 85 *Under our standing hypothesis,*

(i) The units of $D[X]$ are precisely the units of D.

(ii) Any $q \in F[X]$ can be written as $q = \gamma p$ with $\gamma \in F$ and p a primitive polynomial in $D[X]$.

(iii) If p, p' are primitive polynomials in $D[X]$ and $\gamma p = \gamma' p'$ for some $\gamma, \gamma' \in F$ then p and p' are associates in $D[X]$, that is there exists a unit $\epsilon \in D$ such that $p = \epsilon p'$.

Lemma 86 *Under our standing hypothesis, if p and q are primitive polynomials in $D[X]$ so is pq.*

Lemma 87 (Gauss' lemma) *Under our standing hypothesis, a polynomial $p \in D[X]$ is irreducible if and only if it is either (a) an irreducible element of D or (b) it is primitive in $D[X]$ and irreducible in $F[X]$.*

Theorem 88 *Under our standing hypothesis, $D[X]$ is a unique factorisation domain.*

Theorem 88 is just Theorem 79 so we are done. We cease working under the standing hypothesis.

The reader may suspect that it is hard to establish if a particular polynomial is irreducible. She is right[1]. One useful tool is due to Eisenstein. We give it for $\mathbb{Q}[X]$ though it can be generalised.

Lemma 89 (Eisenstein's criterion) *Suppose that*

$$P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

is a polynomial in $\mathbb{Z}[X]$ (i.e. P has integral coefficients). If there exists a prime number p such that $p \not| a_n$, $p|a_{n-1}$, $p|a_{n-2}$, ..., $p|a_0$ but $p^2 \not| a_0$, then P is irreducible over $\mathbb{Q}[X]$.

As an example of how it used consider the following.

Lemma 90 *If p is prime then $1 + X + X^2 + \cdots + X^{p-1}$ is irreducible over $\mathbb{Q}[X]$.*

The trick here is to make the substitution $Y = X - 1$ and to base our algebra on the recollected formula

$$1 + x + x^2 + \cdots + x^{p-1} = \frac{x^n - 1}{x - 1} = \frac{(y-1)^p - 1}{y},$$

from the days before we did abstract algebra.

The formula

$$\begin{aligned}(X-1)(X^3 + X^2 + X + 1) &= X^4 - 1\\ &= (X^2-1)(X^2+1)\\ &= (X-1)(X+1)(X^2+1)\end{aligned}$$

shows us that $(X^3 + X^2 + X + 1) = (X+1)(X^2+1)$ and suggests how to prove the converse.

Lemma 91 *If n is composite then $1 + X + X^2 + \cdots + X^{n-1}$ is not irreducible over $\mathbb{Q}[X]$.*

[1] At least, as far as human beings are concerned. There is an algorithm which will always work and computer algebra programs can handle quite complicated cases.

7 Fields and their simple extensions

We already know that it may be useful to embed a field in a larger field. Not all polynomials are soluble in $\mathbb{R}$ but they are in the larger field $\mathbb{C}$. In this section we study other extensions.

We begin with a couple of examples.

Example 92 *Consider $\mathbb{Q}$ as a subfield of $\mathbb{C}$.*

(i) Let γ be a transcendental number. Then $\mathbb{Q}(\gamma)$ the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and γ is isomorphic to the field of fractions of $\mathbb{Q}[X]$.

(ii) Let ω be a root of $z^2 + z + 1 = 0$ in $\mathbb{C}$. Then each element of $\mathbb{Q}(\omega)$ the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and ω may be written in exactly one way as $a + b\omega$ with $a, b \in \mathbb{Q}$.

Of course, we may not be in the happy position of Example 92 and find our extension 'ready made' as a subfield of some larger field.

Definition 93 *We say that L is an* extension *of a field K if there is an injective homomorphism $\phi : K \to L$ (i.e. if K is isomorphic to a subfield of L).*

Having made this definition we shall usually ignore it and treat K as a subfield of L with the natural identification $k = \phi(k)$ for $k \in K$. However, there are one or two points where we need to act more cautiously.

Definition 94 *We say that L is a* simple extension *of a field K if we can find an element $u \in L$ such that u and K generate L. We write $L = K(u)$.*

We now see that choices of Example 92 are, in some sense, typical.

Definition 95 *Suppose that L is a simple extension of K with $L = K(u)$. If u satisfies a polynomial equation*

$$u^n + a_{n-1}u^{n-1} + a_{n-2}u^{n-2} + \cdots + a_0 = 0$$

with $a_j \in K$ we say that u is algebraic *and that $L = K(u)$ is an* algebraic extension *of K. If not we say that u is* transcendental *and that $L = K(u)$ is a* transcendental extension *of K.*

Lemma 96 *If $K(u)$ is a transcendental extension of a field K then $K(u)$ is isomorphic to $\mathbf{k}$ the field of fractions of $K[X]$ under the natural isomorphism $\theta : \mathbf{k} \to K(u)$ which has $\theta(a) = a$ for all $a \in K$ and $\theta(X) = u$.*

The more interesting case of algebraic extension is dealt with in a series of simple but important lemmas.

Lemma 97 *If $K(u)$ is an algebraic extension of a field K then u is the zero of one and only one monic irreducible polynomial p in $K[X]$. If $q \in K[X]$ and $q(u) = 0$ then $q = hp$ for some $h \in K[X]$ (that is q is in the ideal (p) generated by p).*

Definition 98 *With the notation and hypotheses of Lemma 97 we say that p is the* minimal polynomial *of u. If p has degree n we say that u has* degree over K *of value n. We also write $[u : K] = n$.*

Lemma 99 *With the notation and hypotheses of Lemma 97 the mapping $\phi : K[X] \to K(u)$ given by*

$$\phi(f) = f(u)$$

is a surjective homomorphism with kernel (p).

Lemma 100 *With the notation and hypotheses of Lemma 97 $K(u)$ is isomorphic to $K[X]/(p)$. Thus every algebraic extension of K is isomorphic to the quotient of $K[X]$ by the ideal generated by some irreducible polynomial.*

It is interesting to ask what happens to p when we factorise it in $K(u)[X]$. Since $p(u) = 0$ we know that $X - u$ is a factor of $p(X)$ (by the 'remainder theorem' Lemma 70 (i)) so p will have linear factors. We shall discuss this in detail in the next section but for the moment we just give an example to show that, even in $K(u)[X]$, p may not factorise completely into linear factors.

Example 101 *Consider $\mathbb{Q}$ as a subfield of $\mathbb{C}$. Let $p \in \mathbb{Q}[X]$ be given by $p(X) = X^4 - 3$.*

(i) The polynomial p is monic and irreducible over $\mathbb{Q}[X]$.

(ii) If L is the field generated by $\mathbb{Q}$ and $3^{1/4}$ (the positive fourth root of 3 then $L = \mathbb{Q}(3^{1/4})$ and $p(3^{1/4}) = 0$. In $\mathbb{Q}(3^{1/4})$, p factors into irreducibles as

$$p(X) = (X - 3^{1/4})(X + 3^{1/4})(X^2 + 3^{1/2}).$$

(iii) If L is the field generated by $\mathbb{Q}$ and $3^{1/4}i$ then $L = \mathbb{Q}(3^{1/4}i)$ and $p(3^{1/4}i) = 0$. In $\mathbb{Q}(3^{1/4}i)$, p factors into irreducibles as

$$p(X) = (X - 3^{1/4}i)(X + 3^{1/4}i)(X^2 - 3^{1/2}).$$

We complete the unstarred part of this section with another simple but useful observation.

Lemma 102 *(i) If K is a subfield of L then L can be considered as a vector space of K in a natural manner.*

(ii) If L is a transcendental extension of K then L is infinite dimensional as a vector space over K.

(iii) If $L = K(u)$ and u is algebraic of degree n then L has dimension n as a vector space over K. The elements 1, u, $\dots$, u^{n-1} form a basis for L.

If K is a subfield of L we write $[L : K]$ for the dimension (possibly ∞) of L as a vector space over K. We call $[L : K]$ the degree of L over K.

Lemma 103 (Tower law) *If K is a subfield of L and L is a subfield of M then $[M : K] = [M : L][L : K]$.*

The rest of this section is not on the syllabus and, even if time allows to be covered, will only be sketched. Details may be found in the opening chapters of most texts on Galois theory (e.g. Chapter 6 of [4]).

We know that we stand on the shoulders of giants. The only question to be answered is whether we see any further. Our work so far enables us to solve two geometric problems that the Greeks were unable to solve. Both deal with ruler and compass constructions. The Greeks asked which constructions were possible with a ruler and compass alone. More prosaicly, but essentially equivalently we ask which points $(x, y) \in \mathbb{R}^2$ can be constructed starting from $(0, 0)$ and $(0, 1)$ using ruler and compass alone.

Lemma 104 *Consider a ruler and compass construction starting from $(0, 0)$ and $(0, 1)$ in which the point (x_j, y_j) is obtained at the jth step. If we write $R_0 = \mathbb{Q}$ and $R_j = R_{j-1}(x_j)(y_j)$ (that is R_j is the smallest subfield of $\mathbb{R}^2$ containing R_{j-1}, x_j and y_j) then $[R_j, R_{j-1}]$ takes the value 1 or 2. Thus, by the tower law, $[R_j, \mathbb{Q}] = 2^r$ for some integer $r \geq 0$.*

Theorem 105 (The Delian problem) *(i) The polynomial $X^3 - 2 = 0$ is irreducible over $\mathbb{Q}$.*

(ii) If in Lemma 104 we have $(x_j, , y_j) = (0, 2^{1/3})$ then $[R_j, \mathbb{Q}]$ must be divisible by 3.

(iii)We can not construct the point $(0, 2^{1/3})$ by ruler and compass construction starting from $(0, 0)$ and $(0, 1)$.

(iv) It is impossible using ruler and compass alone to construct a cube whose volume is double that of a cube of given edge.

There are many people for whom only the useful is worthwhile. Hogben dismisses Plato, Eudoxus and Euclid as men who who treated ‘mathematics as a respectable form of relaxation for the opulently idle’. Even Kline in his magisterial history *Mathematical Thought from Ancient to Modern Times* [6] sometimes reminds one of a school teacher in charge of a class of brilliant pupils who will persist in chasing the butterflies of pure mathematics rather than applying themselves to the stern task of understanding the real world. ‘Come on master Gauss stop looking at those cyclotomic polynomials — you have three orbits to compute before bedtime!’ Even if they understand the thrill of seeing a problem solved that has baffled mankind for 2000 years they see that thrill as a sinful diversion.

According to a story current in antiquity the Delians, suffering from pestilence, sent to the oracle who told them to double the size of a particular cubic altar to Apollo. They did as they were told by doubling the length of each of its sides. When the plague continued they consulted Plato who explained that the god wished his altar doubled in volume (preserving the cubic shape). The god, continued Plato, demanded this not because he wanted or needed such an altar but in order to censure the Greeks for their indifference to mathematics and lack of respect for geometry. The gods no longer punish societies which reject the pursuit of knowledge for its own sake quite so directly but perhaps such societies punish themselves.

Theorem 106 (The trisection problem) *(i) We can construct the point* $(\cos \pi/3, \sin \pi/3)$ *by ruler and compass construction starting from* $(0,0)$ *and* $(0,1)$.

(ii) If we could trisect every angle by ruler and compass construction we could construct $(\cos \pi/9, \sin \pi/9)$ *by ruler and compass construction starting from* $(0,0)$ *and* $(0,1)$.

(iii) If $\gamma = \cos \pi/9$ *then* $4\gamma^3 - 3\gamma - \frac{1}{2} = 0$. *If* $\tau = 2\gamma$ *then* $\tau^3 - 3\tau - 1 = 0$.

(iv) The polynomial $X^3 - 3X - 1 = 0$ *is irreducible over* $\mathbb{Q}$.

(v)We can not trisect every angle by a ruler and compass construction.

The credit for these two theorems goes to Wantzel. Possibly if he had done something romantic like being killed in a duel mathematicians would have had the courtesy to attach his name to his theorems.

Suppose we could prove the following theorem.

Theorem 107 (Lindeman) *The number* π *is transcendental.*

Then we would be able to solve a third great problem of antiquity.

Theorem 108 (Impossibility of circle squaring) *(i) It is impossible that* $\pi \in R_j$.

(ii) We cannot construct a square of area equal to a given circle by a ruler and compass construction.

There are now fairly short proofs of Theorem 107 (see, for example, Ian Stewart's beautiful *Galois Theory* [8] Chapter 6) but, so far as I know, no easy ones.

If we consider a regular polygon with n sides inscribed in the unit circle in such a way that one vertex is at $(0, 1)$ we see that the vertices are at points (x_r, y_r) given by $x_r + iy_r = \omega^r$ where $\omega = \exp(2\pi i/n)$ (so the ω^r are the rth roots of unity. The constructibility of a regular polygon with n sides by a ruler and compass construction is thus closely linked to the polynomial

$$X^n - 1 = (X - 1)(1 + X + X^2 + \cdots + X^{n-1})$$

and so to the *cyclotomic polynomial* $1 + X + X^2 + \cdots + X^{n-1}$. In particular, though we shall not do it, it is not hard to get from Lemma 90 to the statement that the regular p-gon (with p a prime) is only constructible by a ruler and compass construction if $p - 1$ is a power of 2. As a very young man, Gauss showed the reverse (if $p - 1$ is a power of 2 the regular p-gon is constructible). It is said that it was this discovery that decided him on a mathematical career. The details of the mathematics involved may be found in [8], Chapter 17.

8 Splitting fields of polynomials

In Lemmas 96 to 100 we derived the properties of simple extensions but took the simple extensions as given. Clearly, there always exists a transcendental extension of a given field K since the field of fractions of $K[X]$ is such an extension. Moreover, Lemma 96 tells us that (up to isomorphism) this extension is unique. Does there always exist an algebraic extension corresponding to a given irreducible polynomial and is it unique (up to isomorphism)?

The obvious way forward is pointed out by Lemma 100.

Lemma 109 *If K is a field and p is irreducible in $K[X]$ then $L = K[X]/(p)$ is a field containing (an isomorphic copy of) K. We can find $u \in L$ such that $L = K(u)$ is simple algebraic extension of K and $X - u$ is a factor of $p(X)$ in L.*

The only problem here is to show that $K[X]/(p)$ is a field and this follows from the analogue of Bezout's theorem for principal ideal domains. Uniqueness is simple.

Lemma 110 *Suppose that K is a field and p is irreducible in $K[X]$. If $K(u_1)$ and $K(u_2)$ are simple algebraic extensions of K such that $X - u_j$ is a factor of $p(X)$ in $K(u_j)$ then there is an isomorphism $\theta : K(u_1) \to K(u_2)$ with $\theta(a) = a$ for $a \in K$ and $\theta(u_1) = u_2$.*

Thus in Example 101 we know without further computation that $\mathbb{Q}(3^{1/4}) \cong \mathbb{Q}(3^{1/4}i)$.

Repeated use of Lemma 109 gives the theorem which the last section lead us to expect.

Theorem 111 *If K is a field and $p \in K[X]$ there exists a field L containing (an isomorphic copy of) K such that $[L : K] < \infty$ and we can find $A \in K$, $\alpha_1, \alpha_2, \dots, \alpha_n \in L$ such that*

$$p(X) = A(X - \alpha_1)(X - \alpha_2) \dots (X - \alpha_n)$$

We say that p *splits* over L. In order to obtain a uniqueness result we need to tighten up the conditions of the theorem.

Definition 112 *If K is a subfield of the field L and $p \in K[X]$ we say that L is a splitting field for p over K if*

(i) p factorises into linear factors

$$p(X) = A(X - \alpha_1)(X - \alpha_2) \dots (X - \alpha_n)$$

over L.

(ii) If p factorises into linear factors over a subfield L' of L then $L' = L$.

Observe that condition (ii) can be replaced by the statement $L = K(\alpha_1, \alpha_2, \dots, \alpha_n)$ the field generated by K, α_1, α_2, $\dots \alpha_{n-1}$ and α_n.

The uniqueness theorem is now easy to state.

Theorem 113 *Suppose that K is a field and $p \in K[X]$. If L and L' are splitting fields of p then there is an isomorphism $\theta : L \to L'$ with $\theta(a) = a$ for all $a \in K$.*

There may be many different ways to go from K to a splitting field by adjoining roots and Theorem 113 is slightly harder to prove than might be expected.

The following lemma contains the key idea.

Lemma 114 *Let K be a field, $p \in K[X]$ and let L be a splitting field for p over K. Suppose that L' is a field containing a subfield K' isomorphic to K under the isomorphism i such that $i(p)$ splits in L'. (Here, if $p(X) = \sum_{r=0}^{n} a_r X^r$ we write $i(p)(X) = \sum_{r=0}^{n} i(a_r) X^r$.) Then there is an injective homomorphism $j : L \to L'$ such that $j|_K = i$.*

This is as far as we shall go with the study of splitting fields but the following remark (which is not on the syllabus) seems worth making. We need results on countability from course C3.

Lemma 115 *(i) If K is a countable subfield of L and $[L : K] < \infty$ then L is countable.*

(ii) If K is a countable field we can find a countable field L containing (an isomorphic copy of) K such that every polynomial $p \in K[X]$ splits in $L[X]$.

(iii) If K is a countable field we can find countable fields K_j with

$$K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots$$

with K_{j-1} a subfield of K_j such that every polynomial $p \in K_{j-1}[X]$ splits in $K_j[X]$.

(iv) If K is a countable field we can find a countable field L containing (an isomorphic copy of) K such that every polynomial $p \in L[X]$ factors completely into linear factors.

The same idea gives the following more striking result.

Lemma 116 *There is a countable subfield $\mathbb{F}$ of $\mathbb{C}$ with $\mathbb{F} \supseteq \mathbb{Q}$ such that every polynomial in $\mathbb{F}[X]$ has a root in $\mathbb{F}$.*

Thus, from the point of view of a dyed in the wool algebraist, the construction of the uncountable field $\mathbb{C}$ in order to have the fundamental theorem of algebra is a reckless extravagance.

9 Finite fields

In this short but interesting section we find all finite fields explicitly.

Our first step is already substantial.

Lemma 117 *If $\mathbb{F}$ is a finite field then $\mathbb{F}$ has characteristic p a prime (that is $\mathbb{F}$ has prime field (an isomorphic copy of) $\mathbb{Z}_p$). The field $\mathbb{F}$ has p^n elements where $[\mathbb{F} : \mathbb{Z}_p] = n$.*

The second step is also remarkable.

Lemma 118 *Let $(\mathbb{F}, +, .)$ be a field. If G is a finite subgroup of the multiplicative group $(\mathbb{F} \setminus \{0\}, .)$ then G is a cyclic group.*

Notice that this result applies to general fields. The reader should identify all possible G in the cases $\mathbb{F} = \mathbb{C}$ and $\mathbb{F} = \mathbb{R}$. Our proof depends on a simple result from the theory of commutative groups.

Lemma 119 *If G is a finite Abelian group there exists an integer N and an element h such that*

(i) $g^N = e$ for all $g \in G$,

(ii) h has order exactly N.

Combining Lemmas 117 and 118, we see that all finite fields have a very simple structure.

Theorem 120 *If $\mathbb{F}$ is a finite field then $\mathbb{F}$ is (isomorphic to) the splitting field of $X^{p^n-1}-1$ over $\mathbb{Z}_p$ for some prime p and some integer $n \geq 1$.*

(We can refer to *the* splitting field since Theorem 113 tells us that splitting fields are unique up to isomorphism.)

Theorem 120 tells us the structure of a given finite field, if it exists, but does not tell us if such a field exists. To obtain existence results we need to investigate the polynomial $X^{p^n-1}-1 \in Z_p[X]$. We use a general result on repeated roots.

Lemma 121 *Let K be a field. Suppose that $p(X) = \sum_{j=0}^{n} a_j X^j \in K[X]$ splits over K. Then p has $(X-a)^2$ as a factor for some $a \in K$ if and only if the formal derivative*

$$p'(X) = \sum_{j=1}^{n} j a_j X^j$$

and $p[X]$ have a non-trivial common factor.

Lemma 122 *Let K be a field of characteristic p a prime. If $X^{p^n-1}-1$ splits over K then all the linear factors are distinct.*

Our results look nicer when stated in terms of $X^{p^n} - X$.

Theorem 123 *If p is a prime and n an integer the splitting field of $X^{p^n}-X$ over $\mathbb{Z}_p$ contains p^n elements consisting of the p^n distinct roots of $X^{p^n}-X$.*

We have now proved existence and uniqueness so we may make the following definition.

Definition 124 *The finite field of order p^n (p prime, $n \geq 1$) is called the* Galois field *of order p^n and written $\mathbf{GF}(p^n)$.*

This triumph completes that part of this section which is on the syllabus. However (strictly off the syllabus) we must admit that the triumph is not quite as complete as it appears. Observe that Lemma 118 tells us that the non-zero elements of $\mathbf{GF}(p^n)$ form a cyclic group generated by a single element x say. As temporary notation let us call x a multiplicative generator of $\mathbf{GF}(p^n)$. Surely, we can not claim to understand $\mathbf{GF}(p^n)$ unless we have some short algorithm for finding a multiplicative generator for it. So far as I know, no such algorithm has been found.

Of course since $\mathbf{GF}(p^n)$ is finite, exhaustive search will eventually turn up such a generator. We note also that quite a large proportion of the elements of $\mathbf{GF}(p^n)$ must be multiplicative generators (can you make this statement more precise?) so properly random trial and error[2] will rapidly find a multiplicative generator x with arbitrarily low probability of failure. Let us choose a basis $u_1, u_2, \dots, u_n$ for $\mathbf{GF}(p^n)$ as a vector space over $\mathbb{Z}_p$. We then have

$$x^r = a_1(r)u_1 + a_2(r)u_2 + \cdots + a_n(r)u_n.$$

The n-tuple in

$$\mathbf{a}_r = (a_1(r), a_2(r), \dots, a_n(r))$$

thus runs through each element of $\mathbb{Z}_p \setminus \{\mathbf{0}\}$ exactly once as r runs from 0 to $p^n - 1$.

In a telepathy experiment, Albert and Bertha are placed in separate sealed rooms. The experiment has already been running for a time $5N$ minutes where N is unknown to them. A bell rings each 5 minutes and (supposing it to be $5r$ minutes since they entered the room) they are asked to guess an n-tuple of integers $\mathbf{a}_{r+N} = (a_1(r+N), a_2(r+N), \dots, a_n(r+N))$ with $0 \leq a_j(r+N) \leq p-1$. If one of them guesses right he or she is told so and presented with a paper star. Bertha has the advantage that she knows how $\mathbf{a}_r$ is constructed and in particular knows x. It is easy to see that initially Albert and Bertha can only guess at random but that *once Bertha has guessed right* she can lock in and give the correct answer each time.

One way of trying to hide a radio signal is to spread it as a large number of weak signals at different frequencies and to change the choice of frequencies at regular intervals. Of course the enemy may make a lucky choice of listening frequencies and catch a brief part of the signal but the change of frequencies should stymie him. On the other hand, our own side may not be able to keep their timekeepers sufficiently synchronised with the transmitter during long periods of silence. We begin to see how military men and others might develop a deep interest in Galois fields.

[2] The ghastly modern educationalist's jargon seeks to replace 'trial and error' by 'trial and improvement' but here the failure of a guess results in no improvement.

10 Modules

The theory of vector spaces is a well developed and powerful one. We have seen examples of its use in this course in Lemma 117 which helped us classify finite fields and in the definition of the degree $[L : K]$ of an extension which helped resolve the classical ruler and compass problems. From time to time we come across structures like the 'lattice' $\mathbb{Z}^2$ which have a vector space 'flavour' without being vector spaces. It is thus natural to seek a theory which generalises the notion of a vector space though though we may expect the development of such a theory to be more intricate and the general results to be less neat.

We proceed in the obvious way by replacing 'field' by 'ring' in the definition of a vector space.

Definition 125 *Let R be a ring. We say that $(M, R, +, .)$ is a module over R if the following is conditions hold.*

(i) $(M, +)$ is an Abelian group.

(ii) There is a map $\theta : R \times M \to M$ written $\theta(r, \mathbf{m}) = r\mathbf{m}$ such that

(a) $r(\mathbf{m}_1 + \mathbf{m}_2) = r\mathbf{m}_1 + r\mathbf{m}_2$,

(b) $(r_1 + r_2)\mathbf{m} = r_1\mathbf{m} + r_2\mathbf{m}$,

(c) $(r_1r_2)\mathbf{m} = (r_1(r_2\mathbf{m})$,

(d) $1\mathbf{m} = \mathbf{m}$,

for all $r, r_1, r_2 \in R$ and $\mathbf{m}, \mathbf{m}_1, \mathbf{m}_2 \in M$.

We say that M is a module over R. Since the syllabus requires it to be explicitly stated, we remark that a vector space over a field $\mathbb{F}$ is automatically a module over $\mathbb{F}$.

We have an immediate pleasant surprise.

Lemma 126 *Let $(G, +)$ be a commutative group. If we write*

$$na = \underbrace{a + a + \cdots + a}_{n},$$

$(-n)a = -na$ and $0a = a$ $[a \in G]$ then G is a module over $\mathbb{Z}$.

However, this example shows us that the behaviour of modules, even over very nice rings, is very different from that of vector spaces. (Precisians will worry that not all the terms in the next example have been defined, everybody else will welcome early warning of trouble.)

Example 127 *Let C_6 be the cyclic group generated by $[1]$ and write $n[1] = [n]$. Then if we take C_6 as a module over $\mathbb{Z}$, $\{[1]\}$ is a minimal generating set but so is $\{[2], [3]\}$.*

This should be contrasted with the theory of finite dimensional vector spaces where every minimal generating set (in the language of Course P1, every minimal spanning set) has the same number of elements. The reader may care to reflect on the importance of *division* in the proof of the Steinitz replacement lemma. For the moment we note that results which involve the notion of basis or dimension explicitly or implicitly are unlikely to carry over from vector spaces to general modules.

Our next example is not surprising.

Lemma 128 *If S is a subring of a ring R then R is a module over S with module multiplication defined to be ring multiplication in R. In particular R is a module over itself.*

Our final introductory example may seem a little strange but much of the strangeness will vanish on reflection.

Lemma 129 *Let V be a vector space over a field $\mathbb{F}$ and let α be an endomorphism of V (that is a linear map from V to V). Then V is a module over the ring of polynomials $\mathbb{F}[X]$ with module multiplication defined by the following rule.*

If $p(X) = \sum_{j=0}^{n} a_j X^j$ and $v \in V$ then $pv = p(\alpha)v$, that is

$$pv = a_0 v + a_1\alpha(v) + a_2\alpha^2(v) + \cdots + a_n\alpha^n(v).$$

The reader should note the implied convention $\alpha^0 = \iota$. She should then examine the definition when $\mathbb{F} = \mathbb{C}$ and α is the linear map given, in turn, by the matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

If the situation described in Lemma 129 holds we talk of *the $\mathbb{F}[X]$ module constructed from V via α.*

It is natural to ask whether a concept so general that it includes both Abelian groups and the effect of polynomials of a given endomorphism on a vector space is not too general to produce interesting mathematics. The object of the last part of this course is to produce a theorem on modules (Theorem 171) so powerful that it gives both a complete classification of finite Abelian groups and of endomorphisms on finite dimensional vector spaces over $\mathbb{C}$.

Before moving directly to this topic we first produce some standard algebraic definitions, theorems and constructions parallelling those already produced in our studies of groups, vector spaces and rings.

Definition 130 *If M and N are modules over a ring R we say that $\phi : M \to N$ is a* (module) homomorphism *if*

$$\phi(r_1m_1 + r_2m_2) = r_1\phi(m_1) + r_2\phi(m_2)$$

for all $r_1, r_2 \in R$ and $m_1, m_2 \in M$. If ϕ is a bijection we say that it is a (module) isomorphism *and that M and N are* isomorphic*.*

Definition 131 *If $(M, R, +, .)$ is a module over a ring R we say that a subset N of M is a* submodule *if N is a subgroup of $(M, +)$ and $rn \in N$ whenever $r \in R$ and $n \in N$.*

The process of quotienting is familiar from our work with rings in Section 2 which we shall follow almost exactly. Since N is a subgroup of $(M, +)$ we may work with cosets $u + N$ of N.

Lemma 132 *Let N be a submodule of a module M over a ring R. Then*
(i) $\bigcup_{u\in M}(u + N) = M$.
(ii) If $u, v \in M$ then either $(u + N) \cap (v + N) = \emptyset$ or $u + N = v + N$.

Lemma 133 *If N is a submodule of a module M over a ring R and*

$$u_1 + N = u_2 + N,\ v_1 + I = v_2 + I$$

then

$$(u_1 + v_1) + I = (u_2 + v_2) + I,\ ru_1 + I = ru_2 + I$$

for all $r \in R$.

Definition 134 *If N is a submodule of a module M over a ring R we write M/N for the set of cosets of N and define addition and multiplication on M/N by*

$$(u + N) + (v + N) = (u + v) + N,\ r(u + N) = ru + N.$$

Lemma 135 *If N is a submodule of a module M over a ring R then M/N with module addition and multiplication as in the previous definition is a module over R.*

We call M/N a quotient module.

We continue along the sequence of Section 2.

Definition 136 *If M and N are modules over a ring R and $\phi : M \to N$ is a homomorphism we write*

$$\ker \phi = \phi^{-1}(0) = \{r \in R : \phi(r) = 0\}$$

and call $\ker \phi$ *the kernel of* ϕ*.*

Lemma 137 *If M and N are modules over a ring R and $\phi : M \to N$ is a homomorphism then*

(i) $\ker \phi$ *is a submodule of* M*.*

(ii) $\phi(u) = v$ *has a solution* $u \in M$ *if and only if* $v \in \operatorname{im} \phi$*.*

(iii) If $\phi(u) = v$ *then* $\phi(u') = v$ *if and only if* $u' \in u + \ker \phi$*.*

Lemma 138 *Let N be an submodule of a module M over a ring R. Then the map $\pi : M \to M/N$ given by*

$$\pi(u) = u + N$$

is a homomorphism with kernel N*.*

Theorem 139 (The isomorphism theorem for modules) *Suppose that M and N are modules over a ring R and $\phi : R \to S$ is a homomorphism. Then*

$$R/\ker \phi \cong \operatorname{im} \phi.$$

We have followed the same path to obtain the same isomorphism theorem for rings and modules. There is a similar result for groups (but the key notion is that of a *normal subgroup* that is of a subgroup H of a group G such that $g^{-1}Hg = H$ for all $g \in G$). Clearly we ought to seek some 'master theorem' from which all these results could be derived. Such concerns are the subject of *Universal Algebra* and its younger cousin *Category Theory.* In the context of the present course, most readers will find the generalisation of vector spaces to modules sufficiently hard without seeking to study a concept of 'algebraic system' which will include objects with a single non-commutative multiplication (groups), objects with two commutative multiplications linked by a distributive law (rings) and products of such objects with further links (modules).

The research supervisor of the great probabilist Feller told him that the best mathematics consists of the general embedded in the concrete. Feller claimed that it was some years before he realised this was not an anti-militarist slogan. Most mathematicians would agree with Feller's supervisor. Unfortunately they would differ widely on the proportion of general and concrete required and still more widely on what, precisely, is general and what concrete.

11 Linear relations in modules

So far, the results we have proved on modules have had a general algebraic flavour. However, we deliberately chose the axioms for modules to echo those for vector spaces and from now on we shall try to exploit that fact.

Lemma 140 *If M is a module over a ring R and A a non empty subset of M then the set N of elements*

$$\sum_{j=1}^{k} r_j a_j$$

with $r_j \in R$, $a_j \in A$ and k a positive integer is a submodule of M. If N' is any submodule of M with $N' \supseteq A$ then $N' \supseteq N$.

We call N the submodule generated by A.

Definition 141 *If M is a module over a ring R and M generated by a single element m we say that M is a* cyclic module *and write $M = Rm$.*

If M_1, M_2, …, M_n are submodules of module M we write $M_1 + M_2 + \cdots + M_n$ for the submodule generated by $\bigcup_{r=1}^{n} M_r$. We recall from vector space theory that direct sums are more useful than sums.

Definition 142 *If M is a module over a ring R and M_1, M_2, …, M_n are submodules of M we say that $M_1 + M_2 + \cdots + M_n$ is a* direct sum *(more specifically an* internal direct sum*) of M_1, M_2, …, M_n and write*

$$M_1 + M_2 + \cdots + M_n = M_1 \oplus M_2 \oplus \cdots \oplus M_n$$

if the only solution to the equation $m_1 + m_2 + \cdots + m_n = 0$ with $m_j \in M_j$ $[j = 1, 2, \ldots n]$ is $m_j = 0$ $[j = 1, 2, \ldots n]$.

Lemma 143 *Let M be a module over a ring R and M_1, M_2, …, M_n submodules of M. The following conditions are equivalent.*
(i) $M_1 + M_2 + \cdots + M_n$ is a direct sum.
(ii) $(\sum_{i \neq j} M_i) \cap M_j = \{0\}$ for each $1 \leq j \leq n$.
(iii) Each $m \in M_1 + M_2 + \cdots + M_n$ can be written in only one way as $m = \sum_{j=1}^{n} m_j$ with $m_j \in M_j$ $[j = 1, 2, \ldots n]$.

We can also define an *external direct sum* (analogous to the direct sum of rings in Lemma 5).

Lemma 144 *Let M_1, M_2, …, M_n be modules over a ring R. If we define addition and module multiplication on $\prod_{j=1}^n M_j$ by*

$$(m_1, m_2, \ldots, m_n) + (m_1', m_2', \ldots, m_n') = (m_1 + m_1', m_2 + m_2', \ldots, m_n + m_n')$$
$$r(m_1, m_2, \ldots, m_n) = (rm_1, rm_2, \ldots, rm_n)$$

for $m_j, m_j' \in M_j$, $r \in R$ then $\prod_{j=1}^n M_j$ is a module over R.

We write $M_1 \oplus M_2 \oplus \cdots \oplus M_n$ for the ring just defined and call it the *external direct sum.* If the M_j are all submodules of the same module M then there is a natural isomorphism between the internal and external direct sums and no problems arise if we identify the two objects.

We shall need the following simple result.

Lemma 145 *If M_1 and M_2 are submodules of a module M over a ring R and $M_1 + M_2$ is a direct sum then*

$$(M_1 \oplus M_2)/M_2 \cong M_1.$$

Our programme in the final part of the course is to show that reasonably well behaved modules M over reasonably well behaved rings can be written as the direct sum $M_1 \oplus M_2 \oplus \cdots \oplus M_n$ of submodules M_j each of which is well behaved (in particular cyclic, so that $M_j = Rm_j$). To place this programme in context, note that that one of the fundamental theorems of vector space theory can be written as follows.

Theorem 146 *If V is module over a field $\mathbb{F}$ generated by a finite set then*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_n$$

where each submodule V_j is cyclic (and is isomorphic to $\mathbb{F}$ as a module over $\mathbb{F}$). Further, the number n is an invariant of V (that is, every such decomposition requires exactly n submodules of the stated type).

We cannot evade consideration of one of the most striking ways that a module like $\mathbb{Z}_7$ or $\mathbb{Z}_{21}$ over $\mathbb{Z}$ differs from a vector space.

Lemma 147 *Let M be a module over a ring R. If $m \in M$ the set*

$$\mathbf{o}(m) = \{r \in R : rm = 0\}$$

is an ideal of R.

Definition 148 *(i) We call the ideal* $\mathbf{o}(m)$ *defined in Lemma 147 the* order ideal *of* m.

(ii) If $\mathbf{o}(m) \neq \{0\}$ *we say that* m *is a* torsion element.

(iii) If a module has no non-zero torsion elements we say that it is torsion free.

Lemma 149 *If* M *is a module over* R *and* T *is the set of torsion elements in* M *then* T *is a submodule of* M *and* M/T *is a torsion free module.*

We adopt a definition of linear independence which is taken directly from vector spaces.

Definition 150 *If* M *is a module over* R *we say that elements* m_1, m_2, $\dots$, m_n *are* linearly independent *if the equation*

$$\sum_{j=1}^{n} r_j m_j = 0$$

with $r_j \in R$ $[j = 1, 2, \dots, n]$ *only has the solution* $r_1 = r_2 = \cdots = r_n = 0$.

The next definition parallels the idea of a basis for a vector space

Definition 151 *If* M *is a module over* R *generated by linearly independent elements* m_1, m_2, $\dots$, m_n *we say that the elements form a* basis *for* M *and that they* generate M freely*. We say that* M *is a* finitely generated free module.

(More generally M is *freely generated* if it has a subset X which generates M and is such that any non-empty finite subset of X is linearly independent. We shall not make use of this idea.)

Lemma 152 *If* M *is a module over* R *and* m_1, m_2, $\dots$, $m_t \in M$ *the following four statements are equivalent.*

(i) The elements m_1, m_2, $\dots$, m_t *form a basis for* M.

(ii) Any element m *of* M *can be written in one and only one way as*

$$m = \sum_{j=1}^{t} r_j m_j$$

with $r_j \in R$.

(iii) The elements m_1, m_2, $\dots$, m_t *generate* M *and the following condition holds. If* N *is an* R *module and* $n_j \in N$ *then there exists a homomorphism* $\phi : M \to N$ *with* $\phi(m_j) = n_j$ $[1 \leq j \leq t]$.

(iv) Each m_j *is torsion free (i.e. not a torsion element) and*

$$M = m_1 R \oplus m_2 R \oplus \cdots \oplus m_t R.$$

Algebraists would prefer to use condition (iii) or something like it as the definition of freely generated since it chimes in with their predeliction for *universal objects*.

The following remark is more or less obvious.

Lemma 153 *The module M over a ring R is freely generated by t elements if and only if*

$$M \cong \underbrace{R \oplus R \oplus R \oplus \cdots \oplus R}_{t}.$$

The next remark is almost as obvious but will play a key role in the proof of our module decomposition theorem (Theorem 171).

Lemma 154 *If a module M over a ring R is finitely generated then we can find a finitely generated free module F and an injective homomorphism $\phi : F \to M$. (In other words, every finitely generated module is the image of some finitely generated free module.)*

In the case of a cyclic module, Lemma 154 can be sharpened.

Lemma 155 *Suppose M is a cyclic module over a ring R generated by m. Then*

$$M \cong R/\mathbf{o}(m).$$

In particular two cyclic modules over R are isomorphic if and only if their generating elements have the same order ideal.

Thus if M is a cyclic module generated by m it is natural to call $\mathbf{o}(m)$ the order ideal of M.

12 Matrices and modules

There is no problem in extending the notion of an $r \times s$ matrix together with the definitions of matrix addition, matrix multiplication and so forth from fields to rings.

Lemma 156 *Let M and N be finitely generated free modules over a ring R. Suppose that M has basis m_1, m_2, $\dots$, m_r and that N has basis n_1, n_2, $\dots$, n_s. Then there is bijection $\alpha \leftrightarrow A$ between homomorphisms $\alpha : M \to N$ and $r \times s$ matrices $A = (a_{ij})$ over R given by*

$$\alpha(m_j) = \sum_{i=1}^{s} a_{ij} n_i.$$

I repeat my warning that generalising results from vector spaces is the natural way forward but that we must act as though we were walking on eggs. The care required may not be obvious to the reader who looks only at the theorems we *do prove* but will be obvious to anyone who asks about the theorems we *do not prove.*

> ‘Is there any other point to which you would wish to draw my attention?’ ‘To the curious incident of the dog in the night-time.’ ‘The dog did nothing in the night-time.’ ‘That was the curious incident.’ remarked Sherlock Holmes.

Definition 157 *We say that an $s \times s$ matrix A over R is invertible if there exists an $s \times s$ matrix $\tilde{A}$ with $A\tilde{A} = \tilde{A}A = I$.*

The standard uniqueness argument shows that $\tilde{A}$, if it exists, is unique.

Lemma 158 *The product of $s \times s$ invertible matrices is itself invertible. (Thus the $s \times s$ invertible matrices over R form a group.)*

Example 159 *The matrix*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

over $\mathbb{Z}$ is invertible if and only if $ad - bc = \pm 1$.

Lemma 160 *Suppose that M is a finitely generated free module over a ring R and that M has basis m_1, m_2, …, m_s. If $A = (a_{ij})$ is an $s \times s$ invertible matrix A over R and*

$$m_j^* = \sum_{i=1}^{s} a_{ij} m_i$$

then m_1^*, m_2^*, …, m_s^* is also a basis for M.

When we dealt with matrices over fields we used elementary row and column operations and their associated matrices. We can do the same thing here. First let us set out the corresponding elementary $s \times s$ matrices.

(i) F_{ij} is the matrix obtained from the identity matrix by interchanging row i and row j.

(ii) $G_i(u)$ is the matrix obtained from the identity matrix by multiplying row i by the unit u.

(iii) $H_{ij}(r)$ is the matrix obtained from the identity matrix by adding r times row j to row i $[i \neq j, r \in R]$.

(iv) $\bar{H}_{ij}(r)$ is the matrix obtained from the identity matrix by adding r times column j to column i $[i \neq j, r \in R]$.

We shall not use $G_i(u)$ but we include it for completeness. Observe that $\bar{H}_{ij}(r) = H_{ji}(r)$. Exactly as in the field case we have the following easy remarks.

Lemma 161 *(i) The effect of pre-multiplying a matrix of the appropriate size*

(1) by F_{ij} is to interchange row i and row j,

(2) by $G_i(u)$ is to multiply row i by the unit u,

(3) by $H_{ij}(r)$ is to add r times row j to row i.

(ii) The effect of post-multiplying a matrix of the appropriate size

(1) by F_{ij} is to interchange column i and column j,

(2) by $G_i(u)$ is to multiply column i by the unit u,

(3) by $\bar{H}_{ij}(r)$ is to add r times column j to column i.

(iii) The matrices F_{ij}, $G_i(u)$, $H_{ij}(r)$ and $\bar{H}_{ij}(r)$ are all invertible.

When we worked over fields we where able to reduce matrices to very special forms by pre- and post-multiplication by invertible matrices.

Definition 162 *Let A and B be $s \times t$ matrices over a ring R. We say that A and B are equivalent if we can find an invertible $s \times s$ matrix P and an invertible $t \times t$ matrix Q such that $B = PAQ$.*

Lemma 163 *(i) Equivalence of matrices is an equivalence relation.*

(ii) Let M and N be finitely generated free modules over a ring R. Suppose that M has basis m_1, m_2, …, m_s and that N has basis n_1, n_2, …, n_t. Suppose that the homomorphism $\alpha : M \to N$ corresponds to the matrix A for these bases. If A is equivalent to B then we can find bases m_1^*, m_2^*, …, m_s^* for M and n_1^*, n_2^*, …, n_t^* for N such that $\alpha : M \to N$ corresponds to the matrix B for these bases.

We can not do very much over general rings but we can do a great deal over Euclidean domains.

Lemma 164 *If A is a non-zero $s \times t$ matrix over a Euclidean domain we can find a sequence of elementary row and column operations which reduce A to a matrix B with $b_{i1} = 0$ for $2 \leq i \leq s$, $b_{1j} = 0$ for $2 \leq j \leq t-1$ and b_{11} dividing every element b_{ij} of B.*

Lemma 165 *If A is a $s \times t$ matrix over a Euclidean domain we can find a sequence of elementary row and column operations which reduce A to a matrix D with $d_{ij} = 0$ for $i \neq j$ (that is D is diagonal) and $d_{ii} | d_{(i+1)(i+1)}$ for all $1 \leq i \leq \min(s,t) - 1$.*

We restate Lemma 165 as a theorem.

Theorem 166 *If A is a $s \times t$ matrix over a Euclidean domain then A is equivalent to a diagonal matrix D with $d_{ii}|d_{(i+1)(i+1)}$ for all $1 \leq i \leq \min(s,t)-1$.*

This result is ultimately due to Henry Smith who proved it for integer valued matrices. Smith was a major pure mathematician at a time and place (19th century Oxford) not particularly propitious for such a talent. He seems to have been valued more as a good College and University man than for anything else[3].

In the next section we obtain the module decomposition theorem (Theorem 171) as a direct consequence of Theorem 166 but for the moment we just note a simple corollary.

Lemma 167 *Let M be a finitely generated free module over a Euclidean domain then all bases of M contain the same number of elements.*

We call the number of elements in a basis of M the *rank* of M.

There are two important remarks to make.

(1) The results which we obtain for Euclidean domains can be extended with a little more work to principal ideal domains. The details are given in [5] Chapters 7 and 8. However all our applications will be to Euclidean domains. (I remarked earlier on the difficulty of finding simple examples of principal ideal domains which are not Euclidean.)

There is a further point. Our applications will be to modules over a domain R where R is $\mathbb{Z}$ and $\mathbb{C}[X]$. For both of these the Euclidean function ϕ is such that given $a \in R$ and a non-zero $b \in R$ there is an *algorithm* for finding $c, r \in R$ such that $a = cb + r$ and $\phi(r) < \phi(b)$. The proof of Theorem 166 is thus *algorithmic*, that is we can actually *calculate* P, Q invertible and D of the correct form such that $PAQ = D$. We are thus not doing abstract algebra but concrete algebra which can be (and is) programmed for electronic computers.

(2) We shall not give general uniqueness theorems corresponding to our general decomposition theorems. Such results will again be found in [5] Chapters 7 and 8. They are not very hard but a 24 hour course cannot contain everything. In the concrete examples that we give uniqueness will be more or less obvious.

[3] He even supervised on Sunday afternoon, telling his students that 'It was lawful on the Sabbath day to pull an ass out of the ditch'.

13 The module decomposition theorems

We are now within sight of our module decomposition theorems. We need three preliminary lemmas. The first is a simple consequence of Lemma 147.

Lemma 168 *If M is a cyclic module over a principal ideal domain D then $M \cong D/(d)$ for some $d \in D$. If $D/(d) \cong D/(d')$ then d and d' are associates.*

We say that M is *of order d*.

The second requires a little work.

Lemma 169 *Every submodule G of a finitely generated free module F over a principal ideal domain D is itself a finitely generated free module. The rank of G is no greater than the rank of F.*

The third is routine abstract algebra.

Lemma 170 *Let M be the internal direct sum*

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_s$$

of submodules M_i. Suppose N_i is a submodule of N_i for each i and $N = N_1 + N_2 + \cdots + N_s$. If ν is the natural homomorphism $\nu : M \to M/N$ then

$$M/N = \nu(M) = \nu(M_1) \oplus \nu(M_2) \oplus \cdots \oplus \nu(M_s)$$

and $\nu(M_i) \cong M_i/N_i$.

Theorem 166 now gives us our first decomposition theorem.

Theorem 171 (Basic module decomposition theorem) *If M is a finitely generated module over a Euclidean domain D then M may be written as an internal direct sum*

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_s$$

where M_i is a non-trivial cyclic submodule of order d_i $[1 \leq i \leq s]$ and $d_i | d_{i+1}$ $[1 \leq i \leq s-1]$.

Let us note the following consequences.

Lemma 172 *If M is a finitely generated module over a Euclidean domain D then $M = T \oplus F$ where T is the torsion submodule and F is a finitely generated free module.*

Lemma 173 *If M is a finitely generated torsion free module over a Euclidean domain D then M is a finitely generated free module.*

Turning from the general to the concrete we obtain a structure theorem for finitely generated Abelian groups.

Theorem 174 *If G is a finitely generated Abelian group then (as a group)*

$$G \cong \mathbb{Z}_{d_1} \oplus \mathbb{Z}_{d_2} \cdots \oplus \mathbb{Z}_{d_r} \oplus \mathbb{Z}^t$$

where $d_i | d_{i+1}$ $[1 \leq i \leq r-1]$.

(Note that this result can be stated entirely in group theoretic terms.)

Lemma 175 *(We work with groups and group isomorphism.)*
(i) If $\mathbb{Z}^t \cong \mathbb{Z}^{t'}$ then $t = t'$.
(ii) If

$$\mathbb{Z}_{d_1} \oplus \mathbb{Z}_{d_2} \cdots \oplus \mathbb{Z}_{d_r} \cong \mathbb{Z}_{d'_1} \oplus \mathbb{Z}_{d'_2} \cdots \oplus \mathbb{Z}_{d'_r}$$

with $d'_i | d'_{i+1}$ $[1 \leq i \leq r'-1]$ and $d_i | d_{i+1}$ $[1 \leq i \leq r-1]$ then $r = r'$ and $d'_i = d_i$ for $1 \leq i \leq r$.
(iii) The decomposition in Theorem 174 is unique.

Notice that we have provided an algorithm which presented with generators for an Abelian group together with relations between them can decide if the largest group compatible with these relations is finite or infinite. It has been shown (though the proof is book length) that no such algorithm can exist for the non-Abelian case that is there exists no computer program which presented with generators for a group together with relations between them can decide if the largest group compatible with these relations is finite or infinite. (This subject is known as the *word problem for groups*.)

Of course, the group $\mathbb{Z}_6$ can be decomposed still further as $\mathbb{Z}_6 = \mathbb{Z}_2 \oplus \mathbb{Z}_3$. This fact suggests that we develop our decomposition theorem, Theorem 171, as follows. Recall that a Euclidean domain is a principal ideal domain and so a unique factorisation domain. Our main result echos the Chinese remainder theorem.

Lemma 176 *Let M be a cyclic module of order d over a principal ideal domain D. If d has the prime factorisation $d = up_1^{\alpha_1} p_2^{\alpha_2} \ldots p_s^{\alpha_s}$ with u a unit, the p_i non-associate primes and $\alpha_i \geq 1$ then*

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_s$$

where M_j is cyclic of order $p_j^{\alpha_j}$.

Cyclic modules of order p^α with p a prime are called primary modules.

Theorem 177 (Primary decomposition theorem) *If M is a finitely generated module over a Euclidean domain D then M may be written as an internal direct sum*

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_s$$

where M_i are primary modules or free cyclic modules.

It is worth noting that no further splitting is possible.

Definition 178 *A non-trivial module over a ring R is called indecomposable if whenever $M = M_1 \oplus M_2$ with M_1, M_2 submodules then either $M_1 = \{0\}$ or $M_2 = \{0\}$.*

Lemma 179 *(i) A primary module over a principal ideal domain is indecomposable.*

(ii) A free cyclic module over an integral domain is indecomposable.

Theorem 177 immediately gives a structure theorem for finitely generated Abelian groups.

Theorem 180 *If G is a finitely generated Abelian group then (as a group)*

$$G \cong \mathbb{Z}_{p_1^{\alpha_1}} \oplus \mathbb{Z}_{p_2^{\alpha_2}} \cdots \oplus \mathbb{Z}_{p_r^{\alpha_r}} \oplus \mathbb{Z}^t$$

where p_i is a prime and $\alpha_i \geq 1$ $[1 \leq i \leq r]$. If we add the condition $p_i^{\alpha_i} \leq p_{i+1}^{\alpha_{i+1}}$ the decomposition is unique.

The following example shows that things are not so simple for non-finitely generated Abelian groups (and so, certainly, for modules in general) as one might at first imagine.

Example 181 *(i) Consider the Abelian group $\mathbb{Q}$. Any non-trivial finitely generated subgroup is generated by a single element and is thus isomorphic to $\mathbb{Z}$. However $\mathbb{Q}$ is not finitely generated.*

(ii) Consider the Abelian group $\mathbb{Q}/\mathbb{Z}$. Any non-trivial finitely generated subgroup is a finite cyclic group. However $\mathbb{Q}/\mathbb{Z}$ is not finitely generated.

Since a finite Abelian group is automatically finitely generated we have a complete classification of all finite Abelian groups.

Theorem 182 *If G is a finite Abelian group then*

$$G \cong \mathbb{Z}_{p_1^{\alpha_1}} \oplus \mathbb{Z}_{p_2^{\alpha_2}} \cdots \oplus \mathbb{Z}_{p_r^{\alpha_r}}$$

where p_i is a prime and $\alpha_i \geq 1$ $[1 \leq i \leq r]$. If we add the condition $p_i^{\alpha_i} \leq p_{i+1}^{\alpha_{i+1}}$ the decomposition is unique.

The rest of this section is very much off the syllabus but gives a striking application of Theorem 182. We write

$$\mathbb{T} = \{\lambda \in \mathbb{C} : |\lambda| = 1\}$$

and note that $\mathbb{T}$ is an Abelian group under multiplication. We write D_n for the subgroup of $\mathbb{T}$ defined by

$$D_n = \{\omega \in \mathbb{T} : \omega^n = 1\}.$$

(Thus D_n is the multiplicative group of nth roots of unity.) We observe that D_n is group isomorphic to $\mathbb{Z}_n$.

Definition 183 *If G is a finite Abelian group and $\chi : G \to \mathbb{T}$ is a group homomorphism we say that χ is a* character *of G.*

Lemma 184 *The collection $\hat{G}$ of characters of a finite group G form an Abelian group under the multiplication rule $(\chi_1\chi_2)(g) = \chi_1(g)\chi_2(g)$.*

We call $\hat{G}$ the *dual group* of G. Once we have the classification theorem for finite Abelian groups we can use the following easy result to give a corresponding classification for their dual groups.

Lemma 185 *If*

$$G = \mathbb{Z}_{p_1^{\alpha_1}} \oplus \mathbb{Z}_{p_2^{\alpha_2}} \cdots \oplus \mathbb{Z}_{p_r^{\alpha_r}}$$

with p_s is a prime and $\alpha_s \geq 1$ $[1 \leq s \leq r]$ then we may identify $\hat{G}$ with $D_{p_1^{\alpha_1}} \oplus D_{p_2^{\alpha_2}} \oplus \cdots \oplus D_{p_r^{\alpha_r}}$ as follows. If

$$\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_r) \in D_{p_1^{\alpha_1}} \oplus D_{p_2^{\alpha_2}} \oplus \cdots \oplus D_{p_r^{\alpha_r}}$$

and

$$\mathbf{n} = (n_1, n_2, \dots, n_r) \in \mathbb{Z}_{p_1^{\alpha_1}} \oplus \mathbb{Z}_{p_2^{\alpha_2}} \cdots \oplus \mathbb{Z}_{p_r^{\alpha_r}}$$

then

$$\boldsymbol{\omega}(\mathbf{n}) = \prod_{s=1}^{r} \omega_s^{n_s}.$$

We shall only make use of the following consequences.

Lemma 186 *If G is a finite Abelian group then $\hat{G}$ is a finite group with the same number of elements. If $g \in G$ there exists a $\chi \in \hat{G}$ with $\chi(g) \neq 0$.*

Lemma 187 *Let G be a finite Abelian group with N elements. If $g \in G$ but $g \neq e$ then*

$$\sum_{\chi \in \hat{G}} \chi(g) = 0.$$

If $g = e$

$$\sum_{\chi \in \hat{G}} \chi(g) = N.$$

If G is a finite Abelian group let us write $C(G)$ for the set of functions $f : G \to \mathbb{C}$. If $f, h \in C(G)$ we write

$$\langle f, h \rangle = |G|^{-1} \sum_{x \in G} f(x) h(x)^*$$

where $|G|$ is the number of elements of G and z^* denotes the complex conjugate of z.

Lemma 188 *If G is a finite Abelian group then $C(G)$ equipped with the usual pointwise addition and scalar multiplication is a vector space over $\mathbb{C}$. The operation $\langle\ ,\ \rangle$ is an inner product on $C(G)$. The characters of G form an orthonormal basis for G.*

It is thus natural to write

$$\hat{f}(\chi) = \langle f, h \rangle,$$

and call $\hat{f} : \hat{G} \to \mathbb{C}$ the *Fourier transform* of an $f \in C(G)$. Lemma 187 gives us the required representation theorem.

Lemma 189 *If G is a finite Abelian group and $f \in C(G)$ then*

$$f = \sum_{\chi \in \hat{G}} \hat{f}(\chi) \chi.$$

This small but perfectly formed Fourier theory for finite Abelian groups is used in number theory and machine computation. Even more importantly it suggests that we should look at Fourier theory in the context of groups and this gives rise to *representation theory* both for finite non-Abelian groups and for infinite groups satisfying reasonable 'continuity' conditions.

14 Applications to endomorphisms

Throughout this section V will be a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V. We recall from Lemma 129 that V is a module over the ring of polynomials $\mathbb{F}[X]$ with module multiplication defined by $pv = p(\alpha)v$. We observe that $\mathbb{F}[X]$ is a Euclidean domain. Further if V has basis e_1, e_2, $\dots e_n$ as vector space then e_1, e_2, $\dots e_n$ generate V as a module (though, of course they may not be linearly independent when V is considered as a module). We note that every $v \in V$ is a torsion element. We can thus apply our module decomposition theorem (Theorem 171). Translated into the language of vector spaces it takes the following form.

Lemma 190 *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V. Then V may be expressed a the direct sum of subspaces*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

where each V_i is associated with a monic polynomial $P_i \in \mathbb{F}[X]$ of degree n_i and a vector v_i as follows.

(i)$'$ Vectors of the form $\alpha^k v_i$ span V_i.

(ii) We have $P_i(\alpha)(v) = 0$ for all $v \in V_i$.

(iii) If $P \in \mathbb{F}[X]$ and $P(\alpha)(v) = 0$ for all $v \in V_i$ then $P_i | P$.

(iv) $P_i | P_{i+1}$ for all $1 \leq i \leq s-1$.

We can immediately improve the form of this result.

Theorem 191 *As for Lemma 190 but with (i)$'$ replaced by*

(i) v_i, $\alpha(v_i)$, $\alpha^2(v_i)$ $\dots$, $\alpha^{n_i-1}(v_i)$ is a basis for V_i.

It is worth noting that the subspace V_i are *invariant* in the sense that $\alpha(V_i) \subseteq V_i$.

Forgetting about modules for the moment and working entirely in standard vector space theory we can translate Theorem 191 into statement about matrices by choosing the obvious basis for V.

Theorem 192 (Rational canonical form) *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V. Then there is basis for V such that α has matrix A (relative to this basis) which consists of zeros except for s blocks consisting of $n_i \times n_i$ square matrices A_i $[1 \leq i \leq s]$ down* *the diagonal satisfying the following conditions. Each*

$$A(i) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 & -a_0(i) \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & -a_1(i) \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & -a_2(i) \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & -a_{n_i-2}(i) \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -a_{n_i-1}(i) \end{pmatrix}$$

and, if we write

$$P_i(X) = X^{n_i} + \sum_{k=0}^{n_i-1} a_k X^k,$$

we have $P_i | P_{i+1}$ for all $1 \leq i \leq s-1$.

Definition 193 *An $n \times n$ matrix of the form*

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -a_{n-1} \end{pmatrix}$$

is called the companion matrix *of the monic polynomial*

$$p(X) = X^n + \sum_{k=0}^{n-1} a_k X^k.$$

Lemma 194 *If A is the companion matrix of a monic polynomial P then*

$$P(X) = \det(XI - A).$$

(N.B. we have here a relation between coefficients with X an indeterminate.) In other words P is the characteristic polynomial of A.

We can now grasp some of the implications of our results on endomorphisms.

Theorem 195 *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V. Let p_i $[1 \leq i \leq s]$ be the polynomials which appear in Theorems 191 and 192.*

(i) $\prod_{i=1}^{s} p_i$ is the characteristic polynomial of A (and so of α).

(ii) The polynomial p_s is the minimal polynomial (more exactly the minimal annihilating polynomial) of α (and so of A). In other words $p_s(\alpha) = 0$ and if $q \in \mathbb{F}[X]$ satisfies $q(\alpha) = 0$ then $p_s | q$.

Incidentally we have proved the Cayley Hamilton theorem for general fields. (The proof via triangular matrices in Course P1 only works for $\mathbb{C}$ though we can deduce the result for real matrices by considering them as complex matrices.)

Theorem 196 (Cayley Hamilton) *If V is a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V then α satisfies its own characteristic equation.*

We can also prove that the rational canonical decomposition is indeed canonical.

Lemma 197 *The matrix in Theorem 192 is uniquely determined by the given conditions.*

What about the Primary Decomposition Theorem (Theorem 177)? Working along the same lines as Theorem 191, we obtain the following result.

Lemma 198 *Let V be a finite dimensional vector space over a field $\mathbb{F}$ and α an endomorphism of V. Then V may be expressed a the direct sum of subspaces*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

where each V_i is associated with a monic polynomial $P_i \in \mathbb{F}[X]$ of degree n_i and a vector v_i as follows.

(i) v_i, $\alpha(v_i)$, $\alpha^2(v_i)$ $\ldots$, $\alpha^{n_i-1}(v_i)$ is a basis for V_i.

(ii) We have $P_i(\alpha)(v) = 0$ for all $v \in V_i$.

(iii) If $P \in \mathbb{F}[X]$ and $P(\alpha)(v) = 0$ for all $v \in V_i$ then $P_i|P$.

(iv) P_i is a power of an irreducible polynomial (that is $P_i = Q_i^{m_i}$ where Q_i is irreducible and $m_i \geq 1$).

Even such a simple field as $\mathbb{R}$ has both linear and quadratic irreducible polynomials and even rather weighty tomes on algebra do not seek any use for Lemma 198 in this case. However if the field is *complete*, that is every polynomial has a root, then the only irreducible polynomials are linear and we get a relatively simple result.

Theorem 199 *Let V be a finite dimensional vector space over a complete field $\mathbb{F}$ and α an endomorphism of V. Then V may be expressed a the direct sum of subspaces*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$$

where each V_i is associated with a polynomial $(X - \lambda_i)^{n_i} \in \mathbb{F}[X]$ and a vector w_i as follows.

(i) w_i, $(\alpha - \lambda_i\iota)w_i$, $(\alpha - \lambda_i\iota)^2 w_i$, $\dots$, $(\alpha - \lambda_i\iota)^{n_i - 1}w_i$ *form a basis for* V_i.
(ii) We have $(\alpha - \lambda_i\iota)^{n_i}(v) = 0$ *for all* $v \in V_i$.
(iii) If $P \in \mathbb{F}[X]$ *and* $P(\alpha)(v) = 0$ *for all* $v \in V_i$ *then* $(X - \lambda_i)^{n_i} | P$.

As with Theorem 191 the obvious choice of basis for V gives us a theorem about matrices. We write $J(\lambda, n)$ for the $n \times n$ matrix with λ's down the diagonal, 1's immediately below and zero every where else so that

$$J(\lambda, n) = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \lambda \end{pmatrix}.$$

We call $J(\lambda, n)$ a Jordan matrix.

Theorem 200 (Jordan normal form) *Let* V *be a finite dimensional vector space over a complete field* $\mathbb{F}$ *and* α *an endomorphism of* V*. Then there is basis for* V *such that* α *has matrix* A *(relative to this basis) which consists of zeros except for Jordan matrices* $J(\lambda_i, n_i)$ $[1 \leq i \leq s]$ *down the diagonal.*

Standard vector space techniques complete the result.

Lemma 201 *The matrix associated with* α *in Theorem 200 is unique up to reordering the diagonal blocks.*

Two and a half years ago in Course C1 we noted that the matrix

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

showed that not all square matrices are diagonalisable even over $\mathbb{C}$. By ad hoc techniques we showed that any 2×2 matrix was conjugate to a matrix of the form

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \text{ or } \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

We noted that these forms were particularly useful in the study of differential equations.

We complete the course by giving the full solution of the problem of classifying square matrices under conjugation over any complete field and, in particular, over $\mathbb{C}$.

Theorem 202 (Jordan normal form for matrices) *If A is a $n \times n$ matrix over a complete field then we can find an invertible $n \times n$ matrix P such that $J_A = PAP^{-1}$ consists of zeros except for Jordan matrices $J(\lambda_i, n_i)$ $[1 \leq i \leq s]$ down the diagonal. The matrix J_A so associated with A is unique up to reordering the diagonal blocks.*

15 Reading and further reading

Not all theorems in mathematics are hard to prove though some are. I would hope that the reader will be able to prove many of the results in the notes as exercises. Where she cannot, the results on rings, integral domains and factorisation (sections 1 to 5) will be found in the standard algebra texts in her College library and in the book of Hartley and Hawkes *Rings, Modules and Linear Algebra* [5]. I have tried (but not very hard and with only partial success) to follow the notation of Hartley and Hawkes. Whichever text she follows she should note that our decision to use ring to mean *commutative ring with 1* is not standard.

The material in sections 6 to 9 belong to Galois theory. Garling's *A Course in Galois Theory* [4] is, not surprisingly, very much in tune with the approach adopted in Cambridge but, again, most of the standard algebra texts cover the material. The book of Hartley and Hawkes covers the remainder of the course on modules and their decomposition theorems. (Since we aim to get to the decomposition theorems as fast as possible and we do not deal with uniqueness, Hartley and Hawkes contains somewhat more material. Since most British algebraists under the age of 50 learnt their module theory from Hartley and Hawkes the close relation between book and syllabus is no accident.)

Turning specifically to some of the more general algebra texts we note that Volume 1 of Cohn's *Algebra* [2] covers most of the material including modules is a typically efficient manner. Those who like to proceed from the general to the particular will find their tastes catered for in MacLane and Birkhoff's *Algebra* [7]. Those who prefer the other direction will also prefer Birkhoff and MacLane's *Introduction to Modern Algebra* [1] but this covers much less of the course. The syllabus also commends Fraleigh's *A First Course in Modern Algebra* [3] but, I must confess that like many American textbooks it reminds me of the vegetables in an American supermarket, whose splendid appearance does not compensate for their bland taste. In any case the reader will do better to browse through several general texts rather than concentrate on one.

The reader who wants to learn more about the topics treated in the course is in an unusually fortunate position. Most mathematicians treat textbook writing in the same way that lawyers treat drafting legal documents and believe that, once they have covered every possible contingency in the most precise manner possible, their job is done. However, Ian Stewart's (yes, the man you saw on TV) first book *Galois Theory* [8] is a brilliantly written text on a fascinating subject and a pleasure to read. He joined David Tall to write *Algebraic Number Theory* [9] which gives the concrete number theory which partners our abstract treatment of factorisation. Kline's *Mathematical Thought from Ancient to Modern Times* provides a picture of mathematical progress from antiquity and thus a context for this course, and indeed the whole Tripos.

References

[1] G. Birkhoff and S. MacLane *A Survey of Modern Algebra* (3rd Ed) Macmillan, New York, 1965.

[2] P. M. Cohn *Algebra* (2nd Ed) Vol 1, Wiley, 1982.

[3] J. B. Fraleigh *A First Course in Modern Algebra* (5th Ed) Addison-Wesley, 1989.

[4] D. J. H. Garling *A Course in Galois Theory* CUP, 1986.

[5] B. Hartley and T. O. Hawkes *Rings, Modules and Linear Algebra* Chapman and Hall, 1970.

[6] M. Kline *Mathematical Thought from Ancient to Modern Times* OUP, 1972.

[7] S. MacLane and G. Birkhoff *Algebra* (2nd Ed) Macmillan, New York, 1979.

[8] I. Stewart *Galois Theory* Chapman and Hall, 1973.

[9] I. Stewart and D. Tall *Algebraic Number Theory* Chapman and Hall, 1979.