Lecture Notes

for

Algebra 2B

Gunnar Traustason
(Autumn 2012)

1 Rings

I. Definitions and basic properties

Informally the idea behind a ring is of a set equipped with a "sensible" addition and multiplication. We would like the definition to be broad enough to include examples like the $n \times n$ matrices over a fixed field with the usual matrix addition and multiplication, the polynomials with coefficients in some fixed field with the usual polynomial addition and multiplication and the integers. At the same time we want the definition to be somewhat restricted so that we can build a general useful theory that deals with all these examples at once. Before recalling the formal definition of a ring as well as the definition of a group, we first introduce binary operations.

Definition. Let S be a set. A binary operation on S is a function

$$f : S \times S \to S$$

Remark. The binary operations that will crop up here, will usually be referred to as an addition denoted by $+$ or a multiplication denoted by $\cdot$. Instead of writing $+(a, b)$ or $\cdot(a, b)$ one writes then normally $a + b$ and $a \cdot b$.

Definition. A *group* is a pair $(G, *)$, where G is a set, $*$ is a binary operation on G and the following axioms hold:

(a) (The associative law)

$$(a * b) * c = a * (b * c) \text{ for all } a, b, c \in G.$$

(b) (Existence of an identity) There exist an element $e \in G$ with the property that

$$e * a = a \text{ and } a * e = a \text{ for all } a \in G.$$

(c) (The existence of an inverse) For each $a \in G$ there exists an element $b \in G$ such that

$$a * b = b * a = e.$$

Remarks.(1) Recall that the identity e is the unique element in G with the property given in (b). If we had another element f with this property, then

$$f = e * f = e$$

where the first identity follows from the fact that e satisfies the property and the latter from the fact that f satisfies the property.

(2) Recall that for a given $a \in G$, the element $b \in G$ as in (c) is unique. If we had another such element c then

$$b = b * e = b * (a * c) = (b * a) * c = e * c = c.$$

We call this unique element b, the inverse of a. It is often denoted a^{-1} (or $-a$ when the group operation is commutative).

(3) If it is clear from the context what the group operation $*$ is, one often simply refers to the group G rather then the pair $(G, *)$.

Definition. We say that a group $(G, *)$ is *abelian* or *commutative* if $a * b = b * a$ for all $a, b \in G$.

Definition. If $(G, *)$ is a group then a subset H of G is said to be a subgroup of G if the following three properties hold.

(1) $a * b \in H$ for all $a, b \in H$.
(2) $e \in H$ where e is the group identity of G.
(3) $a^{-1} \in H$ for all $a \in H$, where a^{-1} is the inverse of a in G.

Remark. It is not difficult to see that one could equivalently say that H is a subgroup of G if and only if $(H, *)$ is a group. So subgroups are groups contained within G that inherit the multiplication from G.

Definition. A *ring* is a triple $(R, +, \cdot)$, where we have a set R and two binary operations $+$ (addition) and $\cdot$ (multiplication) on R such that the following axioms hold.

a) $(R, +)$ is an abelian group. The additive identity is denoted 0 and the additive inverse of a is denoted $-a$. So we have

$$(a + b) + c = a + (b + c) \text{ for all } a, b, c \in R$$
$$a + b = b + a \text{ for all } a, b \in R$$
$$a + 0 = a \text{ for all } a \in R$$
$$a + (-a) = 0 \text{ for all } a \in R.$$

b) $(R, \cdot)$ satisfies the associative law and has a multiplicative identity denoted 1. That is we have

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \text{ for all } a, b, c \in R$$
$$a \cdot 1 = 1 \cdot a = a \text{ for all } a \in R.$$

c) R satisfies the distributive laws:

$$\begin{aligned} a \cdot (b + c) &= (a \cdot b) + (a \cdot c) \\ (b + c) \cdot a &= (b \cdot a) + (c \cdot a) \end{aligned}$$

for all $a, b, c \in R$.

Remark. As $(R, +)$ is a group we know that 0 is the unique additive identity. We also have that 1 is the unique multiplicative identity. The same arguement as before works. If $\bar{1}$ was another multiplicative identity, then $\bar{1} = \bar{1} \cdot 1 = 1$.

Remark. We often omit $\cdot$ and write ab instead of $a \cdot b$. For simplicity it is also useful to avoid brackets when there is no ambiguity. Here the same conventions hold as for real numbers and we assume that $\cdot$ has priority over $+$. For example

$$ab + ac$$

stands for $(a \cdot b) + (a \cdot c)$ and not $(a \cdot (b + a)) \cdot c$. One also writes a^2 for $a \cdot a$ and $2a$ for $a + a$ and so on.

Definition. A ring $(R, +, \cdot)$ is said to be *commutative* if

$$a \cdot b = b \cdot a$$

for all $a, b \in R$.

Lemma 1.1 *In any ring $(R, +, \cdot)$, we have*

a) $a \cdot 0 = 0$ and $0 \cdot a = 0$ for all $a \in R$.
b) $a \cdot (-b) = -(a \cdot b)$ and $(-a) \cdot b = -(a \cdot b)$.

Proof a) Using the fact that 0 is an additive identity and one of the distributive laws, we have

$$a \cdot 0 + 0 = a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0.$$

Adding $-(a \cdot 0)$ on the left on both sides gives

$$-(a \cdot 0) + (a \cdot 0 + 0) = -(a \cdot 0) + (a \cdot 0 + a \cdot 0).$$

Because of the associative law, this implies that

$$(-(a \cdot 0) + a \cdot 0) + 0 = ((-a \cdot 0) + a \cdot 0) + a \cdot 0$$

and thus $0 + 0 = 0 + a \cdot 0$ or $0 = a \cdot 0$. The second identity is proved similarly.

b) We have $a \cdot b + a \cdot (-b) = a \cdot (b + (-b)) = a \cdot 0 = 0$. Hence $a \cdot (-b)$ is the additive inverse of ab. In other words $a \cdot (-b) = -(a \cdot b)$. Similarly for the second identity. □

Definition. An element $a \in R$ is called a *unit* if it has a multiplicative inverse. That is, if there exists $b \in R$ such that $a \cdot b = b \cdot a = 1$.

Remark. The inverse b of a, if it exists, is unqiue. If c were another inverse then $c = 1 \cdot c = (b \cdot a) \cdot c = b \cdot (a \cdot c) = b \cdot 1 = b$. We will denote the inverse by a^{-1}.

Definition. Let R be a ring. The set R^* consisting of all the units of R is called the *group of units*.

Lemma 1.2 *R^* is a group with respect to the ring multiplication.*

Proof. See exercise 3 on sheet 1.

Examples. (1) $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$.
(2) $\mathbb{Z}^* = \{1, -1\}$.

Remark. If 0 is a unit then

$$1 = 0 \cdot 0^{-1} \overset{L\ 1.1}{=} 0$$

and so for all $a \in R$, we have $a = a \cdot 1 = a \cdot 0 = 0$. Hence we must have $R = \{0\}$. We usually try to avoid this ring.

Definition. We say that a commutative ring R is a field if $R \neq \{0\}$ and every $0 \neq a \in R$ has a multiplicative inverse. (In other words if R is commutative and $R^* = R \setminus \{0\}$).

II. Examples of rings.

By definition, every field is an commutative ring. In particular we have that $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are all rings with respect to the usual addition and multiplication.

An example of a commutative ring that is not a field is $\mathbb{Z}$, the ring of integers. Let K be a field then the set, $M_n(K)$, of all $n \times n$ matrices over K is a ring with respect to the usual matrix addition and multiplication (see exercise 2 on sheet 1). This ring is usually not commutative.

The ring End(V). Let V be a finite dimensional vector space over a field K. Let $\mathrm{End}(V)$ be the set of all linear operators $\alpha : V \to V$ (also called endomorphisms on V). We associate with $\mathrm{End}(V)$ an addition and a multiplication as follows. For $\alpha, \beta \in \mathrm{End}(V)$ we let $[\alpha + \beta] : V \to V$ be the map that takes v to $\alpha(v) + \beta(v)$. Notcie that this map is linear as

$$\begin{aligned} [\alpha + \beta](v + w) &= \alpha(v + w) + \beta(v + w) \\ &= \alpha(v) + \alpha(w) + \beta(v) + \beta(w) \\ &= (\alpha(v) + \beta(v)) + (\alpha(w) + \beta(w)) \\ &= [\alpha + \beta](v) + [\alpha + \beta](w). \end{aligned}$$

and

$$[\alpha + \beta](\lambda v) = \alpha(\lambda v) + \beta(\lambda v) = \lambda\alpha(v) + \lambda\beta(v) = \lambda(\alpha(v) + \beta(v)) = \lambda[\alpha + \beta](b).$$

We also define the multiplication $\cdot$ on $\mathrm{End}(V)$ to be the composition of maps. Thus if $\alpha, \beta \in \mathrm{End}(V)$, then

$$\alpha \cdot \beta(v) = \alpha(\beta(v).$$

We know from a previous algebra courses that $\alpha \cdot \beta$ is then also linear and thus in $\mathrm{End}(V)$.

On exercise 1 on sheet 1 we show that $\mathrm{End}(V)$ is a ring with respect to this addition and multiplication. This ring is usually not commutative.

Remark. Let R be a ring an $z \in R$ an element that commutes with all the elements of R. Let

$$a_0, a_1, a_2, \ldots, \ b_0, b_1, b_2, \ldots$$

be two sequences of elements in R, where all but finitely many elements are equal to 0.

Using the distributive laws, commutative law for the addition, and the associative laws for addition and multiplication, we see that

$$\begin{aligned}(\sum_{r=0}^{\infty} a_r z^r) \cdot (\sum_{s=0}^{\infty} b_s z^s) &= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} a_r z^r b_s z^s \\ &= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} a_r b_s z^{r+s} \\ &= \sum_{n=0}^{\infty}(\sum_{r+s=n} a_r b_s) z^n\end{aligned}$$

The ring $R[x]$ of polynomials with coefficients in R. Let R be a ring and let x be a variable. A polynomial f over R is a formal expression

$$f = \sum_{k=0}^{\infty} a_k x^k$$

with $a_k \in R$ and all but finitely many elments in the sequence (a_k) equal to 0. The degree of a polynomial p is the largest n such that $a_n \neq 0$. (If there is no such n, that is if all the coefficients are 0 then the degree is $-\infty$). We then often write

$$f = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

(and $f = 0$ if all the coefficents are 0). We define addition and multiplication on $R[x]$ in the usual way by setting

$$\sum_{k=0}^{\infty} a_k x^k + \sum_{k=0}^{\infty} b_k x^k = \sum_{k=0}^{\infty} (a_k + b_k) x^k$$

and

$$(\sum_{k=0}^{\infty} a_k x^k) \cdot (\sum_{k=0}^{\infty} b_k x^k) = \sum_{k=0}^{\infty} c_k x^k$$

where

$$c_n = \sum_{i+j=n} a_i b_j$$

(Notice that the remark above motivates the definition of the multiplication). Clearly the polynomials 0 and 1 are the additive and multiplicative identities respectively. As R is an abelian group with respect to the ring addition it follows readily that $(R[x], +)$ is an abelian group. To see that $R[x]$ is a ring with respect to this polynomial addition

and multiplication it remains to see that the multiplication is associative and that the distributive laws hold. Let

$$f = \sum_{k=0}^{\infty} a_k x^k, \quad g = \sum_{k=0}^{\infty} b_k x^k, \quad h = \sum_{k=0}^{\infty} c_k x^k$$

be polynomials in $R[x]$. The reason why the multiplication is associative is that the nth coefficent of $(fg)h$ is

$$\sum_{i+j+k=n} (a_i b_j) c_k$$

which (as the field multiplication is associative) is the same as

$$\sum_{i+j+k=n} a_i (b_j c_k),$$

the nth coefficient of $f(gh)$. Finally we check the distributive laws. The nth coefficent of $f(g+h)$ is

$$\sum_{i+j=n} a_i (b_j + c_j) = \sum_{i+j=n} a_i b_j + \sum_{i+j=n} a_i c_j$$

which is the nth coefficient of $fg + fh$. Hence $f(g+h) = fg + fh$. Similary one proves that $(g+h)f = gf + hf$.

Remark. For a given polynomial $f = \sum_{k=0}^{\infty} a_k x^k \in R[x]$ one can assocate a function $\bar{f} : R \to R$ that maps z to $f(z) = \sum_{k=0}^{\infty} a_k z^k$. One should be warned however to confuse the two together. For example if R is a finite then there are only finitely many functions from R to R but infintely many polynomials in $R[x]$. So different polynomials will give rise to the same function.

Remark. The variable x is really superfluous. The polynomial $\sum_{k=0}^{\infty} a_k x^k$ depends only on the sequence (a_k) and two polynomials $\sum_{k=0}^{\infty} a_k x^k$ and $\sum_{k=0}^{\infty} b_k x^k$ are the same if and only if $(a_k) = (b_k)$. One could just as well have defined the polynomial to be the sequence (and this is sometimes done). However it is convenient to introduce the variable x as we use polynomials often to define functions.

III. Subrings and quotient rings

In this section we consider two important ways of deriving other ring structures from a given ring that are in some sense dual to each other.

A. Subrings

Notation. We will use the following standard short hand notation for iterated sums:

$$\begin{aligned} na &= \underbrace{a + \cdots + a}_{n} \\ (-n)a &= -(na) \\ 0a &= 0_R. \end{aligned}$$

for any positive integer n and the integer 0. Here o_R is the additive identity of R.

Remark. This is just a notation and has nothing to do with the ring multiplication. Notice hat $0_R \cdot a = 0_R$ is a fact that we can prove but $0a = 0_R$ is just a natural notation when 0 is the zero integer.

Definition (Subring). A subset S of a ring R is said to be a *subring*, if the following holds

a) $1 \in S$;
b) If $a, b \in S$ then $a + b, a \cdot b, -a \in S$.

Remarks. (1) Let S be a subring of R. Then $0_R = 1_R + (-1_R) \in S$. Hence S always contains 0_R.
(2) Every subring S of a ring R must contain the multiplicative identity 1_R. As S is closed under addition and taking additive inverses it is clear that S must then contain $\mathbb{Z}1_R = \{n1_R : n \in \mathbb{Z}\}$. Also $\mathbb{Z}1_R$ is a subring of R. This is the case since $1_R = 11_R \in \mathbb{Z}1_R$,

$$n1_R + m1_R = (n+m)1_R, \quad (n1_R) \cdot (m1_R) = nm1_R$$

which shows that $\mathbb{Z}1_R$ is closed under addition and multiplication, and $-n1_R = (-n)1_R$ which shows that $\mathbb{Z}1_R$ is also closed under taking additive inverses. From this discussion it is clear that $\mathbb{Z}1_R$ is the smallest subring of R.
(3) R is a subring of R.
(4) Notice that $\{0\}$ is usually not a subring of the ring R. Although $\{0\}$ is a ring it does not have the same multiplicative identity as R (except if $R = \{0\}$).

It is not difficult to see (exercise sheet 2), that S is a subring of R if and only if $(S, +, \cdot)$ is a ring with the same multiplicative identity as R.

Examples. (1) $\mathbb{Z}$ is a subring of $\mathbb{Q}$.
(2) $\mathbb{Z} + \mathbb{Z}i$ is a subring of $\mathbb{C}$ (see sheet 2).

B. Congruences and quotient rings

Definition Let R be a ring and let $\simeq$ be an equivalence relation on R. We say that $\simeq$ is a congruence if

$$\begin{aligned} a \simeq b \text{ and } c \simeq d &\Rightarrow a + c \simeq b + d \\ a \simeq b \text{ and } c \simeq d &\Rightarrow a \cdot c \simeq b \cdot d. \end{aligned}$$

For each $a \in R$, we let $[a]$ be the equivalence class containing a.

Lemma 1.3 *Let R be a ring with a congruence $\simeq$ and let $I = [0]$. Then I has the following properties*

$$\begin{aligned} a, b \in I &\Rightarrow a + b \in I \\ a \in I, r \in R &\Rightarrow r \cdot a, a \cdot r \in I. \end{aligned}$$

Furthermore $b \simeq a$ if and only if $b + (-a) \in I$ and $[a] = a + I$ for all $a \in R$.

Proof We have $a \simeq 0$ and $b \simeq 0$. As $\simeq$ is a congruence it follows that $a + b \simeq 0 + 0 = 0$, $r \cdot a \simeq r \cdot 0 = 0$ and $a \cdot r \simeq 0 \cdot r = 0$. Hence $a + b, ra, ar \in I = [0]$. Again, as $\simeq$ is a congruence, we have that if $b \simeq a$ then $b + (-a) \simeq a + (-a) = 0$ and thus $b + (-a) \in I$. Conversely if $b + (-a) \in I$ then $b + (-a) \simeq 0$ and thus $b = b + (-a) + a \simeq 0 + a = a$ which shows that $b \simeq a$. Finally $b \in [a]$ if and only if $b \simeq a$ if and only if $b + (-a) \in I$ if and only if $b \in a + I$. Hence $[a] = a + I$. $\square$

Definition. Let R be a ring. A non-emtpy subset I of R is called an *ideal* if it satisfies the following properties:

$$\begin{aligned} a, b \in I &\Rightarrow a + b \in I \\ a \in I,\, r \in R &\Rightarrow r \cdot a, a \cdot r \in I. \end{aligned}$$

The sets $a + I$ where a runs through R are called the *cosets* of I in R.

Remarks. (1) Let I be an ideal of R. As I is non-emtpy, it contains some element a. But then I contains $0_R = 0_R \cdot a$. This shows that all ideals contain 0_R.
(2) Let R be a ring. The subsets $\{0\}$ and R are always ideals of R.
(3) Let R be a commutative ring and $a \in R$. Then Ra is an ideal of R. To see this notice first that $0 = 0 \cdot a \in I$ and thus $I \neq \emptyset$. It remains to see that I is closed under addition and multiplication from R. But this follows from $r \cdot a + s \cdot a = (r+s) \cdot a$ and $s \cdot (r \cdot a) = (rs) \cdot a$.

The next lemma can be seen as a converse to Lemma 1.3.

Lemma 1.4 *Let R be a ring with an ideal I. Define a relation $\simeq$ on R by letting*

$$b \simeq a \text{ if and only if } b + (-a) \in I.$$

Then $\simeq$ is a congruence and $[a] = a + I$. In particular $[0] = I$.

Proof We first show that $\simeq$ is an equivalence relation. As $0 \in I$ we have $a + (-a) \in I$ and thus $a \simeq a$. Hence $\simeq$ is reflexive. Next, if $a \simeq b$ then $a + (-b) \in I$. But then the additive inverse of this, namely $b + (-a) = (-1_R \cdot (a + (-b))$ is also in I and thus $b \simeq a$. This shows that $\simeq$ is symmetric. Finally if $a \simeq b$ and $b \simeq c$ then $a + (-b), b + (-c) \in I$. As I is closed under addition, it follows that $a + (-b) + b + (-c) = a + (-c)$ is in I and thus $a \simeq c$. Hence $\simeq$ is transitive and thus we have shown that $\simeq$ is an equivalence relation.

We show next that his equivalence relation is a congruence. Suppose $a \simeq b$ and that $c \simeq d$. Then $a + (-b), c + (-d) \in I$. As I is an ideal we then have

$$(a + c) + (-(b + d)) = a + (-b) + c + (-d) \in I$$

and

$$ac + (-bd) = ac + (-ad) + ad + (-bd) = a(c + (-d)) + (a + (-b))d \in I.$$

Hence $a + c \simeq b + d$ and $ac \simeq bd$.

Finally $b \in [a]$ if and only if $b + (-a) \in I$ if and only if $b \in a + I$. This shows that $[a] = a + I$. $\square$

Remark. Let R be a ring. According to Lemmas 1.3 and 1.4, there is a one-to-one correspondence between congruences $\simeq$ on R and ideals I of R.

The quotient ring R/I. Let R be a ring with an ideal I and corresponding congruence $\simeq$. Let

$$R/I = \{[a] = a + I : a \in R\}$$

be the collection of all the congruence classes (that is the cosets of I in R). We define addition and multiplication on R/I as follows:

$$\begin{aligned} [a] + [b] &= [a + b] \\ [a] \cdot [b] &= [ab]. \end{aligned}$$

(The addition and multiplication do not depend on which elements a, b we pick from these two equivalence classes and thus these two operations are well defined. To see this suppose $\tilde{a} \simeq a$ and $\tilde{b} \simeq b$. Since $\simeq$ is a congruence we have $\tilde{a} + \tilde{b} \simeq a + b$ and $\tilde{a} \cdot \tilde{b} = a \cdot b$. Thus $[\tilde{a} + \tilde{b}] = [a + b]$ and $[\tilde{a} \cdot \tilde{b}] = [a \cdot b]$).

We next show that $(R/I, +, \cdot)$ is a ring with an additive identity $[0]$ and a multiplicative identity $[1]$. We need to check that all the axioms hold.

Addition. We have

$$([a] + [b]) + [c] = [a + b] + [c] = [(a + b) + c] = [a + (b + c)] = [a] + [b + c] = [a] + ([b] + [c]),$$

$$[a] + [b] = [a + b] = [b + a] = [b] + [a],$$

and $[a] + [0] = [a + 0] = [a]$. Also as $[a] + [-a] = [a + (-a)] = [0]$, we have that $[-a]$ is the additive identity of $[a]$.

Multiplication. We have

$$([a] \cdot [b]) \cdot [c] = [ab] \cdot [c] = [(ab)c] = [a(bc)] = [a] \cdot [bc] = [a] \cdot ([b] \cdot [c]),$$

and $[a] \cdot [1] = [a \cdot 1] = [a]$ and $[1] \cdot [a] = [1 \cdot a] = [a]$.

Distributive laws. We have

$$\begin{aligned} [c] \cdot ([a] + [b]) &= [c] \cdot [a + b] \\ &= [c(a + b)] \\ &= [ca + cb] \\ &= [ca] + [cb] \\ &= [c] \cdot [a] + [c] \cdot [b] \end{aligned}$$

and

$$([a] + [b]) \cdot [c] = [a + b] \cdot [c]$$

$$\begin{aligned} &= [(a+b)c] \\ &= [ac+bc] \\ &= [ac]+[bc] \\ &= [a]\cdot[c]+[b]\cdot[a]. \end{aligned}$$

So we have shown that R/I is a ring.

Examples.

(1) The ring of integers modulo n, $\mathbb{Z}_n = \mathbb{Z}/\mathbb{Z}n$, where n is a positive integer.

Notice that $a \simeq b$ if and only if $b + (-a) \in \mathbb{Z}n$ if and only if n divides $b - a$. Any integer m can be written of the form

$$m = nr + s, \quad 0 \leq s < n$$

for a unique s. Thus $[m] = [s]$ for a unique s such that $0 \leq s < n$. It follows that

$$\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}.$$

As an example we have that $\mathbb{Z}/\mathbb{Z}3$ has three elements $[0], [1]$ and $[2]$. The addition and multiplication tables are

$+$	$[0]$	$[1]$	$[2]$
$[0]$	$[0]$	$[1]$	$[2]$
$[1]$	$[1]$	$[2]$	$[0]$
$[2]$	$[2]$	$[0]$	$[1]$

$\cdot$	$[0]$	$[1]$	$[2]$
$[0]$	$[0]$	$[0]$	$[0]$
$[1]$	$[0]$	$[1]$	$[2]$
$[2]$	$[0]$	$[2]$	$[1]$

Notice that any element $a \neq [0]$ has a multiplicative inverse so this is a field. As you have probably seen before (and we will see again later), $\mathbb{Z}_n$ is a field if and only if n is a prime.

(2) The ring $\mathbb{R}[x]/\mathbb{R}[x]x^2$.

Here we have that $f \simeq g$ if and only if $f - g \in \mathbb{R}[x]x^2$ if and only if x^2 divides $f - g$. Any polynomial f can be written of the form

$$f = gx^2 + ax + b$$

for some unique $a, b \in \mathbb{R}$. Thus $[f] = [ax + b]$ for some unique $a, b \in \mathbb{R}$. We thus have

$$\mathbb{R}[x]/\mathbb{R}[x]x^2 = \{[ax + b] : \, a, b \in \mathbb{R}\}.$$

Here

$$[ax + b] + [cx + d] = [(a + c)x + (b + d)]$$

and

$$[ax + b] \cdot [cx + d] = [acx^2 + (ad + bc)x + bd] = [(ad + bc)x + bd].$$

Remark. Notice that informally the situation is as follows. In example (1) we add and multiply like we were adding and multiplying integers and then modify the result using the fact that $[3] = [0]$. In example (2) we similarly add and multiply like we were adding and multiplying polynomials and then modify the result using the fact that $[x^2] = [0]$.

C. The characteristic of a ring.

Definition. Let R be a ring. The *characteristic* of R, denoted $\text{char}(R)$, is a non-negative integer defined as follows. If there is a positive integer m such that $m1_R = 0_R$ then $\text{char}(R)$ is the smallest such positive integer. If there is on the other hand no such positive integer we say that $\text{char}(R) = 0$.

Examples. (1) $R = \{0\}$ is the only ring where $\text{char}(R) = 1$.
(2) For any positive integer n, we have that $\text{char}(\mathbb{Z}_n) = n$.
(3) We have that $\text{char}(\mathbb{Z}) = 0$.

Remarks. (1) Suppose $\text{char}(R) = 0$. Then

$$\mathbb{Z}1_r = \{\cdots, (-2)1_R, -1_r, 0_R, 1_R, 2_R, \cdots\}.$$

Notice also that the elements are distinct, that is $n1_R \neq m1_R$ if $n \neq m$. To see this, we argue by contradiction and suppose that $n1_R = m1_R$ where $n > m$. But then $(n-m)1_R = 0_R$ that contradicts the assumption that $\text{char}(R) = 0$. In fact the subring $\mathbb{Z}1_R$ is just like the ring of integers $\mathbb{Z}$. We have

$$\begin{aligned} n1_R + m1_R &= (n+m)1_R \\ n1_R \cdot m1_R &= nm1_R \\ -n1_R &= (-n)1_R. \end{aligned}$$

(2) Suppose $\text{char}(R) = n$. Let m be any integer and suppose $m = nr + s$ with $0 \leq s < n$. Then

$$m1_R = rn1_R + s1_R = s1_R.$$

If follows that

$$\mathbb{Z}1_R = \{0_R, 1_R, 21_R, \ldots, (n-1)1_R\}.$$

Notice that the elements that are written down are distinct. To see this we argue by contradiction and suppose that $r1_R = s1_R$ where $0 \leq s < r \leq n-1$. Then $(r-s)1_R = 0_r$ and $0 < r - s < n$. But this contradicts the fact that n (the characteristic of R) is the smallest positive integer such that $n1_R = 0_R$. In fact $\mathbb{Z}1_R$ is just like Z_n. The addition and the multiplication is the usual addition and multiplication with the modifcation that $n1_R = 0$. For example we have the following table when the characteristic is 3.

$+$	0_R	1_R	21_R
0_R	0_R	1_R	21_R
1_R	1_R	21_R	0_R
21_R	21_R	0_R	1_R

$\cdot$	0_R	1_R	21_R
0_R	0_R	0_R	0_R
1_R	0_R	1_R	21_R
21_R	0_R	21_R	1_R

which is just like the addition and multiplication tables for $\mathbb{Z}_3$

(3) Suppose that R has characteristic $n > 0$. For all $a \in R$, we have

$$n \cdot a = \underbrace{a + \cdots + a}_{n} = (\underbrace{1_R \cdot a + \cdots + 1_R \cdot a}_{n}) = (\underbrace{1_R + \cdots + 1_R}_{n}) \cdot a = 0_R \cdot a = 0_R.$$

IV. Homomorphisms and isomorphisms

In this section we will deal with ring homomorphims which are for rings what linear maps are for vector spaces.

Definition. Let R, S be rings. A map $\phi : R \to S$ is said to be a *ring homomorphism* if:

$$\begin{aligned} \phi(a+b) &= \phi(a) + \phi(b) \\ \phi(ab) &= \phi(a) \cdot \phi(b) \\ \phi(1_R) &= 1_S \end{aligned}$$

If ϕ is bijective, then we say that ϕ is a *ring isomorphism*. If there exists a ring isomorphisms from R to S then we say that R is *isomorphic* to S and write $R \cong S$.

Lemma 1.5 *Let $\phi : R \to S$ and $\psi : S \to T$ be ring homomorphisms. Then $\psi \circ \phi : R \to T$ is also a ring homomorphism. Furthermore if ϕ is an isomorphism then ϕ^{-1} is also a ring isomorphism.*

Proof (See sheet 3).

Remarks. (1) If there is an isomorphism from R to S, then there is no structural difference between the two rings. The ring S can be thought of as a copy of R.

(2) Let R be a ring. The map $\mathrm{id} : R \to R$ is then obviously a ring isomorphism.

(3) Using (2) and last lemma, one sees readily (sheet 3) that $R \cong R$, that $R \cong S$ iff $S \cong R$ and that if $R \cong S$ and $S \cong T$ then $R \cong T$.

(4) Warning. The map $\phi : R \to S,\ a \mapsto 0$ is normally NOT a ring homomorphism. This is only the case when $1_S = \phi(1_R) = 0_S$, that we know happens only when $S = \{0\}$.

Lemma 1.6 *If $\phi : R \to S$ is a ring homomorphism then*

a) $\phi(0_R) = 0_S$
b) $\phi(-a) = -\phi(a)$ for all $a \in R$.

Proof (a) We have $\phi(0_R) + 0_S = \phi(0_R) = \phi(0_R + 0_R) = \phi(0_R) + \phi(0_R)$. Cancellation by $-\phi(0_R)$ on both sides gives $\phi(0_R) = 0_S$.

(b) We have $\phi(a) + \phi(-a) = \phi(a + (-a)) = \phi(0_R) = 0_S$. Hence $\phi(-a)$ is the additive inverse of $\phi(a)$. That is $\phi(-a) = -\phi(a)$. $\square$.

Examples.(1) Consider the following map.

$$\phi : R \to R/I,\ a \mapsto [a] = a + I.$$

This is homomorphism since $\phi(a+b) = [a+b] = [a] + [b] = \phi(a) + \phi(b)$, $\phi(ab) = [ab] = [a] \cdot [b] = \phi(a) \cdot \phi(b)$ and $\phi(1) = [1]$ that is the multiplicative identity of R/I.

(2) Let S be a commutative ring with a subring R and let $z \in S$. For each polynomial $f = \sum_{k=0}^{\infty} a_k x^k$, we can associate the value

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

Consider the map

$$\phi : R[x] \to S,\ f \mapsto f(z).$$

We will show that this is a ring homomorphism. Clearly $\phi(1) = 1$ and it remains to see that ϕ preserves addition and multiplication. Let $f = \sum_{k=0}^{\infty} a_k x^k$ and $g = \sum_{k=0}^{\infty} b_k x^k$. Then

$$\begin{aligned}
\phi(f+g) &= \phi(\sum_{k=0}^{\infty}(a_k + b_k)x^k) \\
&= \sum_{k=0}^{\infty}(a_k + b_k)z^k \\
&= \sum_{k=0}^{\infty} a_k z^k + \sum_{k=0}^{\infty} b_k z^k \\
&= \phi(\sum_{k=0}^{\infty} a_k x^k) + \phi(\sum_{k=0}^{\infty} b_k x^k) \\
&= \phi(f) + \phi(g)
\end{aligned}$$

and for (c_k) with $c_k = \sum_{i+j=k} a_i b_j$, we have

$$\begin{aligned}
\phi(fg) &= \phi(\sum_{k=0}^{\infty} c_k x^k) \\
&= \sum_{k=0}^{\infty} c_k z^k \\
&= \sum_{k=0}^{\infty}(\sum_{i+j=k} a_i z^i b_j z^j) \\
&= (\sum_{i=0}^{\infty} a_i z^i) \cdot (\sum_{j=0}^{\infty} b_j z^j) \\
&= \phi(\sum_{i=0}^{\infty} a_i x^i) \cdot \phi(\sum_{j=0}^{\infty} b_j x^j) \\
&= \phi(f) \cdot \phi(g).
\end{aligned}$$

So ϕ is a homomorphism. In fact the addition and multiplication in $R[x]$ was defined such that this would be come a homomorphism.

(3) Let V be an n-dimensional vector space over a field K with a basis $(v_1, \ldots, v_n)$. Consider the linear map

$$\alpha : K^n \to V, \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \mapsto a_1 v_1 + \cdots + a_n v_n.$$

Recall from an earlier algebra unit that for this fixed basis $(v_1, \ldots, v_n)$, we can associate to each matrix $A \in M_n(K)$ a unique linear map $\phi_A \in \mathrm{End}\,(V)$ as follows

$$\phi_A(a_1 v_1 + \cdots + a_n v_n) = b_1 v_1 + \cdots + b_n v_n$$

if and only if

$$A \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Considering the matrix as the linear operator from K^n to itself as indicated above, we see that

$$\phi_A = \alpha A \alpha^{-1}$$

as

$$a_1 v_1 + \cdots + a_n v_n \overset{\alpha^{-1}}{\to} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \overset{A}{\to} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \overset{\alpha}{\to} b_1 v_1 + \cdots + b_n v_n.$$

We show that $M_n(K) \simeq \mathrm{End}(V)$ by showing that the map

$$\phi : M_n(K) \to \mathrm{End}\,(V),\ A \mapsto \phi_A$$

is a ring isomorphism. Firstly as I represents the identity map from K^n to itself, we have $\phi(I) = \phi_I = \alpha I \alpha^{-1} = \alpha\alpha^{-1} = \mathrm{id}$. So the multiplicative identity of $M_n(K)$ maps to the multiplicative identity of $\mathrm{End}\,(V)$ as required. Then

$$\phi(A + B) = \alpha(A + B)\alpha^{-1} = \alpha A \alpha^{-1} + \alpha B \alpha^{-1} = \phi(A) + \phi(B)$$

and

$$\phi(AB) = \alpha AB\alpha^{-1} = (\alpha A \alpha^{-1})(\alpha B \alpha^{-1}) = \phi(A)\phi(B).$$

Thus ϕ is a homomorphism and as the map is bijective (the inverse is given by $\beta \mapsto \alpha^{-1}\beta\alpha$) it is an isomorphism.

Definition. Let $\phi : R \to S$ be a ring homomorphism. The set

$$\ker \phi = \{a \in R :\ \phi(a) = 0\}$$

is called the *kernel* of ϕ and the set

$$\mathrm{im}\,\phi = \{\phi(a) : a \in R\}$$

is called the image of ϕ.

Lemma 1.7 *Let $\phi : R \to S$ be a ring homomorphism. We have that $\ker\phi$ is an ideal of R and that $\mathrm{im}\,\phi$ is a subring of S. Furthermore ϕ is injective if and only if $\ker\phi = \{0\}$.*

Proof Let us first see that $\ker\phi$ is an ideal of R. Firstly $\phi(0_R) = 0_S$ and thus $0_R \in \ker\phi$ that shows that $\ker\phi \neq \emptyset$. To see that all the closure requirements hold, let $a, b \in \ker\phi$ and $r \in R$. Then $\phi(a+b) = \phi(a)+\phi(b) = 0+0 = 0$, $\phi(ra) = \phi(r)\phi(a) = \phi(r)\cdot 0 = 0$ and $\phi(ar) = \phi(a)\phi(r) = 0\cdot\phi(r) = 0$. This shows that $a+b, ra, ar \in \ker\phi$. Hence $\ker\phi$ is an ideal of R. We next turn to $\mathrm{Im}\,\phi$. Firstly $1_S = \phi(1_R)$ shows that $1_S \in \mathrm{im}\,\phi$. The closure requirements follow from $\phi(a)+\phi(b) = \phi(a+b)$, $\phi(a)\phi(b) = \phi(ab)$ and $-\phi(a) = \phi(-a)$. Hence $\mathrm{im}\,\phi$ is a subring of S.

Now if ϕ is injective then in particular we have that $0 \in R$ is the only element that maps to 0_S. Hence $\ker\phi = \{0\}$. Conversely suppose that $\ker\phi = \{0\}$. If $\phi(a) = \phi(b)$ then $\phi(a+(-b)) = \phi(a)+\phi(-b) = \phi(a)+(-\phi(b)) = 0$. As $\ker\phi = \{0\}$ it follows that $a+(-b) = 0$ and thus $a = b$. □

Theorem 1.8 *(The fundamental Isomorphism Theorem). Let $\phi : R \to S$ be a homomorphism. Then*

$$R/\ker\phi \simeq \mathrm{im}\,\phi.$$

Proof Consider the map $\Psi : R/\ker\phi \to \mathrm{im}\,\phi$, $[a] \mapsto \phi(a)$.

The map Ψ is well defined and injective. We have

$$[a] = [b] \Leftrightarrow a-b \in \ker\phi \Leftrightarrow 0 = \phi(a-b) = \phi(a)-\phi(b) \Leftrightarrow \phi(a) = \phi(b).$$

The map Ψ is surjective. This is clear as for any $\phi(a) \in \mathrm{im}\,\phi$, we have $\phi(a) = \Psi([a])$.

Ψ is a homomorphism. Firstly $\Psi([1]) = \phi(1) = 1_S$ and then

$$\Psi([a]+[b]) = \Psi([a+b]) = \phi(a+b) = \phi(a)+\phi(b) = \Psi([a])+\Psi([b])$$

and

$$\Psi([a]\cdot[b]) = \Psi([ab]) = \phi(ab) = \phi(a)\cdot\phi(b) = \Psi([a])\cdot\Psi([b]).$$

This finishes the proof. □

Theorem 1.9 *Let R be a ring. If the characteristic of R is 0 then the subring $\mathbb{Z}1_R$ is isomorphic to $\mathbb{Z}$ and if the characteristic is the positive integer n then $\mathbb{Z}1_R$ is isomorphic to $\mathbb{Z}/\mathbb{Z}n$.*

Proof. Consider the map $\phi : \mathbb{Z} \to R$ that maps n to $n1_R$. Let us first see that this is a ring homomorphism. Firstly $\phi(1) = 1_R$. Then

$$\phi(n+m) = (n+m)1_R = n1_R + m1_R = \phi(n)+\phi(m)$$

and

$$\phi(nm) = nm1_r = n1_R \cdot m1_R = \phi(n)\cdot\phi(m).$$

Thus ϕ is a homomorphism.

Now suppose that $\operatorname{char} R = 0$. Then $\phi(n) = n1_R$ is 0_R if and only if $n = 0$. Hence $\ker \phi = \{0\}$ and by Lemma 1.7 it follows that ϕ is injective. Thus the map

$$\mathbb{Z} \to \mathbb{Z}1_R, n \mapsto n1_R$$

is an isomorphism.

Then suppose that $\operatorname{char} K = n$ for some positive integer n. Then $\phi(m) = m1_R = 0$ if and only if $n|m$. Hence $\ker \phi = \mathbb{Z}n$ and as $\operatorname{im} \phi = \mathbb{Z}1_R$, we have by the Fundamental Isomorphism Theorem that

$$\mathbb{Z}/\mathbb{Z}n \cong \mathbb{Z}1_R.$$

This finishes the proof. □

We end this section by looking at finite rings and thinking about the problem of classifying them. The following lemma is very useful here. For a given ring R we will denote by $|R|$, the number of elements in R.

Lemma 1.10 *Let R be a finite ring then $\operatorname{char}(R)$ divides $|R|$.*

Proof Notice first that if $\operatorname{char}(R) = n$ then $\mathbb{Z}1_R \cong \mathbb{Z}_n$. Hence $|\mathbb{Z}1_R| = n = \operatorname{char}(R)$.

As $\mathbb{Z}1_R$ is a subgoup of R with respect to addition, it follows immediately from Lagrange's Theorem in group theory that $\operatorname{char}(R) = |\mathbb{Z}1_R|$ dividies $|R|$. An alternative way of seeing this is to consider the map

$$R \to R,\ r \mapsto r + 1_R.$$

This is a bijection (with inverse $r \mapsto r + (-1_R)$). Thus $R = \{r + 1_R : r \in R\}$. Adding all the elments in R gives

$$\sum_{r \in R} r = \sum_{r \in R} (r + 1_R) = (\sum_{r \in R} r) + m1_R$$

where $m = |R|$. Cancelling on both sides by $\sum_{r \in R} r$, gives $m1_R = 0$ and thus $n = \operatorname{char}(R)$ divides m. □

Theorem 1.11 *Let R be a finite ring such that $|R| = p$ is a prime number. Then $R \cong \mathbb{Z}_p$.*

Proof Consider the subring $\mathbb{Z}1_R$. As $|R| \geq 2$, we have that $1_R \neq 0_R$ and thus $\mathbb{Z}1_R$ has at least two elements, $0_R, 1_R$. Thus $|\mathbb{Z}1_R| = \operatorname{char}(R)$ is at least 2 and divides $|R|$ by Lemma 1.10. As $|R|$ is a prime number it follows that $|\mathbb{Z}1_R| = p$ and thus $R = \mathbb{Z}1_R$. By Theorem 1.9 we then have that $R \cong \mathbb{Z}_p$. □

2 Factorization in rings

I. Integral domains and principal ideal domains

A Integral domains

Definition. An *integral domain* (ID) is a commutative ring $R \neq \{0\}$ that satisfies

$$ab = 0 \Rightarrow a = 0 \text{ or } b = 0.$$

Examples. (1) Any field K is an integral domain. If $ab = 0$ and $a \neq 0$ then $b = a^{-1}ab = a^{-1} \cdot 0 = 0$.

(2) We have that $\mathbb{Z}$ is an integral domain that is not a field.

(3) Consider the ring $\mathbb{Z}_4 = \{[0], [1], [2], [3]\}$. Then

$$[2] \cdot [2] = [4] = [0]$$

but $[2] \neq [0]$. Hence $\mathbb{Z}_4$ is not an ID.

Lemma 2.1 *Let $R \neq \{0\}$ be a commutative ring. The following are equivalent.*

1) R is an ID
2) R satisfies the cancellation property: if $ab = ac$ and $a \neq 0$ then $b = c$.

Proof (1) $\Rightarrow$ (2). Suppose $ab = ac$ and $a \neq 0$. Then

$$0 = ab + (-ac) = ab + a(-c) = a(b + (-c))$$

As R is an ID, it follows that $b + (-c) = 0$, that is $b = c$.

(2) $\Rightarrow$ (1). Suppose that R satisfies the cancellation property. If $ab = 0$ and $a \neq 0$ then

$$a \cdot b = a \cdot 0 \;\Rightarrow\; b = 0.$$

This finishes the proof. □

Theorem 2.2 *The characteristic of an ID is always a prime number or* 0.

Proof Let R be an ID. Notice first that as $R \neq \{0\}$, we have that $\text{char}\,(R) \neq 1$. If the characteristic is neither 0 nor a prime it would then have to be a composite. We argue by contradiction and suppose that the characteristic is a composite $n = r \cdot s,\ 1 < r, s < n$. Then

$$0 = n \cdot 1_R = rs \cdot 1_R = (r \cdot 1_R) \cdot (s \cdot 1_R)$$

i.e.

$$0 = \underbrace{1_R + \cdots + 1_R}_{rs} = (\underbrace{1_R + \cdots + 1_R}_{r}) \cdot (\underbrace{1_R + \cdots + 1_R}_{s}).$$

However $r \cdot 1_R, s \cdot 1_R \neq 0$ which contradicts our assumption that R is an ID. □

Theorem 2.3 *Every finite integral domain is a field.*

Proof (See sheet 4)

B. Closure properties for ideals

Let R be any ring and let I and J be ideals of R. The subset

$$I + J = \{a + b : a \in I, b \in J\}$$

is called the sum of the ideals I and J. Let IJ be the set consisting of all finite sums

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n$$

where n is allowed to vary and can be any integer $n \geq 1$ and $a_1, \ldots, a_n \in I$, $b_1, \ldots, b_n \in J$. This subset is called the product of the ideals I and J.

Remark. Warning. We do NOT have in general that $IJ = \{ab : a \in I \text{ an } b \in J\}$. The reason why we have chosen to define the product differently is that the set $\{ab : a \in I \text{ and } b \in J\}$ is not in general closed under addition. The smallest ideal containing all ab, $a \in I$ and $b \in J$ is our product of ideals.

Lemma 2.4 *Let R be a ring with ideals I, J. Then $I \cap J, I + J$ and IJ are also ideals of R. Furthermore $IJ \subseteq I \cap J \subseteq I + J$.*

Proof $\underline{I \cap J \text{ is an ideal of } R}$. As $0 \in I \cap J$, we have that $I \cap J \neq \emptyset$. Let $a, b \in I \cap J$ and let $r \in R$. As I, J are ideals of R, it follows that $a + b, ra, ar \in I$ and also that these elements are in J. Thus $a + b, ra, ar \in I \cap J$ and $I \cap J$ is an ideal of R.

$\underline{I + J \text{ is an ideal of } R}$. Firstly $0 = 0 + 0 \in I + J$ and thus $I + J \neq \emptyset$. Let $a_1, a_2 \in I$, $b_1, b_2 \in J$ and $r \in R$. As I, J are ideals, we have that $a_1 + a_2, ra_1, a_1r \in I$ and $b_1 + b_2, rb_1, b_1r \in J$. Thus

$$(a_1 + b_1) + (a_2 + b_2) = (a_1 + a_2) + (b_1 + b_2) \in I + J,$$

$r(a_1 + b_1) = ra_1 + rb_1 \in I + J$ and $(a_1 + b_1)r = a_1r + b_1r \in I + J$. This shows that $I + J$ is an ideal of R.

$\underline{IJ \text{ is an ideal of } R}$. As $0 = 0 \cdot 0 \in IJ$, we have that $IJ \neq \emptyset$. Let $a_1, \ldots, a_n, c_1, \ldots, c_m \in I$, $b_1, \ldots, b_n, d_1, \ldots, d_m \in J$ and $r \in R$. As I and J are ideals, we have that $ra_i, a_ir \in I$ and $rb_i, b_ir \in J$ for $i = 1, \ldots, n$. It follows that

$$(a_1b_1 + \cdots + a_nb_n) + (c_1d_1 + \cdots + c_md_m) \in IJ,$$

$$r(a_1b_1 + \cdots + a_nb_n) = (ra_1)b_1 + \cdots + (ra_n)b_n \in IJ$$

and

$$(a_1b_1 + \cdots + a_nb_n)r = a_1(b_1r) + \cdots + a_n(b_nr) \in IJ.$$

Hence IJ is also an ideal.

For the inclusions, notice first that $I \cap J \subseteq I + J$ as any element $a \in I \cap J$ can be written as $a = a + 0 \in I + J$. Then suppose $a_1, \ldots, a_n \in I$ and $b_1, \ldots, b_n \in J$. As both I and J are ideals we have that $a_i b_i \in I$ and $a_i b_i \in J$. Thus $a_1 b_1, \ldots, a_n b_n \in I \cap J$ and as $I \cap J$ is an ideal we have that the sum $a_1 b_1 + \cdots + a_n b_n$ is also in $I \cap J$. This shows that $IJ \subseteq I \cap J$. □

C. Principal ideal domains

Definition. Let R be a commutative ring. An ideal I of R is said to be a *principal ideal* if $I = Ra$ for some $a \in R$.

Remark. Recall that we have seen earlier that any subset of the form Ra is an ideal of R.

Definition. An integral domain is said to be a *principal ideal domain* (PID), if all the ideals of R are principal ideals.

Theorem 2.5 *The ring $\mathbb{Z}$ is a principal ideal domain. Also $K[x]$ is a principal ideal domain for any field K.*

Proof We start with the ring $\mathbb{Z}$. Let I be an ideal of $\mathbb{Z}$. If $I = \{0\}$ then $I = \mathbb{Z} \cdot 0$ is a principal ideal. So we can suppose that $I \neq \{0\}$. If $a \in I$ then also $-a \in I$. Hence I contains positive integers. Let n be the smallest positive integer in I. As I is closed under multiplication from $\mathbb{Z}$, we know that $\mathbb{Z}n \subseteq I$. We show that $I \subseteq \mathbb{Z}n$. Let $m \in I$. Division with n by remainder gives

$$m = nr + s$$

with $0 \leq s < n$. As $m, n \in I$ we have that $s = m + (-r)n \in I$. As n was the smallest positive integer of I we can't have that $s > 0$. Hence $s = 0$ and $m = rn \in \mathbb{Z}n$. Hence $I = \mathbb{Z}n$ is a principal ideal.

The proof that $K[x]$ is a principal ideal domain is similar. Let I be an ideal of $K[x]$. If $I = \{0\}$ then $I = K[x] \cdot 0$ is a principal ideal. So we can suppose that $I \neq 0$. Let f be a non-zero polynomial in I of smallest possible degree. As I is closed under multiplication from $K[x]$, we have that $K[x]f \subseteq I$. It remains to show that $I \subseteq K[x]f$. Let g be any polynomial in $K[x]$. From a previous unit you know that division by f with remainder gives

$$g = fh + k$$

where the degree of k is less than the degree of f. As $f, g \in I$ we have that $k = g + (-h)f \in I$ and as f was a non-zero polynomial in I of smallest degree, we can't have that k is non-zero. Hence $k = 0$ and $g = hf \in K[x]f$. This shows that $I = K[x]f$ is a principal ideal. □

Lemma 2.6 *Let $R \neq \{0\}$ be a commutative ring. Then R is a field if an only if the only ideals of R are $\{0\}$ and R.*

Proof Suppose first that R is a field. Let I be an ideal such that $I \neq \{0\}$. Let $0 \neq a \in I$ and let $b \in R$. Then $b = (ba^{-1})a \in I$ and thus $R \subseteq I$ that implies that $R = I$.

Conversely suppose that $R \neq \{0\}$ is a commutative ring such that $\{0\}$ and R are the only ideals. Let $0 \neq a \in R$. Then the ideal Ra contains $a = 1a$ and thus $Ra \neq \{0\}$. By

assumption $Ra = R$. In particular $1 = ba$ for some $b \in R$ and thus a has a multiplicative inverse. This shows that R is a field. $\square$

Corollary 2.7 *Every field is a principal ideal domain.*

Proof Let R be a field. We have already seen that R is an integral domain. By Lemma 2.6, the only ideals of R are $\{0\} = R \cdot 0$ and $R = R \cdot 1$. Hence all ideals are principal ideals. $\square$.

II. Factorization in integral domains

Let R be an ID. We write

$$a|b$$

and say that a *divides* b (or that b is divisible by a) if there exists $c \in R$ such that $b = ac$.

If $a|b$ and $b|a$ then we say that a and b are *associated* and write $a \sim b$.

Remarks. 1) Notice that $a|a$, $1|a$ and $a|0$ for all $a \in R$.

2) If $a|b$ and $b|c$ then $a|c$. To see this notice that if $b = ar$ and $c = bs$ then $c = ars$.

3) We have that $\sim$ is an equivalence relation on R. Clearly $a \sim a$ and we also have that $b \sim a$ if and only if $a \sim b$. Let us see why $\sim$ is transitive. Suppose that $a \sim b$ and $b \sim c$. Then as $a|b$ and $b|c$, part 1) tells us that $a|c$. Also as $c|b$ and $b|a$ we see from part 1) again that $c|a$. Thus $a \sim c$.

4) We have that $a \in R$ is a unit if and only if $a|1$ if and only if a divides all $b \in R$. (Notice that $a|1$ and $1|b$ implies that $a|b$). As 1 always divides a, notice also that $a|1$ if and only if $a \sim 1$.

Lemma 2.8 *Let R be an integral domain and let R^* be the group of units. We have that b and a are associated if and only if $b \in aR^*$. In other words the equivalence class of a with respect to $\sim$ is aR^*.*

Proof Notice first that $0|a$ if and only if $a = 0$ and thus the equivalence class containing 0 is $\{0\} = 0 \cdot R^*$. Now suppose that a is non-zero and that $b \sim a$. Then $b = ar$ and $a = bs$ for some $r, s \in R$. Thus

$$a \cdot 1 = a = bs = ars$$

and as $a \neq 0$, cancellation (Lemma 2.1) gives $rs = 1$ and thus r is a unit and $b = ar \in aR^*$. Conversely if $b = ar \in aR^*$ where $s \in R$ is the multiplicative inverse of r. Then $b = ar$ and $a = ars = bs$ gives that $b \sim a$. Hence the equivalence class containing a is aR^*. $\square$

Examples (1) We have that $\mathbb{Z}^* = \{1, -1\}$ and the equivalence class of n is thus $n\mathbb{Z}^* = \{n, -n\}$. So each equivalence class has two elements apart from $o\mathbb{Z}^* = \{0\}$.

(2) Let K be a field. Notice that if $fg = 1$ for some polynomials $f, g \in K[x]$ then f and g must be non-zero constants. Thus $K[x]^* = K^* = K \setminus \{0\}$. So $fK[x]^* = f(K \setminus \{0\})$.

Notice that for each non-zero polynomial f there is exactly one monic polynomial (that is a polynomial where the nonzero coeffient of the highest term is 1) in $fK[x]^*$.

Lemma 2.9 *We have*

$$\begin{aligned} a|b &\Leftrightarrow Rb \subseteq Ra \\ a \sim b &\Leftrightarrow Rb = Ra \\ a \in R^* &\Leftrightarrow Ra = R. \end{aligned}$$

Proof To see the first property note that if $b = ar$ the $Rb = Rra \subseteq Ra$ and conversely if $Rb \subseteq Ra$ then $b \in Ra$ and thus $a|b$. The second property follows from this observation since if $a \sim b$ then both $a|b$ and $b|a$ which implies that $Rb \subseteq Ra$ and $Ra \subseteq Rb$. Finally a is a unit if and only if $a|1$. As we always have $1|a$ we thus have that a is a unit if and only if $a \sim 1$. By the second property this happens if and only if $Ra = R1 = R$. □

Definition

(1) An element $p \in R$ is said to be *irreducible* if $p \neq 0$, p is not a unit and

$$\text{if } p = ab \text{ then either } a \text{ or } b \text{ is a unit.}$$

(2) An element $p \in R$ is a *prime* if $p \neq 0$, p is not a unit and

$$\text{if } p|ab \text{ then } p|a \text{ or } p|b.$$

Remark. Suppose that p is irreducble and $p = ab$. If b is a unit then $a = Pb^{-1}$. A different way of saying that p is irreducible is therfore to say that

$$a|p \Rightarrow a \sim 1 \text{ or } a \sim p.$$

Proposition 2.10 *Every prime is irreducible.*

Proof Suppose p is a prime. If $p = ab$ then

$$p|ab \;\Rightarrow\; p|a \text{ or } p|b$$

Without loss of generality we can suppose that $p|a$, say $a = pc$. It follows that

$$p \cdot 1 = p = ab = pcb$$

and cancellation (Lemma 2.1) gives $cb = 1$. Hence b must be a unit. This finishes the proof. □

Remark. The converse is not true in general. See sheet 5.

Now take any integral domain R. The equivalence relation $\sim$ partitions the irreducibles in to equivalence classes. From each equivalance class pick one element and denote the

set of these by $\mathcal{P}$. We would like to have the following two properties

(1) Every non-zero element $a \in R$ has a *factorization* into a product of irreducibles

$$a = up_1 \cdots p_n$$

with $u \in R^*$ and $p_1, \ldots, p_r \in \mathcal{P}$.

(2) The factorization is unique, in the sense that u is unique and that $p_1, \ldots, p_n$ are unique up to order.

Remark. The convention is to think of every unit $a = u$ as being factorizable with factorization $a = u$ (so the number of irreducibles is $n = 0$).

Examples (1) For $R = \mathbb{Z}$, we can choose $\mathcal{P}$ to be the set of all the (positive) prime numbers. Then every integer has a factorization. For example

$$35 = 5 \cdot 7, \quad -28 = (-1) \cdot 2 \cdot 2 \cdot 7.$$

(2) Let $R = \mathbb{C}[x]$. By the Fundamental Theorem of Algebra, every non-zero polynomial can be written as a product of polynomials of degree 1. So here we can take

$$\mathcal{P} = \{x - a : \, a \in \mathbb{C}\}.$$

For example

$$4x^2 + 4 = 4(x + i)(x - i).$$

Remark. In both these examples we have that every non-zero element can be factorised into a product of irreducibles and we will see later that the factorization is unique. There are however integral domains where the factorization (when it exists) is not always unique (sheet 5). We will also see shortly that there are even integral domains with non-zero elements that can not be factorized into a product of irreducibles.

Example (Non-examinable). We give now an example of an ID that is not a field but with no irreducibles. So no non-zero element that is not a unit can be factorised into a product of irreducibles.

Let $x_0, x_1, \ldots$ be variables and for each of these we form the polynomial ring $\mathbb{C}[x_i]$. The map

$$\mathbb{C}[x_i] \to \mathbb{C}[x_{i+1}], \; f \mapsto f(x_{i+1}^2)$$

is an injective ring homomorphism. So we can identify $\mathbb{C}[x_i]$ with the image. In other words we can let $x_i = x_{i+1}^2$. So we have

$$x_0 = x_1^2 = x_2^4 = x_3^8, \ldots$$

Notice that $\mathbb{C}[x_0] \subseteq \mathbb{C}[x_1] \subseteq \mathbb{C}[x_2] \subseteq \ldots$ Let

$$R = \bigcup_{k=0}^{\infty} \mathbb{C}[x_i].$$

Let us firstly see that R is a ring. Let $f, g, h \in R$. Then there exists large enough ingteger i such that $f, g, h \in \mathbb{C}[x_i]$. Now as $\mathbb{C}[x_i]$ is a ring we have that $(f + g) + h =$

$f+(g+h), f+h=h+f, (fg)h=f(gh), f(g+h)=fg+fh$. Clearly the polynomial 1 is a multiplicative identity, 0 a additive identity and $-f$ is the additive inverse of f. The ring R is also an integral domain. To see this suppose $f, g \in R$ with $fg=0$. Pick i large enough such that $f, g \in \mathbb{C}[x_i]$. As $\mathbb{C}[x_i]$ is an integral domain we get that one of f, g must be zero. Let us then determine the units of R. Let $f \in R$ be a unit. Then $fg=1$ for some $g \in R$. Pick i large enough such that $f, g \in \mathbb{C}[x_i]$. Then f is a unit in $\mathbb{C}[x_i]$ but $\mathbb{C}[x_i]^* = \mathbb{C}^*$. This shows that $R^* = \mathbb{C} \setminus \{0\}$.

Let us see that there are no irreducibles in R. We argue by contradiction and suppose that f is an irreducible in R. Then $f \in \mathbb{C}[x_i]$ for some i and must then in particular be irreducible in $\mathbb{C}[x_i]$. By the Fundamental Theorem of Algebra we must have that f is of degree 1 and without loss of generality we can suppose that it is monic. So $f = x_i - a$ for some $a \in \mathbb{C}$. But then

$$f = x_i - a = x_{i+1}^2 - a$$

that is not irreducible in $\mathbb{C}[x_{i+1}]$ and thus not irreducible in R. By this contradiction we have seen that R has no irreducucible elements.

III. Unique factorization in principal ideal domains

Definition Let $a, b \in R$.

1) An element $d \in R$ is called a *highest common factor* (hcf) of a and b if:

(i) $d|a$ and $d|b$
(ii) if $e|a$ and $e|b$ then $e|d$.

2) An element $c \in R$ is called *a least common multiple* (lcm) of a and b if:

(i) $a|c$ and $b|c$
(ii) if $a|e$ and $b|e$ then $c|e$.

Remark 1) If d_1, d_2 are both hcf's of a and b then the second property implies that $d_1|d_2$ and $d_2|d_1$. Hence $d_1 \sim d_2$ and $d_2 = u \cdot d_1$ for some unit $u \in R$. (Notice that we are assuming throughout this chapter that R is an ID).

2) If c_1, c_2 are lcm's of a and b then similarly $c_1 \sim c_2$ and $c_2 = u \cdot c_1$ for some unit $u \in R$.

Lemma 2.11 *Let R be a PID and $a, b \in R$.*

a) $Ra + Rb = Rd$ where d is a hcf of a and b.
b) $Ra \cap Rb = Rc$ where c is a lcm of a and b.

Proof a) As R is a PID we know that $Ra + Rb = Rd$ for some $d \in R$. It remains to see that d is a hcf of a and b. As both $a, b \in Ra + Rb = Rd$ it is clear that d is a factor of both a and b. Now suppose that e is another factor, i.e. $e|a$ and $e|b$. Then both a and b are multiples of e and since $d \in Ra + Rb$ it follows that d is also a multiple of e. Hence $e|d$ and we have shown that d is the hcf of a and b.

b) Again as R is a PID, we know that $Ra \cap Rb = Rc$ for some $c \in R$. We want to show that c is a lcm of a and b. Firstly since $c \in Ra \cap Rb$ we have that c is both a multiple of a and b. Now suppose that e is another element that is a common multiple of a and b. Then $e \in Ra \cap Rb = Rc$ and is therefore a multiple of c. Hence $c|e$ and we have shown that c is a lcm of a and b. $\square$

Definition. Let R be a PID. Let $a, b \in R$. We say that a and b are *coprime* if 1 is a highest common factor of a and b (or equivalently if any highest common factor of a and b is a unit).

Remark. Let R be a PID and suppose c is a highest common factor of a and b. By last lemma, we know that

$$c = ra + sb$$

for some $r, s \in R$. In particular, if a and b are coprime then

$$1 = ra + sb$$

for some $r, s \in R$.

Proposition 2.12 *Let R be a principal ideal domain. Every irreducible $p \in R$ is a prime.*

Proof Suppose that $p|ab$ but $p \not| a$. We want to show that $p|b$. Let c be a common divisor of a and p. As p is irreducible either $c = pu$ or $c = u$ for some unit u. However pu does not divide a so we must have that c is a unit. This shows that a, p are coprime and thus

$$1 = ra + sp$$

for some $r, s \in R$. It follows that

$$b = 1 \cdot b = (ra + sp) \cdot b = rab + psb$$

and as ab is divisible by p it follows that b is divisible by p. This finishes the proof. $\square$

Our aim is to prove a unique factorization theorem for PID's. First let us see that any non-zero element in a PID can be written as a product of irreducibles.

Theorem 2.13 *Let R be a PID and $0 \neq a \in R$. Then a can be expressed as a product of irreducibles in R.*

Proof We argue by contradiction and suppose that some $0 \neq a$ cannot be written as a product of irreducibles. In partcular a is not a unit and not irreducible and thus

$$a = a_1 b_1$$

for some $a_1, b_1 \in R$ where neither a_1 nor b_1 is a unit. If both a_1 and b_1 can be expressed as products of irreducibles then so can a. Hence one of these can not be expressed as a product of irreducibles. Without loss of generality we can suppose that this is a_1. Notice that as b_1 is not a unit we don't have that a divides a_1 (ottherwise $a \sim a_1$ that forces b_1 to be a unit) and thus $Ra \subset Ra_1$. The same argument shows there exists $a_2 \in R$ such that

$Ra_1 \subset Ra_2$ and where a_2 cannot be expressed as a product of irreducibles. Continuing in this manner we get a strictly increasing chain of ideals

$$Ra \subset Ra_1 \subset Ra_2 \subset \cdots$$

Let

$$I = Ra \cup Ra_1 \cup Ra_2 \cup \cdots$$

We next show that I is an ideal. Firstly $0 \in Ra \subseteq I$ and thus $I \neq \emptyset$. Now suppose that $a, b \in I$ and that $r \in R$. For some i we have that both $a, b \in Ra_i$. But then $a + b$ and ra are in $Ra_i \subseteq I$. Hence I is an ideal. Since R is a principal ideal domain we have that $I = Rb$ for some $b \in R$. Then $b = 1 \cdot b \in I$ and thus $b \in Ra_i$ for some i. It follows that

$$Ra_{i+1} \subseteq I = Rb \subseteq Ra_i$$

and we get the contradiction that $Ra_i = Ra_{i+1}$. $\square$

We have now got the tools to prove the main result of this chapter. Let R be a PID and consider the set of all primes/irreducibles of R. As we have seen, these partition into equivalence classes with respect to the equivalence relation $a \sim b$ iff a and b are associated, i.e. $a|b$ and $b|a$. From each equivalence class we pick one representive. We refer to these primes as being *prime representatives*.

Theorem 2.14 *(Unique factorization theorem for PID's)*

Let R be a PID and take any set of prime representatives $\mathcal{P}$. Any $0 \neq a$ can be written in the form

$$a = up_1 \cdots p_r$$

where u is a unit and where $p_1, \ldots, p_r \in \mathcal{P}$. Moreover the unit u is unique and the primes $p_1, \ldots, p_r$ are unique up to order.

Proof (Existence). By Theorem 2.13 we have that a can be written as a product of primes. If a is a unit then we are done. Now suppose that a is a a product of primes of length $r \geq 1$, say

$$a = q_1 \cdots q_r.$$

Then $q_i = u_i p_i$ for some $p_1, \ldots, p_r \in \mathcal{P}$ and units $u_1, \ldots, u_r$. So

$$a = up_1 \cdots p_r$$

where $u = u_1 \cdots u_r$.

(Uniqueness). Suppose that we also have

$$a = vq_1 \cdots q_s$$

for some unit v and $q_1, \ldots, q_s \in \mathcal{P}$. We show by induction on r that $s = r$ and that the primes $q_1, \ldots, q_s$ the same as $p_1, \ldots, p_r$ up to order. If $r = 0$ then $a = u$ is a unit and hence s must be zero (otherwise $q_s|a$ and $a|1$ implies that $q_s|1$ and q_s is a unit). Then also $u = a = v$. This gives us the induction basis. Now for the induction step suppose that $r \geq 1$ and we know that the result holds when the value of r is smaller. As a is not

a unit we must have $s \geq 1$ (otherwise $p_r | v$ and p_r is a unit). As p_r is a prime that divides $a = vq_1 \cdots q_s$, it must divide q_i for some $i \in \{1, \ldots, s\}$. Without loss of generality we can suppose that p_r divides q_s. Since q_s is irreducible we must have $q_s = p_r w$ for some unit $w \in R$. Then q_s and p_r are associated and since $\mathcal{P}$ contains only one element from each equivalence class it follows that $q_s = p_r$. Using the fact that R is an ID and

$$up_1 \cdots p_r = vq_1 \cdots q_s,$$

we can cancel by $p_r = q_s$. This gives us

$$up_1 \cdots p_{r-1} = vq_1 \cdots q_{s-1}.$$

By the induction hypothesis we must have that $u = v$, $r - 1 = s - 1$ and that $p_1, \ldots, p_{r-1}$ are the same primes (up to order) as $q_1, \ldots, q_{s-1}$. This finishes the inductive proof. □

Remark (Not covered in lectures). The hcf and lcm have a natural description in principal ideal domains. Suppose that

$$\begin{aligned} a &= up_1^{r_1} \cdots p_n^{r_n} \\ b &= vp_1^{s_1} \cdots p_n^{s_n} \end{aligned}$$

where u, v are units and $p_1, \ldots, p_n$ are distinct elements in $\mathcal{P}$.

The hcf of a and b. Clearly

$$d = p_1^{\min(r_1, s_1)} \cdots p_n^{\min(r_n, s_n)}$$

divides both a and b. Now suppose we have any common factor e of a and b. By the Unique Factorization Theorem we have that p_i occurs as a factor of e at most $\min(r_i, s_i)$ times. Hence it follows that $e|d$ and d is thus the highest common factor of a and b.

The lcm of a and b. Clearly

$$c = p_1^{\max(r_1, s_1)} \cdots p_n^{\max(r_n, s_n)}$$

is divisble by both a and b. Now let e be any element in R that is divisble by both a and b. By the Unique Factorization Theorem we have that p_i coccurs as a factor of e at least $\max(r_i, s_i)$ times. Hence it follows that $c|e$ and c is a lowest common multiple of a and b.

Notice that $Rab = Rcd$ and thus $hcf(a, b) \cdot lcm(a, b) \sim ab$. In particular if a and b are coprime then ab is a least common multiple.

IV. Quotient rings of principal ideal domains

In this section we will be looking at quotient rings of PID's. Let R be a PID. Recall that $Ra = Rb$ if and only if $a \sim b$ (i.e. $a|b$ and $b|a$). So for each equivalence class with respect to $\sim$ there is a unique ideal I and thus a unique quotient ring R/I.

We now want to understand better the structure of the quotient rings.

Theorem 2.15 *Let R be a commutative ring and let $a \in R$. We have that $[b] \in R/Ra$ is a unit in R/Ra if and only if $Rb + Ra = R$.*

Proof ($\Rightarrow$) Let $[c]$ be the multiplicative inverse of $[b]$. Then $[1] = [c][b] = [cb]$ and thus $1 - cb \in Ra$, say $1 = cb + da$. Thus $1 \in Ra + Rb$ and as $Ra + Rb$ is and ideal we have $R = R \cdot 1 \subseteq Ra + Rb$.

($\Leftarrow$) Then $1 \in R = Ra + Rb$, say $1 = da + cb$. It follows that $[1] = [da + cb] = [cb] = [c] \cdot [b]$ and $[b]$ is a unit with inverse $[c]$. $\square$

Corollary 2.16 *Let R be a PID and $0 \neq a \in R$. Then R/Ra is a field if and only if a is irreducible (or equivalently a prime).*

Proof ($\Rightarrow$). Suppose that a is not irreducible. There are then two cases to consiser. We could have that a is a unit in which case $R/Ra = R/R$ that has only one element and is thus the zero ring $\{0\}$. By definition this is not a field.

The other possibility is that a is a composite, say $a = bc$ where neither b nor c is a unit. Notice that in this case a does not divide b (otherwise $a \sim b$ and c is a unit) and a does not divide c. So $[b], [c] \neq [0]$ whereas $[b] \cdot [c] = [a] = [0]$. Thus R/Ra is not an ID and therefore not a field in particular. $\square$

($\Leftarrow$). Suppose that a is irreducible and let $[0] \neq [b] \in R/Ra$. We want to show that $[b]$ has a multiplicative inverse. By Theorem 2.15 it suffices to show that a and b are coprime. Let e be a common divisor a and b. As a is irreducible we must that $e \sim a$ or $e \sim 1$. But a does not divide b and thus we are only left with the possibility that $e \sim 1$. This shows that the only common divisors of a and b are the units and thus a and b are coprime.

Definition. Let R and S be any rings. The *direct product* of R and S is the ring that consists of the set $R \times S = \{(r, s) : r \in R,\ s \in S\}$ and where the addition and multiplication are given by

$$\begin{aligned} (a, b) + (c, d) &= (a + c, b + d) \\ (a, b) \cdot (c, d) &= (ac, bd). \end{aligned}$$

Clearly $(0_R, 0_S)$ is the additive inverse and $(1_R, 1_S)$ is the multiplicative identity. The additive inverse of (a, b) is $(-a, -b)$. All the algebraic laws hold in $R \times S$ since they hold for both the coordinates. Thus we have a ring.

Remark. Notice that $(a, b) \in R \times S$ is unit if and only if a is a unit in R and b is a unit in S. Thus $(R \times S)^* = R^* \times S^*$.

Proposition 2.17 *(The chinese remainder theorem) Let R be a PID and suppose that $a, b \in R$ are coprime. Then*

$$R/Rab \cong R/Ra \times R/Rb.$$

Proof For $x, c \in R$ and we let $[x]_c$ be the congruence class of x in R/Rc. Define a map

$$\phi : R/Rab \to R/Ra \times R/Rb,\ [x]_{ab} \mapsto ([x]_a, [x]_b).$$

This map is well defined since if $[x]_{ab} = [y]_{ab}$ then ab divides $x-y$ and thus both a, b divide $x-y$ that gives that $[x]_a = [y]_a$ and $[x]_b = [y]_b$. To see that the map is a homomorphism notice that $\phi([1]_{ab}) = ([1]_a, [1]_b)$, that

$$\begin{aligned}
\phi([x]_{ab} \cdot [y]_{ab}) &= \phi([xy]_{ab}) \\
&= ([xy]_a, [xy]_b) \\
&= ([x]_a \cdot [y]_a, [x]_b \cdot [y]_b) \\
&= ([x]_a, [x]_b) \times ([y]_a, [y]_b) \\
&= \phi([x]_{ab}) \cdot \phi([y]_{ab}),
\end{aligned}$$

and that

$$\begin{aligned}
\phi([x]_{ab} + [y]_{ab}) &= \phi([x+y]_{ab}) \\
&= ([x+y]_a, [x+y]_b) \\
&= ([x]_a + [y]_a, [x]_b + [y]_b) \\
&= ([x]_a, [x]_b) + ([y]_a, [y]_b) \\
&= \phi([x]_{ab}) + \phi([y]_{ab}).
\end{aligned}$$

Next we see that the map is injective. Suppose that $\phi([x]_{ab}) = ([0]_a, [0]_b)$. This is the same as saying that $[x]_a = [0]_a$ and $[y]_b = [0]_b$. But then $a|x-0 = x$ and $b|x$ and as a and b are coprime it follows (from the unique factorization theorem) that ab divides x. That is $[x]_{ab} = [0]_{ab}$.

It remains to see that ϕ is surjective. Let $([x]_a, [y]_b)$ be any element in $R/Ra \times R/Rb$. As a and b are coprime we have $1 = ra + sb$ for some $r, s \in R$. Consider $z = sbx + ray$. Then

$$[z]_a = [sbx]_a = [(sb + ra)x]_a = [x]_a$$

and

$$[z]_b = [ray]_b = [(sb + ra)y]_b = [y]_b.$$

Hence $\phi([z]_{ab}) = ([z]_a, [z]_b) = ([x]_a, [y]_b)$. $\square$.

The two (for us) most important PID's are the rings $\mathbb{Z}$ and $K[x]$ where K is a field. We will now look at these.

A. Quotient rings of $\mathbb{Z}$ and the Euler function

The main aim of this section will be to show that if K is a finite field then the group of units K^* is a cyclic group. In order to prove this we will study finite cyclic groups and obtain a criterion for a given group to be cyclic. First we look at the Euler function which will play a crucial role here.

Definition. The function $\phi : \mathbb{Z}_+ \to \mathbb{Z}_+$ where $\phi(n) = |\mathbb{Z}_n^*|$, is called the *Euler function.*

Remark. From Theorem 2.15 we know that $[r]$ is a unit in $\mathbb{Z}_n$ if and only r, n are coprime. This means that $\phi(n)$ is the number of elements in $\{0, 1, \ldots, n-1\}$ that are coprime to n.

Example. For example, among $0, 1, 2, 3, 4, 5$, only 1 and 5 are coprime to 6. Hence $\phi(6) = 2$.

If K is a finite field, then K^* is a finite abelian group. We will thus assume that all groups we are working with are finite and abelian.

Definition. Let G be a finite group and $a \in G$. The *cyclic subgroup* generated by a is $\langle a \rangle = \{a^n : n \in \mathbb{Z}\}$.

Remark. We have that $1 = a^0 \in \langle a \rangle$. We also have that $\langle a \rangle$ is closed under the group multiplication and taking inverses since $a^n \cdot a^m = a^{n+m}$ and $(a^n)^{-1} = a^{-n}$. Hence $\langle a \rangle$ is a subgroup of G.

As the group G is finite we must have a repetition in the sequence

$$1, a, a^2, \cdots$$

Suppose that $a^r = a^s$ for $r < s$ then $a^{s-r} = 1$.

Definition. Let G be a finite group and $a \in G$. The *order*, $o(a)$, of a in G is the smallest positive intger n such that $a^n = 1$.

Example. The element $[2] \in \mathbb{Z}_5^*$ has order 4 since $[2]^4 = [16] = [1]$ wheras none of $[2], [2]^2 = [4], [2]^3 = [3]$ is $[1]$.

Remark. Let G be a finite abelian group and let $a \in G$ be an element such that $o(a) = n$. The following properties hold:

(1) $n = |\langle a \rangle|$.
(2) $a^m = 1$ iff $n|m$.
(3) $a^r = a^s$ iff $n|(r - s)$.
(4) $a^{|G|} = 1$.

To see that (2) holds, suppose first that $n|m$, say $m = nr$, then $a^m = (a^n)^r = 1^r = 1$. Conversely suppose $a^m = 1$ and that $m = nr + s$ with $0 \leq s < n$. Then $a^s = a^{m-rn} = a^m(a^n)^{-r} = 1$. As n is the smallest positive integer such that $a^n = 1$ we then must have $s = 0$ and hence $n = mr$.

(3) We have that $a^r = a^s$ iff $a^{r-s} = 1$. By (1) this holds if and only if $n|r - s$.

(1) Notice that it follows from (3) that there is a 1-1 correspondence between elements in $\langle a \rangle$ and the congruence classes of the integers modulo n. Hence $\langle a \rangle = \{1 = a^0, a, a^2, \ldots, a^{n-1}\}$ and $|\langle a \rangle| = n$.

(4) By Lagrange's Theorem we have that $n = |\langle a \rangle|$ divides $|G|$ and thus by (2) it follows that $a^{|G|} = 1$. Alternatively one can argue as follows. Consider the map $G \to G, x \mapsto xa$. This map is a bijection with inverse $x \mapsto xa^{-1}$. It follows that $G = Ga$. So if $\{a_1, \ldots, a_m\} = G = Ga = \{a_1 a, \ldots, a_m n\}$. Then

$$a_1 \cdots a_m = (a_1 a) \cdots (a_m a) = a_1 \cdots a_m a^m.$$

Cancellation gives $1 = a^m = a^{|G|}$. $\square$

Remarks. (1) In particular as $|\mathbb{Z}_n^*|$ has $\phi(n)$ elements we have $[a]^{\phi(n)} = [1]$ for all $[a] \in \mathbb{Z}_n^*$.

(2) If p is a prime, then $1, 2, \ldots, p-1$ are all coprime to p. Thus $[a]$ is a unit if $[a] \neq [0]$ and $\mathbb{Z}_p$ is a field. In particular $\phi(p) = p - 1$ and from part (1), we have that $[a]^{p-1} = [1]$ whenever $[a] \neq [0]$. It follows that $[a]^p = [a]$ for all $[a] \in \mathbb{Z}_p$. Thus p divides $a^p - a$ for all intgegers a. (The Little Fermat Theorem).

Definition. We say that a group G is cyclic if $G = \langle a \rangle$ for some $a \in G$. Any $a \in G$ such that $G = \langle a \rangle$ is called a generator for a.

Lemma 2.18 *Let $G = \langle a \rangle$ be a finite cyclic group with n elements.*

(1) $o(a^r) = n/d$ where $d = hcf(r, n)$.
(2) The number of generators for G is $\phi(|G|)$.

Proof (1) We have that $a^{rm} = 1$ if and only if $n|rm$ if and only if n/d divides $(r/d)m$. As n/d and r/d have no common factors, it follows that $n|rm$ if and only if n/d divides m. By the remark above this means that $o(a^r) = n/d$.

(2) a^r is a generator for G if and only if $o(a^r) = n$. By (1) we have that $o(a^r) = n/d$. Thus $o(a^r) = n$ if and only if r and n are coprime. The number of generators for G is thus the number of $a^r,\ 0 \leq r \leq n-1$ where r and n is coprime. But by Theorem 2.15 this is the same as the number of units in $\mathbb{Z}_n$. This number is therefore $\phi(n)$. $\square$

Lemma 2.19 *Let $G = \langle a \rangle$ be a cyclic group with n elements.*

(1) Every subgroup of G is cyclic of order dividing n.
(2) Conversely we have exactly one cyclic subgroup of order d for any divisior d of n.

Proof (See sheet 7)

Notation. We will write $H \leq_c G$ if H is a cyclic subgroup of G.

Lemma 2.20 *Let G be a finite group with n elements. Then*

(a) $n = \sum_{H \leq_c G} \phi(|H|)$.
(b) $n = \sum_{d|n} \phi(d)$.

Proof (a) Let $X = \{(a, H) : a \in G,\ H \leq_c G \text{ and } H = \langle a \rangle\}$. We count the number of pairs in X in two different ways. Firstly as for each $a \in G$ there is clearly exactly one cyclic subgroup H such that $H = \langle a \rangle$, we have that the number of pairs is $|G| = n$. Now take the set of all cyclic subgroups of G. Each such subgroup H is generated by $\phi(|H|)$ elements by Lemma 2.18. Adding up gives thus the right hand side of the equation in (a)

(b) Now suppose furthermore that G is cyclic. By Lemma 2.20 we have that for each

$d|n$ there is exactly one cyclic subgroup with d elements. Hence the RHS of (a) in this case becomes $\sum_{d|n} \phi(d)$. $\square$

Proposition 2.21 *Let G be a finite group of order n. G is cyclic if and only if for each divisor d of n there is at most one cyclic subgroup of of order d.*

Proof ($\Rightarrow$) This follows from Lemma 2.19.

($\Leftarrow$) By Lemma 2.20 we have that

$$n \overset{(a)}{=} \sum_{d|n} \epsilon_d \phi(d) \overset{*}{\leq} \sum_{d|n} \phi(d) \overset{(b)}{=} n$$

where ϵ_d is either 1 or 0 depending on whether or not we have a cyclic subgroup of order d. But we must have equality at (*). Thus in particular $\epsilon_n = 1$ and there is a cyclic subgroup of order n which is then G itself. $\square$

Theorem 2.22 *If K is a finite field, then $K^* = K \setminus \{0\}$ is a cyclic group.*

Proof Let d be any divisor of $|K^*|$. By Proposition 2.21, it suffices to show that there is at most one cyclic subgroup of K^* of order d. If there is none, we are done! Now suppose that we have at least one subgroup $H = \{a_1, \ldots, a_d\}$ of order d. By Lagrange's Theorem we know that the order of a_i divides $|H| = d$ and thus

$$a_1^d = a_2^d = \ldots = a_d^d = 1$$

that implies that $a_1, \ldots, a_d$ are roots of $x^d - 1$, i.e.

$$x^d - 1 = (x - a_1) \cdots (x - a_d).$$

As the roots are unique there is only one possible H. $\square$

B. Quotient rings of $K[x]$ and K-algebras

In this section we will consider quotient rings of $K[x]$. It turns out that these can be viewed not only as rings but also as vector spaces over K with some additional properties. We will refer to these as K-algebras. In away this is not so surprising as $K[x]$ itself is a vector space over K.

Definition. Let V be a vector space over a field K equipped with a multiplication so that $(V, +, \cdot)$ is a ring and furthermore

$$(\lambda u) \cdot v = u \cdot (\lambda v) = \lambda(u \cdot v)$$

for all $\lambda \in K$ and $u, v \in V$. We then say that V is a *K-algebra.*

Remark. (1) Let $x \in V$. Then

$$(\alpha u + \beta v) \cdot x = (\alpha u) \cdot x + (\beta v) \cdot x = \alpha(u \cdot x) + \beta(v \cdot x).$$

The multiplication by x from the right is thus a linear map. Similarly the multiplication from the left by x is a linear map.

(2) Suppose that $(v_i)_{i\in I}$ is a basis for the K-algebra V. In order to determine the multiplication $\cdot$ on V, it suffices to know the values of $v_i \cdot v_j$ for all $i, j \in I$. The reason for this is the bilinearly of the multiplication, so

$$(\sum_{i\in I} \alpha_i v_i) \cdot (\sum_{j\in I} \beta_j v_j) = \sum_{i\in I j\in J} (\alpha_i\beta_j)(v_i v_j).$$

Examples. (1) Let K be a field. Then $K = K \cdot 1$ is a K-algebra of dimension 1.

(2) $\mathbb{C} = \mathbb{R} + \mathbb{R}i$ is a $\mathbb{R}$-algebra that is a 2-dimensional vector space over $\mathbb{R}$.

(3) (The quoternions or the Hamiltonian numbers). This is vector space

$$\mathbb{H} = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$$

of dimension 4 over $\mathbb{R}$ with basis $1, i, j, k$. The multiplication is determined from

$$i^2 = j^2 = k^2 = -1, \;\; ij = k, \; jk = i, \; ki = j, \; ji = -k, \; kj = -i, \; ik = -j.$$

Remark. Let 1_V be the identity of the ring V. By the additional property above, we have

$$(\lambda 1_V) \cdot v = \lambda(1_V \cdot v) = \lambda v.$$

For this reason we can identify the field K with the subalgebra $K1_V$ and think of K as being a subalgebra of V.

Fields as K-algebras

Definition. Let F be a field. A subring K of F is said to be a *subfield* if $a^{-1} \in K$ for all $0 \neq a \in K$.

Now take some field F and a subfield K of F. We want to see that F has the structure of a K algebra. First we see that F is a vector space. This is the case as

(1) $(F, +)$ is an abelian group (as $(F, +, \cdot)$ is a ring).
(2) $1_K \cdot v = v$ for all $v \in F$ (notice that $1_K = 1_F$ as K is a subring of F).
(3) $a(bv) = (ab)v$ for all $a, b \in K$ and $v \in F$ (as the multiplication in F is associative).
(4) $(a + b)v = av + bv$ and $a(v + w) = av + aw$ for all $a, b \in K$ and $v, w \in F$ (as the distributive laws hold in F).

Furthermore F is a K-algebra as

(5) $(av)w = v(aw) = a(vw)$ for all $a \in K$ and $v, w \in F$ (as the multiplication in F is associative and commutative).

Now suppose we have some $a \in F$ that is a root of a non-zero polynomial in $K[x]$. Consider the ring homomorphism

$$\phi_a : K[x] \to F, \; f \mapsto f(a).$$

Proposition 2.23 *Suppose* $ker\,\phi_a = K[x]g$.

(1) g *is irreducible and if* $h \in ker\,\phi_a$ *is irreducible then* $g \sim h$.
(2) $K[a] = im\,\phi_a = \{f(a) : f \in K[x]\}$ *is a subfield of* F.
(3) If $deg\,(g) = n$, *then* $(1, a, \ldots, a^{n-1})$ *is a basis for* $K[a]$.

Proof (1) If g was not irreducible, we would have $g = fh$ for some polynomials f, h of smaller degree than g. But as $f(a)h(a) = g(a) = 0$, we must then have $f(a) = 0$ or $h(a) = 0$. Without loss of generality we can suppose that $f(a) = 0$. Then $f \in \mathrm{Ker}\,\phi_a = K[x]g$ and $g|f$ that is absurd as f is a non-zero polynomial of smaller degree than g.

Now suppose that $h \in \ker\phi_a = K[x]g$ is irreducible. Now h is irreducible and divisible by g. As g is not a unit we must have $g \sim h$.

(2) By Lemma 1.7 we know that $K[a] = \mathrm{im}\,\phi_a$ is a subring of F. It remains to show that any $0 \neq f(a) \in K[a]$ has a multiplicative inverse in $K[a]$. As $f(a) \neq 0$ we have that $f \notin \ker\phi_a = K[x]g$ and thus g does not divide f. As g is irreducible, it follows that f and g are coprime, and thus $1 = rf + sg$ for some $r, s \in K[x]$. Then

$$1 = r(a)f(a) + s(a)g(a) = r(a)f(a) + s(a) \cdot 0 = r(a)f(a).$$

Thus $r(a) \in K[a]$ is the inverse of $f(a)$.

(3) $K[a]$ is generated by $1, a, \ldots, a^{n-1}$. To see this let $f(a)$ be an arbitrary element of $K[a]$. Division by g with remainder gives

$$f = gr + s,\ \deg(s) < \deg(g) = n,$$

say $s = b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$. Then

$$f(a) = g(a)r(a) + s(a) = s(a) = b_0 \cdot 1 + b_1a + \cdots + b_{n-1}a^{n-1}.$$

Thus $f(a)$ is in the linear span of $1, a, \ldots, a^{n-1}$.

$1, a, \ldots, a^{n-1}$ are linearly independent. To see this suppose $c_0 \cdot 1 + c_1a + \cdots + c_{n-1}a^{n-1} = 0$. Then $h = c_0 + c_1x + \cdots + c_{n-1}x^{n-1} \in \ker\phi_a = K[x]g$ and $g|h$. As $deg(h) < deg(g)$, this can only happen if $h = 0$, that is only if $c_0 = c_1 = \cdots = c_{n-1} = 0$. □

Examples. (1) We have that $\mathbb{R} \subseteq \mathbb{C}$ and that $i \in \mathbb{C}$ is a root of the irreducible polynomial $x^2 + 1 \in \mathbb{R}[x]$. Here $\mathbb{R}[i] = \mathbb{R} + \mathbb{R}i = \mathbb{C}$ has basis $(1, i)$.

(2) We have that $\mathbb{Q} \subseteq \mathbb{R}$ and that $\sqrt[3]{2}$ is a root of the irreducible polynomial $x^3 - 2 \in \mathbb{R}[x]$. Here

$$\mathbb{Q}[\sqrt[3]{2}] = \mathbb{Q} + \mathbb{Q}\sqrt[3]{2} + \mathbb{Q}(\sqrt[3]{2})^2.$$

has basis $(1, \sqrt[3]{2}, (\sqrt[3]{2})^2)$.

Remark. Loosely speaking, Proposition 2.23 tells us among other things that if $a \in F$ is a root of some irreducible polynomial in $K[x|$ of degree n then $K[a]$ (the collection of all polynomial expressions in a over K) is a field. Now suppose that we only have the field K and some irreducible polynomial over K. We will see that we can construct a larger

field F that contains a root t of that irreducible polynomial and where $F = K[t]$.

The K-algebra $K[x]/K[x]f$.

Let f be a polynomial in $K[x]$ and consider the quotient ring $V = K[x]/K[x]f$. There is a natural scalar multiplication on V given by

$$\lambda[g] = [\lambda g].$$

(Notice that this is well define since $[g] = [h] \Leftrightarrow f|(g-h) \Leftrightarrow f|\lambda(g-h) \Leftrightarrow [\lambda g] = [\lambda h]$).

Let us see briefly that this turns V into a vector space over K. Firstly, as V is a ring, we have that $(V, +)$ is an abelian group. The extra conditions for a vector space hold since

$$1 \cdot [g] = [1g] = [g],$$

$$\alpha(\beta[g]) = \alpha[\beta g] = [\alpha\beta g)] = (\alpha\beta)[g],$$
$$(\alpha + \beta)[g] = [(\alpha + \beta)g] = [\alpha g + \beta g] = \alpha[g] + \beta[g]$$

and

$$\alpha([g] + [h]) = \alpha([g + h]) = [\alpha(g + h)] = [\alpha g + \alpha h] = \alpha[g] + \alpha[h].$$

So V is a vector space over K. Then we also have

$$(\lambda[g]) \cdot [h] = [g] \cdot (\lambda[h]) = \lambda([g] \cdot [h])$$

as all of these are equal to $[\lambda gh]$. Hence $V = K[x]/K[x]f$ is a K-algebra.

Suppose now that f is a monic polynomial of degree at least 1. We next turn to the dimension of $K[x]/K[x]f$. Suppose

$$f = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n.$$

Theorem 2.24 *Let f be as above. We have*

$$K[x]/K[x]f = K \cdot 1 + Kt + \cdots + Kt^{n-1}$$

with $t = [x]$ and where $1, t, \ldots, t^{n-1}$ is a basis for $K[x]/K[x]f$. Also $f(t) = 0$.

Proof Notice that $1_V = [1]$. So we want to show that $[1], [x], \ldots, [x]^{n-1}$ is a basis for $K[x]/K[x]f$.

Linear independence. If

$$[0] = \alpha_0[1] + \alpha_1[x] + \cdots + \alpha_{n-1}[x]^{n-1} = [\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^{n-1}],$$

then f divides $\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^n$. But as f is of degree n, this can ony happen if the latter polynomial is the zero polynomial, that is we must have $\alpha_0 = \alpha_1 = \ldots = \alpha_{n-1} = 0$.

Spanning set. Let $[g]$ be any elment in $K[x]/K[x]f$. Using division by f with remainder we have

$$g = fr + s$$

where s is a polynomial of degree less than or equal to $\deg f = n$,say

$$s = \beta_0 + \beta_1 x + \cdots + \beta_{n-1}x^{n-1}.$$

Then $[g] = [f][r] + [s] = [0] + [s] = [s] = [\beta_0 + \beta_1 x + \cdots + \beta_{n-1}x^{n-1}] = \beta_0[1] + \beta_1[x] + \cdots + \beta_{n-1}[x]^{n-1}$.

Finally, notice that $f(t) = a_0 + a_1 t + \cdots + a_{n-1}t^{n-1} + t^n = a_0[1] + a_1[x] + \cdots + a_{n-1}[x]^{n-1} + [x]^n = [a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n] = [f] = [0]$. □

Corollary 2.25 *Let K be a field and f be a polynomial in $K[x]$ of degree at least 1. There exists a field L with $K \subseteq L$ and such that f can be written as product of polynomials of degree 1 in $L[x]$.*

Proof (See exercise 4 on sheet 7).

Example. (1) Consider the polynomial $x^2 + 1 \in \mathbb{R}[x]$. This polynomial is irreducible in $\mathbb{R}[x]$ and by Theorem 2.24 we have that there is a root t in the field

$$\mathbb{R}[x]/\mathbb{R}[x](x^2 + 1) = \mathbb{R} + \mathbb{R}t$$

where $t = [x]$. Now $t^2 + 1 = 0$ and thus $t^2 = -1$. This field is therefore isomorphic to $\mathbb{C} = \mathbb{R} + \mathbb{R}i$.

(2) Consider the polynomial $x^2 - 2 \in \mathbb{Q}[x]$. This is an irreducible polynomial in $\mathbb{Q}[x]$ and by Theorem 2.24, we have that there is a root t in the field

$$\mathbb{Q}[x]/\mathbb{Q}[x](x^2 - 2) = \mathbb{Q} + \mathbb{Q}t$$

where $t = [x]$. This field is isomorphic to the subfield $\mathbb{Q} + \mathbb{Q}\sqrt{2}$ of $\mathbb{R}$.

(3) Consider the polynomial $f = x^2 + x + 1$ in $\mathbb{Z}_2[x]$. If the polynomial were not irreducible we would have to have a linear factor in $\mathbb{Z}_2[x]$. That is either x or $x + 1$ would have to divide $x^2 + x + 1$. But as $f(0) = f(1) = 1$ this is not the case. Hence f is irreducible and has then a root $t = [x]$ in the field

$$L = \mathbb{Z}_2[x]/\mathbb{Z}_2 f = \mathbb{Z}_2 + \mathbb{Z}_2 t.$$

Notice that the new field L has $2^2 = 4$ elements.

We end this chapter by looking at special types of $\mathbb{R}$-algebras.

C. Normed $\mathbb{R}$-algebras

Definition. A *normed* $\mathbb{R}$-algebra, is an $\mathbb{R}$-algebra V equipped with an inner product such that for the corresponding norm we have

$$\|u \cdot v\| = \|u\| \cdot \|v\|.$$

Remark. (1) We have that $\|1_V\| = \|1_V \cdot 1_V\| = \|1_V\| \cdot \|1_V\|$ and thus $\|1_V\| = 1$.

Examples. (1) The struture of an $\mathbb{R}$-algebra V is determined by its dimension and the products of the basis elements with respect to some chosen basis. If $\dim V = 1$, then $V = \mathbb{R}1_V$ and as $1_V \cdot 1_V = 1_V$ there is (up to isomorphism) only one $\mathbb{R}$-algebra $\mathbb{R}$. This can be thought of as an normed $\mathbb{R}$-algebra where the norm is the absolute value.

(2) We can think of $\mathbb{C}$ as a normed $\mathbb{R}$-algebra with an orthonormal basis $(1, i)$.

We are going to see that there are only three normed $\mathbb{R}$-algebras: $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$. The main ingredients come from the following lemma.

Lemma 2.26 *Let V be a normed $\mathbb{R}$-algebra.*

(1) If $t \in V$ is orthogonal to 1 and $\|t\| = 1$, then $t^2 = -1$.
(2) If $i, j, 1 \in V$ are pairwise orthogonal and $\|i\| = \|j\| = 1$, then ij is orthogonal to $i, j, 1$ and $ji = -ij$.

Proof (1) We have

$$\|t^2 + (-1)\| = \|(t-1)(t+1)\| = \|t-1\| \cdot \|t+1\| = \sqrt{2}\sqrt{2} = 2 = 1 + 1 = \|t^2\| + \|-1\|.$$

(Notice that $\|t^2\| = \|t\|^2 = 1$), According to the triangle inequality we should only get equality here if t^2 is a positive multiple of -1 and, as $\|t^2\| = 1$, this can only happen if $t^2 = (-1)$.

(2) We have that $\frac{i+j}{\sqrt{2}}$ is orthogonal to 1 and of length 1. By part (1), it follows that

$$-1 = (\frac{i+j}{\sqrt{2}})^2 = \frac{i^2 + j^2 + ij + ji}{2} = \frac{(-1) + (-1) + ij + ji}{2} = -1 + \frac{ij + ji}{2}.$$

Hence $ji = -ij$.

Notice now that

$$\|ij + (-i)\|^2 = \|i(j-1)\|^2 = \|i\|^2 \cdot \|j-1\|^2 = 1 \cdot 2 = 1 + 1 = \|ij\|^2 + \|-i\|^2.$$

(Notice that $\|ij\| = \|i\| \cdot \|j\| = 1$). As the pythogoras theorem holds it follows that ij is orthogonal to i. Similarly,

$$\|ij + (-j)\|^2 = \|(i-1)j\|^2 = \|i-1\|^2 \cdot \|j\|^2 = 2 \cdot 1 = \|ij\|^2 + \|-j\|^2,$$

gives that ij is orthogonal to j. Finally

$$\|ij - 1\|^2 = \|ij + i^2\|^2 = \|i(j+i)\|^2 = \|i\|^2 \cdot \|j+i\|^2 = 1 \cdot 2 = 2 = \|ij\|^2 + \|-1\|^2,$$

that gives that ij is orthogonal to 1 as well. □

Theorem 2.27 *Up to isomorphism, there are exactly three normed $\mathbb{R}$-algebras: $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$.*

Proof We have already seen that $\mathbb{R}$ is the unique normed $\mathbb{R}$-algebra of dimension 1. By part (1) of last lemma, $\mathbb{C}$ is the unique normed $\mathbb{R}$-algebra of dimension 2 (if V has orthonormal basis $(1, i)$ then we must have $i^2 = -1$).

From part (2) of last lemma we have that if $\dim V \geq 3$ then we must have $\dim V \geq 4$ (as we get from $1, i, j$ a new element ij that is orthogonal to $1, i, j$ and thus in particular $1, i, j, ij$ are linearly independent).

From part (2) of last Lemma we also see that if we have a four dimensional normed algebra then the structure is uniquely determined. Let us have a closer look at this. According to part (2) of last Lemma, we have that there is an orthonormal basis $(1, i, j, k)$ where $k = ij$. We know that $ji = -ij$ and by symmetry $ik = -ki$ and $kj = -jk$. By part (1) of last lemma we also have $i^2 = j^2 = k^2 = -1$. Then $jk = jij = -j^2 i = -(-i) = i$ and $ki = iji = -i^2 j = -(-j) = j$ and the product of the basis elements, and hence the structure of the algebra, are then uniquely determined.

There is no normed $\mathbb{R}$-algebra of dimension higher than 4. To see this we argue by contradiction and suppose that we have a normed algebra V of dimension at least 5. Using last lemma we get a subspace $\mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}ij$ (taking any i, j such that $1, i, j$ are pairwise orthogonal). Now pick $e \in V$ that is orthogonal to $1, i, j, ij$. Using Lemma 2.26, we have

$$ije = -e(ij) = iej = -ije$$

and thus we get $ije = 0$ but $\|ije\| = \|i\| \cdot \|j\| \cdot \|e\| = 1$ so this is absurd. $\square$

Remark. We haven't yet shown the existence of $\mathbb{H}$. One way of doing this is simply to take the structure as it comes to us and check that it satisfies all the algebraic laws needed for it to be an $\mathbb{R}$-algebra. This is very cumbersome as one needs to check a number of things. A neater approach, given in exercise 4 on sheet 8, is to construct the quaternions as a subring of $\mathrm{End}\,(V)$ as then we get all the algebraic ring laws of $\mathrm{End}\,(V)$ for free.

On exercise sheet 7 we see how the quaternions are linked with geometry, where we see how the inner product and cross product in three dimensions can be interpreted in terms of the quaternions. Now we are going to look at another very beautful application in Number Theory.

Consider the subring $\mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ of the quaternions. For

$$z = x_1 + x_2 i + x_3 j + x_4 k, \quad w = y_1 + y_2 i + y_3 j + y_4 k,$$

we have

$$\begin{aligned} zw &= (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4) + (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3)i \\ &\quad +(x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2)j + (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1)k. \end{aligned}$$

Now since $\|z\|^2\|w\|^2 = \|zw\|^2$, we get

$$\begin{aligned} &(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = \\ &\quad (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4)^2 + (x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3)^2 \\ &\quad +(x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2)^2 + (x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1)^2. \end{aligned} \tag{1}$$

It follows that a if we have two sums of four squares, then their product is also a sum of four squares. We are now going to prove that every natural number can be written as sum of four integer squares. Notice first that $1 = 1^2+0^2+0^2+0^2$ and that $2 = 1^2+1^2+0^2+0^2$. Since the set consisting of sum of four squares is closed under multiplication, it follows that it now suffices to show that every odd prime can be written as a sum of four squares. We need first a lemma.

Lemma 2.28 *If p is an odd prime, then there exists integers x, y and m such that*

$$1 + x^2 + y^2 = mp$$

and $0 < m < p$.

Proof We calculate modulo p. Notice that $[0]^2, [1]^2, \ldots, [\frac{p-1}{2}]^2$ are distinct. (If $[x]^2 = [y]^2$ for some $0 \leq y < x \leq (p-1)/2$, then $p|(x^2-y^2) = (x-y)(x+y) \Rightarrow p|(x-y)$ or $p|(x+y)$. But as $1 \leq x-y, x+y \leq p-1$ this doesnt happen). It follows from this that we get two lists

$$[1 + x^2], \quad 0 \leq x \leq (p-1)/2$$

and

$$[-y^2], \quad 0 \leq y \leq (p-1)/2$$

each of which has $(p+1)/2$ distinct values. In total we then have $p+1 > p$ values so the two lists must have a value in common, say $[1 + x^2] = [-y^2] \Rightarrow [1 + x^2 + y^2] = [0]$. Hence

$$1 + x^2 + y^2 = pm$$

for some integer m. Now $pm = 1 + x^2 + y^2 \leq 1 + (\frac{p-1}{2})^2 + (\frac{p-1}{2})^2 < 1 + 2(p/2)^2 < p^2$ and thus $m < p$. □

Theorem 2.29 *(Lagrange's four square theorem) Every natural number can be written as a sum of four integer squares.*

Proof We have already seen that it suffices to prove that every odd prime p can be written as a sum of four squares. We choose the smallest positive integer m such that

$$pm = x_1^2 + x_2^2 + x_3^2 + x_4^2. \tag{2}$$

By Lemma 2.28, such an integer m exists and we know that $0 < m < p$. The aim is to show that $m = 1$. We argue by contradiction and suppose that $m > 1$.

Step 1. m is odd. Otherwise an even number of x_1, x_2, x_3, x_4 are odd. By rearranging the order of terms if needed we can assume that both x_1, x_2 are even/odd and both x_3, x_4 are even/odd. Hence $x_1 + x_2, x_1 - x_2, x_3 + x_4, x_3 - x_4$ are all even. It follows that

$$\begin{aligned} p(m/2) &= \frac{2(x_1^2 + x_2^2 + x_3^2 + x_4^2)}{4} \\ &= (\frac{x_1 - x_2}{2})^2 + (\frac{x_1 + x_2}{2})^2 + (\frac{x_3 - x_4}{2})^2 + (\frac{x_3 + x_4}{2})^2, \end{aligned}$$

that contradicts the minimality of m. Hence m is odd.

Step 2. We do not have $[x_1]_m = [x_2]_m = [x_3]_m = [x_4]_m = [0]_m$. Otherwise m would divide all of $x_1, \ldots, x_4$. From equation (2) we would then have that the right hand side is divisble by m^2 and thus $m|p$ and as $m < p$ this would imply that $m = 1$ contracting our assumption that $m > 1$.

We can now finish the proof of the Theoem. For each $i \in \{1, 2, 3, 4\}$ we pick a representative y_i such $-(m-1)/2 \leq y_i \leq (m-1)/2$ and $[y_i] = [x_i]$. As the sum $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is the same as the sum $y_1^2 + y_2^2 + y_3^2 + y_4^2$ modulo m and as the former sum is divisible by m, the latter sum is also divisble by m. It follows that

$$mr = y_1^2 + y_2^2 + y_3^2 + y_4^2 \tag{3}$$

for some r where

$$mr \leq 4(\frac{m-1}{2})^2 = (m-1)(m-1).$$

Notice that this implies that $r < m$. By step 2 we know that not all of the $[y_i]_m = [x_i]_m$ are $[0]_m$ and thus in particular the rhs of (3) is non-zero that implies that $r > 0$. Thus $0 < r < m$.

Now mulitplying together (2) and (3) gives

$$\begin{aligned} prm^2 &= (x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + (-y_2)^2 + (-y_3)^2 + (-y_4)^2) \\ &= (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3)^2 \\ &\quad +(-x_1y_3 + x_2y_4 + x_3y_1 - x_4y_2)^2 + (-x_1y_4 - x_2y_3 + x_3y_2 + x_4y_1)^2. \end{aligned}$$

Now if we calculate modulo in $\mathbb{Z}_m$ we see that (since $[y_i]_m = [x_i]_m$)

$$[x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4]_m = [x_1^2 + x_2^2 + x_3^2 + x_4^2]_m = [pm]_m = [0]_m$$

and thus $m|(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)$. Also

$$[-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3]_m = [-x_1x_2 + x_2x_1 - x_3x_4 + x_4x_3]_m = [0]_m,$$

$$[-x_1y_3 + x_2y_4 + x_3y_1 - x_4y_2]_m = [-x_1x_3 + x_2x_4 + x_3x_1 - x_4x_2]_m = [0]_m$$

and

$$[-x_1y_4 - x_2y_3 + x_3y_2 + x_4y_1]_m = [-x_1x_4 - x_2x_3 + x_3x_2 + x_4x_1]_m = [0]_m.$$

As all these integers are divisible by m, we get

$$\begin{aligned} pr &= (\frac{x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4}{m})^2 + (\frac{-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3}{m})^2 \\ &\quad (\frac{-x_1y_3 + x_2y_4 + x_3y_1 - x_4y_2}{m})^2 + (\frac{-x_1y_4 - x_2y_3 + x_3y_2 + x_4y_1}{m})^2. \end{aligned}$$

As $r < m$, we get a contradiction about our minimality assumption on m. It follows that the smallest m given in (1) must be 1 and thus p is a sum of integer squares. □

3 The structure of linear operators

Let V be an n-dimensional vector space over a field K with basis $\mathcal{V} = (v_1, v_2, \ldots, v_n)$. Let $\alpha : V \to V$ be a linear operator in $\mathrm{End}\,(V)$ and let A be the matrix representing α with respect to the basis $\mathcal{V}$.

Aim. To find a basis $\mathcal{V}$ such that A looks simple.

I. Minimal polynomials

Recall that the rings $\mathrm{End}\,(V)$ and $M_n(K)$ are isomorphic and are vector spaces over K of dimension n^2. It follows that the matrices

$$I, A, A^2, \ldots, A^{n^2}$$

(or equivalently the linear operators $\mathrm{id}, \alpha, \alpha^2, \ldots, \alpha^{n^2}$) are linearly dependent. Take $a_0, \ldots, a_{n^2} \in K$ (not all zero) such that

$$a_0 I + \cdots + a_{n^2} A^{n^2} = 0$$

then $f(A) = 0$, where f is the polynomial $a_0 + a_1 t + \cdots + a_{n^2} t^{n^2}$.

Consider the ring homomorphism

$$K[t] \to M_n(K),\ f \mapsto f(A).$$

We have seen that the kernel of this map is not $\{0\}$ and as $K(t)$ is a principal ideal domain, we know that the kernel is of the from $K[t]m(t)$ for some monic polynomial $m(t)$ of degree at least 1. Notice that we also have a ring homomorphism

$$K[t] \to \mathrm{End}\,(V),\ f \mapsto f(\alpha)$$

with the same kernel $K[t]m(t)$. Notice that $m(t)$ is the unique monic polynomial of smallest degree such that $m(A) = 0$ ($m(\alpha) = 0$).

Definition. 1)The *minimal polynomial* of the linear operator $\alpha : V \to V$ is the monic polynomial, $m_\alpha(t)$ of lowest degree such that $m_\alpha(\alpha) = 0$.

2) The minimal polynomial of the $n \times n$ matrix A is the monic polynomial $m_A(t)$ of smallest degree such $m_A(A) = 0$.

Examples. 1)If $\alpha = \lambda \mathrm{id}$ then $p(\alpha) = 0$ where $p(t) = t - \lambda$. Clearly, thus $m_\alpha(t) = t - \lambda$.

2) If $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, then $A^2 = I$ and $p(A) = 0$ where $p(t) = t^2 - 1$. As A is not a diagonal matrix, we have that $q(A) \neq 0$ for any $q = t - \lambda$. Hence $m_A(t) = t^2 - 1$.

Recall from Algebra 1B. (1) The *characteristic polynomial* of A is $\Delta_A(t) = \det\,(A - tI)$.

If A is a matrix representing a linear operator α with respect to some basis then we define the characteristic polynomial of α to be $\Delta_\alpha(t) = \det(\alpha - t\mathrm{id}) = \det(A - tI)$. This doesn't depend on the matrix that represents α and thus this is well defined.

(2) Recal that the *algebraic multiplicity*, $\mathrm{am}(\lambda)$, of an eigenvalue λ is the multiplicity of λ as a root of $\Delta_A(t)$ (or $\Delta_\alpha(t)$). The *geometric multiplicity* of λ is the dimension of the eignspace $E_A(\lambda)$ (or $E_\alpha(\lambda)$). We know that we always have $\mathrm{am}(\lambda) \geq \mathrm{gm}(\lambda)$.

Example. In the example above we have

$$\Delta_A(t) = \det(A - tI) = \begin{vmatrix} -t & 1 \\ 1 & -t \end{vmatrix} = t^2 - 1 = (t-1)(t+1) = m_A(t).$$

We will later see that minimal polynomial and the characteristic polynomial are strongly related and that the latter is always a multiple of the minimal polynomial. Here $\mathrm{am}(1) = \mathrm{am}(-1) = \mathrm{gm}(1) = \mathrm{gm}(-1) = 1$.

Lemma 3.1 *Let p be a polynomial such that $p(\alpha) = 0$ then every eigenvalue of α is a root of p. In particular every eigenvalue of α is a root of m_α.*

Proof Let $v \neq 0$ be an eigenvector with respect to λ and suppose $p(t) = a_0 + a_1 t + \ldots + a_k t^k$. Then $p(\alpha) = 0$ gives us

$$\begin{aligned} 0 &= p(\alpha)\, v \\ &= (a_0 \mathrm{id} + a_1\alpha + \cdots + a_k\alpha^k)v \\ &= (a_0 + a_1\lambda + \cdots + a_k\lambda^k)v \\ &= p(\lambda)v. \end{aligned}$$

As $v \neq 0$ it follows that $p(\lambda) = 0$. $\square$

We now turn to a remarkable fact. It turns out that any linear operator $\alpha : V \to V$ satisfies the characteristic polynomial $\Delta_\alpha(t)$.

Theorem 3.2 *(Cayley-Hamilton). For any $n \times n$ matrix A we have $\Delta_A(A) = 0$. Equivalently, for any linear $\alpha : V \to V$ we have $\Delta_\alpha(\alpha) = 0$.*

Remark. Some warning before we give the proof. One can't simply argue as follows. $\det(A - AI) = \det(0) = 0$ and thus $\Delta_A(A) = 0$. What calculating $\Delta_A(A)$ means is to calculate the determinant

$$\begin{pmatrix} a_{11} - A & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - A & \ldots & a_{2n} \\ & & \vdots & \\ a_{n1} & a_{(n-1)2} & \ldots & a_{nn} - A \end{pmatrix}$$

where we treat A as a scalar and come up with a polynomial expression in A that turns out to be the zero matrix. This is not at all the same as calculating the determinant of

the matrix $A - AI = 0$ over K. The following proof does however in a sense carry through the elementary spirit of this false approach.

Proof Suppose

$$\Delta_A(t) = \det(A - tI) = a_0 + a_1 t + \cdots + a_n t^n.$$

We must show that $\Delta_A(A) = a_0 I + a_1 A + \cdots + a_n A^n = 0$ as a matrix. We will apply the adjugate formula. Notice that

$$\operatorname{adj}(A - tI) = B_0 + B_1 t + \cdots + B_{n-1} t^{n-1},$$

where each B_i is an $n \times n$ matrix. The adjugate formula tells us that

$$\operatorname{adj}(A - tI)(A - tI) = \det(A - tI) I = \Delta_A(t) I.$$

That is

$$(B_0 + B_1 t + \cdots + B_{n-1} t^{n-1})(A - tI) = (a_0 + a_1 t + \cdots + a_n t^n) I.$$

This formula simply tells us that coefficient to t^i on the lhs has to be the square $n \times n$ matrix $a_i I$. Now replace t by any square $n \times n$ matrix T that commutes with A. What we get is that when expanded out the lhs becomes again a polynomial expression, this time in T, where the coefficient to T^i is the same as before (that is $a_i I$). Thus the same formula

$$(B_0 + B_1 T + \cdots + B_{n-1} T^{n-1})(A - TI) = (a_0 + a_1 T + \cdots + a_n T^n) I$$

as before holds. In particular letting $T = A$, we get

$$\Delta_A(A) = a_0 I + a_1 A + \cdots a_n A^n = (B_0 + B_1 A + \cdots + B_{n-1} A^{n-1})(A - A) = 0.$$

This finishes the proof. □.

Remark. It follows from the Cayley-Hamilton Theorem that $m_\alpha(t)$ divides $\Delta_\alpha(t)$. Next we are going to see that these have the same roots.

Proposition 3.3 . *The roots of m_α are precisely the eigenvalues of α.*

Proof By the remark above, we have that m_α divides Δ_α and thus every root of m_α is a root of Δ_α and therefore an eigenvalue of α. The converse follows from Lemma 3.1. □

Remark. It follows from this last propositon and the Cayley-Hamilton Theorem that, over $\mathbb{C}$, if $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of λ and

$$\Delta_\alpha(t) = (\lambda_1 - t)^{r_1} \cdots (\lambda_k - t)^{r_k},$$

then

$$m_\alpha(t) = (t - \lambda_1)^{s_1} \cdots (t - \lambda_k)^{s_k}$$

with $1 \leq s_i \leq r_i$ for all $1 \leq i \leq k$.

II. Invariant subspaces and primary decompositions

A. Invariant subspaces

Definition Let $\alpha : V \to V$ be a linear operator. We say that a subspace W of V is α-*invariant* if $\alpha(W) \subseteq W$.

If W is α-invariant, then the *restriction* of α to W is the linear operator

$$\alpha|_W : W \to W : w \mapsto \alpha(w).$$

Examples. (1) The subspaces $\{0\}$ and V are always α-invariant.

(2) Let λ be an eigenvalue of α and v is an eigenvectors with respect to λ then the one dimensional subspace Kv is α-invariant. This is because $\alpha(rv) = r\alpha(v) = r\lambda v \in Kv$.

(3) Let $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator that rotates every vector 30 degrees around the z-axis (counter clockwise). Here $\mathbb{R}e_3$ and $\mathbb{R}e_1 + \mathbb{R}e_2$ are α-invariant.

Now suppose that

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

where $V_1, \ldots, V_k$ are α-invariant subspaces. Let $\alpha_i = \alpha|_{V_i}$ be the restriction of α to V_i. Then $\alpha_i \in \mathrm{End}\,(V_i)$. In this situation we often write

$$\alpha = \alpha_1 \oplus \cdots \oplus \alpha_k.$$

Notice that if $v = v_1 + \cdots + v_k$, with $v_i \in V_i$, then

$$\begin{aligned} \alpha(v) &= \alpha(v_1) + \cdots + \alpha(v_k) \\ &= \alpha_1(v_1) + \cdots + \alpha_k(v_k). \end{aligned}$$

Pick a basis $\mathcal{V}_i$ for V_i and let A_i be the matrix representing α_i with respect to the basis $\mathcal{V}_i$. Then the matrix representing α with respect to the basis $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \cdots \cup \mathcal{V}_k$ is the matrix

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}.$$

that we often write as $A = A_1 \oplus \cdots \oplus A_k$.

Example. Suppose $V_1 = Kv_1 \oplus Kv_2$ and $V_2 = Kv_3 \oplus Kv_4$. Suppose furthermore that the linear operators $\alpha_1 : V_1 \to V_1$ and $\alpha_2 : V_2 \to V_2$ are defined by

$$\begin{array}{ll} \alpha_1(v_1) = 2v_1 + v_2 & \alpha_1(v_2) = v_1 - v_2 \\ \alpha_2(v_3) = v_4 & \alpha_2(v_4) = v_3. \end{array}$$

Then $\alpha_1 \oplus \alpha_2 : V_1 \oplus V_2 \to V_1 \oplus V_2$ has matrix

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} = A_1 \oplus A_2$$

where

$$A_1 = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The aim is to break V into a direct sum $V_1 \oplus V_2 \oplus \cdots \oplus V_k$ where k is as big as possible.

Remark. Notice that if f is any polynomial in $K[t]$ then

$$f(A) = \begin{pmatrix} f(A_1) & & & \\ & f(A_2) & & \\ & & \ddots & \\ & & & f(A_k) \end{pmatrix} = f(A_1) \oplus \cdots \oplus f(A_k).$$

Thus $f(A) = 0$ if and only if $m_{A_i} | f$ for all $i = 1, \ldots, k$. This implies that the m_A is the least common multiple of $m_{A_1}, \ldots, m_{A_k}$. Equivalently m_α is the least common multiple of $m_{\alpha_1}, \ldots, m_{\alpha_k}$. In particular if $m_{\alpha_1}, \ldots, m_{\alpha_k}$ are pairwise comprime, then

$$m_\alpha = m_{\alpha_1} \cdots m_{\alpha_k}.$$

Thus a decomposition of V into α-invariant subspaces leads to a factorization of the minimal polynomial. Our next aim is to see that one can reverse this procedure so a factorization of the minimal polynomial leads to a decomposition of V into α-invariant subspaces.

B. Primary Decompositions

Lemma 3.4 *Suppose $\alpha, \beta : V \to V$ are linear operators such that $\alpha\beta = \beta\alpha$. Then $\ker \beta$ and $\operatorname{im} \beta$ are α-invariant.*

Proof If $w \in \ker \beta$ then

$$\beta(\alpha(w)) = \alpha(\beta(w)) = \alpha(0) = 0.$$

hence $\alpha(w) \in \ker \beta$. This shows that $\ker \beta$ is α-invariant. To see that $\operatorname{im} \beta$ is α-invariant, notice that if $v = \beta(u)$ then $\alpha(v) = \alpha(\beta(u)) = \beta(\alpha(u)) \in \operatorname{im} \beta$. □

Lemma 3.5 *Let $\alpha : V \to V$ be a linear operator whose minimal polynomial has a factorization*

$$m_\alpha(t) = p_1(t)p_2(t)$$

where p_1 and p_2 are monic polynomials that are coprime. Let $V_1 = im\, p_2(\alpha)$ and $V_2 = im\, p_1(\alpha)$. Then

(1) The subspaces V_1 and V_2 are α-invariant.
(2) $V = V_1 \oplus V_2$.
(3) The minimal polynomial of $\alpha_i = \alpha|_{V_i}$ is $p_i(t)$.
(4) $V_1 = ker\, p_1(\alpha)$ and $V_2 = ker\, p_2(\alpha)$.

Proof (1) Notice that $p_1(\alpha)$ and $p_2(\alpha)$ commute with α and thus V_1, V_2 are α-invariant by Lemma 3.4.

(2) As p_1 and p_2 are coprime, there are polynomials $a_1, a_2 \in K[t]$ such that $1 = a_1(t)p_1(t) + a_2(t)p_2(t)$. Hence

$$\mathrm{id} = p_2(\alpha)a_2(\alpha) + p_1(\alpha)a_1(\alpha)$$

Thus for any $v \in V$, we have

$$v = \mathrm{id}(v) = [p_2(\alpha)a_2(\alpha)](v) + [p_1(\alpha)a_1(\alpha)](v) \in \mathrm{im}\, p_2(\alpha) + \mathrm{im}\, p_1(\alpha) = V_1 + V_2.$$

This shows that $V = V_1 + V_2$. To see that the sum is direct, suppose $v \in V_1 \cap V_2$, say $v = p_2(\alpha)(v_2) = p_1(\alpha)(v_1)$. Then

$$\begin{aligned} v &= a_1(\alpha)(p_1(\alpha)(v) + a_2(\alpha)p_2(\alpha)(v) \\ &= [a_1(\alpha)p_1(\alpha)p_2(\alpha)](v_2) + [a_2(\alpha)p_2(\alpha)p_1(\alpha)](v_1) \\ &= [a_1(\alpha)m_\alpha(\alpha)](v_2) + [a_2(\alpha)m_\alpha(\alpha)](v_1) \\ &= 0. \end{aligned}$$

Hence $V_1 \cap V_2 = \{0\}$ and $V = V_1 \oplus V_2$.

(3) We have that $f(\alpha_1) = 0$ if and only if $f(\alpha)(v) = 0$ for all $v \in V_1$. As $V_1 = \mathrm{im}\, p_2(\alpha)$ this happens if and only if $[f(\alpha)(p_2(\alpha)](v) = 0$ for all $v \in V$. As m_α is the minimal polynomial for α, this happens if and only if $m_\alpha = p_1p_2$ divides fp_2. But this happens if and only if $p_1|f$. Hence p_1 is the minimal polynomial of α_1. Similarly p_2 is the minimal polynomial of α_2.

(4) As $p_1(\alpha)p_2(\alpha)(v) = m_\alpha(v) = 0$ for all $v \in V$, it is clear that $V_1 = \mathrm{im}\, p_2(\alpha) \subseteq \ker p_1(\alpha)$. To get equality we just need to show that the dimensions are the same. But this follows from

$$\dim V = \dim V_1 + \dim V_2 = \dim \mathrm{im}\, p_2(\alpha) + \dim \mathrm{im}\, p_1(\alpha)$$

and (using the nullity rank theorem from year 1)

$$\dim V = \dim \ker p_1(\alpha) + \dim \mathrm{im}\, P_1(\alpha).$$

comparing the two equations we see that $\dim \ker p_1(\alpha) = \dim \mathrm{im}\, p_2(\alpha) = \dim V_1$. Similarly one shows that $V_2 = \ker p_2(\alpha)$. □

Now let $\mathcal{P}$ be the set of all irreducibles in $K[t]$ that are monic. We have seen earlier that these form a set of prime representatives for $K[t]$. Using Lemma 3.5 and induction on k, we get one of the main results about the structure of linear operators.

Theorem 3.6 *(Primary Decomposition) Let $\alpha : V \to V$ be a linear operator whose minimal polynomial has a factorization*

$$m_\alpha(t) = p_1(t)^{n_1} \cdots p_k(t)^{n_k}$$

where the $p_1, \ldots, p_k$ are distinct primes in $\mathcal{P}$. Let $q_i = p_i^{n_i}$ and let $V_i = ker\, q_i(\alpha)$.

(1) The subspaces $V_1, \ldots, V_k$ are α-invariant,
(2) $V = V_1 \oplus \cdots \oplus V_k$,
(3) the minimal polynomial of $\alpha_i = \alpha|_{V_i}$ is $q_i = p_i^{n_i}$.

Diagonalisable linear operators. Suppose we have a diagonalisable linear operator $\alpha : V \to V$ with a basis $\mathcal{V} = (v_1, \ldots, v_n)$ of eigenvectors. Notice that

$$V = Kv_1 \oplus \cdots \oplus Kv_n$$

where $Kv_1, Kv_2, \ldots, Kv_n$ are α-invariant. If the corresponding eigenvalues are $\lambda_1, \ldots, \lambda_n$ then the matrix for A for α with respect to $\mathcal{V}$ is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

It is not difficult to see (sheet 9) that

$$m_\alpha(t) = (t - \mu_1)(t - \mu_2) \cdots (t - \mu_k)$$

where $\mu_1, \ldots, \mu_k$ are the DISTINCT eigenvalues. The next result shows that the converse is also true.

Theorem 3.7 *The linear map $\alpha : V \to V$ is diagonalisable iff*

$$m_\alpha(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_k)$$

for some distinct $\lambda_1, \ldots, \lambda_k \in K$.

Proof By the remark above we have that the minimal polynomial of a diagonalisable linear map is a product of distinct linear factors. For the converse we make use of the Primary Decomposition Theorem. According to it we have that

$$\begin{aligned} V &= \ker(\alpha - \lambda_1 \mathrm{id}) \oplus \cdots \oplus \ker(\alpha - \lambda_k \mathrm{id}) \\ &= E_\alpha(\lambda_1) \oplus \cdots \oplus E_\alpha(\lambda_k). \square \end{aligned}$$

III. Linear operators over $\mathbb{C}$

In this section we will focus on the case when $k = \mathbb{C}$.

A. Generalised eigenvectors

In $\mathbb{C}[x]$, all polynomials can be factorized as a product of polynomials of degree 1. Now suppose that the linear operator $\alpha : V \to V$ has minimial polynomial

$$m_\alpha(t) = (t - \lambda_1)^{r_1} \cdot (t - \lambda_2)^{r_2} \cdots (t - \lambda_k)^{r_k}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of α (recall the roots of m_α are exactly the eigenvalues of α). According to the Primary Decomposition Theorem we get the following decompostion into a direct sum of α-invariant subsplaces

$$V = \ker(\alpha - \lambda_1 \mathrm{id})^{r_1} \oplus \cdots \oplus \ker(\alpha - \lambda_k \mathrm{id})^{r_k}.$$

Definition (1) Let $\alpha : V \to V$ be a linear map with eigenvalue λ. We say that $0 \neq v \in V$ is a *generalised eigenvector* with respect to the eigenvalue λ if

$$(\alpha - \lambda \mathrm{id})^r v = 0$$

for some positive integer r.

(2) The *generalised λ-eigenspace* of V is

$$G_\alpha(\lambda) = \{v \in V : (\alpha - \lambda\,\mathrm{id})^t v = 0 \text{ for some positve integer } r\}.$$

Remark. We have $E_\alpha(\lambda) \subseteq G_\alpha(\lambda)$. Notice also that

$$\ker(\alpha - \lambda \mathrm{id}) \subseteq \ker(\alpha - \lambda \mathrm{id})^2 \subseteq \ker(\alpha - \lambda \mathrm{id})^3 \subseteq \cdots$$

As V has finite dimension, this chain must become constant at some point. The next Lemma tells us when.

Lemma 3.8 *Suppose that the multiplicity of the eigenvalue λ as a root of $m_\alpha(t)$ is s. Then*

$$G_\alpha(\lambda) = ker(\alpha - \lambda\, id)^t$$

for any $t \geq s$.

Proof (Non-examinable). That we have $\supseteq$ is obvious. We need only to show $\subseteq$.

Suppose that $m_\alpha(t) = (t - \lambda_1)^{s_1}(t - \lambda_2)^{s_2} \cdots (t - \lambda_k)^{s_k}$. By the Primary Decomposition Theorem we have that

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k,$$

where $V_i = \ker(\alpha - \lambda_i \mathrm{id})^{s_i}$. Also we know from the same theorem that the minimal polynomial of $\alpha_i = \alpha|_{V_i}$ is $(t - \lambda_i)^{s_i}$. Now suppose that $\lambda = \lambda_i$. For $j \neq i$ we have that α_j only has the eigenvalue λ_j. Hence $\ker(\alpha_j - \lambda_i \mathrm{id}) = \{0\}$ and $\alpha_j - \lambda_i \mathrm{id}$ is a bijective linear operator on V_j. Now let

$$v = v_1 + v_2 + \cdots + v_k$$

be any element in $G_\alpha(\lambda)$ with $v_i \in V_i$. Suppose that $(\alpha - \lambda_i \mathrm{id})^t v = 0$. Then

$$\begin{aligned} 0 &= (\alpha - \lambda_i \mathrm{id})^t v \\ &= (\alpha_1 - \lambda_i \mathrm{id})^t v_1 + \cdots + (\alpha_k - \lambda_i \mathrm{id})^t v_k. \end{aligned}$$

This happens if and only if $(\alpha_j - \lambda_i \mathrm{id})^t v_j = 0$ for all $j = 1, \ldots, k$. As $(\alpha_j - \lambda_i \mathrm{id})^t$ is bijective if $j \neq i$, we must have that $v_j = 0$ for $j \neq i$. Hence $v = v_i \in V_i = \ker(\alpha - \lambda_i)^{s_i}$.

This shows that $G_\alpha(\lambda_i) \subseteq \ker(\alpha - \lambda_i \mathrm{id})^{s_i}$ and as $(\alpha - \lambda_i \mathrm{id})^{s_i} v = 0$ clearly implies that $(\alpha - \lambda_i \mathrm{id})^t v = 0$ for any $t \geq s_i$, it follows that

$$G_\alpha(\lambda_i) \subseteq \ker(\alpha - \lambda_i \mathrm{id})^t.$$

This finishes the proof. $\square$.

Remark. This last lemma implies in particular that

$$G_\alpha(\lambda) = \ker(\alpha - \lambda\,\mathrm{id})^s$$

which we need for the next result. But we also have

$$G_\alpha(\lambda) = \ker(\alpha - \lambda \,\mathrm{id})^r$$

where r is the algebraic multiplicity of λ. This is useful for calculating $G_\alpha(\lambda)$ as it is often easier to determine $\Delta_\alpha(t)$ than $m_\alpha(t)$.

We can now state the following special important case of the Primary Decomposition Theorem.

Theorem 3.9 *(Jordan Decomposition) Suppose that*

$$\begin{aligned} \Delta_\alpha(t) &= (\lambda_1 - t)^{r_1} \cdots (\lambda_k - t)^{r_k} \\ m_\alpha(t) &= (t - \lambda_1)^{s_1} \cdots (t - \lambda_k)^{s_k}. \end{aligned}$$

Then

$$V = G_\alpha(\lambda_1) \oplus \cdots \oplus G_\alpha(\lambda_k).$$

For the corresponding decomposition of α

$$\alpha = \alpha_1 \oplus \cdots \oplus \alpha_k,$$

we have $\Delta_{\alpha_i}(t) = (\lambda_i - t)^{r_i}$ *and* $m_{\alpha_i}(t) = (t - \lambda_i)^{s_i}$.

Proof Almost everything follows directly from the Primary Decomposition Theorem and Lemma 3.8. The only thing that remains to be proved is that $\Delta_{\alpha_i}(t) = (\lambda_i - t)^{r_i}$. To see this notice first that by Proposition 3.3 we have that the roots of m_{α_i} are exactly the eigenvalues of α_i. Hence $\Delta_{\alpha_i}(t) = (\lambda_i - t)^{t_i}$ for some positive integer t_i. As $\alpha = \alpha_1 \oplus \cdots \oplus \alpha_k$ it follows that

$$(\lambda_1 - t)^{t_1} \cdots (\lambda_k - t)^{t_k} = \Delta_\alpha(t) = (\lambda_1 - t)^{r_1} \cdots (\lambda_k - t)^{r_k}.$$

Hence $t_i = r_i$ for $i = 1, \ldots, k$. $\square$.

Remarks (1) Our study of the structure of α thus reduces to understanding $\alpha_1, \ldots, \alpha_k$. So we are left with the situation

$$\begin{aligned} \Delta_\alpha(t) &= (\lambda - t)^r \\ m_\alpha(t) &= (t - \lambda)^s \end{aligned}$$

where $1 \leq s \leq r$.

(2) Let us consider the matrix version of Theorem 3.9. Suppose that A_i is a matrix for α_i then

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}$$

is a matrix for α where for each A_i we have $(A_i - \lambda_i I)^{s_i} = 0$. If we let $N_i = A_i - \lambda_i I$ then

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix} = \begin{pmatrix} \lambda_1 I & & & \\ & \lambda_2 I & & \\ & & \ddots & \\ & & & \lambda_k I \end{pmatrix} + \begin{pmatrix} N_1 & & & \\ & N_2 & & \\ & & \ddots & \\ & & & N_k \end{pmatrix}.$$

Notice that $N_i^{s_i} = 0$. A matrix N with the property that $N^s = 0$ for some postive integer s is said to be nilpotent. We have written A as a sum of a diagonal matrix and a nilpotent matrix. This can be used for calculations of powers and exponential expressions (see sheet 9 and 10).

B. Cyclic α-invariant subspaces and the Jordan normal form.

In this section $\alpha : V \to V$ is a linear operator such that

$$\begin{aligned} \Delta_\alpha(t) &= (\lambda - t)^r \\ m_\alpha(t) &= (t - \lambda)^s \end{aligned}$$

where $1 \leq s \leq r$.

Definition. Let $v \in V$. The *cyclic α-invariant subspacee generated by v* is the subspace

$$K[\alpha]v = \{p(\alpha)v : \, p \in \mathbb{C}[t]\}.$$

Remark. Suppose that $p, q \in \mathbb{C}[t]$ and $\lambda \in \mathbb{C}$. Notice that $p(\alpha)v + q(\alpha)v = r(\alpha)v$ where r us the polynomial $p+q$ and $\lambda p(\alpha)v = s(\alpha)v$ where s is the polynomial λp. Hence $K[\alpha]v$ is a subspace of V. It is also α-invariant since $\alpha p(\alpha)v = u(\alpha)v$ where u is the polynomial $tp(t)$.

Example. If $v \in E_\alpha(\lambda)$, i.e. $\alpha(v) = \lambda v$, then $K[\alpha]v = Kv$. So for every eigenvector v we have that Kv is the cyclic α-invariant subspace generated by v.

Notice that any polynomial $p(t) \in K[t]$ can be written as $q(t - \lambda)$ for some other polynomial q. Thus every $p(\alpha)$ can be written as $q(\alpha - \lambda\mathrm{id})$. Let $v \in V$. As $m_\alpha(t) = (t-\lambda)^s$, we have that $(\alpha - \lambda\mathrm{id})^s v = 0$. Now let e be the smallest integer such that $(\alpha - \lambda\mathrm{id})^e v = 0$. Then every element in $K[\alpha]v$ is of the form

$$a_0 v + a_1(\alpha - \lambda\mathrm{id})v + \cdots + a_{e-1}(\alpha - \lambda\mathrm{id})^{e-1}v.$$

This shows that $v, \ (\alpha - \lambda\mathrm{id})v, \ \cdots, (\alpha - \lambda\mathrm{id})^{e-1}v$ span $K[\alpha]v$. In fact these form a basis for this subspace (see exercise 4(b) on sheet 9).

Lemma 3.10 *Let $0 \neq v \in V$ and let e be the smallest positive integer such that $(\alpha - \lambda id)^e v = 0$. Let*

$$v_1 = (\alpha - \lambda\, id)^{e-1}v, \ v_2 = (\alpha - \lambda\, id)^{e-2}, \ \ldots, \ v_{e-1} = (\alpha - \lambda id)v, \ v_e = v.$$

The matrix for β, the restriction of α on $K[\alpha]v$, with repsect to the basis $(v_1, v_2, \ldots, v_e)$ is the $e \times e$ matrix

$$J(\lambda, e) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

and $E_\beta(\lambda) = Kv_1$. Also $m_\beta(t) = (t - \lambda)^e$ and $\Delta_\beta(t) = (\lambda - t)^e$.

Proof. Notice that

$$\alpha(v_1) = \lambda\, v_1 + (\alpha - \lambda \mathrm{id})v_1 = \lambda\, v_1 + (\alpha - \lambda\, \mathrm{id})^e v = \lambda\, v_1$$

and for $2 \leq i \leq e$ we have

$$\alpha(v_i) = \lambda\, v_i + (\alpha - \lambda\, \mathrm{id})v_i = \lambda\, v_i + v_{i-1} = v_{i-1} + \lambda\, v_i$$

the matrix for α with respect to the basis $v_1, \ldots, v_e$ is therefore $J(\lambda, e)$. We have seen in exercise 2 on sheet 8 that $m_J(t) = (t-\lambda)^e$, that $\Delta_\beta(t) = (\lambda - t)^e$ and that $E_\beta(\lambda) = Kv_1$. $\square$

Remark. $J(\lambda, 1) = (\lambda)$ and $J(\lambda, 2) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

Definition. We call $J(\lambda, e)$ the *Jordan block* of degree e with eigenvalue λ.

Example. Consider the linear operator $\alpha : \mathbb{C}^2 \to \mathbb{C}^2,\ v \mapsto Av$ where

$$A = \begin{pmatrix} 3/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}.$$

The characteristic polynomial is $(3/2 - t)(1/2 - t) + 1/4 = 1 - 2t + t^2 = (1-t)^2$. As the matrix A is not the unit matrix the minimal polynomial is $(t-1)^2 = \Delta_\alpha(t)$. The situation is thus like in Lemma 3.10 with $e = 2$. We follow the recepi given there and look for an vector v such that $(A - I)v \neq 0$. For example the vector $v = (0\ 2)^T$ works. Then we let $v_1 = (A - I)v = (1\ \ -1)^T$ and $v_2 = v$. The matrix for α with respect to (v_1, v_2) is then $J(1, 2)$.

Remark. Suppose that $V = V_1 \oplus \cdots \oplus V_K$ where each summand is α-invariant. Let $\alpha = \alpha_1 \oplus \cdots \oplus \alpha_k$ be the corresponding decompositon of α. Now let $v = v_1 + \cdots + v_k \in V$. We have that if $\alpha(v) = \lambda v$ then

$$\alpha_1(v_1) + \cdots + \alpha_k(v_k) = \lambda v_1 + \cdots + \lambda v_k$$

and thus $\alpha_i(v_i) = \lambda v_i$ for $i = 1, \ldots, k$. This shows that

$$E_\alpha(\lambda) = E_{\alpha_1}(\lambda) \oplus \cdots \oplus E_{\alpha_k}(\lambda).$$

The next result is one of the main results in this unit.

Theorem 3.11 *(Jordan normal form) Let $\alpha : V \to V$ be any linear map such that*

$$\begin{aligned} \Delta_\alpha(t) &= (\lambda - t)^r \\ m_\alpha(t) &= (t - \lambda)^s \end{aligned}$$

There exists a basis for V such that the matrix for α with respect to this basis is

$$A = \begin{pmatrix} J(\lambda, s_1) & & & \\ & J(\lambda, s_2) & & \\ & & \ddots & \\ & & & J(\lambda, s_k) \end{pmatrix} = J(\lambda, s_1) \oplus \cdots \oplus J(\lambda, s_k).$$

Furthermore

(a) The number of Jordan blocks is $k = gm(\lambda)$.
(b) $s = max\{s_1, \ldots, s_k\}$.
(c) $s_1 + \cdots = s_k = r$.

Proof (Non-examinable). From Lemma 3.10 we know that this is the same as showing that there exist some non-zero $v_1, \ldots, v_k \in V$ such that

$$V = K[\alpha]v_1 \oplus \cdots \oplus K[\alpha]v_k, \tag{4}$$

where the dimension of $K[\alpha]v_i$ is s_i. Suppose that we have already established this. Let α_i be the restriction of α on $K[\alpha]v$. By the remark made before the statement of this theorem we know that

$$E_\alpha(\lambda) = E_{\alpha_1}(\lambda) \oplus \cdots \oplus E_{\alpha_k}(\lambda).$$

By Lemma 3.10, $\dim E_{\alpha_i}(\lambda) = 1$. Hence $\dim E_\alpha(\lambda) = \dim E_{\alpha_1}(\lambda) + \cdots + E_{\alpha_k}(\lambda) = k$. This proves (a).

Now $m_\alpha(t)$ is the least common multiple of $m_{\alpha_1}(t), \ldots, m_{\alpha_k}(t)$ and (b) follows from this. Also (c) is simply saying that $\dim V = \dim K[\alpha]v_1 + \cdots + \dim K[\alpha]v_k$.

It thus just remains to show that (4) holds. We prove this by induction on s. If $s = 1$, then $\alpha = \lambda\mathrm{id}$. Pick any basis $v_1, \ldots, v_r$ for V and $V = Kv_1 \oplus \cdots \oplus Kv_r = K[\alpha]v_1 \oplus \cdots \oplus K[\alpha]v_r$. This deals with the induction basis.

Now suppose that $s \geq 2$ and that the claim holds for smaller values of s. Now consider the subspace $W = (\alpha - \lambda\mathrm{id})V$. As $\alpha - \lambda\mathrm{id}$ commutes with α we know from Lemma 3.4 that W is α-invariant. Now

$$(\alpha - \lambda\mathrm{id})^{s-1}w = 0$$

for all $w \in W$ and the minimal polynomial of $\alpha|_W$ is $(t - \lambda)^{s-1}$. By the induction hypothesis, there exist non-zero $(\alpha - \lambda\mathrm{id})v_1, \ldots, (\alpha - \lambda\mathrm{id})v_e \in W$ such that

$$W = (\alpha - \lambda\mathrm{id})V = K[\alpha](\alpha - \lambda\mathrm{id})v_1 \oplus \cdots \oplus K[\alpha](\alpha - \lambda\mathrm{id})v_e.$$

Let β_i be the restriction of α on $K[\alpha]v_i$. We know then from Lemma 3.10 that $E_{\beta_i}(\lambda)$ has dimension 1 and that there is a basis vector $w_i = (\alpha - \lambda\mathrm{id})^{e_i}v_i$ for $E_{\beta_i}(\lambda)$ where $e_i \geq 1$. Notice that it follows that $w_i \in K[\alpha](\alpha - \lambda\mathrm{id})v_i$. Thus $(w_1, \ldots, w_e)$ is a basis for $E_{\alpha|_W}(\lambda)$. Extend this to a bases $(w_1, \ldots, w_e, v_{e+1}, \ldots, v_{e+f})$ for $E_\alpha(\lambda)$. I claim that

$$V = K[\alpha]v_1 \oplus \cdots \oplus K[\alpha]v_e \oplus K[\alpha]v_{e+1} \oplus \cdots \oplus K[\alpha]v_{e+f}.$$

Notice that, as $v_{e+1}, \ldots, v_{e+f}$ are eigenvectors, this is the same as saying that $V = K[\alpha]v_1 \oplus \cdots \oplus K[\alpha]v_e \oplus (Kv_{e+1} \oplus \cdots \oplus Kv_{e+f})$.

First we show that $V = K[\alpha]v_1 + \cdots + K[\alpha]v_e + Kv_{e+1} + \cdots + Kv_{e+f}$. Take any $v \in V$ then $(\alpha - \lambda\mathrm{id})v \in W$ and thus

$$(\alpha - \lambda\mathrm{id})v = p_1(\alpha)(\alpha - \lambda\mathrm{id})v_1 + \cdots + p_e(\alpha)(\alpha - \lambda\mathrm{id})v_e$$

for some $p_1, \ldots, p_e \in K[t]$. It follows that

$$(\alpha - \lambda \mathrm{id})(v - (p_1(\alpha)v_1 + \cdots + p_e(\alpha)v_e)) = 0$$

and thus $v - (p_1(\alpha)v_1 + \cdots + p_e(\alpha)v_e) \in E_\alpha(\lambda) \subseteq W + Kv_{e+1} + \cdots + Kv_{e+f}$. Hence $v = p_1(\alpha)v_1 + \cdots + p_k(\alpha)v_k + (v - (p_1(\alpha)v_1 + \cdots + p_k(\alpha)v_k)) \in K[\alpha]v_1 + \cdots + K[\alpha]v_e + Kv_{e+1} + \cdots + Kv_{e+f}$.

Next we show that the sum is direct. Suppose

$$0 = p_1(\alpha)v_1 + \cdots + p_e(\alpha)v_e + a_{e+1}v_{e+1} + \cdots + a_{e+f}v_{e+f}.$$

Applying $\alpha - \lambda \mathrm{id}$ to both sides, gives

$$0 = p_1(\alpha)(\alpha - \lambda \mathrm{id})v_1 + \cdots + p_e(\alpha)(\alpha - \lambda \mathrm{id})v_e.$$

Since W was a direct sum it follows that $(\alpha - \lambda \mathrm{id})p_i(\alpha)v_i = 0$ for $i = 1, \ldots, e$. So $p_i(\alpha)v_i$ is an eigenvector belonging to $K[\alpha]v_i$ and thus a multiple of w_i. Since $w_1, \ldots, w_e$ are linearly independent, it follows that $p_i(\alpha)v_i = 0$ for $i = 1, \ldots, e$. Hence

$$0 = a_{e+1}v_{e+1} + \cdots + a_{e+f}v_{e+f}$$

and as $v_{e+1}, \ldots, v_{e+f}$ are linearly independent, it follows that $a_{e+1} = \ldots = a_{e+f} = 0$. This finishes the proof. □

Remarks (1) The matrix A in Theorem 3.11 is called a *Jordan Normal Form* for α often denoted $\mathrm{JNF}(\alpha)$. One can show that the Jordan blocks in $\mathrm{JNF}(\alpha)$ are unique up to order.

(2) This can be generalised. If $\Delta_\alpha(t) = (\lambda_1 - t)^{r_1} \cdots (\lambda_m - t)^{r_m}$ and

$$V = G_\alpha(\lambda_1) \oplus G_\alpha(\lambda_2) \oplus \cdots \oplus G_\alpha(\lambda_m)$$

with the corresponding decomposition of α

$$\alpha = \alpha_1 \oplus \cdots \oplus \alpha_m,$$

then we let $\mathrm{JNF}(\alpha) = \mathrm{JNF}(\alpha_1) \oplus \cdots \oplus \mathrm{JNF}(\alpha_m)$.

Example. Suppose that $\alpha : V \to V$ is a linear map where $m_\alpha(t) = (t-5)^2$ and $\Delta_\alpha(t) = (t-5)^4$. What are the possible JNF's for α.

Solution. As the degree of $m_\alpha(t)$ is 2, we must have at least one largest block $J(5,2)$ and to complete we need to add two more dimensions. So the possibilities are

$$J(5,2) \oplus J(5,2), \quad J(5,2) \oplus J(5,1) \oplus J(5,1).$$

If we furthermore know that $\mathrm{gm}(5) = 3$ then we must have three blocks and thus only the second possibility applies.

The problem of finding the basis that gives us $\mathrm{JNF}(\alpha)$ is something we will not get

much into here. There is a example on sheet 10 with more than one eigenvalue. Here we give on example where there is exactly one eigenvalue.

Example. Consider the linear operator $\alpha : \mathbb{C}^3 \to \mathbb{C}^3,\ v \mapsto Av$ where

$$A = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 0 & 3/2 & 1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix}.$$

Step 1. We find Δ_α, m_α and $\mathrm{JNF}(\alpha)$.

The characteristic polynomial is $(1-t)[(3/2-t)(1/2-t)+1/4) = (1-t)(1-2t+t^2) = (1-t)^3$. As the matrix A is not the unit matrix the minimal polynomial is either $(t-1)^2$ or $(t-1)^3$. Let us first check if it is the first one. We have

$$(A-I)^2 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & -1/2 & -1/2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & -1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

So the minimal polynomial is $(t-1)^2$. It follows that there is a Jordan block $J(1,2)$ of degree 2 and as the algebraic multiplicity is 3 we have another Jordan block $J(1,1)$ of degree 1. Hence $\mathrm{JNF}(\alpha) = J(1,2) \oplus J(1,1)$.

Step 2. Find the basis for V that gives us $\mathrm{JNF}(\alpha)$.

We look for vectors v and w such that

$$V = K[\alpha]v \oplus K[\alpha]w.$$

The first summand should correspond to $J(1,2)$ and thus be of dimension 2 whereas the 2nd one corresponds to $J(1,1)$ and should therefore be of dimension 1. We therefore need

$$K[\alpha]v = K(A-I)v + Kv$$

where $(A-I)v \neq 0$ and

$$K[\alpha]w = Kw$$

where $(A-I)w = 0$ or equivalently $Aw = w$.

We pick any vector v that is not an eigenvector. For example $v = (0,0,2)^T$. We have $(A-I)v = (1,1,-1)^T$. We need an eigenvector w that is linearly independent to v and $(A-I)v$. Here we can take $w = (1,0,0)^T$. From Lemma 3.10 we know that the matrix with repect to the basis $v_1 = (1,1,-1)^T, v_2 = (0,0,2)^T, v_3 = (1,0,0)^T$ is $J(1,2) \oplus J(1,1)$.