MATH 42041/62041: NONCOMMUTATIVE ALGEBRA

UNIVERSITY OF MANCHESTER, AUTUMN 2018

NOTES BY TOBY STAFFORD, MODIFIED BY MIKE PREST

Contents

0. Introduction.

Noncommutative algebra is a very rich subject with a great many different types of rings, so here I will give an informal introduction to a few of them, and use these examples to illustrate the main ideas of the course.

Among commutative rings one has the ring of integers, fields like the rationals $\mathbb{Q}$ the reals $\mathbb{R}$ and the complexes $\mathbb{C}$, and (factor rings of) polynomial rings like $\mathbb{C}[x,y]$ or $\mathbb{Z}[x,y]$. Examples of noncommutative rings include the full ring of $n \times n$ matrices $M_n(k)$ over a field k; these have no ideals. (See the next chapter for the formal definition, but in brief an *ideal* is an abelian subgroup closed under left and right multiplication by elements of the ring. We will also want to consider *left ideal*s - abelian subgroups closed under left multiplication by elements of the ring.)

Here is an idea of the proof of the above statement in the special case of $M_2(k)$. So, if P is a nonzero ideal, pick a nonzero element $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in P. For simplicity assume that $a \neq 0$ (the case $a = 0$ is left as an exercise). Then $P \ni \begin{pmatrix} a^{-1} & 0 \\ 0 & 0 \end{pmatrix} \cdot \alpha \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and hence

$$P \ni \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Adding these two elements shows that $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in P$, hence $P = R$. This implies that (0) is a maximal ideal of $R = M_2(k)$, so the factor ring $R = R/(0)$ contains zero-divisors. In contrast, recall the basic fact from commutative ring theory that an ideal I of a commutative ring C is maximal $\Leftrightarrow$ C/I is a field. So, we have to regard rings of matrices as analogues of fields!! In fact matrix rings are basic examples of *simple Artinian* rings, meaning rings with no nonzero proper ideals, and where any descending chain of left (or right) ideals $I_1 \supseteq I_2 \supseteq \cdots$ is eventually stationary.

One of the starting points of the subject is to understand such rings. And the answer is nice. Recall that a **division ring** is a ring in which every nonzero element has an inverse. Then the main result of Chapter 4 shows that the simple Artinian rings are just matrix rings over division rings. This generalises considerably; for example one gets nice structure theorems about arbitrary Artinian rings. This has some interesting applications, not only to the structure of various classes of rings, but also to properties of groups (though the group ring construction).

The proof of these results use the notion of a *left module*. This is something on which the ring acts by left multiplication, just as an $n \times n$ matrix ring acts on n-dimensional column vectors by left multiplication or the ring of integers acts by multiplication on an additive abelian group (how?). The theory of modules will be described in Chapter 2 after a preliminary chapter of basic concepts and examples.

In Chapter 5 we will classify finitely generated modules over PIDs (principal ideal domains). This includes the classification, which has been mentioned in Math 32001, of finitely generated abelian groups. More subtly it also shows that matrices, over an algebraically closed field, have Jordan canonical forms.

Another ring that will keep reappearing is the *(first) Weyl algebra* $A_1(\mathbb{C})$ or *ring of differential operators on the affine (complex) line*. We will define this more carefully in the next chapter but, in brief, these differential operators are represented by expressions of the form $\sum_{n=0}^{r} f_n(x)\partial^n$ where the coefficients of the operator $\partial = \frac{d}{dx}$ are from the polynomial ring $\mathbb{C}[x]$. This is the simplest example of the kind of ring that arises in the study of differential operators. We will see a little about the structure of division rings and the connection with group representation theory. All this uses the idea of a module - clearly the ring $A_1(\mathbb{C})$ of differential operators acts not just on $\mathbb{C}[x]$ but also on $\mathbb{C}(x)$ and on the group (indeed ring) of all power series $\mathbb{C}[[x]]$ in x - all these are modules over the ring $A_1(\mathbb{C})$. More subtle is that (solutions of) differential equations also correspond to modules (see e.g. S.C. Coutinho, A Primer of Algebraic D-Modules).

0.1. **The quaternions.** Let me finish by giving explicitly one basic example of a noncommutative ring - the quaternions. As you may know (or perhaps can see why), the field of complex numbers is the unique finite-dimensional field extension of the reals. However, if you allow the extension "field" to be noncommutative you get one extra example—the quaternions $\mathbb{H}$ (and if you also drop associativity you get another called the octonions, but that is another course!) The quaternions can be described as the ring that is a 4-dimensional

real vector space, with basis $1, i, j, k$ and multiplication defined by

$$\begin{aligned} i^2 = j^2 = k^2 &= -1 \\ ij &= -ji = k \\ jk &= -kj = i \\ ki &= -ik = j. \end{aligned} \tag{0.1}$$

This definition is not brilliant, since one has to *prove* that it is a ring and also that it is 4-dimensional (it must have dimension at most 4 but conceivably the relations could force some collapsing of dimension); in particular one has to prove that multiplication is associative. To do that directly would be a rather tedious exercise. Fortunately there is a clever, lazy way to check this - we can define the quaternions in a way that makes such properties obvious. This we do by defining $\mathbb{H}$ as a subring of $M_2(\mathbb{C})$. Inside that ring take the matrices

$$I = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \text{and} \qquad K = IJ = -JI = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

Now, inside $M_2(\mathbb{C})$ take the *real* subspace spanned by $1, I, J, K$. It is an exercise of moments to check that this really is a ring and our basis elements satisfy the rules (0.1) of the quaternions[1].

Exercise 0.1. Inside $M_2(\mathbb{R})$ show that the identity matrix 1 and $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ generate a ring isomorphic to $\mathbb{C}$. Inside $M_4(R)$ find 4 elements that generate a copy of $\mathbb{H}$.

Can you find such a ring inside $M_2(\mathbb{R})$ that is isomorphic to $\mathbb{H}$?

Just as the complex numbers are closely related to rotations of the plane, so the quaternions are very useful for describing and manipulating 3D rotations—look at the book "On quaternions..." by John Conway in the library. They have many other uses in physics and even in computer animation (have a look on the web for material about this).

[1]It seems that we made two definitions of the quaternions - do they give the same ring? The answer is yes, but can you see how to argue that?

1. **Preliminaries and examples.**

This section introduces basic examples (matrix rings (full, triangular, subrings of these); quaternions; polynomial rings; group rings; first Weyl algebra) and definitions (division ring; ring homomorphism; k-algebra; (right/left/2-sided ideal); simple rings; factor rings; generation of and operations on ideals) and results (first isomorphism theorem for rings; Zorn's Lemma).

Definition 1.1. *A* **ring** R *is a set on which two binary operations are defined, addition (denoted by* $+$*) and multiplication (denoted by* $.$ *or* $\times$ *or juxtaposition), such that*

(i) R *forms an abelian group under addition,*

(ii) multiplication is associative; $(rs)t = r(st)$ *for all* $r, s, t \in R$*,*

(iii) the operations satisfy the distributive laws, i.e.

$$a(b+c) = ab + ac \qquad \text{and} \qquad (a+b)c = ac + bc$$

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

(iv) R *contains an identity element* 1 *for multiplication: that is,* $1.r = r.1 = r$ *for all* $r \in R$*.*

Comments: Part (i) implies that $R \neq \varnothing$.

If we're checking that a subset R of a given ring S is itself a ring using the addition and multiplication of S, then saying that the operations on R are "defined" means that $a, b \in R \implies a+b,\ ab \in R$ (i.e. R is *closed* in S under addition and multiplication).

Strictly speaking, a ring is a triple $(R, +, .)$. But it is convenient to simply speak of "the ring R".

The identity element of a ring R is unique. (Why? Hint: consider two candidates and multiply them together.)

Some people do not make the existence of an identity element part of the axioms—and there are certainly situations where it would consider rings without a 1.

When more than one ring is under discussion, the identity of a ring R is sometimes denoted by 1_R.

Definition 1.2. *A ring* R *such that* $ab = ba$ *for all* $a, b \in R$ *is said to be* **commutative**. *The* **trivial ring** *or* **zero ring** *is the set* $R = \{0\}$ *with the only possible* $+$ *and* $\times$*. In this ring* $0 = 1$ *(conversely, this equation means that we have the trivial ring).*

We will almost always deal with nonzero rings—note that these are precisely the rings S for which $1_S \neq 0_S$. Similarly, we shall mainly be concerned with **noncommutative** rings, i.e. rings that are not necessarily commutative. (It's convenient to allow the general term "noncommutative rings" to include "commutative rings" but, if you say that a *particular* ring is noncommutative you probably mean that it's not commutative!)

Examples 1.3. (1) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}_n$, where n is an integer greater than 1, all are rings.

(2) Let R be a ring. The set $R[X]$ of all polynomials in the indeterminate X with coefficients from R forms a ring under the usual operations of addition and multiplication. Here X is required to commute with every element of R: $rX = Xr$ for all $r \in R$. (If you make no such assumption on X then the resulting ring, usually denoted $R\langle X\rangle$ is *much* more complicated.)

(3) Let R be a ring and n a positive integer. The set $M_n(R)$ of all $n \times n$ matrices with entries from R forms a ring under the usual matrix operations.

(4) A **domain** D is a nonzero ring such that, for all $a, b \in D$,

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

(In the commutative case one tends to say "integral domain" rather than "domain" but I will not be consistent about this.)

(5) A **division ring** (sometimes called a **skew field**) D is a nonzero ring such that, for all nonzero $a \in D$, there exists $b \in D$ such that $ab = ba = 1$. In other words, $D \setminus \{0\}$, the set of nonzero elements of D, forms a group under multiplication. So the element b is uniquely determined by a. It is denoted by a^{-1}.

$\mathbb{Z}$ is an integral domain, but not a division ring. On the other hand the quaternions form a division ring (and hence a domain) that is not commutative.

(6) A **field** is a commutative division ring. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Z}_p$ with p a prime are fields, but a zero ring is not a field.

Definition 1.4. *Let R, S be rings. A* **ring homomorphism** *from R to S is a map $\theta : R \to S$ such that $\theta(r_1 + r_2) = \theta(r_1) + \theta(r_2)$ and $\theta(r_1 r_2) = \theta(r_1)\theta(r_2)$ for all $r_i \in R$ and such that $\theta(1_R) = 1_S$. If we drop the last requirement, then we use the term* **non-unital ring homomorphism** *- these arise when we consider direct products of rings.*

An injective (1-1) homomorphism is also referred to as a **monomorphism**[2]*. An* **isomorphism** *means a bijective homomorphism. If $R = S$ then θ is called a (***ring***) endomorphism** *of R. (In fact, in this course it will be the endomorphisms of R regarded as a module over itself that come up.)*

Definition 1.5. *A* **subring** *S of a ring R is a subset of R which forms a ring under the operations inherited from R.*

Lemma 1.6. *A subset S of a ring R is a subring if and only if:*

(i) $a, b \in S \implies a - b \in S$ *(S is closed under subtraction),*

(ii) $a, b \in S \implies ab \in S$ *(S is closed under multiplication),*

(iii) $1_R \in S$ *(and hence R and S have the same identity element).*

[2]The term "epimorphism" means something different (weaker) in the context of rings.

Proof. Make sure that you can prove this. If you have not seen it before it is very similar to the analogous result about when a subset of a group is actually a subgroup. □

Remark: In many books you will find that a homomorphism does not need to send 1 to 1 and that a subring can have a different 1. This is really personal preference and/or context, but it is important to be consistent. Note that with the definition given above, if one ring R sits inside another one S then the inclusion $R \hookrightarrow S$ is a homomorphism $\Leftrightarrow$ R is a subring of S. *Prove this!*

However, for example, given a direct product of rings $R = R_1 \otimes R_2$ then, with this definition, the R_i are not subrings of R (they are "non-unital" subrings). But life is never perfect.

Examples 1.7. (1) $\mathbb{Z}$ is a subring of $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.

(2) $\mathbb{Q}$ is a subring of $\mathbb{R}$.

(3) $\mathbb{Z}[\sqrt{2}]$ is the set obtained by substituting $\sqrt{2}$ for X in $\mathbb{Z}[X]$. In fact

$$\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\}.$$

This is a subring of $\mathbb{R}$.

(4) Similarly,

$$\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$$

(here $i^2 = -1$). This is called the set of **Gaussian integers** and is a subring of $\mathbb{C}$.

The last two examples above are important in Algebraic Number Theory.

(5) Let R be a ring and n a positive integer. Let $U_n(R)$ be the set of **upper triangular matrices** over R, i.e. the subset of $M_n(R)$ consisting of all matrices which have 0 as every entry below the main diagonal, and $L_n(R)$ the set of **lower triangular matrices** over R. Then $U_n(R)$ and $L_n(R)$ are both subrings of $M_n(R)$. For example,

$$U_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} : a, b, c \in \mathbb{Z} \right\}.$$

(6) One can also take the "top left corner" of $M_2(\mathbb{C})$:

$$R = \left\{ \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} : a \in \mathbb{C} \right\} \subset M_2(\mathbb{C}).$$

This is easily seen to be a ring with identity element $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. However, it is not a (unital) subring of $M_2(\mathbb{C})$.

In all these examples, one could prove that the given set is indeed a ring by explicitly checking the axioms. However this is a tedious exercise—so try to avoid it. In each case one can indeed avoid it by using Lemma 1.6. For example, in parts (3) or (4), use the lemma to check quickly that $\mathbb{Z}[\sqrt{2}]$, respectively $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. Then, of course, the lemma says it is actually a ring! In this way you avoid having to prove associativity and distributivity—which are usually very tedious. This was also the way we proved that the quaternions were a ring in the introduction.

(7) Let R be a ring. Then we can define $R[X,Y]$ as the set of polynomials in Y with coefficients from $R[X]$, i.e.

$$R[X,Y] = R[X][Y].$$

Alternatively,

$$R[X,Y] = R[Y][X].$$

That these are the same ring reflects the fact that a typical element $f \in R[X,Y]$,

$$f = \sum_{i=0}^{m} \sum_{j=0}^{n} a_{ij} X^i Y^j$$

for some nonnegative integers m, n, where each $a_{ij} \in R$. can be rewritten either as a polynomial in Y with coefficients from $R[X]$, or as a polynomial in X with coefficients from $R[Y]$.

Similarly, we define

$$R[X_1, X_2, \ldots, X_n] = R[X_1, X_2, \ldots, X_{n-1}][X_n]$$

for any integer $n > 1$.

(8) If R and S are rings then so is the Cartesian product $R \times S$ with addition $(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$ and multiplication $(r_1, s_1)(r_2, s_2) = (r_1 r_2, s_1 s_2)$ for all $r_j \in R$ and $s_k \in S$. The identity element is $(1_R, 1_S)$.

I will assume that you have seen Examples 7 and 8 before, so will not prove that they are rings. However, in both cases there is not any (obvious) over-ring containing the given ring, so one cannot easily use Lemma 1.6 to prove *that the given set is a ring. And in fact, the proof you have surely seen that $R[x]$ is a ring is pretty tedious.*

Definition 1.8. *Let K be a field. An* **algebra** *A over K is a set on which three operations are defined, addition, multiplication, and multiplication by scalars from K such that*

(i) A is a ring under addition and multiplication,

(ii) A is a vector space over K under addition and multiplication by scalars,

(iii) for all $\alpha \in K$, $x, y \in A$,

$$(\alpha x)y = x(\alpha y) = \alpha(xy).$$

It is said to be **finite-dimensional** *(***f.d.***) if it is finite-dimensional as a vector space.*

The **centre** *of a ring R is the set*

$$Z(R) = \{z \in R : zr = rz \text{ for all } r \in R\}$$

of all elements of R that commute with all the other elements of R.

Exercise: If A is a nonzero K-algebra, check that $\{\alpha 1_A : \alpha \in K\}$ is a subring of A isomorphic to K. Since we have

$$\alpha x = (\alpha 1_A)x,$$

for $x \in A$, $\alpha \in K$, we often identify this subring with K and write α instead of $\alpha 1_A$. (Note that $1_K 1_A = 1_A$.)

Generalising this slightly one sees that R is a K-algebra if and only if K (or rather a copy of K) sits inside the centre of R.

Many rings are K-algebras for some field K, for example $\mathbb{H}$ is a $\mathbb{R}$-algebra. However it is *not* a $\mathbb{C}$-algebra as $\mathbb{C}$ is not central in $\mathbb{H}$.

Examples 1.9. (1) For any field K, $K[X]$ is an algebra over K. It is not finite-dimensional.

(2) For any field K and $n \in \mathbb{N}$, $M_n(K)$ is an algebra over K. It is finite-dimensional. The K-vector space dimension is n^2. Moreover, in this case we can identify $M_n(K)$ as the set of all K-linear transformations from $V = K^{(n)}$ to itself, at least once we have chosen a basis for V as a vector space over K. In notation introduced in the next chapter, this will be written $M_n(K) = \mathrm{End}_K(V)$ where the latter denotes the ring of endomorphisms of V regarded as a module over K. You should check that the centre of $M_n(K)$ is just the set of scalar matrices.

(3) That generalizes. Suppose, now that V is an infinite dimensional K-vector space, and again write $\mathrm{End}_K(V)$ for the set of all K-linear transformations from V to itself. *A basic fact from linear algebra is that this really is a ring.*

For notational convenience I will assume that V is countable dimensional, say with basis $\{v_1, \ldots, \}$. If we write elements of V as column vectors, and write linear transformations as acting from the left, then the set of all linear transformations can be naturally identified with the set of all **column finite matrices.** These are the infinite matrices

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & \cdots & \\ \vdots & & \ddots & \end{pmatrix}$$

where in each *column* there are only finitely many nonzero entries. The fact that we can write endomorphisms as matrices is presumably familiar to you. The fact that they are column finite comes from the fact that if one multiplies out

$$\theta \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

then this gives gives the first column of θ thought of as a matrix. On the other hand it must give a column vector with only finitely many entries. So only finitely many a_{j1} can be nonzero. The same argument works for the other columns.

Definition 1.10. The Weyl Algebra. *We begin with the polynomial ring $\mathbb{C}[x]$ (actually $\mathbb{C}$ can be replaced by any field of characteristic zero). We regard this as a $\mathbb{C}$-vector space, for which we have two obvious linear transformations. The first, written x, is multiplication by x, while the second, written ∂, is differentiation. We use $*$ to denote their operation in order to distinguish this from multiplication inside $\mathrm{End}_{\mathbb{C}}(\mathbb{C}[x])$; so their actions on a function $f(x) \in \mathbb{C}[x]$ are given by*

$$x * f(x) = xf(x) \qquad \text{and} \qquad \partial * f(x) = \frac{df}{dx}.$$

The first Weyl algebra $A_1 = A_1(\mathbb{C})$ *is defined to be the set of all linear differential operators with polynomial coefficients; that is, the set*

$$A_1 = \left\{ \sum_{i=0}^{n} f_i(x)\partial^i : f_i(x) \in \mathbb{C}[x],\ n \in \mathbb{N} \right\}$$

of $\mathbb{C}$-vector space endomorphisms of $\mathbb{C}[x]$. Note that, as operators

$$(\partial \cdot x) * f(x) = \frac{d}{dx}(xf) = f + x\frac{df}{dx} = (1 + x\partial) * f.$$

So, as operators, we have

$$\partial \cdot x = 1 + x \cdot \partial. \tag{1.1}$$

Exercise: Write out the operators x and ∂ as column finite matrices.

Intuitively, if we multiply out two elements of A_1, say $(\sum f_i(x)\partial^i)(\sum g_j(x)\partial^j)$, then we can use rule (1.1) to move the ∂'s to the right of the x's and hence show that A_1 is at least closed under multiplication. However, since the formulae are useful anyway, let's do it carefully. We start with:

Lemma 1.11. *For all $n, m \geq 0$ we have the following formulae relating operators:*

(1) $\partial^n \cdot x = n\partial^{n-1} + x \cdot \partial^n$.

(2) $\partial \cdot x^m = mx^{m-1} + x^m \cdot \partial$.

(3) *(Leibniz's Rule)*

$$\partial^n x^m = x^m \partial^n + \lambda_{m-1} x^{m-1}\partial^{n-1} + \lambda_{m-2} x^{m-2}\partial^{n-2} + \cdots$$

The scalars λ_j can easily be computed—they are appropriate binomial coefficients. The $\cdots$ means you keep getting smaller and smaller exponents until one of $(m-j)$ or $(n-j)$ hits zero.

(4) *For $f(x) \in \mathbb{C}[x]$, we have (again multiplying in the ring of differential operators) $\partial^m f = f\partial^m + \sum_{i=0}^{m-1} g_i \partial^i$ where $g_i = g_i(x) \in \mathbb{C}[x]$ is such that $\deg g_i < \deg f$ for all i.*

Proof. (1) We prove this by induction on n, using 1.1 both as the base case and for the induction step.

(2) This is left as an exercise - it is also a special case of (3).

(3) We prove this by induction on m, with the case $m = 1$ being part (1). So, if we take the given equation and multiply on the right by x we get

$$\partial^n x^{m+1} = x^m \partial^n x + \sum_{j=1}^{\min(n,m)} \lambda_{m-j} x^{m-j} \partial^{n-j} x$$

In each term we can use part (1) to "move" the ∂^j through the final x; this gives

$$\partial^n x^{m+1} = x^m (x\partial^n + n\partial^{n-1}) + \sum \lambda_{m-j} x^{m-j} \Big((n-j)\partial^{n-j-1} + x\partial^{n-j} \Big).$$

Now collect terms. Of course, one can use the same argument to work out formulæ for the scalars λ_j.

(4) Use (3) and collect terms. □

Corollary 1.12. *The Weyl algebra A_1 is indeed a subring of $\mathrm{End}_{\mathbb{C}}(\mathbb{C}[x])$. In particular it is a ring. It is also a $\mathbb{C}$-algebra.*

Proof. We first prove that A_1 is closed under multiplication. First, by definition, A_1 is spanned as a $\mathbb{C}$-vector space by the $x^n\partial^m$ for $n, m \geq 0$.

So we need to prove that any product $\left(\sum \lambda_{ij} x^i \partial^j\right)\left(\sum \mu_{k\ell} x^k \partial^\ell\right)$, where the λ_{ij} and the $\mu_{k\ell}$ are scalars from $\mathbb{C}$, can be rewritten as $(\sum \nu_{uv} x^u \partial^v)$ for some scalars ν_{uv}. By distributivity it suffices to prove this for $\left(x^i\partial^j\right)\left(x^k\partial^\ell\right)$. But as linear transformations do satisfy associativity, this equals $x^i\left(\partial^j x^k\right)\partial^\ell$. By Leibnitz's rule this equals $(\sum \nu_{uv} x^u \partial^v)$ for some scalars ν_{uv}.

Thus, A_1 is closed under multiplication. From its definition A_1 is a vector subspace of $\mathrm{End}_{\mathbb{C}}(\mathbb{C}[x])$ (that is, it is closed under addition and multiplication by scalars from $\mathbb{C}$). Similarly by definition $1 = x^0\partial^0 \in A_1$. Thus, A_1 is a subring of $\mathrm{End}_{\mathbb{C}}(\mathbb{C}[x])$ by Lemma 1.6. It is clearly a $\mathbb{C}$-algebra. □

Corollary 1.13. *As a $\mathbb{C}$-vector space, A_1 has basis $\{x^i\partial^j : 0 \leq i, j < \infty\}$.*

Proof. By construction $\{x^i\partial^j : 0 \leq i, j < \infty\}$ spans A_1. But if these elements are not independent then there exists some $\theta = \sum_{i=n}^{i=m} f_i(x)\partial^i$ with $f_n(x) \neq 0$ such that θ is zero as an element of A_1. However, as $\partial^j x^n = 0$ when $j > n$ we find that $0 = \theta * x^n = f_n(x)\frac{d^n(x^n)}{dx^n} = (n!)f_n(x)$. This is absurd. □

Corollary 1.14. *$A_1(\mathbb{C})$ is a domain.*

Proof. Suppose that α and β are two nonzero elements of A_1 and write them out as $\alpha = \sum_{i=0}^{n} f_i(x)\partial^i$ and $\beta = \sum_{j=0}^{m} g_j(x)\partial^j$ where n, m are chosen such that $f_n \neq 0 \neq g_m$. Then Leibniz's rule says that $\alpha\beta = f_n g_m \partial^{n+m} +$ terms of lower order in ∂. This leading term $f_n g_m \partial^{n+m}$ is nonzero - use Corollary 1.13 - and then the same corollary implies that the whole expression for $\alpha\beta$ is nonzero. □

Exercise: Prove that the centre $Z(A_1(\mathbb{C})) = \mathbb{C}$.

Remark 1.15. (1) These corollaries imply that we can alternatively define the Weyl algebra abstractly, as the $\mathbb{C}$-algebra generated by two elements x, ∂ which satisfy the relation (1.1). For, the corollaries imply that the obvious ring homomorphism from the Weyl algebra defined in this abstract way to the Weyl algebra defined as a ring of operators is injective and surjective, hence an isomorphism.

(2) In the definition one can replace the complex field $\mathbb{C}$ by any field of characteristic 0 (the action of ∂ on polynomials being formal differentiation), with no change in the arguments or results.

If, however, a field of characteristic $p > 0$ is used, then, although the last three results still are true, the proofs become harder: notice that in this case $\partial^p * x^p = p! = 0$.

(3) Even though the ring looks fairly easy there are still some things we do not know—and which will make you famous if you solve them. For example, no-one knows whether there exists a $\mathbb{C}$-algebra endomorphism of A_1 that is not an automorphism! Here a $\mathbb{C}$-algebra endomorphism means a ring endomorphism θ that satisfies $\theta(\lambda) = \lambda$ for all scalars $\lambda \in \mathbb{C}$.

Example 1.16. (Group rings) *Let G be a finite (multiplicative) group and K a field. Suppose that $G = \{g_1, g_2, \ldots, g_n\}$, where $g_1 = 1_G$ and $n = |G|$. The* **group algebra** *KG is the n-dimensional vector space over K with $\{g_1, g_2, \ldots, g_n\}$ as a basis and multiplication defined as follows. Let*

$$x = \sum_{i=1}^{n} a_i g_i, \qquad y = \sum_{j=1}^{n} b_j g_j,$$

where $a_i, b_j \in K$ $(i, j = 1, 2, \ldots, n)$. Then

$$xy = \sum_{i,j=1}^{n} (a_i b_j)(g_i g_j) \ = \sum_{k=1}^{n} c_k g_k,$$

where

$$c_k = \sum_{g_i g_j = g_k} a_i b_j;$$

in other words c_k equals the sum of the $a_i b_j$ for all i, j for which $g_i g_j = g_k$.

Comments (1) Don't try to interpret the product ag, where $a \in K$, $g \in G$. It is purely formal.

(2) We leave as a simple exercise the fact that KG is a ring—the fact that it is associative, for example, is something one has to multiply out explicitly so is a bit tedious. Or, can you find a cunning way of making KG into a subring of some matrix ring that makes this obvious?

(3) We can extend the definition of KG to infinite groups. In this case, elements have the form

$$\sum_{g \in G} a_g g,$$

where all but a finite number of the coefficients $a_g \in K$ are zero.

Example 1.17. *Let us work out what is the group ring in a (very) special case.*

First, we take $G = C_2 = \langle \sigma : \sigma^2 = 1 \rangle$. *Thus, as vector spaces,* $KG = K \cdot 1 + K\sigma$ *with multiplication*

$$(a + b\sigma)(c + d\sigma) = (ac + bd) + (ac + bd)\sigma$$

where we have used the fact that, as $\sigma^2 = 1$ *we also have* $(a\sigma)(b\sigma) = ab\sigma^2 = ab$. *So,* KC_2 *is a 2-dimensional* K*-vector space that is also a commutative ring. There are not many such examples; indeed just* $K \oplus K$ *and* $K[x]/(x^2)$ *if we assume that* K *is algebraically closed. (We do not need it, but can you prove this assertion?)*

So which one is KG*? In fact it depends upon the characteristic of* K*:*

(1) *If char* $K = 0$ *(or indeed if char* $K \neq 2$*) then* $KC_2 \cong K \oplus K$ *as rings.*

(2) *If char* $K = 2$ *then* $KC_2 \cong K[x]/(x^2)$.

First Proof: Since we just have a two-dimensional vector space, we can write down the multiplication tables for the two sides of the isomorphism (or to be more precise, we can write down the multiplication table for carefully chosen bases of the two sides) and it will then be obvious that we have the required isomorphism.

In case (1) set $e_1 = \frac{1}{2}(1 + \sigma)$ and $e_2 = 1 - e_1 = \frac{1}{2}(1 - \sigma)$. Then $e_i^2 = e_i$ from which it follows that the two vector spaces e_iK are subrings of KC_2. (Sorry, this is a case where one does get lazy—they are not subrings as defined before since they do not contain 1 - you could say that they are *non-unital* subrings; each e_iK is a ring under the multiplication induced from that in KC_2.) Also, as vector spaces certainly $KC_2 = Ke_1 \oplus Ke_2$ simply because both are 2-dimensional vector spaces. But we need to check that as rings we get $KC_2 = Ke_1 \oplus Ke_2$. So write down the multiplication table:

$\times$	e_1	e_2
e_1	e_1	0
e_2	0	e_2

Of course, this is the same as the multiplication table for $K \oplus K$ with basis $e_1' = (1, 0)$ and $e_2' = (0, 1)$. Thus we get a ring isomorphism by sending $e_i \mapsto e_i'$.

Now suppose that *char* $K = 2$. Set $r = 1 + \sigma$ and notice that $KC_2 = K \cdot 1 + K \cdot r$, but now our multiplication rule is $r^2 = (1 + \sigma)^2 = 1 + 2\sigma + 1 = 0$; thus our multiplication table is in this case:

$\times$	1	r
1	1	r
r	r	0

which is exactly the same as the multiplication table for $K[x]/(x^2)$ (with $r = x$) and so again we get the desired isomorphism. □

Clearly this sort of argument is useless when the rings get bigger, since it is impractical to write out multiplication tables for big-dimensional rings. But we can give more elegant proofs if we use a bit of the theory about ideals and factor rings. Recall the definitions.

Definition 1.18. *Let R be a ring. An* **ideal** *I of R is a subset of R such that*

(i) I is an abelian subgroup of R under addition (thus in particular $0_R \in I$).

(ii) $a \in I$, $r \in R \implies ar \in I$,

(iii) $a \in I$, $r \in R \implies ra \in I$.

If just (i) and (ii) hold, then I is called a **right ideal of** *R while if just (i) and (iii) hold, then I is called a* **left ideal of** *R.*

Note that, since $-1 \in R$, to check condition (i) in the presence of conditon (iii) (for either right or left multiplication) it is enough to check that I is closed under addition and contains 0_R. s

Notation. We sometimes write $I \trianglelefteqslant R$ to indicate that I is an ideal of R. We write $I \underset{\text{r}}{\trianglelefteqslant} R$, $I \underset{\text{l}}{\trianglelefteqslant} R$ to indicate that I is a right ideal, left ideal of R, respectively. Note that

$$I \trianglelefteqslant R \iff I \underset{\text{r}}{\trianglelefteqslant} R \text{ and } I \underset{\text{l}}{\trianglelefteqslant} R.$$

Of course, in a commutative ring, ideals, right ideals and left ideals coincide.

Examples (1) $\{0\}$ and R are ideals of any ring R. Most ideals of R are not subrings of R since we insist that all rings have an identity.

(2) If a right or left ideal I contains 1, or contains any invertible element, then it must be the whole ring R.

(3) If $R = M_n(\mathbb{C})$ then a column (or several columns) of R forms a left ideal, while a row (or several rows) gives a right ideal. For example

$$I = \left\{ \begin{pmatrix} 0 & a_{12} & a_{13} & 0 & 0 & \cdots \\ 0 & a_{22} & a_{23} & 0 & \cdots & \\ 0 & a_{32} & a_{33} & 0 & \cdots & \\ 0 & \vdots & \vdots & & & \end{pmatrix} \right\}$$

is a left ideal of R. These are probably the best examples to use to picture what a left or right ideal in a ring really looks like.

(4) The kernel $Ker(\theta)$ of a ring homomorphism $\theta : R \to S$ is an ideal of R, while the image $Im(\theta)$ is a subring of S. (Note that the axiom that $\theta(1_R) = 1_S$ for homomorphisms is really the same as insisting that a subring A of a ring B has $1_A = 1_B$.)

As you have seen in earlier courses, we have:

Lemma 1.19. *(1) Let I be an ideal of a ring R and let $x \in R$. The* coset *of I represented by x is the set $[x + I]$, where*

$$[x + I] = \{x + a : a \in I\}.$$

Then R/I denotes the set of all cosets of I in R. Addition and multiplication of cosets are defined (consistently) by

$$(x+I)+(y+I)=(x+y)+I,$$
$$(x+I)(y+I)=xy+I$$

for all $x, y \in R$.

Under these operations of addition and multiplication, R/I forms a ring called the **factor ring** *of R by I.* □

It is often messy to work with factor rings, but you can usually make life easier for yourself by using the

Theorem 1.20. (First Isomorphism Theorem For Rings) *If $\theta : R \to S$ is a ring homomorphism then $Im(\theta) \cong R/Ker(\theta)$.*

Proof. You will have seen this before, but make sure you remember how to prove it! □

We return to Example 1.17 and see that it gets a little less *ad hoc* using these results. We begin with:

Lemma 1.21. *Suppose that $R \supset S$ are commutative rings and that $r \in R$. Then:*

(1) *There exists a ring homomorphism $\phi : S[x] \to R$ defined by $\phi(x) = r$ and $\phi(s) = s$ for all $s \in S$. Thus $\phi(\sum s_i x^i) = \sum s_i r^i$ for $s_i \in S$.*

(2) *If, in (1), there exists an inverse $r^{-1} \in R$ then there exists a ring homomorphism $\phi : S[x, x^{-1}] \to R$ defined by $\phi(x) = r$ and $\phi(s) = s$ for all $s \in S$.*

(3) *If each element $t \in R$ can be written as $t = \sum s_i r^i$ for some $s_i \in S$ then ϕ is surjective. Moreover $R \cong S[x]/Ker\,\theta$.*

Proof. (1) This is *the* basic property of polynomial rings. The rule $\phi(\sum s_i x^i) = \sum s_i r^i$ certainly defines a well-defined map from $S[x]$ to R. So, now check that it is a ring homomorphism—this is yet another easy exercise. (2) is left as an exercise.

(3) This is obvious from (1) and the First Isomorphism Theorem for rings. □

Now go back to Example 1.17, with the notation from there in the case when *char* $K = 2$; thus we took $r = 1 + \sigma$ and noticed that certainly $KC_2 = K.1 + K.r$. So by the lemma, $KC_2 \cong K[x]/I$ for some ideal I. However, as $r^2 = 1 + \sigma^2 + 2\sigma = 2 + 2\sigma = 0$ we see that $x^2 \in I$. Since $\dim_K K[x]/(x^2) = 2 = \dim_K KC_2 = \dim_K K[x]/I$, this implies that $I = (x^2)$, as required. □

Exercises: (1) Show that $KC_3 \cong K \oplus K \oplus K$ in characteristic zero but that $KC_3 \cong K[x]/(x^3)$ in characteristic 3.

(2) Much harder is to prove that for the symmetric group $G = S_3$ one has $\mathbb{C}S_3 \cong M_2(\mathbb{C}) \oplus \mathbb{C} \oplus \mathbb{C}$. (This becomes very easy in Chapter 4, which is maybe too long to wait.)

(3) If G is the infinite cyclic group generated by X prove that $KG = K[X, X^{-1}]$.

In contrast to Lemma 1.19, there is no sensible way to define a ring structure on R/I if I is only a left ideal of R. The problem is that if I is not a two-sided ideal then there exists $a \in I$ and $r \in R$ such that $ar \notin I$. Hence in the factor abelian group R/I we would have

$$0 \cdot r = [a + I]r = [ar + I] \neq 0,$$

which would be a bit upsetting.

Examples 1.22. (1) Let $a \in R$. Then $aR \underset{\text{r}}{\trianglelefteqslant} R$ and $Ra \underset{\text{l}}{\trianglelefteqslant} R$, where

$$aR = \{ax : x \in R\}, \qquad Ra = \{xa : x \in R\}.$$

Observe that $a \in aR$ since $a = a1$, and any right ideal that contains a must also contain aR. So aR is the smallest right ideal of R that contains a. It is called the **right ideal generated by** a. Similarly, Ra is the smallest left ideal that contains a. It is called the **left ideal generated by** a.

(2) Let $n \in \mathbb{N}$. Then $n\mathbb{Z} = \{nz : z \in \mathbb{Z}\} \trianglelefteqslant \mathbb{Z}$. If $n > 1$, then $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ since, for $a \in \mathbb{Z}$, we have $[a] = a + n\mathbb{Z}$.

(3) Suppose that $I \underset{\text{r}}{\trianglelefteqslant} R$, $J \underset{\text{r}}{\trianglelefteqslant} R$. Then we define

$$I + J = \{x + y : x \in I,\ y \in J\}.$$

Check that $I + J \underset{\text{r}}{\trianglelefteqslant} R$. So if $a, b \in R$ then $aR + bR \underset{\text{r}}{\trianglelefteqslant} R$, where $aR + bR = \{ax + by : x, y \in R\}$. Thus $aR + bR$ is the right ideal generated by a and b (the smallest right ideal that contains a and b).

(4) Suppose that I_1, I_2, I_3, ... is a (finite or infinite) collection of right ideals. Then we define

$$I_1 + I_2 + \cdots + I_n = \{x_1 + x_2 + \cdots + x_n : x_i \in I_i \ (i = 1, 2, \ldots, n)\}.$$

Once again

$$I_1 + I_2 + \cdots + I_n \underset{\text{r}}{\trianglelefteqslant} R.$$

Also

$$a_1R + a_2R + \cdots + a_nR$$

is the right ideal generated by $a_1, a_2, \ldots, a_n \in R$.

We also define

$$\sum_{i=1}^{\infty} I_i \qquad \text{or} \qquad \sum_{i \in \mathbb{N}} I_i$$

to be the set of elements of the form

$$\sum_{i=1}^{\infty} x_i$$

where, for each i, $x_i \in I_i$, and all but a finite number of the x_i are 0. In other words, we can express $x \in \sum_{i=1}^{\infty} I_i$ in the form

$$x = x_1 + x_2 + \cdots + x_m$$

for some positive integer m (depending on x), where $x_i \in I_i$ $(i = 1, 2, \ldots, m)$. When the range of i is clear, we just write $\sum_i I_i$.

For $x, y \in \sum_i I_i$, we can always find m such that

$$x = x_1 + x_2 + \cdots + x_m \qquad \text{and} \qquad y = y_1 + y_2 + \cdots + y_m$$

with $x_i, y_i \in I_i$ $(i = 1, 2, \ldots, m)$. But it is convenient to write $x = \sum_{i=1}^{\infty} x_i$.

In particular, for $a_1, a_2, a_3, \ldots \in R$, $\sum_i a_i R$ is the set of elements of the form $\sum_i a_i x_i$, where $x_1, x_2, x_3, \ldots \in R$ and all but a finite number of the x_i are 0.

We can generalise this further and consider $\sum_{\lambda \in \Lambda} I_\lambda$, where I_λ $(\lambda \in \Lambda)$ is any collection of right ideals. Any nonempty set Λ may be used as an index set. Of course, there are corresponding versions of (3) and (4) for left ideals.

Definition 1.23. *A ring R is* **simple** *if it has no 2-sided ideals other than R and 0.*

Examples 1.24. (a) If R is a commutative ring then R is simple $\Leftrightarrow$ R is a field.

Proof. $\Rightarrow$ Given $0 \neq r \in R$, then Rr is a nonzero ideal of R and hence $Rr = R$. Thus $1 = ar$ for some $a \in R$ and R is a field. The other direction is obvious. □

(b) The rings $M_n(\mathbb{C})$, or $M_n(k)$ for any n and any field k are simple. (We did the case of $M_2(K)$ in the introduction. The general case is an exercise.)

(c) $A_1(\mathbb{C})$ is simple.

Proof. Let I be a nonzero ideal of $A_1 = A_1(\mathbb{C})$. Pick a nonzero element $\alpha \in I$ and, by Corollary 1.13 write $\alpha = \sum_{i=0}^{n} f_i(x)\partial^i$, where $f_i \in \mathbb{C}[x]$ and $f_n \neq 0$. Choose such an $\alpha \in I$ with n as small as possible. If $n > 0$ then, using Lemma 1.11, we see that

$$\begin{aligned} I \ \ni \ -x\alpha + \alpha x \ &= \ \sum_{i=0}^{n} f_i(x)(-x\partial^i + \partial^i x) \\ &= \sum_{i=0}^{n} f_i(x)(i\partial^{i-1}) \ = \ n f_n(x)\partial^{n-1} + \cdots \ \neq \ 0, \end{aligned}$$

where the fact that it is nonzero comes from Corollary 1.13. This contradicts the minimality of n and forces $n = 0$. Thus we can write $\alpha = \sum_{j=0}^{m} \lambda_i x^i$, for some $\lambda_i \in \mathbb{C}$ and we can pick m so that $\lambda_m \neq 0$. Now we note that

$$I \ \ni \ \partial\alpha - \alpha\partial = \sum_{j=0}^{m} \lambda_i i x^{i-1}.$$

Repeating m times (or using induction) we see that $I \ni \lambda_m(m!)$. Thus, $I \ni 1$ and $I = A_1$. □

One should note, however, that $A_1 = A_1(\mathbb{C})$ has lots of complicated *left* ideals. Certainly one has the strict descending chain

$$A_1 \supsetneq A_1x \supsetneq A_1x^2 \supsetneq A_1x^3 \supsetneq \cdots$$

(Why are the inequalities strict? As a hint, remember that A_1 is a domain.)

But there are more subtle examples, like the following:

Example 1.25. *Let $I = A_1\partial^2 + A_1(x\partial - 1) = \{a\partial^2 + b(x\partial - 1) : a, b \in A_1\}$. We claim that $I \neq A_1$ for the proof of which we think about differential operators acting on polynomials. Recall that we write $\theta * f$ for the action of $\theta \in A_1$ acting on a polynomial (or any differentiable function) $f = f(x)$. Then*

$$(a\partial^2 + b(x\partial - 1)) * x \;=\; a * (\partial^2 * x) + b * ((x\partial - 1) * x) \;=\; 0 + b * (x - x) \;=\; 0.$$

So, if $1 \in I$ then $1 * x = 0$, which is clearly absurd! □

Hard Exercise: Show that I also cannot be written as a cyclic left ideal $I = A_1\beta$ for any element $\beta \in I$.

The Axiom of Choice and Zorn's Lemma.

When dealing with large (uncountable) sets, there are lots of foundational problems (though that is a different subject) and we will always assume that **the Axiom of Choice** holds. Any discussion about this—or indeed about the precise axioms for a "set"—is something that should be left to a course on set theory. For completeness we note that the Axiom of Choice states that given any family $\mathcal{F}$ of subsets of a set X then we can form a set containing just one element from each set in $\mathcal{F}$. [3]

We will need Zorn's Lemma, which is equivalent to the Axiom of Choice and for which we need some definitions. So, let $(X, \leq)$ be a partially ordered set. A *chain* is a totally ordered subset Y of X; thus if $x, y \in Y$ then either $x \leq y$ or $y \leq x$. The set X is *inductive* if every chain Y in X has an *upper bound*, meaning an element $c \in X$ such that $y \leq c$ for all $y \in Y$.

Theorem 1.26. (Zorn's Lemma) *Every inductive, partially ordered set X has a maximal element; in other words there exists $c \in X$ such that if $x \geq c$ for $x \in X$, then $x = c$.* □

Zorn's Lemma is equivalent to the Axiom of Choice, so we cannot prove it (and we will not prove its equivalence to the Axiom of Choice); rather we will regard it as one of the axioms of mathematics. However, in using it, it is important to prove that all chains, rather than just the ascending chains indexed by natural numbers, have an upper bound. Here is probably its most important application within ring theory. A left ideal I of a ring R is *proper* if $I \neq R$. Define a left ideal of a ring R to be *maximal* if it is maximal among the set of proper ideals of R. The notions of proper and maximal ideals are defined similarly.

[3] Apparently Bertrand Russell characterised it by saying that if you have infinitely many pairs of shoes and socks, then you can pick one shoe from each pair—for example just take each left shoe. But to pick one sock from each pair you need the axiom of choice!

Corollary 1.27. *Let R be a ring. Then R has maximal left ideals and maximal ideals.*

Proof. Let X be the set of all proper left ideals of R ordered by inclusion. It is important to notice that a left ideal I is proper $\Leftrightarrow 1 \notin I$. Now suppose that we are given a chain $Y = \{I_\lambda\}_\lambda$ of left ideals of R. As Y is a chain it is routine to see that $I = \bigcup I_\lambda$ is a left ideal of R (the crucial point is that if $x, y \in I$ then $x \in I_\lambda$ and $y \in I_\mu$ for some λ, μ. Now either $I_\lambda \subseteq I_\mu$ or $I_\mu \subseteq I_\lambda$; assume the former. Then $x + y \in I_\mu \subseteq I$.) Finally as $1 \notin I_\nu$ for any ν it follows that $1 \notin I = \bigcup I_\nu$. Thus $I \neq R$.

Therefore, by Zorn's Lemma, X has a maximal element - that is, a maximal left ideal. To prove the corresponding result for ideals, just delete the word "left" throughout the proof. □

It is a fact that this result can fail if one does not assume the Axiom of Choice. It also fails for rings without identity. For example, let G be the additive abelian group of all rational numbers of the form $2^{-n}b$ with $b \in \mathbb{Z}$ and $n \geq 0$. Then $\mathbb{Z} \subseteq G$ and we take $R = G/\mathbb{Z}$. Make R into a ring by defining $rs = 0$ for all $r, s \in R$. (Note that if one uses the zero multiplication like this, then the axioms involving multiplication are trivial to prove—all axioms reduce to the equation $0 = 0$.) In this case one should check:

(1) an ideal of R is the same as an abelian subgroup;

(2) the only subgroups of R are $0 = \mathbb{Z}/\mathbb{Z}$, R and the $([2^{-n}]) = (2^{-n}\mathbb{Z} + \mathbb{Z})/\mathbb{Z}$.

[Hint: The key point is that if one takes $[2^{-n}a] \in R$, with a odd then by Euclid's algorithm there exists $r_1, r_2 \in \mathbb{Z}$ such that $1 = ar_1 + 2^n r_2$ and hence

$$[2^{-n}ar_1] = [2^{-n}ar_1] + 0 = [2^{-n}ar_1] + [2^{-n}2^n r_2] = [2^{-n}].$$

It follows that $[2^{-n}a]R = [2^{-n}]R$ after which the rest is easy.]

(3) So there is no maximal ideal.

2. **Modules.**

The key notion of a module over a ring R is defined, an R-module being an abelian group (or K-vector space in case R is a K-algebra) on which each element of R acts by a linear map (a "scalar multiplication"). This generalises the idea of a vector space over a field. Over noncommutative rings we have to distinguish between right and left actions, hence right and left modules. The usual concepts - subobjects, factor objects, homomorphisms, kernels of homomorphisms, direct sums and products - and basic results, including the nth isomorphism theorem for various n, hold. There's also the important modular law relating sum and intersection of submodules and, more generally, we make use of the structure of the poset (partially ordered set) of submodules of any given module. Zorn's Lemma gives us existence of maximal submodules in various contexts and these are important because they correspond to simple factor modules which, in turn, are the minimal "components" in various structure theorems about modules.

The endomorphisms of any (nonzero) module form a ring - an important ring, as seen for instance in Section 5.

The Weyl algebra well-illustrates how different the representation theory of noncommutative rings is from that for commutative rings. It has many non-isomorphic simple modules[4] yet its only maximal 2-sided ideal is 0 (which would make a ring a field if it were commutative).

The distinction between finitely generated and infinitely generated modules is important - a lot that works for f.g. modules doesn't work for infinitely generated modules.

We can put modules together by forming direct sums; in the opposite direction we can try to decompose modules into direct sums - which can be useful because those summands/components can be easier to understand than the original module. There's a bunch of helpful, but not very exciting, lemmas about these constructions, including what happens when the ring itself decomposes as the direct sum of (necessarily non-unital) subrings.

The concept of a module includes:

(1) any left ideal I of a ring R;

(2) any abelian group (which will become a $\mathbb{Z}$-module);

(3) an n-dimensional $\mathbb{C}$-vector space (which will become both a module over $\mathbb{C}$ and over $M_n(\mathbb{C})$).

The key point which these have in common is that one can both add elements of the module and multiply elements of the module from the left by elements of the ring. So, we generalize the idea.

Definition 2.1. *Let R be a ring. Then a (***unital***) left R-module M is an additive abelian group together with an operation*

$$R \times M \to M \qquad (r, m) \mapsto rm \tag{2.1}$$

[4] Modules and representations are the same thing; or at most slightly different perspectives on the same thing.

satisfying

(i) $(r+s)m = rm + sm$ *and* $r(m+n) = rm + rn$ *(distributivity)*

(ii) $(rs)m = r(sm)$ *(associativity)*

(iii) $1_R\, m = m$ *(unitarity)*

for all $m, n \in M$ *and* $r, s \in R$.

Remarks (a) In older books you may find that unitarity is not required in the definition.

(b) Similarly, one has the notion of a **right** R**-module** M, where one replaces (2.1) by an operation

$$M \times R \to M : (m, r) \mapsto mr \tag{2.2}$$

and adjusts (i), (ii), and (iii) accordingly.

(c) I will often write "Given ${}_RM$" for "Given a left R-module M".

Examples 2.2. (1) The set of n-dimensional column K-vectors, for a field K, is naturally a left module over $M_n(K)$ (or indeed over K itself).

(2) The set of n-dimensional row K-vectors, for a field K, is naturally a right module over $M_n(K)$.

(3) An abelian group A is a left (or right) $\mathbb{Z}$-module under the standard operation

$$n \cdot a \;=\; \underbrace{a + a + \cdots + a}_{n \text{ times}}.$$

(4) A ring R is always a left (or right) R-module and any left ideal of R is also a left R-module.

(5) Slightly more generally, given a subring S of a ring R then R is a (left or right) S-module.

(6) Let R be a ring. The **zero right** R**-module** is the right R-module $\{0\}$ with just one element 0 such that $0r = 0$ for all $r \in R$. The zero left R-module is defined similarly, and in both cases we just write it as 0.

(7) Over a *commutative* ring R there is no difference between left and right modules—given a left R-module M you get a right module by defining $m * r = rm$ for $m \in M$ and $r \in R$. However, over noncommutative rings, associativity is likely to fail when you do this. So they really are different.

(8) $\mathbb{C}[x]$ is a left module over the first Weyl algebra $A = A_1(\mathbb{C})$, where A acts on $\mathbb{C}[x]$ as differential operators. In this case the fact that we do get a module is just (a special case of) the assertion that, by definition, differential operators are linear operators.

We have the following basic properties of modules.

Lemma 2.3. *Given* ${}_RM$*, then for all* $m \in M$*,* $r \in R$ *we have*

(1) *(a)* $0_R m = 0_M$*, and (b)* $r0_M = 0_M$*, where* 0_R *stands for the zero element of* R *and* 0_M *stands for the zero element of* M*. So, from now on we will just write* $0 = 0_M = 0_R$ *without fear of confusion.*

(2) $(-1) \cdot m = -m$.

Proof. (1) $0_R m + 0_R m = (0_R + 0_R)m = 0_R m$. So, cancelling—which is allowed in an abelian group—gives $0_R m = 0_M$. The proof of (b) is similar starting with $r0_M + r0_M$.

(2) In this case

$$m + (-1)m = 1m + (-1)m = (1 + (-1))m = 0_R m = 0_M.$$

So $(-1)m$ is the additive inverse of m, i.e. $(-1)m = -m$. □

Most definitions from the theory of linear algebra or abelian groups ("subthings", "factor things", "thing" homomorphisms...) have analogues here. So, before reading the next few pages see if you can guess all the relevant definitions.

Definition 2.4. *Let R be a ring and M a left R-module. A* **submodule** *N of M is a subset of M which forms a left R-module under the operations inherited from M. Write $N \leq M$ for "N is a submodule of M"*

(I suppose more formally I should write "left submodule" here, but the second left is almost always superfluous.)

As you have seen with vector spaces and groups, we have the standard way of testing this:

Lemma 2.5. *A subset N of a left R-module M is a submodule $\Leftrightarrow$*

(i) $N \neq \varnothing$ (equivalently, $0_M \in N$)

(ii) $x, y \in N \implies x - y \in N$ (so N is a subgroup under addition)

(iii) $x \in N$, $r \in R \implies rx \in N$.

Proof. Use the proofs you have seen before. □

Note that, in the light of part (2) of the previous lemma the condition (ii) above may be replaced by:
(ii)' $x, y \in N \implies x + y \in N$
(condition (iii) with $r = -1$ takes care of closure under negatives).

Examples 2.6. :

(1) The submodules of a vector space V over a field K are just the subspaces of V.

(2) The submodules of a $\mathbb{Z}$-module A are just the subgroups of A.

(3) As usual $\{0_M\}$ is a submodule of any module M and it will just be written 0. Similarly M is a submodule of M.

(4) For any ring R, the submodules of R_R are just the right ideals of R. Similarly the left ideals of R are just the (left) submodules of ${}_R R$.

In particular, for all $n \in \mathbb{Z}$, $n\mathbb{Z}$ is a submodule of $\mathbb{Z}_{\mathbb{Z}}$.

The module ${}_R M$ is **simple** if $M \neq 0$ and M has no submodules other than M and 0. For example, a vector space over a field is simple as a module if and only if it is 1-dimensional. An abelian group is simple (as a $\mathbb{Z}$-module) $\iff$ it is a simple abelian group.

Definition 2.7. *Given a left R-module M and elements $\{m_i \in M : i \in I\}$, we write $\sum_{i\in I} Rm_i$ for the set of elements*

$$\sum_{i\in I} Rm_i \;=\; \Big\{m = \sum_{i\in I} r_i m_i, \text{ where the } r_i \in R \text{ and only finitely many } r_i \text{ are nonzero}\Big\}.$$

We say that M is **generated** *by $\{m_i \in M : i \in I\}$ if $M = \sum_{i\in I} Rm_i$. We say that M is* **cyclic** *if $M = Rm = \{rm : r \in R\}$ for some $m \in M$, and that M is* **finitely generated** *if $M = \sum_{i=0}^{n} Rm_i$ for some finite set $\{m_i\}_i$ of elements of M.*

Lemma 2.8. *Let $\{m_i : i \in I\}$ be elements of the left R-module M. Then*

(1) *The set $N = \sum_{i\in I} Rm_i$ is a submodule of M.*
(2) *$N = \sum_{i\in I} Rm_i$ is the unique smallest submodule of M containing $\{m_i : i \in I\}$.*
(3) *If M is a finitely generated module, then M has a maximal submodule (meaning a submodule maximal among the submodules $N \neq M$).*

Proof. Part (1) is left as an exercise.

(2) If L is a submodule of M containing all the m_i then it also contains all finite sums $r_1 m_{i_1} + \cdots + r_n m_{i_n}$ and hence $L \subseteq N$. Since N is a submodule we are done.

(3) This is very similar to Corollary 1.27.

First, we can write $M = \sum_{i=1}^{n} Rm_i$ with n as small as possible. The advantage of this is that now $M \neq N = \sum_{i=1}^{n-1} Rm_i$. Let X be the set of all proper submodules of M that contain N and order X by inclusion. It is important to notice that a submodule $M \supseteq I \supseteq N$ is not equal to M if and only if $m_n \notin I$. Now suppose that we are given a chain $Y = \{I_\lambda\}$ of elements of X. As Y is a chain we claim that $I = \bigcup I_\lambda$ is a submodule of M. This is one of the few cases where addition is the subtle point. Indeed, $x, y \in I$ then $x \in I_\lambda$ and $y \in I_\mu$ for some λ, μ. Now either $I_\lambda \subseteq I_\mu$ or $I_\mu \subseteq I_\lambda$; assume the former. Then $x \pm y \in I_\mu \subseteq I$. If $m \in I$ and $r \in R$ then $m \in I_\lambda$ for some λ whence $rm \in I_\lambda \subseteq I$.

Finally as $m_n \notin I_\nu$ for any ν it follows that $m_n \notin I = \bigcup I_\nu$. Thus $I \neq M$.

Thus X is indeed inductive and any maximal element in X - and there is at least one by Zorn's Lemma - is a maximal submodule of M □

As was true of Corollary 1.27, part (3) fails if you do not assume Zorn's Lemma and it also fails if you do not assume that the module is finitely generated. The standard counterexample is $\mathbb{Q}$ as a $\mathbb{Z}$-module. Can you prove this? An easier example is

Exercise 2.9. Let $R = \{\frac{a}{b} \in \mathbb{Q} : b \text{ is odd}\}$.

(1) Prove that R is a ring.
(2) Prove that (apart from 0 and $\mathbb{Q}$) the R-submodules of $\mathbb{Q}$ are just the $\{R2^m = 2^m R : m \in \mathbb{Z}\}$; thus they form a chain:

$$0 \subset \cdots \subset R2^n \subset R2^{n-1} \cdots \subset R2 \subset R \subset \frac{1}{2}R \subset \cdots \subset \frac{1}{2^n}R \subset \cdots \subset \mathbb{Q} \qquad \text{for } n \in \mathbb{N}.$$

(3) Hence $\mathbb{Q}$ has no proper maximal R-submodule

The details of Exercise 2.9 are given in the Second Example Sheet. Instead, here, I will give the details of a variant:

Exercise *Let $\mathbb{Z}[2^{-1}] = \{a2^{-n} : n \geq 1, a \in \mathbb{Z}\} \subset \mathbb{Q}$. Then the $\mathbb{Z}$-submodules of $M = \mathbb{Z}[2^{-1}]/\mathbb{Z}$ are M itself and $[2^{-n}]\mathbb{Z}$, where I have used the short-hand $[x] = [x + \mathbb{Z}]$ for $x \in \mathbb{Z}[2^{-1}]$.*

In particular, N has no maximal submodule.

Proof. Suppose that $0 \neq [q] \in M$ and write $q = 2^{-n}b$ where $n \geq 1$ and b is odd. By Euclid's Algorithm, write $1 = 2^n x + by$ for some integers x and y. Then

$$[2^{-n}] = [x + 2^{-n}by] = [2^{-n}b]y = [q]y \in [q]\mathbb{Z}.$$

In particular, $[2^{-n}]\mathbb{Z} = [q]\mathbb{Z}$.

Now, suppose that N is some $\mathbb{Z}$-submodule of M. Then N is generated by all the elements q in N and so, by the last paragraph, is generated by a set of the form $\{2^{-m_i} : i \in I\}$ for some index set I. There are now two cases: It could be that the m_i are bounded above, in which case they are bounded above by some m_j and then $N = \mathbb{Z}2^{-m_j}$. Or, the m_j have no upper bound. But in this, case for any $n \geq 0$, there exists some $m_i > n$ and hence $2^{-n}\mathbb{Q} \subseteq 2^{m_i}\mathbb{Q} \subseteq N$. Thus $N = M$.

It follows from the last paragraph that the $\mathbb{Z}$-submodules of M form a chain:

$$0 \subset \cdots \subset 2^{-n}\mathbb{Z} \subset 2^{-n-1}\mathbb{Z} \cdots\cdots \subset M.$$

So, there is certainly no maximal R-submodule of $\mathbb{Q}$. □

Note that the definition and properties of modules depend upon the ring concerned—for example $\mathbb{Q}$ is cyclic as a $\mathbb{Q}$-module, but is not even finitely generated as a $\mathbb{Z}$-module. This follows from the last example, but a more direct proof is the following: Suppose that $\mathbb{Q}$ is finitely generated as a $\mathbb{Z}$-module, say by $x_1, ...x_n$. Write the $x_i = a_i/b$ over a common denominator b (thus, for integers a_i, b). Then it is easy to see that $\mathbb{Q} = \sum \mathbb{Z}x_i \subseteq \mathbb{Z}.\frac{1}{b}$. But this is crazy since $\frac{1}{2b}$ is not contained in $\mathbb{Z}\frac{1}{b}$.

Definition 2.10. *Let R be a ring, and let M and N be left R-modules. An R-**module homomorphism** (or R-**homomorphism**) from M to N is a map $\theta : M \to N$ which satisfies*

(i) $\theta(x + y) = \theta(x) + \theta(y)$ (thus θ is a homomorphism of abelian groups),

(ii) $r\theta(x) = \theta(rx)$

for all $x, y \in M$, $r \in R$.

We say that θ is a monomorphism/epimorphism/isomorphism when θ is (1-1), onto and bijective, respectively. If $M = N$ we say that θ is an endomorphism of M and write $\mathrm{End}_R(M)$ for the set of all such endomorphisms. Finally an automorphism is a bijective endomorphism.

Examples: (1) Let K be a field. Then K-module homomorphisms are just linear mappings between vector spaces.

(2) $\mathbb{Z}$-module homomorphisms are just homomorphisms of abelian groups. (Check this.)

(3) Given a homomorphism $\theta : M \to N$ then $\theta(0_M) = 0_N$, and $\theta(-x) = -\theta(x)$, for all $x \in M$.

The **kernel** $\ker\theta$ of an R-module homomorphism $\theta : M \to N$ is the subset of M defined by

$$\ker\theta = \{x \in M : \theta(x) = 0_N\}.$$

Lemma 2.11. *Given a homomorphism $\theta : M \to N$ of left (or right) R-modules then $\ker(\theta)$ is a submodule of M and $\theta(M) = Im(\theta)$ is a submodule of S*

Proof. This is an easy exercise, but for once let me give all the details.

First, $\ker\theta \neq \varnothing$ since $0_M \in \ker\theta$. Suppose that $x, y \in \ker\theta$. Then

$$\theta(x - y) = \theta(x) - \theta(y) = 0_N - 0_N = 0_N.$$

So $x - y \in \ker\theta$.

Now suppose that $x \in \ker\theta$, and $r \in R$. Then

$$\begin{aligned} \theta(rx) &= r\theta(x) \qquad \text{[by homomorphism condition (ii)]} \\ &= r0_N \qquad \text{[as } x \in \ker\theta\text{]} \\ &= 0_N. \end{aligned}$$

So $rx \in \ker\theta$. Hence $\ker\theta$ is a submodule of M.

Now $\operatorname{im}\theta \neq \varnothing$ since $0_N = \theta(0_M) \in \operatorname{im}\theta$. Suppose that $u, v \in \operatorname{im}\theta$. Then $u = \theta(x)$, $v = \theta(y)$ for some $x, y \in M$. Then

$$u - v = \theta(x) - \theta(y) = \theta(x - y) \in \operatorname{im}\theta.$$

Suppose further that $r \in R$. Then

$$ru = r\theta(x) = \theta(rx) \in \operatorname{im}\theta.$$

Hence $\operatorname{im}\theta$ is a submodule of N. □

Factor Modules. Recall that, if I is a left ideal of a ring R that is not a two-sided ideal, then one cannot make the factor abelian group R/I into a factor ring. (See the comments after Lemma 1.19.) However, we *can* make it into a factor module, as we next describe.

For completeness, let us recall the construction of the factor group M/N when $N \subseteq M$ are abelian groups. (As M is an abelian group, N is a normal subgroup of M.) The **cosets** of N in M are the subsets $x + N$ of M with $x \in M$, where $x + N = \{x + u : u \in N\}$. Two cosets $x + N$ and $x' + N$ of N in M are either identical or disjoint, i.e. $(x + N) \cap (x' + N) = \varnothing$. Furthermore, $x + N = x' + N \iff x - x' \in N$. Every element $y \in M$ belongs to some coset of N in M. In particular, $y \in y + N$ since $y = y + 0_N$. The set of cosets of N

in M, which is denoted by M/N, forms a partition of M. We define $(x+N)+(y+N)=(x+y)+N$ for $x, y \in M$. This consistently defines an operation of addition on M/N, which makes M/N into an additive abelian group.

Now suppose that N is a left R-submodule of a module M. Assume that

$$x+N=x'+N$$

for some $x, x' \in M$. Then $x-x' \in N$. Let $r \in R$. Then $rx-rx'=r(x-x') \in N$ since N is a submodule of M. Hence

$$rx+N=rx'+N.$$

This means we can consistently define an operation $R \times M/N \to M/N$ by putting

$$r(x+N)=rx+N \tag{2.3}$$

for all $x \in M$, $r \in R$. We have:

Theorem 2.12. *Let $N \subset M$ be left R-modules. The rule* (2.3) *turns M/N into a left R-module called the* **factor module** *of M by N.*

Proof. As we started by defining M/N as the factor of abelian groups, certainly M/N is an abelian group, and we have explained why we have a consistent multiplication map. To check module condition (i) from Definition 2.1 for M/N, let $x \in M$, $r, s \in R$. Then

$$\begin{aligned}
(r+s)(x+N) &= (r+s)x+N && \text{by (2.3)} \\
&= (rx+sx)+N && \text{as } M \text{ is an } R\text{-module} \\
&= (rx+N)+(sx+N) && \\
&= r(x+N)+s(x+N) && \text{by (2.3).}
\end{aligned}$$

You should check that module conditions (ii), and (iii) also hold. □

As usual, all the results we have proved above for left R-modules also have analogues for right modules. It is a good way to check that you understand these concepts by writing out the analogous results on the right!

THE HOMOMORPHISM THEOREMS FOR MODULES:

You will have seen homomorphism theorems for factor groups and for factor rings. As we see next, almost exactly the same theorems apply for factor modules.

To begin, note that for any R-module homomorphism θ, because it is a homomorphism of abelian groups,

$$\theta \text{ is a monomorphism if and only if } \ker\theta=\{0\}.$$

Let M and N be left (or right) R-modules. If there is an R-module isomorphism $\theta : M \to N$ then M and N are said to be **isomorphic**. We indicate that this is the case by writing $M \cong N$. The inverse mapping $\theta^{-1} : N \to M$ is also an R-module homomorphism. To see this, let $u, v \in N$, $r \in R$, and just for a change, we will prove it for *right* modules. Thus $u = \theta(x)$, $v = \theta(y)$ for some $x, y \in M$. So

$$\begin{aligned}\theta^{-1}(u+v) = \theta^{-1}(\theta(x)+\theta(y)) &= \theta^{-1}(\theta(x+y)) \\ = x+y &= \theta^{-1}(u)+\theta^{-1}(v)\end{aligned}$$

and

$$\begin{aligned}\theta^{-1}(ur) = \theta^{-1}(\theta(x)r) &= \theta^{-1}(\theta(xr)) \\ = xr &= \theta^{-1}(u)r.\end{aligned}$$

Hence, being bijective, θ^{-1} is an isomorphism.

If $\psi : L \to M$ and $\theta : M \to N$ are R-module isomorphisms then so is $\theta \circ \psi : L \to N$. Hence $\cong$ defines an equivalence relation on the collection of all right R-modules. We often use the notation $\theta\psi$ for the composition $\theta \circ \psi$.

Theorem 2.13. (The First Isomorphism Theorem for Modules) *Let R be a ring, M and N right R-modules and $\theta : M \to N$ an R-module homomorphism. Then*

$$M/\ker\theta \cong \operatorname{im}\theta.$$

Proof. Suppose that $x, x' \in M$ and that $x + \ker\theta = x' + \ker\theta$. Then $x - x' \in \ker\theta$. So

$$\theta(x) - \theta(x') = \theta(x - x') = 0_N, \qquad \text{i.e. } \theta(x) = \theta(x').$$

Therefore, we may consistently define a mapping

$$\overline{\theta} : M/\ker\theta \to \operatorname{im}\theta$$

by

$$\overline{\theta}(x + \ker\theta) = \theta(x) \qquad \text{for } x \in M.$$

It is easy to check that $\overline{\theta}$ is an R-module homomorphism. Indeed, from the First isomorphism Theorem for Groups, we know that it is a well-defined group homomorphism, so we need only check multiplication. Thus, suppose that $x + \ker\theta \in M/\ker(\theta)$ and $r \in R$. Then

$$r\overline{\theta}(x + \ker\theta) = r\theta(x) = \theta(rx) = \overline{\theta}(rx + \ker\theta) = \overline{\theta}\left(r(x + \ker\theta)\right),$$

as required.

Let $x \in M$ such that $\overline{\theta}(x + \ker\theta) = 0_N$. By the definition of $\overline{\theta}$, $\theta(x) = 0_N$, i.e. $x \in \ker\theta$. Therefore $x + \ker\theta = \ker\theta = 0_{M/\ker\theta}$. Hence $\ker\overline{\theta} = \{0_{M/\ker\theta}\}$, i.e. $\overline{\theta}$ is a monomorphism.

Now let $u \in \operatorname{im}\theta$. Then $u = \theta(x)$ for some $x \in M$ and so

$$u = \overline{\theta}(x + \ker\theta).$$

Therefore $\overline{\theta}$ is also surjective and hence an isomorphism. □

Theorem 2.14. (The Correspondence Theorem for Modules) *Let M be a left module over a ring R and let N be a submodule of M.*

(i) Every submodule of M/N has the form K/N, where K is some submodule of M with $N \subseteq K$.

(ii) There is a 1–1 correspondence

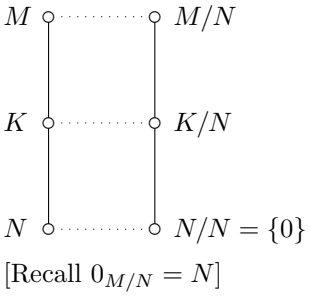
$$K \mapsto K/N$$

between the submodules K of M which contain N and the submodules of M/N. This correspondence preserves inclusions.

(iii) If $M \to N$ is an isomorphism of left R-modules then there is a (1-1) correspondence between submodules of M and N.

Proof. If K is a submodule of M with $N \subseteq K$ then, clearly, N is a submodule of K and K/N is a submodule of M/N since

$$K/N = \{x + N : x \in K\} \subseteq \{x + N : x \in M\} = M/N.$$



[Recall $0_{M/N} = N$]

(i) Let T be any submodule of M/N. Let

$$K = \{x \in M : x + N \in T\}.$$

It is easy to check that K is a submodule of M. For $u \in N$,

$$u + N = N = 0_{M/N} \in T$$

and so $u \in K$. Hence $N \subseteq K$. Furthermore $T = K/N$.

(ii) In part (i), we saw that the mapping from the set of submodules K of M such that $N \subseteq K$ to the submodules of M/N defined by $K \mapsto K/N$ is surjective.

Now suppose that J, K are submodules of M that contain N and $J/N = K/N$. Let $j \in J$. Then $j + N = k + N$ for some $k \in K$. But then $j \in j + N = k + N \subseteq K$. So $J \subseteq K$. Similarly $K \subseteq J$. Hence $J = K$. This shows the mapping is injective. It clearly preserves inclusion.

(iii) Let $\theta : M \to N$ be the isomorphism and recall that $\theta^{-1} : N \to M$ is also an isomorphism from the chat before Theorem 2.13. Now, given a submodule $K \subset M$ then certainly $\theta(M)$ is a submodule of N with $\theta^{-1}(\theta(K)) = K$. Hence the mapping $K \mapsto \theta(K)$ gives a (1-1) correspondence. □

Theorem 2.15. (1) **(The Second Isomorphism Theorem for Modules)** *If A and B are submodules of a left R-module M then $A + B$ and $A \cap B$ are submodules of M and*

$$\frac{A+B}{B} \cong \frac{A}{A \cap B}.$$

(2) **(The Third Isomorphism Theorem for Modules)** *Let N, K, M be right (or left) modules over a ring R such that $N \subseteq K \subseteq M$. Then*

$$\frac{M/N}{K/N} \cong M/K.$$

Proof. (1) Let us first check that $A \cap B$ is a submodule. As usual, it is a sub-abelian group, so it only remains to check that $rx \in A \cap B$ for all $r \in R$ and $x \in A \cap B$. But as $x \in A$ certainly $rx \in A$ and as $x \in B$ similarly $rx \in B$. Thus $rx \in A \cap B$, as required. The proof for $A + B$ is similar. We now define

$$\psi : A \to \frac{A+B}{B} \quad \text{by} \qquad \psi(x) = [x + B] \text{ for } x \in A.$$

This is a surjective homomorphism of abelian groups and trivially $\psi(rx) = [rx + B] = r[x + B] = r\psi(x)$, for any $r \in R$. Hence it is also an R-homomorphism. As abelian groups it has kernel $A \cap B$ and so by Theorem 2.13 we get the desired isomorphism

$$\frac{A}{A \cap B} \;=\; \frac{A}{\ker(\psi)} \;\cong Im(\psi) \;=\; \frac{A+B}{B}.$$

By chasing the maps you see that this isomorphism is also given by $[a + A \cap B] \mapsto [a + B]$ fior all $a \in A$.

(2) We can define a mapping $\theta : M/N \to M/K$ by putting $\theta(x + N) = x + K \qquad (x \in M)$. By the corresponding result for abelian groups, this is an isomorphism of abelian groups. Thus, to prove the stated result, all we need to check is that it is also a homomorphism of R-modules. This is (almost) obvious. □

As usual, everything we have stated for left modules has an analogue on the right.

Remark: One feature of this proof works for many results about modules, especially for factor modules: What we have really done is to observe that the result *does* hold for the factor abelian group. So then really all that is left to do is to check that the given map of groups also preserves the natural R-module structures. This is a valid approach in proving many such results.

Generalizing the observation at the beginning of the Second Isomorphism Theorem, we note that one can consider arbitrary sums and intersections of submodules. To be precise, suppose that N_λ $(\lambda \in \Lambda)$ is a collection of submodules of a right module M over a ring R. (The nonempty index set Λ may be finite or infinite.) Then

$$\bigcap_{\lambda \in \Lambda} N_\lambda$$

(the intersection of all the submodules) is also a submodule of M (see Example Sheet 2).

The **sum** $\sum_{\lambda \in \Lambda} N_\lambda$ of the submodules is defined by

$$\sum_{\lambda \in \Lambda} N_\lambda = \{x_1 + x_2 + \cdots + x_m : m \in \mathbb{N},\ x_i \in N_{\lambda_i},\ \lambda_i \in \Lambda\ (i = 1, 2, \ldots, m)\}.$$

This is also a submodule of M (see Example Sheet 2). It is the smallest submodule of M that contains all N_λ $(\lambda \in \Lambda)$.

In particular, if $N_1, N_2, \ldots, N_n$ are submodules of M, then $\bigcap_{i=1}^{n} N_i$ and $N_1 + N_2 + \cdots + N_n$ are submodules of M, where

$$N_1 + N_2 + \cdots + N_n = \{x_1 + x_2 + \cdots + x_n : x_i \in N_i \ (i = 1, 2, \ldots, n)\}.$$

Let us apply these results to answer the question:

What are the simple modules over a commutative ring R?

Some of the lemmas will also be used for other results.

Lemma 2.16. (1) *If M is a finitely generated left R-module, then M has a simple factor module.*

(2) *If M is a simple left module over any ring R then $M = Rm$ for any $0 \neq m \in M$.*

Proof. (i) By Lemma 2.8(3), M has a maximal submodule, say N. We claim that M/N is a simple module. Indeed, if it has a nonzero submodule $\overline{K}$, then by the Correspondence Theorem, we can write $\overline{K} = K/N$ for some module $N \subseteq K \subseteq M$. The maximality of N then ensures that either $N = K$ or $K = M$. Equivalently, either $\overline{K} = N/N = 0$ or $K = M/N$, as desired.

(ii) Rm is a submodule by Lemma 2.8 and is nonzero as it contains $m = 1 \cdot m$. By simplicity of M, Rm must therefore equal M. □

Lemma 2.17. *If M is a left module over any ring R and $a \in M$ then:*

(1) *There is an R-module homomorphism $\theta : R \to Ra$ given by $\theta(r) = ra$ for $r \in R$.*

(2) *Moreover, $Ra \cong R/I$ where $I = \{r \in R : ra = 0\}$.*

Proof. (1) is routine. For (2) just note that, by the first isomorphism theorem for modules, $Ra \cong R/\ker(\theta)$ and $\ker(\theta) = I$ by definition. □

Corollary 2.18. *If R is a (nontrivial) commutative ring, then simple R-modules are just the factors R/P where P runs through the maximal ideals of R. Moreover, $R/P \ncong R/Q$ for distinct maximal ideals P, Q.*

Proof. As we remarked in Examples 2.2 it does not matter if we work with right or left R-modules in this case, but let's work with left modules for concreteness.

If N is a simple (left) R-module, then: $N = Ra$ for some $a \in N$, by Lemma 2.16 and then $N \cong R/I$ for some ideal I by Lemma 2.17 (as R is commutative, ideals, left ideals and submodules of R are all the same thing). Note that, by definition, $M \neq 0$ and so $I \neq R$. If I is *not* maximal, say $I \subsetneq J \subsetneq R$, then the Correspondence Theorem 2.14 says that J/I is a proper submodule of R/I, in the sense that J/I is a submodule of R/I that is neither zero nor R/I. Hence by Theorem 2.15(iii) $N \cong R/I$ also has a proper submodule.

In order to prove the last part we digress a little. Suppose that M is a left R-module over a possibly noncommutative ring R. Given a subset S of M we write

$$ann_R S = \{r \in R : rm = 0 \text{ for all } m \in S\}$$

for the **annihilator** of S.

Suppose that M is cyclic; say $M = Ra \cong R/I$ where $I = \{r \in R : ra = 0\}$ as in Lemma 2.17. There are several observations to make here. First, the definition of I just says that $I = ann_R(a)$. Secondly, if there is an isomorphism $\theta : M \to N$ then for any $r \in ann_R M$ and $n = \theta(m) \in N$ we have $rn = r\theta(m) = \theta(rm) = \theta(0) = 0$ and so $ann(M) \subseteq ann(N)$. Applying the same argument to $\theta^{-1} : N \to M$ we obtain

(2.4) If M and N are isomorphic modules over (any) ring R then $ann_R M = ann_R N$.

Now return to the special case of a cyclic module $M = Rm$ over a commutative ring R. Then we claim that in fact $ann_R M = ann_R(m)$. The inclusion $\subseteq$ is obvious, so suppose that $r \in ann_R(a)$ and that $m \in M$. Then $m = sa$ for some $s \in R$ and so

$$rm = rsa = sra = s \cdot 0 = 0,$$

as claimed.

Note that the final assertion of the corollary is a special case of these observations: We are given a module $M \cong R/P \cong R/Q$. From (2.4) we see that $ann_R M = ann_R(R/P) = ann_R(R/Q)$. But from the last paragraph we see that $ann_R(R/P) = P$. Thus $P = Q$, as required. □

We proved rather more than was necessary in the last part of the proof of the corollary, but it does show that the concept of annihilators is useful; indeed the concept will return several times in this course.

Exercise 2.19. (i) Show that, if R is a noncommutative ring then the simple left R-modules are the same as the modules R/I where I runs through the maximal left ideals of R. However, the left ideal I will *not* be unique (see Example 2.21 below).

(ii) If M is a left module over a ring R and $m \in R$ show that $ann_R M$ is a *two-sided* ideal of R and that $ann(m)$ is a left ideal of R.

(iii) Show that $ann_R m$ is usually not an ideal of R when M is a module over a non-commutative ring R and $m \in M$. (Consider, for example, one column of $M_2(\mathbb{C})$ as a left $M_2(\mathbb{C})$-module.)

In the noncommutative case, it is very difficult to say much more in general about simple R-modules—except that they are complicated. Let us illustrate this with the Weyl algebra $A_1 = A_1(\mathbb{C})$.

Example 2.20. (A simple module over the Weyl Algebra)

(1) *$\mathbb{C}[x]$ is a left A_1-module where $\alpha = \sum f_i \partial^i \in A_1$ acts on $g(x) \in \mathbb{C}[x]$ as a differential operator*

$$\alpha \cdot g(x) \;=\; \alpha * g(x) \;=\; \sum f_i(x) \frac{d^i g}{dx^i}.$$

The proof of this is almost obvious—A_1 was defined as the set of all differential operators and the fact that they act linearly on functions is really the same as saying that those functions form an A_1-module.

Of course, this same argument means that other spaces of differentiable functions, like $\mathbb{C}(x)$, are left A_1-modules.

(2) $\mathbb{C}[x] = A_1 \cdot 1$ *simply because* $g(x) = g(x) * 1$ *for all* $g(x) \in \mathbb{C}[x]$.

(3) $\mathbb{C}[x] \cong A_1/A_1\partial$.

To see this, take the map $\chi : A_1 \to M = \mathbb{C}[x]$ *given by* $\chi(\alpha) = \alpha * 1$. *Then Lemma 2.17 shows that* $\mathbb{C}[x] \cong A_1/\ker(\chi)$. *Now, clearly* $A_1\partial \in \ker(\chi)$, *so what about the other inclusion? The fact that* A_1 *has* $\mathbb{C}$*-basis* $\{x^i\partial^j\}$ *(see Corollary 1.13) means that any element* $\alpha \in A_1$ *can be written as* $\alpha = \beta\partial + h(x)$, *where* $\beta \in A_1$ *but* $h(x) \in \mathbb{C}[x]$. *Now*

$$\alpha * 1 = \beta * (\partial * 1) + h(x) * 1 = 0 + h(x).$$

Thus, $\alpha \in \ker(\chi) \iff h(x) = 0$, *as required.*

(4) $\mathbb{C}[x]$ *is a* **simple** A_1*-module.*

To see this, suppose that $f(x), g(x)$ *are nonzero polynomials in* $\mathbb{C}[x]$ *with* $\deg f = d$, *say* $f = \lambda x^d + \cdots$ *where* $\lambda \neq 0$. *Then*

$$\frac{1}{\lambda d!} g(x)\partial^d * f = \frac{1}{\lambda d!} g(x)\frac{d^d f}{dx^d} = g(x).$$

So certainly $\mathbb{C}[x]$ *is simple as a left* A_1*-module.*

Exercise. We should note that there are lots of other A_1-modules. For example prove that, as A_1-modules,

$$\mathbb{C}(x) \supset \mathbb{C}[x, x^{-1}] = A_1 \cdot x^{-1}$$

and that

$$A_1x^{-1}/\mathbb{C}[x] \;\cong\; A_1/A_1x$$

is also a simple left A_1-module. (Why is this? It will be explained in more detail in an exercise set.)

Example 2.21. *One complicating feature of noncommutative rings is that lots of different-looking modules can be isomorphic. To see this we again take* $A_1 = A_1(\mathbb{C})$ *and the simple module* $N = \mathbb{C}[x] \cong A_1/A_1\partial$. *As* $\mathbb{C}[x]$ *is a simple* A_1*-module we can, by Lemmas 2.16, and 2.17 write*

$$N \;=\; A_1x \;\cong\; A_1/ann_{A_1}(x).$$

So, what is $ann_{A_1}(x)$*? In fact*

$$ann_{A_1}(x) = A_1\partial^2 + A_1(x\partial - 1)$$

and hence

$$A_1/A_1\partial \;\cong\; N \;\cong\; A_1/(A_1\partial^2 + A_1(x\partial - 1)).$$

Proof. First of all it is easy to check that $\partial^2 * x = 0 = (x\partial - 1) * x$ and hence that $ann_{A_1}(x) \supseteq A_1\partial^2 + A_1(x\partial - 1)$.

In order to prove the other direction, recall from Corollary 1.13 that any element $\alpha \in A_1$ can be written as $\alpha = \sum_{i=0}^{n} f_i(x)\partial^i$ and hence as $\alpha = \beta\partial^2 + f(x)\partial + g(x)$, for $\beta \in A_1$ but $f, g \in \mathbb{C}[x]$. Now suppose that $\alpha \in ann_{A_1}(x)$ and write $\alpha = \beta\partial^2 + f(x)\partial + g(x)$ as above. Rearranging slightly we see that $\alpha = \beta\partial^2 + f_1(x)(x\partial - 1) + \lambda\partial + g_1(x)$ for some $\lambda \in \mathbb{C}$ and $f_1(x), g_1(x) \in \mathbb{C}[x]$.

But $\alpha * x = 0$ and hence $(\lambda\partial + g_1(x)) * x = 0$. This in turn forces

$$0 = (\lambda\partial + g_1(x)) * x = \lambda + g_1(x)x \qquad \textit{and so} \qquad \lambda = 0 = g_1(x).$$

Thus, $\alpha = \beta\partial^2 + f_1(x)(x\partial - 1) \in A_1\partial^2 + A_1(x\partial - 1)$. □

DIRECT SUMS: Just as for groups and rings we have direct sums of modules. We begin by reminding you of the definitions in the first two cases.

Abelian groups: Let A_1 and A_2 be additive abelian groups (with zeros 0_1 and 0_2, respectively). The **(external) direct sum** of A_1 and A_2 denoted by $A_1 \oplus A_2$, is the set of ordered pairs

$$\{(a_1, a_2) : a_1 \in A_1,\ a_2 \in A_2\}$$

made into an additive abelian group by defining

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

for all $a_1, b_1 \in A_1$, $a_2, b_2 \in A_2$. It is straightforward to check that the conditions for an additive abelian group hold. The zero is $(0_1, 0_2)$ and $-(a_1, a_2) = (-a_1, -a_2)$.

Rings: Let R_1 and R_2 be rings (with identities 1_1 and 1_2, respectively). The additive abelian group $R_1 \oplus R_2$ can be made into a ring, also denoted $R_1 \times R_2$, by defining

$$(a_1, a_2)(b_1, b_2) = (a_1b_1, a_2b_2)$$

for all $a_1, b_1 \in R_1$, $a_2, b_2 \in R_2$. The identity is $(1_1, 1_2)$. Note that $R_1 \oplus R_2$ and $R_1 \times R_2$ are different notations for the same set. The second is the standard notation for the product of two rings but, in this course, it is convenient to use the first notation (which is more typical when we forget the multiplication and consider just the underlying abelian group).

Modules: Let R be a ring and M_1, M_2 left R-modules. The additive abelian group $M_1 \oplus M_2$ can be made into a left R-module by defining

$$r(x_1, x_2) = (rx_1, rx_2)$$

for all $x_1 \in M_1$, $x_2 \in M_2$, $r \in R$.

These constructions can be generalised to more than 2 summands.

Definition 2.22. *Given additive abelian groups A_1, A_2, ..., A_t, the* **(external) direct sum** $A_1 \oplus A_2 \oplus \cdots \oplus A_t$ *is the set*

$$\{(a_1, a_2, \ldots, a_t) : a_i \in A_i \ (i = 1, 2, \ldots, t)\}$$

made into an additive abelian group in the obvious way.

If the A_i happen to be (left) modules over a ring R then $A_1 \oplus A_2 \oplus \cdots \oplus A_t$ becomes a left R-module under the natural map

$$r(a_1, a_2, \ldots, a_t) = (ra_1, ra_2, \ldots, ra_t), \qquad \text{for } a_i \in A_i \text{ and } r \in R.$$

Remark: As remarked already, if R_1 and R_2 are rings, then either term direct product or direct sum can be used for the ring defined above. Both are fine, being equivalent for two, or any finitely many, rings R_i. (If there are infinitely many rings R_i to be combined then the direct product is the one which gives a ring with 1, whereas direct sum would give a non-unital ring.)

Exercise 2.23. (i) Check that $A_1 \oplus A_2 \oplus \cdots \oplus A_t$ really is a module in the last definition.

(ii) Suppose that M_i are left modules over a ring R. Then

$$M_1 \oplus M_2 \oplus M_3 \cong (M_1 \oplus M_2) \oplus M_3 \cong M_1 \oplus (M_2 \oplus M_3)$$

under the mappings

$$(x_1, x_2, x_3) \mapsto ((x_1, x_2), x_3) \mapsto (x_1, (x_2, x_3)).$$

Similarly, $M_1 \oplus M_2 \cong M_2 \oplus M_1$ since $(x_1, x_2) \mapsto (x_2, x_1)$ defines an isomorphism.

Internal Direct Sums: Suppose that

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_t,$$

where M_1, M_2, ..., M_t are modules over a ring R. For each $i = 1, 2, \ldots, t$, let

$$N_i = \{(0, 0, \ldots, \underset{\substack{\nwarrow \\ i\text{th entry}}}{x_i}, \ldots, 0) : x_i \in M_i\}. \tag{2.5}$$

Then N_i is a submodule of M and $M = N_1 + N_2 + \cdots + N_t$. Furthermore $M_i \cong N_i$ (as R-modules). The isomorphism is given by

$$x_i \mapsto (0, 0, \ldots, x_i, \ldots, 0).$$

Every element $x \in M$ can be expressed *uniquely* in the form

$$x = \widehat{x_1} + \widehat{x_2} + \cdots + \widehat{x_t},$$

where $\widehat{x_i} \in N_i$ $(i = 1, 2, \ldots, t)$. (In fact, $\widehat{x_i} = (0, 0, \ldots, x_i, \ldots, 0)$.)

We can now turn these observations around and make:

Definition 2.24. *Given any submodules N_1, N_2, ..., N_t of an R-module M, we say that M is the* **(internal) direct sum** *of N_1, N_2, ..., N_t if*

(i) $M = N_1 + N_2 + \cdots + N_t$,

(ii) every element x of M can be expressed uniquely in the form

$$x = x_1 + x_2 + \cdots + x_t, \tag{2.6}$$

where $x_i \in N_i$ $(i = 1, 2, \ldots, t)$.

We immediately get

Lemma 2.25. *(i) Let N_1, N_2, ..., N_t be submodules of an R-module M. If M is the internal direct sum of the N_i then $N_1 \oplus \cdots \oplus N_t \cong M$ under the map $\phi : (n_1, \ldots, n_t) \mapsto n_1 + \cdots n_t$.*

(ii) Conversely if $M = M_1 \oplus M_2 \oplus \cdots \oplus M_t$, is an external direct sum of modules M_i then M is the internal direct sum of the submodules N_i defined by (2.5)

Proof. (i) The definitions of internal and external direct sums ensure that ϕ is an isomorphism of sets and it is then routine to check that it is a module homomorphism.

(ii) See the discussion before Definition 2.24. □

If N_1, N_2, ..., N_t are submodules of an R-module M then we can form their external direct sum (which has underlying set the cartesian product $N_1 \times \cdots \times N_t$ of the sets $N_1, \ldots, N_t$) and their "internal" sum $N_1 + \cdots + N_t$ (which is a subset of M and usually not "direct"). In the special case that M *is* the internal *direct* sum of $N_1, \ldots, N_t$ then, as we have just seen, these modules (M and the module based on the cartesian product) are isomorphic, so we usually omit the words "internal" and "external" and depend on the context to make it clear which we mean (if it matters). We also use the notation

$$M = N_1 \oplus N_2 \oplus \cdots \oplus N_t$$

for the internal direct sum (that is, to emphasise that an internal sum is direct). The condition for an internal sum to be direct is given in the next remark.

Remark 2.26. Since

$$x_1 + x_2 + \cdots + x_t = 0 \iff x_i = -\sum_{\substack{j=1 \\ j \neq i}}^{t} x_j,$$

the uniqueness of (2.6) is also equivalent to the statement:

(ii)$'$ for $i = 1, 2, \ldots, t$,

$$N_i \cap \sum_{\substack{j=1 \\ j \neq i}}^{t} N_j = 0.$$

In the case $t = 2$, the condition (ii)$'$ is simply the assertion that $N_1 \cap N_2 = 0$ (but pairwise intersections being 0 is *not* enough when $t \geq 3$). Since we use the special case $t = 2$ so often I will call it:

Corollary 2.27. *A module M is a direct sum of submodules N_1 and N_2 if and only if*

(i) $M = N_1 + N_2$ *and*

(ii) $N_1 \cap N_2 = 0$.

You have to be a bit more careful when relating the direct sums of rings to direct sums of modules.

Lemma 2.28. *Let R and S be rings.*

(i) If I is a left ideal of R and J is a left ideal of S then $I \oplus J$ is a left ideal of $R \oplus S$.

(ii) If K is a left ideal of $R \oplus S$ then $K = K_1 \oplus K_2$ for some left ideals K_1 of R and K_2 of S.

Proof. (i) This is an easy exercise, but let us give the proof for once.

Assume that $I \underset{\text{l}}{\trianglelefteq} R$, $J \underset{\text{l}}{\trianglelefteq} S$. Then $I \oplus J$ is an abelian subgroup of $R \oplus S$ by standard results for groups. Let $a, b \in I \oplus J$, $x \in R \oplus S$. Then

$$a = (a_1, a_2), \quad b = (b_1, b_2), \quad x = (r, s)$$

for some $a_1, b_1 \in I$, $a_2, b_2 \in J$, $r \in R$, $s \in S$. Then

$$xa = (ra_1, sa_2) \in I \oplus J$$

since $ra_1 \in I$, $sa_2 \in J$. Hence $I \oplus J \underset{\text{l}}{\trianglelefteq} R \oplus S$.

(ii) Assume that $K \underset{\text{l}}{\trianglelefteq} R \oplus S$. Let

$$K_1 = \{a_1 \in R : (a_1, 0_S) \in K\}, \quad K_2 = \{a_2 \in S : (0_R, a_2) \in K\}.$$

It is easy to check that $K_1 \underset{\text{l}}{\trianglelefteq} R$, $K_2 \underset{\text{l}}{\trianglelefteq} S$.

Let $a \in K_1 \oplus K_2$. Then $a = (a_1, a_2)$ for some $a_1 \in K_1$, $a_2 \in K_2$, and so $(a_1, 0) \in K$, and $(0, a_2) \in K$. So $a = (a_1, 0) + (0, a_2) \in K$. Therefore $K_1 \oplus K_2 \subseteq K$.

Conversely let $a \in K$. Then $a = (a_1, a_2)$ for some $a_1 \in R$, $a_2 \in S$, and

$$(a_1, 0) = (1, 0)(a_1, a_2) \in K, \quad (0, a_2) = (0, 1)(a_1, a_2) \in K.$$

So $a_1 \in K_1$, $a_2 \in K_2$. Consequently $a \in K_1 \oplus K_2$ and $K \subseteq K_1 \oplus K_2$. Hence $K = K_1 \oplus K_2$. □

Remark. So, the situation for ideals of a direct sum of rings is very different to that of submodules of a direct sum of modules. For example consider the 2D vector space $M = \mathbb{C} \oplus \mathbb{C}$ as a $\mathbb{C}$-module. Then, for any $\lambda, \mu \in \mathbb{C} \setminus \{0\}$, the module $\mathbb{C}(\lambda, \mu)$ is a one dimensional submodule that certainly does not split as a direct sum of its components.

There are results similar to Lemma 2.28 for right ideals, and hence for ideals. The results can be extended to direct sums of t rings, where $t \in \mathbb{N}$. (See Example Sheet 2.)

Similar remarks apply to direct sums of rings, but there is more to say in connection with multiplication since ideals are not the same as subrings.

Definition 2.29. *Let A, B be nonempty subsets of a ring R and let $x \in R$. Then AB is defined by*

$$AB = \left\{ \sum_{i=1}^{m} a_i b_i : m \in \mathbb{N},\ a_i \in A,\ b_i \in B\ (i = 1, 2, \ldots, m) \right\},$$

i.e. AB is the set of all finite sums of elements of the form ab with $a \in A$, $b \in B$.

If A is closed under addition then we find that $\{x\}A$ is the same as xA and $A\{x\}$ is the same as Ax, where

$$xA = \{xa : a \in A\} \qquad \textit{and} \qquad Ax = \{ax : a \in A\}.$$

Internal Direct Sums of Ideals: We make a few more observations about direct sums of rings.

Suppose that $R_1, \ldots, R_t$ are rings and write

$$R = R_1 \oplus R_2 \oplus \cdots \oplus R_t$$

for their (external) direct sum. Let

$$S_i = \{(0, 0, \ldots, \underset{\substack{\nwarrow \\ i\text{th entry}}}{r_i}, \ldots, 0) : r_i \in R_i\} \qquad (i = 1, 2, \ldots, t).$$

Then S_i is a ring in its own right and $R_i \cong S_i$ (I do not like calling it a subring of R since the identity of S_i, $(0, 0, \ldots, 1_i, \ldots, 0)$, is not the same as the identity of R, $(1_1, 1_2, \ldots, 1_t)$ when $t > 1$). The isomorphism is given by

$$r_i \mapsto (0, 0, \ldots, r_i, \ldots, 0).$$

In addition, S_i is an ideal of R and $S_i S_j = 0$ if $i \neq j$.

We have seen above that every ideal I of R has the form

$$I = I_1 \oplus I_2 \oplus \cdots \oplus I_t,$$

where $I_i \trianglelefteq R_i$ $(i = 1, 2, \ldots, t)$, and every set of this form is an ideal of R. Let

$$J_i = \{(0, 0, \ldots, \underset{\substack{\nwarrow \\ i\text{th entry}}}{a_i}, \ldots, 0) : a_i \in I_i\} \qquad (i = 1, 2, \ldots, t).$$

Then $J_i \trianglelefteq S_i$, and I_i corresponds to J_i under the above isomorphism from R_i to S_i. Also $J_i \trianglelefteq R$ and $J_i J_j = 0$ if $i \neq j$.

Since

$$R = S_1 + S_2 + \cdots + S_t$$

and every element $a \in R$ can be expressed uniquely in the form

$$a = a_1 + a_2 + \cdots + a_t,$$

where $a_i \in S_i$ $(i = 1, 2, \ldots, t)$, we also have

$$R = S_1 \oplus S_2 \oplus \cdots \oplus S_t,$$

an internal direct sum of *ideals* of R (each of which has its own 1), and

$$I = J_1 \oplus J_2 \oplus \cdots \oplus J_t,$$

an internal direct sum of ideals of R.

Conversely, given such ways of representing R and I as internal direct sums (with $J_i \trianglelefteq S_i$ $(i = 1, 2, \ldots, t)$), we can form the external direct sums and

$$J_1 \oplus J_2 \oplus \cdots \oplus J_t \trianglelefteq S_1 \oplus S_2 \oplus \cdots \oplus S_t \cong R$$

just as for modules. Furthermore, $J_1 \oplus J_2 \oplus \cdots \oplus J_t$ corresponds to I under the isomorphism.

Lemma 2.30. (The Modular Law) *Let R be a ring, M be a left R-module and A, B, C be submodules of M with $B \subseteq A$. Then*

$$A \cap (B + C) = B + (A \cap C).$$

Proof. Let $x \in A \cap (B + C)$. Then $x \in A$ and $x = b + c$ for some $b \in B$, $c \in C$. As $B \subseteq A$, we have $c = x - b \in A \cap C$. So $x \in B + (A \cap C)$ and hence $A \cap (B + C) \subseteq B + (A \cap C)$.

Also $B + (A \cap C) \subseteq A$ since $B \subseteq A$, and $B + (A \cap C) \subseteq B + C$ because $A \cap C \subseteq C$. Hence $B + (A \cap C) \subseteq A \cap (B + C)$. Therefore $A \cap (B + C) = B + (A \cap C)$. □

Lemma 2.31. *Let R be a ring and M a right R-module with submodules K, L and N. Suppose that $M = L \oplus K$ and $L \subseteq N$. Then $N = L \oplus (N \cap K)$.*

Proof. By the Modular Law,

$$N = N \cap M = N \cap (L + K) = L + (N \cap K)$$

since $L \subseteq N$. But $L \cap (N \cap K) = 0$ since $L \cap K = 0$. Therefore $N = L \oplus (N \cap K)$ by Corollary 2.27. □

Remark 2.32. There are versions of this lemma where M, K, L, N are either:

(a) left R-modules,

(b) right (or left) ideals of R,

(c) ideals of R.

Version (b) (for left ideals) is just a special case of Lemma 2.31. Version (c) requires further comment. We require M, K, L, N to be ideals of R such that $L \subseteq N \subseteq M$.

Lemma 2.33. *Let I be an ideal of a ring R which is a direct sum*

$$I = J_1 \oplus J_2 \oplus \cdots \oplus J_t \oplus K$$

of ideals J_1, J_2, …, J_t, K of R. Suppose that K is a direct sum

$$K = J_{t+1} \oplus J_{t+2}$$

of ideals J_{t+1}, J_{t+2} *of* R. *Then*

$$I = J_1 \oplus J_2 \oplus \cdots \oplus J_t \oplus J_{t+1} \oplus J_{t+2}.$$

Proof. Clearly,

$$I = J_1 + J_2 + \cdots + J_t + (J_{t+1} + J_{t+2}) = J_1 + J_2 + \cdots + J_{t+2}.$$

Let $a \in I$. Then

$$a = a_1 + a_2 + \cdots + a_t + b$$

for some uniquely determined $a_i \in J_i$ $(i = 1, 2, \ldots, t)$, $b \in K$. Also

$$b = a_{t+1} + a_{t+2}$$

for some uniquely determined $a_{t+1} \in J_{t+1}$, $a_{t+2} \in J_{t+2}$. So

$$a = a_1 + a_2 + \ldots + a_t + a_{t+1} + a_{t+2}.$$

The elements $a_i \in J_i$ $(i = 1, 2, \ldots, t+2)$ are uniquely determined by a. This establishes the result. □

Comments on examinability: All the material in this section (and in Section 1) is basic, important and used in what follows; it's set-up for what is done in Section 3 onwards. There are short arguments which could be asked in an exam but you should concentrate on knowing the definitions, main examples, basic techniques and facts that are used frequently later on.

3. **Chain conditions.**

The ascending chain condition on the poset of submodules of a modules forces it, and all its submodules, to be finitely generated. Both it and the descending chain condition have strong consequences for the structure of a module. Each condition is robust in the sense that it is inherited by submodules and by factor modules and putting together two modules with the condition gives a module with the same condition.

Putting one of these conditions on a ring R, that is, on its poset of right or left ideals, also has strong conditions for R and for R-modules, and also for any ring S which contains R and which is finitely generated as a module over R. If R satisfies one of these conditions then so do the polynomial rings over R in finitely many variables. The Weyl algebra has the ascending chain condition on right and left ideals.

Definition: Let R be a ring and M a left (or right) R-module.

(i) M is said to be **Noetherian** or to satisfy the **ascending chain condition** (ACC) if every ascending chain

$$N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$$

of submodules of M is eventually stationary; thus there exists some integer n_0 such that $N_n = N_{n+1}$ for all $n \geq n_0$.

(ii) M is said to be **Artinian** or to satisfy the **descending chain condition** (DCC) if every descending chain

$$N_1 \supseteq N_2 \supseteq N_3 \supseteq \dots$$

of submodules of M is eventually stationary; thus there exists some integer n_o such that $N_n = N_{n+1}$ for all $n \geq n_0$.

(iii) M is said to satisfy the **maximum condition on submodules** (MAX) if every nonempty set $\mathcal{S}$ of submodules of M contains a maximal member. (By a maximal member of $\mathcal{S}$, we mean a submodule N in $\mathcal{S}$ such that there is no submodule T in $\mathcal{S}$ with $N \subsetneqq T$.)

(iv) M is said to satisfy the **minimum condition on submodules** (MIN) if every non-empty set $\mathcal{S}$ of submodules of M contains a minimal member.

Example: A finite-dimensional vector space satisfies both ACC and DCC. Use the dimensions of subspaces to prove this.

Theorem 3.1. *Let M be a left module over a ring R. Then the following are equivalent:*

(i) M satisfies ACC;

(ii) M satisfies MAX;

(iii) every submodule of M (including M itself, of course) is finitely generated.

Proof. (i) ⇒ (ii). Assume M satisfies ACC. Let $\mathcal{S}$ be a nonempty set of submodules of M. Suppose $\mathcal{S}$ does not have a maximal member. Choose $N_1 \in \mathcal{S}$. Since N_1 is not a maximal member of $\mathcal{S}$, we can choose $N_2 \in \mathcal{S}$ such that $N_1 \subset N_2$. Continuing in this way, we can construct an ascending chain

$$N_1 \subset N_2 \subset N_3 \subset \dots$$

of submodules of M with infinitely many distinct terms. This is a contradiction. So $\mathcal{S}$ must have a maximal member. Hence M satisfies MAX.

(ii) ⇒ (iii). Assume that M satisfies MAX and let N be a submodule of M. Let $\mathcal{S}$ be the set of all finitely generated submodules of N; this is not empty since 0 is certainly a finitely generated submodule of N. So, $\mathcal{S}$ has a maximal element; say $L = \sum_{i=1}^{n} Ra_i$. If $L = N$ we are done. If not then there exists some $a_{n+1} \in N \setminus L$. But then $L \subsetneq L + Ra_{n+1} = \sum_{i=1}^{n+1} Ra_i \subseteq N$, contradicting the maximality of L.

(iii) ⇒ (i). Assume all the submodules of M are f.g. Let

$$N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$$

be an ascending chain of submodules of M. Let $N = \bigcup_{k=1}^{\infty} N_k$. Then N is a submodule of M (see the proof of Lemma 2.8) and so N is f.g. So

$$N = Rx_1 + Rx_2 + \dots + Rx_n$$

for some $n \in \mathbb{N}$, $x_1, x_2, \dots, x_n \in N$. For each $i = 1, 2, \dots, n$, we have $x_i \in N_{k_i}$ for some k_i. Let $m = \max(k_1, k_2, \dots, k_n)$. Then $x_i \in N_m$ since $N_{k_i} \subseteq N_m$ $(i = 1, 2, \dots, n)$. Hence

$$N = Rx_1 + Rx_2 + \dots + Rx_n \subseteq N_m \subseteq N.$$

Therefore $N_m = N$ and so $N_k = N$ when $k \geqslant m$. Hence M has ACC. □

We remark that the above proof uses the Axiom of Choice. Indeed there is a lovely paper

W. Hodges, Six impossible rings. J. Algebra 31 (1974), 218–244.

which shows what happens when you do not assume this axiom.

Theorem 3.2. *Let M be a left module over a ring R. Then the following are equivalent:*

(i) M satisfies DCC;

(ii) M satisfies MIN.

Proof. Adapt the proof of Theorem 3.1. □

Obviously, there are analogous results for right R-modules.

Definition (i) A ring R which is Noetherian as a right R-module is called a **right Noetherian ring**. A **left Noetherian ring** is defined similarly.

(ii) A ring R which is Artinian as a right R-module is called a **right Artinian ring**. A **left Artinian ring** is defined similarly.

(iii) A **Noetherian ring** is one which is both right and left Noetherian. An **Artinian ring** is one which is both right and left Artinian.

In particular, a ring is right Noetherian if and only if all its right ideals are finitely generated.

Examples 3.3. (1) Fields and division rings are both Artinian and Noetherian: they only have two right (or left) ideals.

(2) $\mathbb{Z}$ is Noetherian since its ideals are principal and so finitely generated. But it is not Artinian (see (4)).

(3) Any ring with a finite number of elements is both Artinian and Noetherian, e.g. $\mathbb{Z}/n\mathbb{Z}$ with n a nonzero integer.

(4) Let R be a commutative integral domain. Then R is Artinian if and only if R is a field. Indeed, if R is not a field pick a non-unit, non-zero $x \in R$ and consider the chain of (left) ideals

$$R \supseteq Rx \supseteq Rx^2 \supseteq \cdots .$$

We claim that this is never stationary. Indeed, if $Rx^n = Rx^{n+1}$ then $x^n = rx^{n+1}$ for some $r \in R$ and, as we are in a domain, we can cancel to get $1 = rx$, contradicting the fact that x was not a unit.

(The commutativity assumption was not necessary; the same argument proves that a noncommutative domain that is not a division ring is also not left (or right) Artinian.)

(5) Let $Z[1/2] = \{a/b : a \in \mathbb{Z}, b = 2^m \text{ for some } m \geq 0\} \subset \mathbb{Q}$. We claim:

Claim 1 $M = \mathbb{Z}[1/2]/\mathbb{Z}$ is an Artinian $\mathbb{Z}$-module that is not Noetherian.

Claim 2 The submodules of M are precisely

$$0 = \frac{\mathbb{Z}}{\mathbb{Z}} \subsetneq \frac{1/2\mathbb{Z}}{\mathbb{Z}} \subsetneq \cdots \subsetneq \frac{2^{-m}\mathbb{Z}}{\mathbb{Z}} \subsetneq \cdots \subsetneq M.$$

Proof. This is similar to the proof of Example 2.9. Clearly Claim 1 follows from Claim 2, so it suffices to prove the latter. First, suppose that N is a submodule of M and pick $\alpha = [a2^{-m} + \mathbb{Z}] \in N$ with a odd. By Euclid's Algorithm $1 = xa + y2^m$ for some $x, y \in \mathbb{Z}$ from which we get

$$N \ni x\left[\frac{a}{2^m} + \mathbb{Z}\right] = \left[\frac{xa}{2^m} + \mathbb{Z}\right] = \left[\frac{xa}{2^m} + \frac{y2^m}{2^m} + \mathbb{Z}\right] = \left[\frac{1}{2^m} + \mathbb{Z}\right].$$

Now, N is generated by some collection of its elements (for example all the elements from N). Hence by the last display it is generated by a collection of elements of the form $\{2^{-m_i}\}$ for some collection of natural numbers $\{m_i : i \in I\}$. Now, there are two possibilities. It could be that the integers m_i are unbounded in which case $N = M$. Otherwise the m_i are bounded above, say $m_i \leq \Omega$ for all i. But in this case we have a maximal element $m_\infty = \max\{m_i\} \leq \Omega$ and then $N = \mathbb{Z}2^{-m_\infty}$. □

Here is one case where everything is easy:

Proposition 3.4. *let R be a finite dimensional k-algebra, where k is a field. Then R is both Artinian and Noetherian.*

Proof. Recall that the hypothesis means that R contains a field k inside its centre and that R is finite dimensional as a k-vector space, say $\dim_k R = n$. Now, if I is any left ideal of R then I is closed under multiplication by elements of k and hence is a k-vector space. But if $I \subsetneq J$ are left ideals of R then $\dim_k I < \dim_k J$. Hence any ascending or descending chain of left (or right) ideals has length at most n. □

Example 3.5. *If A, B, C, D are subsets of a ring R then we will use the notation*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a \in A, b \in B, c \in C, d \in D \right\} \tag{3.1}$$

Now let

$$R = \begin{pmatrix} \mathbb{Q} & \mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix} \tag{3.2}$$

Then R is a right noetherian and right artinian ring that is neither left artinian nor left noetherian.

Here is a sketch of the proof.

Claim: The right ideals of R are 0, R and

$$\begin{pmatrix} \mathbb{Q} & \mathbb{C} \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix} \quad \text{and} \quad I_{\lambda,\mu} = \left\{ \begin{pmatrix} 0 & \lambda \\ 0 & \mu \end{pmatrix} \cdot \mathbb{C} \right\}$$

where λ and μ are some complex numbers. The key point here is that the entries from the second column of any right ideal must form a $\mathbb{C}$-sub vector space of $\begin{pmatrix} \mathbb{C} \\ \mathbb{C} \end{pmatrix}$. As such the longest chain of right ideals one can get has length 3. Thus

$$R \text{ is both right Artinian and right Noetherian.}$$

(A more elegant proof of this step also follows from the right-hand version of Corollary 3.9.)

However, on the left for *any* $\mathbb{Q}$-subspace V of $\mathbb{C}$ then the set

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

is a left ideal of R. As such there are infinite ascending and descending chains of left ideals—even chains of uncountable length—and so

$$R \text{ is neither left Artinian nor left Noetherian.}$$

The detailed proofs of these assertions are left to the reader. □

Theorem 3.6. *Let R be a ring, M a left (or right) R-module and N a submodule of M. Then*

$$M \text{ is Noetherian} \iff N \text{ and } M/N \text{ are Noetherian .}$$

Similarly M is Artinian $\iff$ N and M/N are Artinian.

Proof. $\Rightarrow$ Assume that M is Noetherian. Then the Correspondence Theorem 2.14 says that every ascending chain of submodules of M/N can be written in the form

$$L_1/N \subseteq L_2/N \subseteq L_3/N \subseteq \ldots,$$

where

$$L_1 \subseteq L_2 \subseteq L_3 \subseteq \ldots$$

is an ascending chain of submodules of M that contain N. The second chain has only a finite number of distinct terms and hence this is also true for the first chain. So M/N is Noetherian. As every ascending chain of submodules of N is also a chain in M it too must be eventually stationary and so N is Noetherian.

$\Leftarrow$ Assume that N and M/N are Noetherian. Let

$$L_1 \subseteq L_2 \subseteq L_3 \subseteq \ldots$$

be an ascending chain of submodules of M.

Consider the ascending chain

$$L_1 \cap N \subseteq L_2 \cap N \subseteq L_3 \cap N \subseteq \ldots$$

of submodules of N. Since N is Noetherian, there is an integer m such that

$$L_k \cap N = L_m \cap N \quad \text{for all integers } k \geqslant m.$$

Now consider the ascending chain

$$L_1 + N \subseteq L_2 + N \subseteq L_3 + N \subseteq \ldots$$

of submodules of M which contain N. By the Correspondence Theorem,

$$(L_1 + N)/N \subseteq (L_2 + N)/N \subseteq (L_3 + N)/N \subseteq \ldots.$$

Since M/N is Noetherian, there is an integer n such that

$$(L_k + N)/N = (L_n + N)/N \quad \text{for all integers } k \geqslant n,$$

$$\text{i.e. } L_k + N = L_n + N \quad \text{for all integers } k \geqslant n.$$

Put $t = \max(m, n)$. Then, for integers $k \geqslant t$,

$$\begin{aligned} L_k &= L_k \cap (L_k + N) \\ &= L_k \cap (L_t + N) \qquad [\text{since } k \geqslant t \geqslant n] \\ &= L_t + (L_k \cap N) \qquad \text{by the Modular Law 2.30} \\ &= L_t + (L_t \cap N) \qquad [\text{since } k \geqslant t \geqslant m] \\ &= L_t. \end{aligned}$$

So M has ACC and so is Noetherian.

The Artinian case is proved similarly. □

Corollary 3.7. *Let N_1, N_2, ..., N_n be Noetherian submodules of a (left or right) module M over a ring R. Then $N_1 + N_2 + \cdots + N_n$ is also Noetherian.*

(2) Similarly, if the N_i are Artinian then so is $N_1 + \cdots + N_n$.

(3) The free left R-module $M = R^{(n)} = R \oplus \cdots \oplus R$ (n copies) is Noetherian (respectively Artinian) $\iff$ R is likewise.

Proof. (1) Assume that N_1, N_2 are Noetherian. Then, by Theorem 3.6 , $N_1/(N_1 \cap N_2)$ is Noetherian. But

$$(N_1 + N_2)/N_2 \cong N_1/(N_1 \cap N_2)$$

by the Second Isomorphism Theorem 2.15. So $(N_1 + N_2)/N_2$ is Noetherian. Therefore, by Theorem 3.6 again, $N_1 + N_2$ is Noetherian since N_2 is also Noetherian. This means that

$$N_1 + \cdots + N_n \;=\; (N_1 + N_2) + N_3 + \cdots + N_n$$

is now a sum of $n - 1$ Noetherian modules and hence, by induction on n, it is Noetherian.

(2) The Artinian case is exactly the same.

(3) This is a trivial consequence of parts (1) and (2). □

Corollary 3.8. *(1) Let R be a left Noetherian ring and M a left R-module. Then M is Noetherian $\iff$ M is finitely generated as an R-module.*

(2) Let R be a left Artinian ring. Then any finitely generated left R-module is Artinian.

Proof. (1) $\Rightarrow$ This is a special case of Theorem 3.1.

$\Leftarrow$ If M is finitely generated, write $M = Rm_1 + \cdots + Rm_n$ for some $m_i \in M$. By Lemma 2.17 we can write each Rm_i as $Rm_i \cong R/I_i$ for some left ideals I_i of R. Then each such module is Noetherian by Theorem 3.6. By Corollary 3.7 this means that $M = Rm_1 + \cdots + Rm_n$ is also Noetherian.

(2) The proof of $\Leftarrow$ from (1) works without change. □

Obviously the analogue of Corollary 3.8 also holds if we replace "left" throughout by "right".

Remark. It is true that an Artinian module M over an Artinian ring R is finitely generated, but the proof requires more work. This is because there is no analogue for Artinian modules of part (iii) of Theorem 3.1. Indeed Artinian modules need not be finitely generated as we showed in Example 3.3(5). The proof will appear in Chapter 4.

Corollary 3.9. *(1) Suppose that $R \subset S$ are rings where S is finitely generated as a left R-module. If R is left Noetherian then so is S. If R is left Artinian then so is S.*

(2) If A and B are both left Noetherian rings, respectively left Artinian rings then $A \oplus B$ is a left Noetherian ring, respectively left Artinian ring.

Proof. (1) For change we will treat the Artinian case–the Noetherian case is identical.

As S is a finitely generated left R-module it is Artinian as a left R-module by the last corollary. Hence we have DCC for left R-submodules of S. But, any left ideal of S is also closed under left multiplication by elements of the subring R. Hence any left ideal of S is also a left R-submodule of ${}_RS$. Thus these left ideals satisfy DCC.

(2) Use Lemma 2.28. □

This last result is very useful for checking that particular rings are Noetherian or Artinian. The point is that it may be very hard to find all the left ideals of a particular ring–just think about Example 3.5. But the given ring may have a subring we understand well. Here are some typical examples:

Example 3.10. *(1) Using the notation from Example 3.5 let*

$$R = \begin{pmatrix} \mathbb{Z} & 3\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} \end{pmatrix} \qquad \text{and} \qquad R' = \begin{pmatrix} \mathbb{Z} & 3\mathbb{Z} & 6\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} & 2\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} \end{pmatrix}.$$

Then we claim that both R and R' are (left and right) Noetherian rings.

(2) If R is left Noetherian, respectively left Artinian, then so is the matrix ring $M_n(R)$ for any integer n.

(3) $R = M_{n_1}(D_1) \oplus M_{n_2}(D_2) \oplus \cdots M_{n_r}(D_r)$ is an Artuinian ring for any $n_j \in \mathbb{N}$ and division rings D_j.

(4) If G is a finite group and K is a field or the integers then the group ring KG is Artinian and Noetherian.

Proof. (1) First one does need to check that R and R' are rings! As they are subsets of $M_2(\mathbb{Z})$ respectively $M_3(\mathbb{Z})$ that contain 1, if we check that they are closed under addition and multiplication then the rest of the axioms will be automatic. That they are closed under addition is clear from the fact that addition of matrices is componentwise. So this leaves multiplication. I will do the case of R leaving the messier R' to

you. So, suppose that we are given two elements from R, say

$$\alpha = \begin{pmatrix} a & 3b \\ c & d \end{pmatrix} \qquad and \qquad \beta = \begin{pmatrix} e & 3f \\ g & h \end{pmatrix}$$

for some integers $a, \ldots, h$. Then $\alpha\beta = \begin{pmatrix} ae+3bg & 3(af+bh) \\ ce+dg & 3cf+dh \end{pmatrix} \in R$, as required. On the other hand, R contains the ring of scalar matrices

$$Z = \left\{ \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} : n \in \mathbb{Z} \right\} \cong \mathbb{Z}$$

over which it is clearly 4-generated. Thus by Corollary 3.9 it is (left and right) Noetherian.

[And yes, that is an adequate proof if such a ring turned up on an exam!]

(2) In this case $M_n(R)$ contains the set of scalar matrices which is isomorphic as a ring to R itself. Over this subring $M_n(R)$ is a (free) module generated by the n^2 matrix units. So, the same argument works.

(3) Each $M_{n_j}(D_j)$ is Noetherian and Artinian by part (2). So the result follows from Corollary 3.9.

(4) KG is a finitely generated module over K.

□

A simplifying trick: At the end of the proof of part (1) of this example I mean that I could write down 4 explicit generators of R as a left Z-module—which you should do. However, there is a nice way of being lazy, which is wise when one gets to a more complicated ring like R'. The trick (let's do it for R') is to notice that certainly $M_3(\mathbb{Z})$ is a 9-generated left $\mathbb{Z}$-module and hence is a Noetherian left $\mathbb{Z}$-module by Corollary 3.8. Thus the subring R' is also finitely generated as a left $\mathbb{Z}$-module (and then, as in the proof is a left Noetherian ring).

Exercise 3.11. Use the corollary to give a better proof of Example 3.5.

Theorem 3.12. (Hilbert's Basis Theorem) *Let R be a left Noetherian ring and X an indeterminate. Then the polynomial ring $R[X]$ is a left Noetherian ring. The analogous result holds for right Noetherian rings.*

Consequently the polynomial ring $R[x_1, \ldots, x_n]$ is left Noetherian for any n.

Proof. By induction it suffices to prove it for $R[x]$ and it is enough to prove the left Noetherian case. Pick a left ideal I of $R[x]$ and let I_∞ be the set of all leading coefficients of elements of I, regarded as polynomials in x.

Claim: I_∞ is a left ideal of R.

Proof of the claim: Let $r, s \in I_\infty$ and suppose that $\rho = rx^m + r_{m-1}x^{m-1} + \cdots$ and $\sigma = sx^n + s_{n-1}x^{n-1} + \ldots$ are the corresponding elements of I. By symmetry we may as well assume that $m \geq n$ in which case

$$I \ni \rho \pm x^{m-n}\sigma = (r \pm s)x^n + \cdots$$

and so $r \pm s \in I_\infty$ and I_∞ is an abelian group. Multiplication is easier: If $t \in R$ then $I \ni t\rho = (tr)x^n + \cdots$ so $tr \in I$. Thus, the claim is proved.

We return to the proof of the theorem. As I_∞ is a left ideal of R it is finitely generated, say by $r_1, \ldots, r_n$ which are the leading coefficients of $\rho_i \in I$. Let $m = \max\{\deg \rho_i : 1 \leq i \leq m\}$. Then $M = \sum_{\ell=0}^{m} Rx^\ell$ is a finitely generated left R-module, whence is Noetherian by Corollary 3.8. Notice that $M \cap I$ is a left R-module as both M and I are closed under multiplication by elements from R. Thus $I \cap M$ is a finitely generated left R-module, by Theorem 3.1, say $I \cap M = \sum_{i=1}^{p} R\gamma_i$ for some $\gamma_i \in I$.

We now claim that if $J = \sum_{i=1}^{n} R[x]\rho_i + \sum_{i=1}^{p} R[x]\gamma_i$ then $I = J$. To see this, first note that $J \subseteq I$ by construction, so suppose that $I \supsetneq J$ and pick $\beta \in I \smallsetminus J$ of smallest possible degree. We cannot have $d = \deg \beta \leq m$ as this would imply that $\beta \in I \cap M \subset \sum R[x]\gamma_i$. Thus, $d \geq m$. However, now we can subtract of the leading term of β : As $\beta \in I$ its leading coefficient, say b lies in I_∞ and we can write $b = \sum \lambda_i r_i$ for some $\lambda_i \in R$. But now

$$\sum x^{d-\deg \rho_i} \lambda_i \rho_i$$

has the same leading term as β, so

$$\gamma = \beta - \sum x^{d-\deg \rho_i} \lambda_i \rho_i$$

has degree $\deg \gamma < d$. But certainly $\gamma \in I$ and so, by the choice of d, $\gamma \in J$. Hence $\beta \in J$. This contradicts our hypothesis and proves that, in fact, $I = J$ and so I is certainly finitely generated. □

As we have seen, at least at the level of elements, the Weyl algebra looks a bit like a polynomial ring, so one might hope that the proof of the Hilbert Basis Theorem works here. It does!

Theorem 3.13. *The Weyl algebra $A = A_1(\mathbb{C})$ is Noetherian.*

Proof: Pick a nonzero left ideal I of A. Each $\alpha \in I$ can be written as $\alpha = \sum_{i=0}^{n} f_i(x)\partial^i$. If $f_n \neq 0$ then we call f_n *the leading coefficient of* α. Let I_∞ be the set of all leading coefficients of elements of I.

Claim I. *The leading coefficient of $\partial^m \alpha$ equals the leading coefficient of α.*

Proof: If $\alpha = \sum_{i=0}^{n} f_i(x)\partial^i$ with $f_n \neq 0$ then Leibniz's rule (Lemma 1.11) says that

$$\partial\alpha = f_n(x)\partial^{n+1} + \text{terms of degree} \leq n-1 \text{ in } \partial,$$

which certainly has leading coefficient $f_n(x)$. The claim therefore follows by induction on m. □

Claim II: *I_∞ is an ideal of $\mathbb{C}[x]$.*

Proof: Let $r, s \in I_\infty$ and suppose that $\rho = r\partial^m + r_{m-1}\partial^{m-1} + \cdots$ and $\sigma = s\partial^n + s_{n-1}\partial^{n-1} + \ldots$ are the corresponding elements of I. By symmetry we may as well assume that $m \geq n$ in which case Claim I shows that $\partial^{m-n}\sigma = s\partial^m +$ (terms of lower degree). Thus

$$I \ \ni \ \rho - \partial^{m-n}\sigma = (r-s)\partial^n + \cdots$$

and so $r - s \in I_\infty$ and I_∞ is an abelian group. If $t \in \mathbb{C}[x]$ then $I \ni t\rho = (tr)\partial^n + \cdots$ so $tr \in I$. Thus, the claim is proved. □

Proof of the theorem. As I_∞ is an ideal of the PID $\mathbb{C}[x]$ it is cyclic, generated say by r which is the leading coefficient of $\rho \in I$. Let $m = \deg \rho$. Then $M = \sum_{\ell=0}^{m} \mathbb{C}[x]\partial^\ell$ is a finitely generated left $\mathbb{C}[x]$-module, and so is Noetherian by Corollary 3.8. Notice that $M \cap I$ is a left $\mathbb{C}[x]$-module as both M and I are closed under multiplication by elements from $\mathbb{C}[x]$. Thus $I \cap M$ is a finitely generated left $\mathbb{C}[x]$-module; say $I \cap M = \sum_{i=1}^{p} \mathbb{C}[x]\gamma_i$ for some $\gamma_i \in I$.

Claim III: *Suppose that* $J = A\rho + \sum_{i=1}^{p} A\gamma_i \subseteq I$. *Then* $I = J$.

Proof: To see this, suppose that $I \supsetneq J$ and pick $\beta \in I \setminus J$ of smallest possible degree d in ∂. We cannot have $d = \deg \beta \leq m$ as this would imply that $\beta \in I \cap M \subset \sum A\gamma_i$. Thus, $d \geq m$. However, now we can subtract of the leading term of β. Formally, as $\beta \in I$ its leading coefficient, say b, lies in I_∞ and we can write $b = \lambda r$ for some $\lambda \in \mathbb{C}[x]$. But now Claim I implies that

$$\theta = \partial^{d-m}\lambda\rho$$

has the same leading term as β. So $\gamma = \beta - \theta$ has degree $\deg \gamma < d$. But $\gamma \in I$ and so $\gamma \in J$ by the choice of d. Hence $\beta \in J$, a contradiction. Thus I is finitely generated. This proves both the claim and the theorem for left ideals. Exactly the same argument shows that A is right Noetherian. □

Comments re examinability: the proofs of neither 3.12 nor 3.13 are examinable (I only summarised them in lectures); on the other hand, those of 3.1 and 3.6 would make reasonable exam questions (in that both have quite natural proofs and, if you've seen and understood these proofs, then you will have a good chance of being able to reconstruct them in the exam).

4. **The nilradical and nilpotent ideals.**

A 1- or 2-sided ideal is nilpotent if some power of it is 0. The nilradical of a ring - the sum of all nilpotent ideals - turns out to be important for understanding the structure of a ring, allowing the ring to be seen as being composed of a semisimple ring (a ring with nilradical = 0) put on top of a nil ideal. Because we are dealing with noncommutative rings there are subtleties not seen in the commutative case - in particular a nilpotent element does not necessarily generate a nilpotent ideal and the sum of nilpotent elements need not be nilpotent.

Recall that a ring R is *left Artinian* if every descending chain of left ideals is eventually stationary and it is *Artinian* if it is left and right Artinian. The aim of the next two chapters is to study the structure of Artinian rings, and to a lesser extent Noetherian rings. Here is a typical example to illustrate the main results in the Artinian case: Let

$$U \;=\; U_3(\mathbb{C}) \;=\; \begin{pmatrix} \mathbb{C} & \mathbb{C} & \mathbb{C} \\ 0 & \mathbb{C} & \mathbb{C} \\ 0 & 0 & \mathbb{C} \end{pmatrix}$$

be the ring of complex upper triangular matrices. Then the ideal

$$I = \begin{pmatrix} 0 & \mathbb{C} & \mathbb{C} \\ 0 & 0 & \mathbb{C} \\ 0 & 0 & 0 \end{pmatrix}$$

satisfies $I^3 = 0$ and

$$U/I \;\cong\; \begin{pmatrix} \mathbb{C} & 0 & 0 \\ 0 & \mathbb{C} & 0 \\ 0 & 0 & \mathbb{C} \end{pmatrix} \;\cong\; \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \tag{4.1}$$

is just a direct sum of simple rings (even fields in this case). We aim to prove something similar in this chapter—given any left Artinian ring R then R has an ideal N with $N^r = 0$ for some $r \geq 0$ and such that $R/N \cong M_{n_1}(D_1) \oplus M_{n_2}(D_2) \oplus \cdots \oplus M_{n_\ell}(D_\ell)$ is a direct sum of matrix rings over division rings.

Exercise. Prove Equation 4.1. Probably the best way of doing this is to find the (rather obvious!) homomorphism from U to the second ring in (4.1) and check that it really is a homomorphism and that it has kernel I. Now apply the first isomorphism theorem for rings.

We begin with some relevant definitions. First, given abelian subgroups A, B of a ring R then, as in Definition 2.29, we write $AB = \{\sum_i a_i b_i : a_i \in A, b_i \in B\}$. The same definition applies if (say) A is a subgroup of R and B is a subgroup of a left R-module M. The following simple facts about this concept are left as an exercise.

Exercise 4.1. (1) If A is a left ideal and B is a right ideal of a ring R then AB is a two-sided ideal of R. In particular, if A and B are both ideals then so is AB.

(2) if A is an ideal of R and B is a submodule of a left R-module M then AB is also a submodule of B.

(3) Given abelian subgroups A_1, A_2, A_3 of R then

$$(A_1A_2)A_3 \;=\; A_1(A_2A_3) \;=\; \left\{ \sum_{i=1}^{t} a_ib_ic_i : t \in \mathbb{N},\ a_i \in A_1,\ b_i \in A_2,\ c_i \in A_3 \right\}.$$

By part (3) and induction we can also define, without ambiguity, $A_1A_2 \cdots A_r$ for subgroups A_j of R.

Definition 4.2. *Let R be a ring. An element $a \in R$ is* **nilpotent** *if $a^n = 0$ for some $n \geq 1$. An abelian subgroup A of R is* **nilpotent** *if $A^n = 0$ for some n. Note that, by distributivity one has*

$$\text{The subgroup } A \text{ is nilpotent} \iff a_1a_2 \cdots a_n = 0 \ \text{ for all } \ a_j \in A. \tag{4.2}$$

The subgroup A is **nil** *if each element $a \in A$ is nilpotent.*

Examples: (1) 0 is a nilpotent element in any ring.

(2) In an integral domain, 0 is the only nilpotent element.

(3) $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ is a nilpotent element in $M_3(\mathbb{Z})$.

(4) Consider $R = \mathbb{C}[x_1, x_2, \cdots]/(x_1, x_2^2, x_3^3, \dots)$. Then I claim that the maximal ideal $M = (x_1, x_2, x_3, \dots)$ is nil but not nilpotent. In order to prove this claim (which I leave as an exercise) the important thing to prove is that M is nil. To prove this you should first prove part (1) of the next lemma.

Lemma 4.3. *(1) If $a^n = b^m = 0$ in a* commutative *ring R then $(a+b)^{n+m-1} = 0$.*

(2) If I is a nilpotent left ideal in a ring S then I is contained in a nilpotent two-sided ideal.

Proof. Part (1) is an exercise. (If you need a hint, see Lemma 4.5, below.)

For part (2) note that, if $I^n = 0$ then IR is an ideal by Exercise 4.1(1) and so

$$(IR)^n = (IR)(IR)\dots(IR) = I(RI)(RI)\dots RI \subseteq I^n = 0$$

by Exercise 4.1(3) and induction. □

The first fundamental fact about Artinian rings is that, for such rings, the distinction between nil and nilpotent left ideals is illusory.

Theorem 4.4. *If I is a nil left ideal in a left Artinian ring R then I is nilpotent.*

Proof. We have a descending chain of *left* ideals $I \supseteq I^2 \supseteq I^3 \cdots$. As R is left Artinian this chain must be eventually stationary; say $I^n = I^{n+1} = \cdots$. If $I^n = 0$ we are done, so assume not and set $J = I^n$. Then J is still nil but now $J^2 = I^{2n} = I^n = J$.

Now let

$$\mathcal{S} \;=\; \{\text{left ideals } A : A \subseteq J \text{ and } JA \neq 0\}.$$

Note that $\mathcal{S} \neq \emptyset$ as $J \in \mathcal{S}$. Thus by Theorem 3.2 there exists a minimal element $M \in \mathcal{S}$. Notice that this means there exists $x \in M$, so $Rx \subseteq M$, such that $Jx \neq 0$ hence such that $J(Rx) \neq 0$. Thus $M = Rx$ by minimality of M. Also, $J(Jx) = J^2x = Jx \neq 0$ and so $Jx = Rx = M$.

But this means that $x = 1x \in Rx = Jx$; say $x = ax$ for some $a \in J$. This in turn implies that

$$x = ax = a(ax) = a^2x = a^3x = \cdots = a^mx$$

for any $m \geq 1$. As J is nil, $a^m = 0$ for some m and hence $x = xa^m = 0$, giving the required contradiction. □

Lemma 4.5. *Given nilpotent left ideals I_j in a ring R then $\sum_{j=1}^{n} I_j$ is also nilpotent.*

Proof. It suffices to prove the result for two left ideals, say I and J. If $I^n = J^m = 0$ consider $(I+J)^{n+m-1}$. This consists of a sum of sets of the form

$$X = I^{a_1}J^{b_1}I^{a_2}\cdots I^{a_r}J^{b_r} \qquad \text{where } \sum a_j + \sum b_\ell = n+m-1.$$

In each such expression, either $\sum b_\ell \geq m$ in which case $X \subseteq J^{b_1}J^{b_2}\cdots J^{b_r} \subseteq J^m = 0$ or $\sum a_j \geq n$ in which case $X \subseteq I^{a_1}I^{a_2}\cdots I^{a_r}J^{b_r} \subseteq I^nJ^{b_r} = 0$. In either case $X = 0$ and hence so is $(I+J)^{n+m-1}$. □

Note that this lemma fails for elements or indeed abelian subgroups of R; for example take

$$I_1 = \begin{pmatrix} 0 & 0 \\ \mathbb{C} & 0 \end{pmatrix} \quad \text{and} \quad I_2 = \begin{pmatrix} 0 & \mathbb{C} \\ 0 & 0 \end{pmatrix} \quad \text{inside} \quad R = \begin{pmatrix} \mathbb{C} & \mathbb{C} \\ \mathbb{C} & \mathbb{C} \end{pmatrix} = M_2(\mathbb{C}).$$

Then of course I_1 and I_2 are nilpotent but $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \in I_1 + I_2$.

Definition 4.6. *Let R be a ring, and let*

$$N(R) = \sum\{I : I \textit{ is nilpotent ideal of } R\},$$

i.e. let $N(R)$ be the sum of all nilpotent ideals of R. The ideal $N(R)$ is called the nilradical *of R.*

Clearly

$$N(R) \subseteq \sum\{I : I \text{ is nilpotent left ideal of } R\}.$$

But the reverse inclusion also holds since every nilpotent left ideal I is contained in a nilpotent ideal IR (see Lemma 4.3). Hence

$$N(R) \;=\; \sum\{I : I \text{ is nilpotent left ideal of } R\} \;=\; \sum\{I : I \text{ is nilpotent right ideal of } R\}.$$

Proposition 4.7. *Let R be a ring. Then $N(R)$ is a nil ideal of R.*

Proof. Let $x \in N(R)$. Then $x \in I_1 + I_2 + \cdots + I_t$ some *finite* set of nilpotent ideals I_j. By Lemma 4.5 $I_1 + I_2 + \cdots + I_t$ is a nilpotent ideal. So x is nilpotent and $N(R)$ is a nil ideal. □

The example after Definition 4.2 shows that $N(R)$ need not be nilpotent in an arbitrary ring R. However, once again, things are nicer for Artinian rings:

Theorem 4.8. *Let R be a left Artinian ring. Then*

(1) *$N(R)$ is a nilpotent ideal of R,*

(2) *$R/N(R)$ is a left Artinian ring with no nonzero nil or nilpotent left ideals.*

Proof. (1) By Proposition 4.7 and Theorem 4.4, $N(R)$ is a nilpotent ideal.

(2) $R/N(R)$ is a left Artinian ring by Theorem 3.6. Suppose that $R/N(R)$ has a nil left ideal $\overline{I}$. By the Correspondence Theorem for Rings, $\overline{I}$ has the form $I/N(R)$, where I is a left ideal of R containing $N(R)$.

Let $a \in I$. Then there is a positive integer m such that $(a + N(R))^m = 0$; equivalently $a^m \in N(R)$. Since $N(R)$ is nil, there is a positive integer n such that $a^{mn} = (a^m)^n = 0$. Therefore I is a nil left ideal and hence is nilpotent by Theorem 4.4. Thus $I \subseteq N(R)$ and $I/N(R) = 0$. This shows that $R/N(R)$ has no nonzero nil left ideals and hence no nonzero nilpotent left ideals. □

Remark: The theorem shows that, for a left Artinian ring R, $N(R)$ is the unique largest nilpotent ideal of R. By Lemma 4.3 it is therefore also the largest nilpotent right ideal and the largest nilpotent left ideal.

Example 4.9. *(formerly Example 4.10) (1) Consider the ring of upper triangular matrices $U = U_3(\mathbb{C})$ from the beginning of the chapter, with ideal I of strictly upper triangular matrices. Then certainly $I \subseteq N(U)$. If $I \neq N(U)$, then $N(U)/I$ would be a non-zero nil ideal of U/I. On the other hand* (4.1) *shows that U/I has no nilpotent elements, contradicting Theorem 4.8(2). Hence $I = N(U)$.*

(2) Given the ring $\overline{\mathbb{Z}} = \mathbb{Z}/a\mathbb{Z}$ for some $a \in \mathbb{Z}$, write $a = \prod_{i\in I} p_i^{n_i}$ for distinct primes p_i and let $b = \prod_{i\in I} p_i$. Then $a|b^n$ for large n and so $(b\mathbb{Z})^n \subseteq a\mathbb{Z}$; equivalently $(b\overline{\mathbb{Z}})^n = 0$. On the other hand,

$$\overline{\mathbb{Z}}/b\overline{\mathbb{Z}} \;\cong \mathbb{Z}/b\mathbb{Z} \;\cong \bigoplus_{i\in I} \mathbb{Z}/p_i\mathbb{Z}$$

is a direct sum of fields. Thus $\overline{\mathbb{Z}}/b\overline{\mathbb{Z}}$ has no nilpotent elements. It follows (why?) that $b\overline{\mathbb{Z}} \supseteq N(\overline{\mathbb{Z}})$ and hence that $b\overline{\mathbb{Z}} = N(\overline{\mathbb{Z}})$.

(3) Explicitly, in $\mathbb{Z}/200\mathbb{Z}$ the nilradical is $10\mathbb{Z}/200\mathbb{Z}$.

Let us make the example a bit more complicated.

(4) Write $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$. We want to consider $R = \begin{pmatrix} \mathbb{Z}_9 & \mathbb{Z}_3 \\ 0 & \mathbb{Z}_3 \end{pmatrix}$, but we better start by explaining how it is a ring! We do this by saying that by definition *$R = S/I$ where S is the ring $\begin{pmatrix} \mathbb{Z} & \mathbb{Z} \\ 0 & \mathbb{Z} \end{pmatrix}$ and I is the ideal*

$\begin{pmatrix} 9\mathbb{Z} & 3\mathbb{Z} \\ 0 & 3\mathbb{Z} \end{pmatrix}$ *of* S*. Of course, for this to work one does need to check that* I *is indeed an ideal of* S*. And this is perhaps not completely obvious, though it is easy to check (do it!).*

Clearly R *is an Artinian ring, since it is a finite set. However, what is* $N(R)$ *and how does* $R/N(R)$ *decompose?*

Very often the best approach is to "guess" and check what is $N(R)$*; more formally find a nilpotent ideal and then prove that the factor ring has no nilpotence. Let me explain what I mean with this example. It is easy to see that* $J = \begin{pmatrix} 3\mathbb{Z} & \mathbb{Z} \\ 0 & 3\mathbb{Z} \end{pmatrix}$ *is an ideal of* S *with* $J^2 = \begin{pmatrix} 9\mathbb{Z} & 3\mathbb{Z} \\ 0 & 9\mathbb{Z} \end{pmatrix} \subset I$*. Since nothing else comes to mind, we might guess therefore that* $J/I = N(R)$*. So, let's prove it.*

Thus, J/I *is a nilpotent ideal of* R *and* $\overline{J} = J/I \subseteq N(R)$*. Next, by the appropriate isomorphism theorem,* $R/\overline{J} = (S/I)/J/I) \cong S/J$*. As in the proof of Equation 4.1 (which I hope you did) we see that*

$$R/\overline{J} \;\cong\; S/J \;\cong\; \begin{pmatrix} \mathbb{Z}_3 & 0 \\ 0 & \mathbb{Z}_3 \end{pmatrix} \;\cong\; \mathbb{Z}_3 \oplus \mathbb{Z}_3.$$

Thus, $R/\overline{J}$ *has no nonzero nilpotent ideals. But, if* $\overline{J} \subsetneq N(R)$ *then* $N(R)/\overline{J}$ *would be a* nonzero *nilpotent ideal of* $R/\overline{J}$*, giving a contradiction. Hence we conclude that* $\overline{J} = N(R)$ *and that* $R/N(R) \cong \mathbb{Z}_3 \oplus \mathbb{Z}_3$*.*

The remaining material of this chapter is optional, in particular non-examinable. Some of it - which you should try for practice with various concepts - does appear on Examples Sheet 5.

As we next show, both Theorems 4.4 and 4.8 hold with "Artinian" replaced by "Noetherian." Since we will later see (in Example Sheet 7) that any Artinian ring is noetherian, this says that in one sense we did not need to prove Theorems 4.4 and 4.8. However, since we use them to prove that an Artinian ring is noetherian, we do still need them both!

Theorem 4.10. (Levitsky's Theorem) *Let* S *be a left Noetherian ring. Then every nil one-sided ideal of* S *is nilpotent.*

Moreover the nilradical $N(S)$ *is nilpotent and is the unique largest nilpotent ideal in* S*.*

Remark. It might seem more natural to prove the theorem just for left ideals. But strangely, it is easier to prove it for right ideals!

Proof. We start by considering $N(S)$. First let N be maximal among nilpotent ideals of S, say with $N^t = 0$. Note that, by Lemma 4.5, $N = N(S)$. If S/N has a nilpotent ideal, say U/N then some $(U/N)^r = 0$. Hence $U^r \subseteq N$ and so $U^{rt} \subseteq N^t = 0$, giving a contradiction. Thus S/N has no nilpotent ideals.

If S has a nil (left or right) ideal I, then $(I + N)/N$ is still a nil one-sided ideal in S/N. So, if we prove that S/N has no nil one-sided ideals then every nil one-sided ideal of S must be contained in N and hence be nilpotent.

We can therefore replace S by S/N and assume that S has no nilpotent ideals. Suppose, for a contradiction, that S does have a non-zero nil left ideal I. For any $x \in I$ and $r \in S$ then $(xr)^{n+1} = x(rx)^n r = 0$ for $n \gg 0$, and so xS is a nonzero nil *right* ideal.

So S has a non-zero nil right ideal J. For $x \in J$ the left annihilator $l - ann_S(x) = \{r \in S : rx = 0\}$ is a left ideal of S and so, since S is left noetherian, we can pick $0 \neq u \in J$ for which $l - ann(u)$ is maximal among such left annihilators. Given any $r \in S$, then $l - ann(ur) \supseteq l - ann(u)$ and so, by maximality, $l - ann(u) = l - ann(ur)$ for $r \in R$ such that $ur \neq 0$.

We now claim that $usu = 0$ for all $s \in S$. Certainly this is true when $us = 0$, so suppose that $us \neq 0$. Then $us \in J$ and so $(us)^m = 0$ for some m, where we can choose m so that $(us)^{m-1} \neq 0$. As $us \neq 0$, certainly $m > 1$. But now $(us) \in l - ann(us)^{m-1} = l - ann(u)$ by the last paragraph. In other words $usu = 0$, as claimed. In particular this means that $(SuS)^2 = S(uSu)S = 0$, contradicting the fact that S has no nilpotent ideals.

This contradiction proves that S has no nil one-sided ideals, as desired. □

Recall that in a commutative ring C an ideal P is *prime* provided that, whenever $xy \in P$ for $x, y \in C$, then either $x \in P$ or $y \in P$. Crucially, maximal ideals are prime (see below). On the other hand, in the matrix ring $M_2(\mathbb{C})$ the ideal 0 is a maximal ideal, simply because there are no other proper ideals. Yet one certainly has elements $x, y \neq 0$ with $xy = 0$. This suggests that the commutative definition of primality is not quite right in the noncommutative universe. Instead we will use the following definition:

Definition 4.11. *An ideal I of a ring R is* prime *if, whenever A and B are ideals of R with $AB \subseteq I$, then either $A \subseteq I$ or $B \subseteq I$.*

The following exercises will give you some feel for this concept.

Exercises 4.12. (1) The ideal I of R is **prime** if and only if it satisfies the following condition: For all elements $x, y \in R$ if $xRy \subseteq I$ then either $x \in I$ or $y \in I$. In particular, if R is commutative then this reduces to the familiar definition.

[The point is that $xRy \subseteq I \iff (RxR)(RyR) \subseteq I$.]

(2) The following tends to be a very useful refinement of the definition: Prove that an ideal I of a ring R is *prime* if, whenever A and B are ideals of R with $I \subseteq A$ and $I \subseteq B$ but $AB \subseteq I$, then either $A = I$ or $B = I$.

[The point of course is that, in the definition of primality one can always replace A, B by $(A + I)$ and $(B + I)$.

(3) Any maximal ideal in a ring R is prime.

[Use part (2): the only ideal strictly containing a maximal ideal is the ring itself.]

Theorem 4.13. *Let R be a ring for which $N(R/N(R)) = 0$. Then*

$$N(R) \;=\; \bigcap\{\textit{all prime ideals } P \textit{ of } R\}.$$

Remark: (1) We will show eventually that in a left Artinian ring all prime ideals are maximal. Hence this theorem implies that, for a left Artinian ring R, the intersection of the maximal ideals is nilpotent. Once again, in the Noetherian universe there are counterexamples to such a statement; just take the power series ring $R = \mathbb{C}[[x]]$.

(2) The assumption that $N(R/N(R)) = 0$ is annoying but necessary—without the assumption counterexamples exist. These are due to Amitsur and are beyond this course. However, when R is Artinian the assumption is automatically satisfied (see Theorem 4.8) as is the case when R is left noetherian (see the next examples sheet).

Proof of Theorem 4.13. Playing with nilpotent elements is not quite enough to prove this result, so we introduce a stronger concept:

Definition 4.14. *An element α in a ring R is* strongly nilpotent $\iff$ *for all sequences*

$$\alpha_0 = \alpha, \quad \alpha_1 = \alpha_0 r_1 \alpha_0, \quad \alpha_2 = \alpha_1 r_2 \alpha_1, \quad \dots$$

(where the r_j are arbitrary elements of R) we have $\alpha_r = 0$ for all $r \gg 0$.

Obviously any strongly nilpotent element is nilpotent, but the converse fails—just take $\alpha = e_{12} \in M_2(\mathbb{C})$. However, if α belongs to a nilpotent ideal I then certainly the above elements $\alpha_m \in I^{m+1}$ for all m and so α is strongly nilpotent. Consequently, by Lemma 4.5, any $\alpha \in N(R)$ is strongly nilpotent.

We return to the proof of the theorem.

Step I: Set $N'(R) = \bigcap\{\text{all prime ideals } P \text{ of } R\}$. Then we claim that $N'(R)$ is the set of all strongly nilpotent elements in R.

Proof of Step I: If α is not in $N'(R)$ then α is not in some prime ideal P. So, $\alpha R \alpha \notin P$ by the above exercises. Hence there exists $r_1 \in R$ such that $\alpha_1 = \alpha r_1 \alpha \notin P$. Now we can repeat this process and find $r_2 \in R$ such that $\alpha_2 = \alpha_1 r_2 \alpha_1 \notin P$. By induction we obtain an infinite such chain and so α is not strongly nilpotent. Consequently

$$N'(R) \supseteq \{\text{the strongly nilpotent elements of } R\}.$$

Conversely, suppose that α is *not* strongly nilpotent and pick the corresponding chain of **nonzero elements**

$$\mathcal{S} = \{\alpha_0 = \alpha,\ \alpha_1 = \alpha r_1 \alpha,\ \alpha_2 = \alpha_1 r_2 \alpha_1,\ \alpha_3 = \alpha_2 r_3 \alpha_2,\ \cdots\}.$$

Now let $\mathcal{I} = \{\text{ideals } I : I \cap \mathcal{S} = \emptyset\}$; note that $\mathcal{I} \neq \emptyset$. It is immediate that Zorn's Lemma applies, so $\mathcal{I}$ has a maximal element; say I. We claim that I is prime. If not, then by Exercise 4.12 we can find ideals $A, B \supsetneq I$ such that $AB \subseteq I$. But this means that $A \ni \alpha_i$ and $B \ni \alpha_j$ for some $i, j \geq 0$. Notice that if $A \ni \alpha_i$

then certainly $A \ni \alpha_{i+1}$ and hence $A \ni \alpha_{i+m}$ for all m. Thus, if $\ell = \max\{i, j\}$ then $A \cap B \ni \alpha_\ell$. But now $\alpha_{\ell+1} = \alpha_\ell r_{\ell+1} \alpha_\ell \in AB \subseteq I$; a contradiction.

Thus I is indeed prime and hence, as $\alpha \notin I$ this implies that $\alpha \notin N'(R)$. This in turn says that

$$N'(R) \subseteq \{\text{the strongly nilpotent elements of } R\}$$

and completes the proof of Step 1.

Step II: $N(R) = \{\text{the strongly nilpotent elements of } R\}$.

Proof of Step II: As we remarked at the beginning of the proof any element $\alpha \in N(R)$ is strongly nilpotent, so $\subseteq$ holds. So, suppose that $N(R) \subsetneq \{\text{the strongly nilpotent elements of } R\}$.

We consider $\overline{R} = R/N(R)$. Note that $N(\overline{R}) = 0$ by the assumption of Theorem 4.13. Now any strongly nilpotent element of R certainly remains strongly nilpotent in $\overline{R}$ and so $\overline{R}$ contains a *nonzero* strongly nilpotent element, say $\alpha \in \overline{R}$. As $N(\overline{R}) = 0$, it must be that $\overline{R}(\alpha\overline{R}\alpha)\overline{R} = (\overline{R}\alpha\overline{R})^2 \neq 0$ and so we can find $r \in \overline{R}$ such that $\alpha_1 = \alpha r_1 \alpha \neq 0$. But then $(\overline{R}\alpha_1\overline{R})^2 \neq 0$ and so we can repeat the game to find $\alpha_2 = \alpha_1 r_2 \alpha_1 \neq 0$. Inducting on this procedure we find an infinite sequence of nonzero elements $\{\alpha_n = \alpha_{n-1} r_n \alpha_{n-1}\}$ and so α is not strongly nilpotent. This contradiction proves Step II.

Obviously the theorem follows from the results of the two steps. □

Comments re examinability: The proofs of 4.4 and of 4.8(2) are potential exam questions; note that the first is a little tricky, so having understood it might not be enough to (quickly) reconstruct it - remembering some key step would help. I am not listing shorter, "one-(perhaps long)line", exam-suitable arguments like those for 4.3.

Note that the material after Example 4.9 was not covered so is not examinable.

5. **Artinian Rings**

A lot comes together in this section, where the main result is the Artin-Wedderburn Theorem; that identifies the semisimple artinian rings as being exactly the finite products of matrix rings over division rings. Furthermore, over these rings every (finitely generated) module is a direct sum of simple modules, each of which is isomorphic to a minimal left ideal of the ring. So we have a strong structure theorem for the ring and for its modules.

Idempotent elements of a ring - elements e satisfying $e^2 = e$ - play a key role since they induce decompositions of a ring as a direct sum of two right (or left) ideals; if the idempotent is central this is actually a decomposition of the ring as a direct sum of two-sided ideals, hence a decomposition as a direct product of two (non-unital) subrings. (I'm using the terms product and sum interchangably here.)

We now return to the study of Artinian rings. Most of this chapter is involved with the structure theorem for Artinian rings R satisfying $N(R) = 0$ (these are called semisimple Artinian rings). Recall that an **idempotent** is an element e in a ring R such that $e = e^2$. As a bit of notation, a left ideal I of a ring R will be called **a minimal left ideal** if it is not zero but minimal in the collection of all non-zero left ideals. In a general ring (even when $R = \mathbb{Z}$) these need not exist. We sometimes call them simple left ideals—as they are the same as simple submodules of R regarded as a left R-module.

Lemma 5.1. (1) *If R is a left Artinian ring then R has minimal left ideals.*

(2) *If I is a minimal left ideal of any ring R then* either $I^2 = 0$ or $I = Re$ *for some idempotent e.*

(3) *If $J = Re$ for any ring R, with $e = e^2$, then $J = Je$ and $R = J \oplus R(1 - e)$.*

Proof. (1) is obvious.

(2) Assume that $I^2 \neq 0$. Then $Ia \neq 0$ for some $a \in I$ and, since $I \supseteq Ia$, the minimality of I forces $I = Ia$. Thus $a = ea$ for some $e \in I$ and we will show that e is our desired idempotent.

First, let $J = \{f \in I : fa = 0\}$. Then J is a left ideal and $J \subseteq I$ by construction. Since $Ia \neq 0$, we also know that $J \neq I$ and so, by the minimality of I, we have $J = 0$. Now from $a = ea$ we obtain $a = ea = e^2a$ and hence $(e - e^2)a = 0$. Thus $e - e^2 \in J = 0$ and $e = e^2$. Also, as $e.e = e \neq 0$, clearly $Ie \neq 0$ and so, by minimality, $I = Ie$. Consequently $I = Re$.

(3) First note that $Re \cap R(1-e) = 0$; indeed if $xe = y(1-e)$ for some $x, y \in R$ then $xe = xe^2 = y(1-e)e = y(e - e^2) = 0$. On the other hand, clearly $1 \in Re + R(1 - e)$ and so $R = Re \oplus R(1 - e)$ by Corollary 2.27. Now $J = Re = Re^2 = (Re)e = Je$. □

Theorem 5.2. *Let R be a left Artinian ring with $N(R) = 0$. Then $R = I_1 \oplus I_2 \oplus \cdots \oplus I_t$ for some finite set of simple left ideals I_j. More generally, every left ideal J of R can be written as a direct sum of finitely many simple left ideals.*

Proof. If the result fails then there exists a left ideal J minimal with respect to J not being a finite direct sum of simple left ideals. Clearly J contains a minimal left ideal I and, as $I^2 \neq 0$ by hypothesis, Lemma 5.1 implies that $I = Ie$ for some idempotent $e \in I$. Now $R = Re \oplus R(1-e)$ and so, as $J \supseteq Re$, the Modular Law (Lemma 2.30) says that

$$J \;=\; J \cap (Re + R(1-e)) \;=\; Re + (J \cap R(1-e)).$$

Since $Re \cap (J \cap R(1-e)) = 0$, Corollary 2.27 implies that $J = Re \oplus L$ for $L = J \cap R(1-e)$. Clearly $J \supsetneq L$ and so, by the minimality of J, we can write L as a finite direct sum $L = N_1 \oplus \cdots \oplus N_r$ of minimal left ideals. Hence $J = Re \oplus N_1 \oplus \cdots \oplus N_r$ is also such a direct sum, (contradiction) as required. □

We will see later that a left Artinian ring R has $N(R) = 0 \iff R$ can be written as a finite direct sum of simple left ideals, but this will require a discussion about decompositions of modules. We begin by seeing what Theorem 5.2 tells us about the structure of modules over this ring R.

We now want to show that Theorem 5.2 implies that a left artinian ring R with $N(R) = 0$ is automatically a direct sum of matrix rings over division rings (the converse is an easy exercise). The proof of this is largely contained in several results from the exercise sheets, which I will first remind you about.

First to save repetition we make a definition:

Definition 5.3. *A left artinian ring R with $N(R) = 0$ is said to be* semsimple left Artinian.

In fact a semsimple left Artinian is automatically right Artinian (see Corollary 5.15 below) and so these rings are usually called semisimple Artinian. The term "semisimple" really refers to the conclusion of Theorem 5.2 since a module is often called semisimple if it is a direct sum of simple modules. Or it can refer to the fact that the ring is a direct sum of simple artinian rings—it does not matter since, as we will see these conditions are all the same. Indeed, different books may use different facets of these equivalent conditions as the formal definition of the term.

Remark 5.4. Recall that the set of all endomorphisms of an R-module M is a ring under composition and addition of functions (see Example Sheet 2 for the details). One should note that endomorphism rings are unaffected by taking isomorphic modules in the sense that

(5.1) *If $M \cong N$ as R-modules, then* $\mathrm{End}_R(M) \cong \mathrm{End}_R(N)$ *as rings.*

Indeed, suppose that $\theta : M \to N$ is the isomorphism and $\rho \in \mathrm{End}(M)$. Then as the composition of R-module homomorphisms is an R-module homomorphism, we find that

$$\theta \circ \phi \circ \theta^{-1} : N \to M \to M \to N$$

is an R-module homomorphism from N to N. In other words, $\phi \mapsto \theta\phi\theta^{-1}$ defines a map $\chi : \mathrm{End}(M) \to \mathrm{End}(N)$. The map χ is readily checked to be a ring homomorphism and as it has inverse $\phi \mapsto \theta^{-1}\phi\theta$ Equation 5.1 is proved. □

A particular case of (5.1) is that, as rings, $\mathrm{End}_R(M \oplus N) \cong \mathrm{End}_R(N \oplus M)$.

Let us recap three results from the Exercise sheets. First a notational point.

Remember from Example Sheet 2 the notational point that if one is dealing with a right module M then it is natural to write endomorphisms on the left; thus $\theta\phi(m) = \theta(\phi(m))$ for all $\theta, \phi \in \mathrm{End}_R(M)$ and $m \in M$. However, when one works with left modules it is better to write the endomorphisms on the right, say writing m^θ for $\theta(m)$. In this case one naturally uses the opposite convention for products: $m^{\theta\phi} = (m^\theta)^\phi$. As the next lemma shows, this gives the most natural expression for endomorphism rings. It does not really matter which of these two conventions one uses, but it is best to be consistent, since changing from the one to the other will replace an endomorphism ring by its "opposite" ring in the notation of Example Sheet 2.

Lemma 5.5. *If we regard a ring R as a left R-module and write endomorphisms on the right then* $\mathrm{End}({}_RR) \cong R$.

Proof. This is Example sheet 2 number 5, but for the record, here is the proof.

Given $r \in R$ let $\theta_r \in \mathrm{End}({}_RR)$ be the map defined by $s^{\theta_r} = sr$ for $s \in R$. We need to check that this is an R-module endomorphism. It is clearly a map of abelian groups. But $t(s^{\theta_r}) = t(sr) = (ts)r = (ts)^{\theta_r}$ for all $t, s \in R$, as required. We next check that the map $r \mapsto \theta_r$ is a ring homomorphism. Note that $s^{\theta_r\theta_p} = (s^{\theta_r})^{\theta_p} = (sr)^{\theta_p} = srp = s^{\theta_{rp}}$ for $r, p \in R$. Similarly $s^{(\theta_r+\theta_p)} = s^{\theta_r} + s^{\theta_p} = s(r+p) = s^{\theta_{r+p}}$. Finally $s^{\theta_1} = s$ for all $s \in R$ and so $\theta_1 = 1$ and we do indeed have a ring homomorphism.

Conversely, given any endomorphism ϕ of the module R, we map $\phi \mapsto 1^\phi = \phi(1) \in R$. Since $s^{\theta_{\phi(1)}} = s(1^\phi) = (s.1)^\phi = s^\phi$ it follows that the two operations are inverse to each other and we are done. □

Proposition 5.6. *(a)* (Schur's Lemma) *If S is a simple left module over a ring R then* $\mathrm{End}_R(S)$ *is a division ring.*

(b) Suppose that S, T are simple left R-modules with $\mathrm{Hom}_R(S,T) \neq 0$. *Then $S \cong T$.*

Proof. (a) This appeared on Example Sheet 4, but here is the argument.

As above, $\mathrm{End}(S)$ is a ring and if $\theta \in \mathrm{End}(S)$ is nonzero, then $\theta(S)$ is a nonzero submodule hence equal to S. Similarly $\ker(\theta)$ is a submodule of S that cannot equal S as that would force $\theta = 0$. Hence $\ker(\theta) = 0$ and θ is an automorphism. As was observed before Theorem 2.13, θ^{-1} is an R-module endomorphism of S and so inverses exist in $\mathrm{End}(S)$.

(b) Use a similar argument: Suppose that $0 \neq \theta \in Hom_R(S,T)$. Then $Ker(\theta) \neq S$ and so, as $Ker(\theta)$ is a submodule of the simple module S, $Ker(\theta) = 0$. Similarly, as $Im(\theta)$ is a submodule of T, we see that $Im(\theta) \neq 0$ and so $Im(\theta) = T$. Thus $S \cong T$. □

Proposition 5.7. *If $M = S \oplus S \oplus \cdots \oplus S = S^{(r)}$ is the direct sum of r copies of an R-module S, then* $\mathrm{End}_R(M) \cong M_r(D)$ *where $D = \mathrm{End}_R(S)$.*

In particular, if the S_i are simple left R-modules then Schur's Lemma implies that $\mathrm{End}_R(M) \cong M_r(D)$ *for a division ring D.*

Proof. For right modules, this is just the same as the proof of Example Sheet 3, Number 3, but for completeness, here is the proof. We first write $M = S_1 \oplus S_2 \oplus \cdots \oplus S_r$ (where $S_i = S$ for each i) in order to distinguish the different copies of S. In fact the first part of the argument works more generally, so we will write it as

Sublemma 5.8. *Let $S_1, \cdots, S_r$ be right modules over a ring R (where we no longer assume that $S_i \cong S_j$) and set $M = S_1 \oplus \cdots \oplus S_r$. Then we can identify $\mathrm{End}_R(S_1 \oplus S_2 \oplus \cdots \oplus S_r)$ as a "matrix ring"*

$$\mathrm{End}_R(S_1 \oplus S_2 \oplus \cdots \oplus S_r) \;\cong\; \begin{pmatrix} \mathrm{Hom}(S_1,S_1) & \mathrm{Hom}(S_2,S_1) & \cdots & \mathrm{Hom}(S_r,S_1) \\ \mathrm{Hom}(S_1,S_2) & \cdots & & \mathrm{Hom}(S_r,S_2) \\ \vdots & & & \vdots \\ \mathrm{Hom}(S_1,S_r) & \cdots & & \mathrm{Hom}(S_r,S_r) \end{pmatrix} \;=\; X, \tag{5.2}$$

say. The expression on the right has the natural matrix multiplication: If $\Theta = (\theta_{ij})$ and $\Phi = (\phi_{ij})$, then $\Theta\Phi = \Xi = (\xi_{ij})$ where $\xi_{ij} = \sum_k \theta_{ik}\phi_{kj}$.

As usual, the same result holds for left modules, except that matrices act by right multiplication and so (5.2) *gets replaced by its transpose.*

Proof of the Sublemma. Write $m \in M$ as a column vector $m = (s_1, ..., s_r)^T$, for $s_i \in S_i$. Then any $\theta \in End(M)$ satisfies

$$\theta \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{pmatrix} \;=\; \begin{pmatrix} \sum_j \theta_{1j}(s_j) \\ \sum_j \theta_{2j}(s_j) \\ \vdots \\ \sum_j \theta_{rj}(s_j) \end{pmatrix}, \tag{5.3}$$

for some maps $\theta_{ij} : S_j \to S_i$. We first check that this is indeed a (set-theoretic) map from $\mathrm{End}_R(M)$ to X. To see this, multiply (5.3) on the right by some $r \in R$, and apply the rule $\theta(mr) = \theta(m)r$ for $m = (0, 0..., 0, s_j, 0, ..., 0)^T$. This gives:

$$\theta(mr) \;=\; \theta \begin{pmatrix} s_1 r \\ s_2 r \\ \vdots \\ s_r r \end{pmatrix} \;=\; \begin{pmatrix} \sum_j \theta_{1j}(s_j r) \\ \sum_j \theta_{2j}(s_j r) \\ \vdots \\ \sum_j \theta_{rj}(s_j r) \end{pmatrix}.$$

On the other hand,

$$\theta(mr) \;=\; \theta(m)r \;=\; \begin{pmatrix} \sum_j \theta_{1j}(s_j) \\ \sum_j \theta_{2j}(s_j) \\ \vdots \\ \sum_j \theta_{rj}(s_j) \end{pmatrix} r \;=\; \begin{pmatrix} \sum_j \theta_{1j}(s_j) r \\ \sum_j \theta_{2j}(s_j) r \\ \vdots \\ \sum_j \theta_{rj}(s_j) r \end{pmatrix}.$$

Comparing these expressions in the case where just one s_j is nonzero, ensures that each θ_{ij} is an R-module homomorphism. Therefore, the map $\Lambda : \theta \mapsto \Lambda_\theta = (\theta_{ij}) \in X$ provides a natural set-theoretic identification of $\mathrm{End}_R(M)$ with X, as we claimed.

The proof that we have a ring isomorphism is the same argument as was used in Example 3 of Example Sheet 3. To be precise, given $\theta, \phi \in \mathrm{End}(M)$, notice that in matrix notation (5.3) can be expressed as

$$\theta(\underline{s}^T) = \theta \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{pmatrix} = \begin{pmatrix} \sum_j \theta_{1j}(s_j) \\ \sum_j \theta_{2j}(s_j) \\ \vdots \\ \sum_j \theta_{rj}(s_j) \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1r} \\ \theta_{21} & \cdots & & \theta_{2r} \\ \vdots & & & \vdots \\ \theta_{r1} & \cdots & & \theta_{rr} \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{pmatrix} = \Lambda_\theta \cdot \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{pmatrix}, \tag{5.4}$$

where the final product is the usual matrix multiplication. Similarly, write $\phi(\underline{s}^T) = (\phi_{ij})(s_1, \cdots, s_r)^T$. Now multiply out

$$\Lambda_\theta \Lambda_\phi \cdot \underline{s}^T \;=\; \theta\Big(\phi \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{pmatrix}\Big) \;=\; \theta \begin{pmatrix} \sum_j \phi_{1j}(s_j) \\ \sum_j \phi_{2j}(s_j) \\ \vdots \\ \sum_j \phi_{rj}(s_j) \end{pmatrix} \;=\; \begin{pmatrix} \sum_{kj} \theta_{1k}\phi_{kj}(s_j) \\ \sum_{kj} \theta_{2k}\phi_{kj}(s_j) \\ \\ \vdots \\ \\ \sum_{kj} \theta_{rk}\phi_{kj}(s_j) \end{pmatrix} \;=\; \Lambda_{\theta\phi} \cdot \underline{s}^T.$$

Of course this is exactly what one gets by multiplying out the matrices $(\theta_{kj})(\phi_{ji})\underline{s}^T$, which in turn is exactly what we need to prove to show that Θ is a ring homomorphism (as usual the fact that $\Theta(1) = 1$ and that our map respects addition is routine).

The proof for left modules is exactly the same, except that, as one should now write endomorphisms on the right it all looks less natural. Let's at least write it out for once. Now, given $m = (s_1, \ldots, s_r) \in M = S_1 \oplus \cdots \oplus S_r$, for left R-modules S_i and $\theta \in \mathrm{End}(M)$ then we can write $m^\theta = m\theta$ as the vector m multiplied on the right by a matrix: $m^\theta = (s_1, \ldots, s_r)\left(\theta_{ij}\right)$ where now $\theta_{ij} \in \mathrm{Hom}(S_i, S_j)$. So (5.2) gets replaced by its transpose:

$$\mathrm{End}_R(S_1 \oplus S_2 \oplus \cdots \oplus S_r) \;\cong\; \begin{pmatrix} \mathrm{Hom}(S_1, S_1) & \mathrm{Hom}(S_1, S_2) & \cdots & \mathrm{Hom}(S_1, S_r) \\ \mathrm{Hom}(S_2, S_1) & \cdots & & \mathrm{Hom}(S_2, S_r) \\ \vdots & & & \vdots \\ \mathrm{Hom}(S_r, S_1) & \cdots & & \mathrm{Hom}(S_r, S_r) \end{pmatrix} \tag{5.5}$$

Of course, all one is really doing in this second case is applying the rule $(\Theta\underline{v})^T = \underline{v}^T\Theta^T$. □

We return to the proof of the proposition, so now $S_i \cong S_j$ for all i, j. Therefore $\theta_{ij} \in \mathrm{Hom}(S_j, S_i) = \mathrm{Hom}(S, S) = \mathrm{End}(S) = D$. for all i, j and so $\mathrm{End}(S^{(r)}) \cong M_r(D)$. □

Corollary 5.9. *The following are equivalent for a ring R:*

(1) *$R \cong M_n(D)$ for a division ring D.*

(2) $R = S_1 \oplus \cdots \oplus S_n$ *where each* S_j *is a copy of the same simple left* R*-module* S*.*

Proof. The proof is left as an exercise. As a hint, note that R is a direct sum of its columns C_j. So, check that these columns are simple left R-modules. *Right* multiplication by the matrix unit e_{ij} gives the desired isomorphism of left R-modules $C_i \cong C_j$ □

We now want to jazz this proof up to the case coming from Theorem 5.2, where the simple modules are no longer all isomorphic.

Lemma 5.10. *Let* R *be a ring with two* non-isomorphic *simple left* R*-modules* S_1 *and* S_2 *and write* $A = S_1^{(n)}$ *for the direct sum of* n *copies of* S_1 *and* $B = S_2^{(m)}$*, for some* m*. Then,* $\mathrm{Hom}_R(A, B) = 0$*.*

Proof. If $0 \neq \theta \in \mathrm{Hom}_R(A, B)$, then there must exist some $s = (0, \ldots, s_i, 0, \ldots, 0)$, with a non-zero i^{th} entry such that $\theta(s) \neq 0$. But now some entry, say the j^{th} entry of $\theta(s)$ is non-zero. But if π_j denotes the projection onto this j^{th} entry, then $\phi : s_i \mapsto \pi_j\theta((0, 0, \ldots, s_i, 0 \ldots, 0))$ is a non-zero R-module homomorphism $\phi \in \mathrm{Hom}(S_1, S_2)$, contradicting Proposition 5.6(2). □

Theorem 5.11. *Suppose that* $S_1, S_2, \cdots, S_r$ *are pair-wise nonisomorphic simple left modules over a ring* R *and let* $n_1, \cdots, n_r$ *be natural numbers. Then*

$$\mathrm{End}_R(S_1^{(n_1)} \oplus \cdots \oplus S_r^{(n_r)}) \;\cong\; \bigoplus_{i=1}^{r} \mathrm{End}(S_i^{(n_i)}) \;\cong\; \bigoplus_{i=1}^{r} M_{n_i}(D_i),$$

where $D_i = \mathrm{End}_R(S_i)$ *for each* i *and each isomorphism in the display is an isomorphism of rings.*

Proof. The second isomorphism in the display follows from Proposition 5.7 applied to each term separately, so only the first isomorphism needs proof. Here the only difficulty is organising the notation.

Write $A_j = S_j^{(n_j)}$ for each j; thus we are trying to understand $E = \mathrm{End}_R(A_1 \oplus A_2 \oplus \cdots \oplus A_r)$. Now by Sublemma 5.8 in the form (5.5)

$$\mathrm{End}_R(A_1 \oplus A_2 \oplus \cdots \oplus A_r) \cong \begin{pmatrix} \mathrm{Hom}(A_1, A_1) & \mathrm{Hom}(A_1, A_2) & \cdots & \mathrm{Hom}(A_1, A_r) \\ \mathrm{Hom}(A_2, A_1) & \cdots & & \mathrm{Hom}(A_2, A_r) \\ \vdots & & & \vdots \\ \mathrm{Hom}(A_r, A_1) & \cdots & & \mathrm{Hom}(A_r, A_r) \end{pmatrix}.$$

But, by Lemma 5.10, the off-diagonal terms are zero, and we get an isomorphism

$$\mathrm{End}_R(A_1 \oplus \cdots \oplus A_r) \xrightarrow{\sim} \bigoplus_{i=1}^{r} \mathrm{End}(A_i).$$

(in this case the map is the obvious one: $\phi \in \mathrm{End}(A_j)$ from the RHS gets sent to the endomorphism that acts solely on the copy of A_j on the LHS, and so one can also prove this directly without appealing to that sublemma.) □

We are ready to prove the main theorem of this chapter.

Theorem 5.12. (The Artin-Wedderburn Theorem) *The following are equivalent for a ring R:*

(1) *R is semisimple left Artinian (thus, $N(R) = 0$).*

(2) *$R = I_1 \oplus \cdots \oplus I_t$ is a finite direct sum of simple left ideals.*

(3) *As rings, $R \cong M_{n_1}(D_1) \oplus M_{n_2}(D_2) \oplus \cdots \oplus M_{n_r}(D_r)$ for some integers n_j and division rings D_j.*

Moreover, in part (3), and up to reordering, the integers $r, n_1, \ldots, n_r$ are unique and the division rings D_j are unique up to isomorphism.

Proof. We have shown that (1) $\Rightarrow$ (2) in Theorem 5.2.

For (2) $\Rightarrow$ (3), reorganise our direct sum as $R = S_1^{(n_1)} \oplus \cdots \oplus S_r^{(n_r)}$, where the S_j are non-isomorphic simple left R-modules; by (5.1) such a reorganisation does not affect the endomorphism ring. Then combining Lemma 5.5 and Theorem 5.11 shows that $R \cong \mathrm{End}_R(R) \cong \bigoplus M_{n_i}(D_i)$ for $D_i = \mathrm{End}_R(S_i)$. By Schur's Lemma 5.6, the D_j are indeed division rings.

(3) $\Rightarrow$ (1). By Examples 1.24, each $M_{n_i}(D_i)$ is a simple ring, and so certainly $N(M_{n_i}(D_i)) = 0$. Hence $N(R) = 0$ by Example sheet 5, Question 7. Next, a division ring D is (left) Artinian and hence so is $R = M_{n_1}(D_1) \oplus \cdots \oplus M_{n_r}(D_r)$ by combining Parts (1) and (2) of Corollary 3.9.

(3) $\Rightarrow$ (2) (Not that we really need it.) If D is a division ring, check that each column C_ℓ of $M_n(D)$ is a simple left $M_n(D)$-module. Also, note that multiplying on the *right* by the matrix unit e_{ij} gives an isomorphism $C_i \cong C_j$ of left $M_n(D)$-modules. Thus adding these all together gives the required isomorphism $R \cong S_1^{(n_1)} \oplus \cdots \oplus S_r^{(n_r)}$ of left R-modules.

In order to prove uniqueness, we note that the Jordan-Holder Theorem, as given below, shows that the decomposition $R \cong S_1^{(n_1)} \oplus \cdots \oplus S_r^{(n_r)}$ of R into a direct sum of simple modules is unique up to isomorphism. Therefore, so are the numbers n_i and the division rings $D_i = \mathrm{End}_R(S_i)$. □

Theorem 5.13. (The Jordan-Holder Theorem) *Suppose we are given a (left) module M over a ring R with two* composition series

$$M = M_r \supset M_{r-1} \supset \cdots \supset M_0 = 0$$

and

$$M = N_s \supset N_{s-1} \supset \cdots \supset N_0 = 0$$

(thus each factor M_i/M_{i-1} and N_j/N_{j-1} is a simple module). Then $r = s$ and, for some permutation σ of $\{1, ...r\}$, the subfactors $M_i/M_{i-1} \cong N_{\sigma(i)}/N_{\sigma(i)-i}$ are isomorphic.

In particular if an R-module M can be written as a direct sum of simple modules $M = S_1^{(n_1)} \oplus \cdots \oplus S_r^{(n_r)}$, for $S_i \ncong S_j$, then simple modules S_j and the numbers r and n_j in that decomposition are uniquely determined by M.

Remark: Hopefully you have already seen it in a group theory course; the proof is identical.

Proof. We first of all *refine* the two sequences, by defining

$$M_{ij} = M_{i-1} + (M_i \cap N_j) \qquad N_{ij} = N_{i-1} + (N_i \cap M_j)$$

noting that $M_{i-1} = M_{i0} = M_{i-1,s}$, and similarly for N_j. We now have refined series of submodules

$$M = M_r \supseteq \cdots \supseteq M_i = M_{is} \supseteq M_{i,s-1} \supseteq \cdots \supseteq M_{i0} = M_{i-1} \supseteq \cdots \supseteq M_0 = 0, \tag{5.6}$$

and

$$N = N_s \supseteq \cdots \supseteq N_j = N_{jr} \supseteq N_{j,r-1} \supseteq \cdots \supseteq N_{j0} = N_{j-1} \supseteq \cdots \supseteq N_0 = 0. \tag{5.7}$$

Of course, now many of the new factors are likely to be zero, but we do not care. We now claim:

Sublemma 5.14. *In these new series $N_{ji}/N_{j,i-1} \cong M_{ij}/M_{i,j-1}$ as left R-modules.*

Proof of the Sublemma. We use the second isomorphism theorem in the form: If A and $C \subseteq B$ are all submodules of a module L then $\dfrac{A+B}{A+C} \cong \dfrac{B}{(A+C)\cap B} \cong \dfrac{B}{C + B \cap A}$. Thus,

$$\frac{N_{ji}}{N_{j,i-1}} = \frac{N_{j-1} + N_j \cap M_i}{N_{j-1} + N_j \cap M_{i-1}} \cong \frac{N_j \cap M_i}{N_j \cap M_{i-1} + (M_i \cap N_j \cap N_{i-1})} \cong \frac{N_j \cap M_i}{(N_j \cap M_{i-1}) + (M_i \cap N_{i-1})}$$

This final term is symmetric in M_i and N_j and hence $\dfrac{N_{ji}}{N_{j,i-1}} \cong \dfrac{M_{ij}}{M_{i,j-1}}$. □

Returning to the proof of the theorem, this says that, in the two new series (5.6) and (5.7), there is a bijection between the two sets of subfactors, say $X_\ell \mapsto Y_{\sigma(\ell)}$ with $X_\ell \cong Y_{\sigma(\ell)}$ for each ℓ. Of course, many of these subfactors are zero, but that does not matter, since it still implies that the bijection restricts to a corresponding bijection between the *non-zero* subfactors. These are, of course, the subfactors of the original two series. □

We will now use this result to give a more detailed analysis of the structure of semisimple artinian rings. We first note:

Corollary 5.15. *If R is a semisimple left Artinian ring then R is also right Artinian (and so we can just call R* semisimple Artinian *without confusion). It is also left and right noetherian.*

Proof. This is clearly true for the matrix ring $R_i = M_{n_i}(D_i)$ over a division ring and so, by the Artin-Wedderburn Theorem, it is true in general. □

Note that $R = \begin{pmatrix} \mathbb{C} & \mathbb{C} \\ 0 & \mathbb{Q} \end{pmatrix}$ is left but not right Artinian (modify Example 3.5 appropriately), so the corollary definitely fails for non-semsimple Artinian rings.

We begin by seeing what Theorem 5.2 tells us about the structure of modules over this ring R.

Definition: A module M over a ring R is *completely reducible* if, for every submodule $N \subseteq M$ there exists a submodule $L \subseteq M$ such that $M = N \oplus L$.

Proposition 5.16. *Let M be an Artinian left module over a ring R. Then the following are equivalent:*

(1) *M is a sum of simple submodules;*

(2) *M is a direct sum of simple submodules;*

(3) *M is completely reducible.*

Moreover, in (2) the direct sum is necessarily a direct sum of finitely many simple modules.

Remark: The equivalence of (1), (2) and (3) in the theorem does not require M to be Artinian, but the proof is a little more complicated in the general case since one has to replace minimality arguments by applications of Zorn's Lemma. The proofs can be found in any of the recommended texts.

Proof. (2) $\Rightarrow$ (1) is obvious.

(1) $\Rightarrow$ (2) and (1) $\Rightarrow$ (3) and the final assertion of the proposition all follow from:

Sublemma 5.17. *Assume that an artinian left R-module M can be written as $M = \sum_{i\in I} V_i$ for some simple submodules V_i and that W is a submodule of M (possibly with $W = 0$). Then $M = W \oplus (V_{j_1} \oplus V_{j_2} \oplus \cdots \oplus V_{j_n})$ for some finite set of elements $\{j_\ell\}$ from the index set I.*

Proof of Sublemma 5.17. Suppose we have found some finite collection $V_{j_1}, \dots V_{j_m}$ such that

$$W + \sum_{\ell=1}^{m} V_{j_\ell} = W \oplus (V_{j_1} \oplus V_{j_2} \oplus \cdots \oplus V_{j_m}) = Z,$$

say. If $Z = M$ we are done, so suppose not. Then there exists some $V_{j_{m+1}}$ (with $j_{m+1} \in I$) such that $V_{j_{m+1}} \not\subseteq Z$. Now $V_{j_{m+1}} \cap Z$ is a submodule of $V_{j_{m+1}}$ and it is not equal to $V_{j_{m+1}}$ as otherwise $V_{j_{m+1}} \subseteq Z$. So, by the simplicity of $V_{j_{m+1}}$ we conclude that $V_{j_{m+1}} \cap Z = 0$. Hence by Corollary 2.27, yet again, we conclude that $Z + V_{j_{m+1}} = Z \oplus V_{j_{m+1}}$.

Continue in this way. Then either we end up with $M = W \oplus (V_{j_1} \oplus V_{j_2} \oplus \cdots \oplus V_{j_n})$ for some V_{j_ℓ} or we find that M contains the infinite direct sum

$$M \supseteq W \oplus V_{j_1} \oplus V_{j_2} \oplus \cdots \oplus V_{j_n} \oplus V_{j_{n+1}} \oplus \cdots$$

But in this case M also contains the proper infinite descending chain of submodules

$$\begin{aligned} M \quad &\supseteq V_{j_1} \oplus V_{j_2} \oplus\ V_{j_3} \oplus V_{j_4} \oplus \cdots \\ &\supsetneq\ V_{j_2} \oplus\ V_{j_3} \oplus V_{j_4} \cdots \\ &\supsetneq\ \ V_{j_3} \oplus V_{j_4} \cdots \\ &\supsetneq\ \cdots \end{aligned} \quad .$$

This contradiction proves the sublemma. □

We return to the proof of the theorem, for which it remains to show that (3) $\Rightarrow$ (1). As M is artinian, it has a minimal (and hence simple) submodule S. Let M' denote the sum of all the simple submodules of M. If $M' = M$ we are done, so suppose not. Since M is completely reducible, there exists a (necessarily

nonzero) submodule $L \subset M$ such that $M = M' \oplus L$. However, L is artinian as it is a submodule of M. Thus L contains a simple submodule N and hence $L \cap M' \supseteq N \neq 0$, a contradiction. Thus $L = 0$ and $M' = M$.

We remark that one can also prove (3) $\Rightarrow$ (2) directly using a proof very similar to the sublemma. □

Alternative proof: What follows is the proof I gave in lectures. It's not really different in content but organised rather more compactly.

First we prove 5.18 which, note, can be proved right after the definition of "completely reducible"

Lemma 5.18. *If M is a completely reducible Artinian module, then so is every submodule $N \subseteq M$.*

Proof. If $K \subseteq N$ is a submodule then $M = K \oplus T$ for some submodule T. Hence $N = (K + T) \cap N = K + (T \cap N)$ by the Modular Law 2.30. Hence, as $K \cap (T \cap N) \subseteq K \cap T = 0$, Corollary 2.27 implies that $N = K \oplus (T \cap N)$. □

Proof of 5.16

(2) $\Rightarrow$ (1): is immediate.

(1) $\Rightarrow$ (3): Let N be a submodule of M and consider the set of submodules L of M such that the sum $N + L$ is direct (that is, such that $L \cap N = 0$). This set is nonempty (it contains the 0 submodule) and clearly satisfies the condition for Zorn's Lemma. Therefore there is a maximal such submodule, L say.

If $N + L \neq M$ then choose a simple submodule S of M which is not contained in $N + L = N \oplus L$ (if every simple submodule of M were contained in $N + L$ then, since M is a sum of simples, M would be contained in, hence equal to, $N + L$). Since S is simple, it must be that $(N + L) \cap S = 0$ and hence the sum $(N \oplus L) + S$ is direct - equal to $(N \oplus L) \oplus S = N \oplus (L \oplus S)$, contradicting maximality of L[5]. Therefore, $M = N \oplus L$, and M is indeed completely reducible.

(3) $\Rightarrow$ (2): First, note that, since M is artinian, every nonzero submodule of M contains a minimal, hence simple, submodule.

Let $S_1 \leq M$ be a simple submodule of M. Choose a complementary submodule N_1, so $M = S_1 \oplus N_1$. If $N_1 \neq 0$, choose a simple submodule S_2 of N_1 and, by Lemma 5.18, a complement N_2 of S_2 in N_1; so $N_1 = S_2 \oplus N_2$. Then $M = S_1 \oplus S_2 \oplus N_2$. If $N_2 \neq 0$ continue. The process cannot continue indefinitely because the sequence $M > N_1 > N_2 > \dots$ is strictly decreasing and M is artinian. Therefore, $M = S_1 \oplus S_2 \oplus \cdots \oplus S_k$ for some simple modules S_j.

We can now add another equivalent condition to the Artin-Wedderburn Theorem:

Corollary 5.19. *Let R be a ring. Then the following are equivalent:*

(1) *R is left Artinian with $N(R) = 0$;*

(2) *any finitely generated left R-module M can be written as a finite direct sum of simple left submodules.*

[5] If that step is not clear to you: we're saying that $N \cap (L + S) = 0$, which certainly contradicts maximality of L - since if $n = l + s \neq 0$ for some $n \in N$, $l \in L$ and $s \in S$ then, rearranging, we'd get $n - l = s$, but $(N + L) \cap S = 0$, so $n - l = s = 0$, so $n = l$, contradicting $N \cap L = 0$.

Proof. (1) $\Rightarrow$ (2). From Question 3 on Exercise Sheet 3, any finitely generated left R-module M can be written $M = R^{(n)}/J$. By Theorem 5.2 (or the Artin-Wedderburn Theorem) R and hence $R^{(n)}$ is a direct sum of simple modules and hence is completely reducible by Proposition 5.16; say $R^{(n)} = J \oplus T$. Then the surjective map $\theta : R^{(n)} \twoheadrightarrow M$ induces a map $\theta' : T \to M$. Since $\theta(J) = 0$ clearly θ' is surjective. Also, since $T \cap J = 0$ certainly $\ker(\theta') = 0$. So θ' is an isomorphism. Hence $M \cong T$ has complete reducibility by Lemma 5.18.

(2) $\Rightarrow$ (1). Since R is therefore a finite direct sum of simple modules, this follows from the Artin-Wedderburn Theorem. $\square$

Remark: Once again, one can delete the phrase "finitely generated" from part (2) the corollary.

Our next application of the Artin-Wedderburn Theorem is to the structure of group rings, so you should recall the definition from Example 1.16 and the examples thereafter. For this we will need:

Theorem 5.20. The Chinese Remainder Theorem.

Let $P_1, \ldots, P_r$ be ideals of a ring R and suppose that

(*) *For all $1 \leq j \leq r$ we have $P_j + \left[P_1 \cap \cdots \cap P_{j-1} \cap \widehat{P}_j \cap P_{j+1} \cap \cdots \cap P_r\right] = R$.*

Then

$$R/I \cong R/P_1 \oplus R/P_2 \oplus \cdots \oplus R/P_r$$

for $I = P_1 \cap \cdots \cap P_r$.

Remarks In the statement of (*) the notation means that we are omitting P_j from the intersection.

Proof. For $r = 2$ this was proved in Exercise Sheet 1, Question 9, where we used it to understand the group ring kC_2. It is not hard to use that result and induction to prove the general result, but but it is as easy to prove it directly.

We certainly have a ring homomorphism

$$\phi : R \to R/P_1 \oplus \cdots \oplus R/P_r \qquad r \mapsto (\,[r + P_1], \ldots, [r + P_r]\,).$$

Clearly $\ker(\phi) = \bigcap P_i = I$. So it remains to show that (*) forces ϕ to be surjective. To see this we know that $P_1 + (P_2 \cap \cdots \cap P_r) = R$; thus if $r \in R$ we can write $r = r_1 + r'$, where $r_1 \in P_1$ and $r' \in (P_2 \cap \cdots \cap P_r)$. Now,

$$\begin{aligned} \phi(r') = \quad & (r', 0, \ldots, 0) = (r_1 + r', 0, \ldots, 0) \\ & = (r, 0, \ldots, 0). \end{aligned}$$

The same argument works for all the other P_j but let's be formal and give the proof. Sor any j we have $r = r_j + s_j$ where $r_j \in P_j$ and $s_j \in \left[P_1 \cap \cdots \cap P_{j-1} \cap \widehat{P}_j \cap P_{j+1} \cap \cdots \cap P_r\right]$. Thus $\phi(s_j)$ only has a non-zero entry in the j^{th} entry and there we get:

$$\begin{aligned} \phi(s_j) = \quad & (0, \ldots, 0, s_j, 0, \ldots, 0) = (0, \ldots, 0, s_j + r_j, 0, \ldots, 0) \\ & = (0, \ldots, 0, r, 0, \ldots, 0). \end{aligned}$$

Thus ϕ is surjective and hence it is an isomorphism of rings. $\square$

Definition 5.21. *Let R be an algebra over a field k and regard R as a k-vector space, say V. Then left multiplication $\lambda_r : v \mapsto rv$ by elements of R is a k-linear transformation and hence defines a map $\lambda : R \to End_k(V)$ called the* left regular representation*.*

The right regular representation *is defined similarly using the map $\rho_r : v \mapsto vr$.*

Lemma 5.22. *The left regular representation λ is an injective ring homomorphism.*

Proof. For all $r, s \in R$ and $v \in V$ we have $\lambda_r(\lambda_s(v)) = rsv = \lambda_{rs}(v)$ and $(\lambda_r + \lambda_s)(v) = (r+s)v = \lambda_{r+s}(v)$ and $\lambda_1 v = 1v = v$ whence $\lambda_1 = 1$. Thus λ is a ring homomorphism. Since $\lambda_r(1) = r \neq 0$, for all $r \in R$, certainly λ is injective. $\square$

Theorem 5.23. (Maschke's Theorem) *Suppose that G is a finite group and K a field with characteristic either zero or coprime to $|G|$. Then KG is a semisimple Artinian ring.*

Remark. The proof uses some basic facts about eigenvectors, that you should be familiar with for complex matrices, but maybe not for arbitrary fields. If so, just read the proof as if $k = \mathbb{C}$, but accept that it actually works perfectly well in general.

Proof. By definition we need to prove that $N(KG) = 0$, for which we use the regular representation λ. Here $End_K(KG) \cong M_n(K)$, where $n = |G|$. We use the elements $\{g_1 = e, g_2, \ldots, g_n\}$ of G as a basis of $V = KG$. Thus $\lambda_g g_i = g_j \iff gg_i = g_j$ and so, as a matrix, λ_g has a 1 in the $(j,i)^{th}$ entry if $gg_i = g_j$ and zeros elsewhere. Notice that when $g \neq e$ this means that λ_g has no nonzero diagonal entries.

In particular, $\lambda_e = I_n$ has trace $tr(\lambda_e) = n \neq 0$ but $tr(\lambda_g) = 0$ for $g \neq e$. Now, if $\alpha = \sum_{g \in G} \alpha_g g \in KG$ then $tr(\lambda_\alpha) = \sum_{g \in G} tr(\alpha_g g) = n\alpha_e$. Now fix a nilpotent element $\alpha \in KG$. Then λ_α is still nilpotent. Since $tr(\lambda_\alpha)$ is the sum of the eigenvalues of λ_α it will also be zero. In particular $n\alpha_e = 0$ which, since $n \neq 0$ in K, implies that $\alpha_e = 0$.

Finally, let I be a nilpotent ideal of KG and pick $\alpha = \sum_{g \in G} \alpha_g g \in I$. For any $h \in G$

$$I \ni \alpha h^{-1} = \sum_{g \in G} \alpha_g (gh^{-1}) = \sum_{g \in G} \alpha_{gh} g.$$

But $\alpha h^{-1} \in I$ so it is nilpotent and hence, by the last paragraph, the coefficient α_h of e in αh^{-1} must be zero. Since h was arbitrary, $\alpha = 0$ and $I = 0$. $\square$

Remark: The left regular representation works for modules as well as rings and can make some of our earlier proofs a little more conceptual. Here is one such example.

Suppose that R is a k-algebra and that M is a nonzero left R-module that is finite dimensional as a k-vector space. Then $ann_R(M)$ is an ideal of R such that $R/ann_R(M)$ is also finite dimensional.

(Compare this proof to the one given in Example Sheet 2.)

Proof. Let $\dim_k(M) = n$; thus $End_k(M) \cong M_n(k)$. We again have an analogue of the left regular representation by defining a map $\Theta : R \to End_k(M)$ by defining $\Theta(r) = \theta_r$ where $\theta_r(m) = rm$. The proof of Lemma 5.22 shows that each θ_r is a linear transformation of M and that Θ is a ring homomorphism. In this case by the definition of annihilators, $Ker(\Theta) = \{r \in R : \theta_r = 0\} = ann_R(M)$. Thus, by the first isomorphism theorem, $R/ann_R(M) \cong Im(\Theta) \subseteq M_n(k)$. As such, $R/ann_R(M)$ is certainly finite dimensional. □

We now want study the applications of Maschke's Theorem in more detail, for which we need a couple of easy lemmas.

Lemma 5.24. *(1) If e is a central idempotent in a ring R then $R = Re \oplus R(1-e)$ as rings.*

(2) If K is a field of characteristic zero then one summand of KG is eK for the "trivial" idempotent $e = \frac{1}{|G|}\sum_{g\in G} g$.

Proof. (1) Use Theorem 5.20.

(2) First note that for any $h \in G$ we have $eh = \frac{1}{|G|}\sum_{g\in G}(gh) = \frac{1}{|G|}\sum_{k\in G} k$, as gh runs through all the element of G. Hence $eh = e$ and, by summing, $e^2 = \frac{1}{|G|}\sum_{g\in G} e = e$. Since trivially $er = re$ for all $r \in G$ and hence for all $r \in KG$, it follows that e is a central idempotent. Thus eKG is a summand of R. However, since $KG = \sum_{h\in G} Kh$ the rule $eh = e$ implies that $eKG = eK$. □

Lemma 5.25. *Suppose that A is a simple Artinian ring that has an algebraically closed centre K and such that A is finite dimensional as a K-vector space.*

Then $A \cong M_n(K)$ for some n.

Proof. By the Artin-Wedderburn Theorem 5.12 we know that $A \cong M_n(D)$ for some division ring D, so we need to show that $D = K$. However D still contains K (in fact K is the centre of D) and so D is a finite dimensional K-vector space. Now pick any $d \in D$ and think about the ring R generated by d and K. Since K is central in this ring it just consists of all polynomials $\sum \lambda_i d^i$ in d. Any two such polynomials commute since K commutes with everything and d commutes with d.

In other words, R is a commutative, finite dimensional K-algebra and is a domain as it sits inside the domain D. Thus R is a field, and hence equals K since K was algebraically closed. As d was arbitrary this forces $D = K$. □

Example 5.26. *(a) First, consider any finite group G and the group ring $\mathbb{C}G$. By Maschke's Theorem combined with the Artin-Wedderburn Theorem 5.12 $\mathbb{C}G \cong \bigoplus M_{n_i}(D_i)$ for some n_i and division rings D_i. Since each D_i is a finite dimensional $\mathbb{C}$-algebra they must equal $\mathbb{C}$ by Lemma 5.25. Hence $\mathbb{C}G \cong \bigoplus M_{n_i}(\mathbb{C})$.*

(b) If G is abelian then $\mathbb{C}G$ is commutative and so $\mathbb{C}G = \mathbb{C} \oplus \cdots \oplus \mathbb{C}$ ($|G|$ copies) is the only possibility.

(c) Now suppose that $G = S_3$, the symmetric group. As S_3 is not abelian, $\mathbb{C}S_3$ cannot be a direct sum of fields and so at least one matrix ring must occur. By counting dimensions, the only possibility is

$$\mathbb{C}S_3 \;\cong\; M_2(\mathbb{C}) \oplus \mathbb{C} \oplus \mathbb{C}.$$

Notice that one of these copies of $\mathbb{C}$ *is given by the trivial idempotent (see Lemma 5.24). What is the other one? (In representation-theoretic language it corresponds to the "sign" representation but that is another course!)*

(d) Now suppose that $G = D_4$*, the dihedral group of order* 8*. Here, as* D_4 *is not abelian, we must have at least one matrix ring. By counting dimensions, the only possibilities are*

$$\mathbb{C}D_4 \cong M_2(\mathbb{C}) \oplus M_2(\mathbb{C}) \qquad or \qquad \mathbb{C}D_4 \cong M_2(\mathbb{C}) \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}.$$

But, Lemma 5.24 implies that, for any group G*, one summand of* $\mathbb{C}G$ *is* $\epsilon\mathbb{C}G \cong \mathbb{C}$ *for* $\epsilon = \frac{1}{|G|}\sum_{g\in G} g$*. So the first possibility cannot happen and*

$$\mathbb{C}D_4 \cong M_2(\mathbb{C}) \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C} \oplus \mathbb{C}.$$

We end the section by noting that the restriction on the characteristic of K in the proof of Maschke's Theorem is essential.

Proposition 5.27. *Suppose that* K *is a field and* G *a finite group such that the characteristic of* K *divides the order* $|G|$*. Then* KG *is* not *semisimple.*

Proof. The culprit is the element $\eta = \sum_{g\in G} g$. As in the proof of Lemma 5.24(2) one sees that η is central in KG and $\eta h = \eta$ for all $h \in G$.

However, now this implies that $\eta^2 = \sum_{h\in G} \eta h = |G|\eta = 0$ as $|G| = 0$ inside K. Thus, ηKG is a nonzero ideal of KG such that $(\eta KG)(\eta KG) = \eta\eta KG = 0$. □

Comments re examinability: Examples of examinable arguments are 5.1(2) (part(3) is more a "one-(perhaps long)line" argument in the sense of the comments at the end of Section 4), 5.2, 5.6. The arguments involving matrices of endomorphisms are too messy to be good exam questions and the proof for the Artin-Wedderburn Theorem itself (5.12) is not a good exam question since it's not sufficiently "stand-alone" - it depends too much on quoting previously proved technical things. Note that 5.13 and 5.20 were not gone through, so are not examinable. The proof of 5.16 is another proof suitable for examination. We didn't go through the proof of 5.23 but the surrounding shorter arguments and the examples are suitable for examination.

6. **Modules over Principal Ideal Domains.**

In this section we return to commutative domains, specifically principal ideal domains and, in practice, Euclidean domains such as $\mathbb{Z}$ and $k[x]$ where k is a field. It turns out that finitely generated (but not infinitely generated!) modules over these rings can be decomposed into finite direct sums of cyclic modules, and those can be described quite explicitly. Indeed, given a finite presentation of a module, its decomposition(s) can be computed using simple matrix operations.

Comments on examinability: We spent only one week on this section. What we covered - and what you need to know - is the initial material; in particular, the statement of the fundamental theorem 6.2, the set-up for the proof of that theorem ("Step I"), the idea of how the rest of the proof goes, at least to the extent of being able to use the method to compute examples over $\mathbb{Z}$, such as the couple of examples immediately after the proof, and the alternative decomposition given in 6.9. You don't need to be able to deal with examples using $\mathbb{C}[x]$ in place of $\mathbb{Z}$ and nothing from 6.13 on is examinable.

The aim of this chapter is to determine the structure of any finitely generated module over a commutative principal ideal domain R (hereafter abbreviated as **PID**). The answer is very nice; in the case when $R = \mathbb{Z}$ it just says that a finitely generated abelian group G is isomorphic to

$$\mathbb{Z}^r \times Z_{n_1} \times Z_{n_2} \times \cdots \times \mathbb{Z}_{n_m}$$

for some integers r and n_j. Moreover, by being precise about the integers r and n_j, this decomposition can be made unique. This result generalizes to any PID, although we will need some notation to state it precisely. In fact we will only prove the result for Euclidean domains - a special class of PIDs, but these rings include the most important PIDs; namely $\mathbb{Z}$ and $k[x]$ for a field k.

Likely most of you will have seen the basic results about Euclidean domains, but an appendix to this chapter recalls their main properties.

Definition 6.1. *A* Euclidean domain *is a commutative domain with a function* $\chi : R \to \mathbb{N} = \mathbb{Z}_{\geq 0}$ *such that*

(1) *For all* $a, b \in R$ *with* $b \neq 0$ *there exists* $q, r \in R$ *such that* $a = qb + r$ *where* $\chi(r) < \chi(b)$*;*

(2) *For all* $a, b \in R$ *with* $a, b \neq 0$ *one has* $\chi(ab) \geq \chi(a)$*.*

It can be convenient also to assume:

(3) $\chi(0) = 0$ *and hence* $\chi(r) > 0 \iff r \neq 0$*. (This can always be assumed by replacing* χ *by* $\chi = \chi(0)$*, so is harmless, though maybe a bit unnatural in some examples.)*

Our two basic examples are $R = \mathbb{Z}$ with $\chi(a) = |a|$, and $R = k[x]$ for a field k with $\chi(f(x)) = \deg f$, with $\chi(0) = -1$. If one wants to insist on axiom (3) one would take $\chi(f) = 1 + \deg f$ if $f \neq 0$ and $\chi(0) = 0$ in the case $R = k[x]$.

The fundamental property of Euclidean domains is, of course, that one can use Euclid's algorithm to find GCDs (Greatest Common Divisors). See Proposition 6.27 and the discussion before it for the details.

Modules over PIDs. We now come to the main topic of this chapter, which is to describe the structure of finitely generated modules over PIDs. In fact we will only prove it for Euclidean domains, as the proof is easier and it covers the cases that interest us. For the more general result, see, for instance, Cohn's book.

Theorem 6.2. (The Fundamental Theorem for Modules over a PID) *(but stated and proved just for Euclidean domains)*

Let R be a Euclidean domain and M a finitely generated R-module. Then

$$M \cong R/R\alpha_1 \oplus R/R\alpha_2 \cdots \oplus R/R\alpha_s \oplus \underbrace{R \oplus \cdots \oplus R}_{t}, \qquad \text{where } \alpha_1|\alpha_2 \cdots |\alpha_s \text{ are non-units.} \tag{6.1}$$

Moreover s, t are unique and the α_j are unique up to equivalence (meaning up to multiplication by invertible elements - for example 3 and -3 are equivalent in the ring $\mathbb{Z}$).

Proof. I am going to give a somewhat algorithmic proof of the theorem, which will be a bit longer than necessary but it has the advantage that one can reproduce it in examples. The proof has five steps, which we now describe.

Step I. Write M as a factor of a free module, say $M = R^{(n)}/N$ (see Example Sheet 3, Question 3(i)). Equivalently we have a surjective map $\phi : R^{(n)} \to M$. As R is Noetherian, Corollary 3.8 implies that $N = Ker(\phi)$ is also finitely generated and so we can similarly write $N = \theta(R^{(m)})$ for some homomorphism $\theta : R^{(m)} \to R^{(n)}$. We regard this data as a pair of maps or *presentation*

$$R^{(m)} \xrightarrow{\theta} R^{(n)} \xrightarrow{\phi} M \longrightarrow 0 \tag{6.2}$$

Here, just as for linear transformations (see also Example Sheet 3, Question 3(ii)), the action of θ is right multiplication by some $m \times n$ matrix $A = A_\theta$. If you unravel the definitions you will find that A_θ is simple to write down:

The rows of A are given by the generators of N as a submodule of $R^{(n)} = (R, R, ..., R)$. (6.3)

So, for example, if $\theta = (\theta_{ij})$ then θ applied to $e_1 = (1, 0, ..., 0)$ gives $e_1\theta = (\theta_{11}, \theta_{12}, ..., \theta_{1n})$, which is exactly what we want.

Step II. We are allowed to do elementary row and column operations to the matrix A (as in linear algebra). Doing a row operation on A corresponds to left multiplying A by an elementary matrix which in turn is the same as doing an elementary change of generators of R^m; equivalently of generators of N. Similarly an elementary column operation is the same as an elementary change of the generators of M, so is harmless.

Before giving the details of this step, let me state the other steps:

Step III. Using the Euclidean algorithm and Step II we can then reduce A to *Smith normal form*. This means that now

$$A \;=\; \mathrm{diag}\{\alpha_1, \alpha_2, \cdots, \alpha_r\} \;=\; \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & \cdots \\ 0 & \alpha_2 & 0 & 0 & \cdots \\ 0 & & \ddots & 0 & \\ 0 & \cdots & 0 & \alpha_r & \\ & & & & \ddots \end{pmatrix} \quad \text{with } \alpha_i | \alpha_{i+1} \text{ for each } i. \tag{6.4}$$

Here $r = \min\{n, m\}$ and we allow $\alpha_t = 0 = \alpha_{t+1} = \cdots = \alpha_r$ for some t. Also, the matrices we use need not be square, but in equations like (6.4) I mean that the α_i are on the leading diagonal $\{a_{ii}\}$ of the matrix $A = (a_{ij})$ even though this will not exactly end at the bottom right hand corner.

Step IV. Returning to M this just says that

$$M \;=\; R/R\alpha_1 \oplus R/R\alpha_2 \cdots R/R\alpha_n \quad \text{where we put } \alpha_i = 0 \text{ if } r < i \leq n.$$

Step V. Finally, this answer is unique up to replacing the α_j by equivalent elements $\beta_j = \alpha_j u_j$ for units (i.e. invertible elements) u_j.

Let's put in the details of steps II–V.

In **Step II** we allow the usual elementary operations from linear algebra, specifically

(1) add a multiple of one row (or column) to another;

(2) swop two rows (or columns);

(3) multiply a row (or column) by a *unit* of R.

Note that doing an elementary row (respectively column) operation is the same as multiplying on the left (respectively right) by the corresponding elementary matrix (which is obtained by doing the same operation to the appropriately sized identity matrix). For example, if

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} : \mathbb{Z}^{(2)} \to \mathbb{Z}^{(3)},$$

and we do the operation "replace column 2 by column 2 - twice column 1" then we are making the replacement

$$A \;\rightsquigarrow\; \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 3 \\ 4 & -3 & 6 \end{pmatrix}.$$

Write $C = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and let γ denote the corresponding function of "right multiply by C". Of course elementary operations are invertible with their inverses just being the inverse elementary operation, so

if $AC = A'$ corresponds to a map θ' then $\theta = \theta'\gamma^{-1}$. Thus if we think of the original presentation of $M = R^{(n)}/N$ as

$$R^{(m)} \xrightarrow{\theta} R^{(n)} \xrightarrow{\phi} M \longrightarrow 0 \tag{6.5}$$

from (6.2) then we can rewrite this as

$$R^{(m)} \xrightarrow{\theta'} R^{(n)} \xrightarrow{\gamma^{-1}} R^{(n)} \xrightarrow{\phi} M \longrightarrow 0. \tag{6.6}$$

(Remember that when we write functions on the right then fg means do f first and then g.) So, in other words, replacing A by A' really just means that we first act on $R^{(n)}$ by the matrix C—which in turn just means that we are doing the corresponding elementary change of basis to $R^{(n)}$. Another way of thinking of this is that we are replacing the surjection $\phi : R^{(n)} \to M$ by $\gamma^{-1} \circ \phi : R^{(n)} \to M$. This is of course harmless, meaning that the resulting presentation is just a different presentation of the same module M.

If we did a row operation, say multiplying by a matrix D corresponding to map δ, then we get $\theta = \delta^{-1}\theta'$ and (6.5) becomes

$$R^{(m)} \xrightarrow{\delta^{-1}} R^{(m)} \xrightarrow{\theta'} R^{(n)} \xrightarrow{\phi} M \longrightarrow 0 \tag{6.7}$$

So, this means our elementary row operation just corresponds to an elementary change of basis to $R^{(m)}$ and this just gives an elementary change of generators for the module N. Again, this is of course harmless.

So, in summary, in Step II we are allowed to do elementary row and column operations to our matrix A and this just corresponds to harmless changes of generator of M or N.

In **Step III** we need to prove:

Proposition 6.3. *Let A be an $m \times n$ matrix with entries from a Euclidean domain R. Then by doing elementary row and column operations to $A = (a_{ij})$ we can reduce A to* Smith normal form*, that is, a matrix of the following form*

$$diag\{\alpha_1, \alpha_2, \cdots, \alpha_r\} = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & \cdots \\ 0 & \alpha_2 & 0 & 0 & \cdots \\ 0 & & \ddots & 0 & \\ 0 & \cdots & 0 & \alpha_r & \\ & & & & \ddots \end{pmatrix} \quad \text{with } \alpha_i | \alpha_{i+1} \text{ for each } i. \tag{6.8}$$

Here, I allow the possibility that $\alpha_j = 0$ for $j \geq s$ and some integer s.

Proof. We do this algorithmically. Recall that a Euclidean domain has a function $\chi : R \to \mathbb{Z}_{\geq 0}$ normalised so that $\chi(0) = 0$.

If $A = 0$ there is nothing to prove, and if not then we can always swop a couple of rows and then columns to ensure that $a_{11} \neq 0$.

Step III.A. *We can assume that a_{11} divides a_{1j} for all $j \geq 1$. In this process $\chi(a_{11})$ never increases.*

To see this, suppose that $a_{ij} = a_{11}b + r$ for some $0 < \chi(r) < \chi(a_{11})$. Now subtract b times the first column from j^{th} column and then swop the first and j^{th} columns. This replaces a_{11} by r. Since $\chi(r) < \chi(a_{11})$ this process must stop in a finite amount of time.

Step III.B. Now repeat this process for the first column by doing row operations in place of the column operations. Of course in the process you may change elements in the first row, in which case we just repeat step III.A. As $\chi(a_{11})$ decreases at each move (or, formally after each pair of moves), this will stop after finitely many steps. Thus we can reduce to the case where a_{11} *divides all* a_{1j} *and all* a_{j1}.

Step III.C. *We can assume that $a_{1j} = 0 = a_{1j}$ for all $j > 1$.* To do this just subtract $a_{1j}a_{11}^{-1}$ times the first column from the j^{th} column and similarly for rows.

Step III.D. *We can assume that $a_{1j} = 0 = a_{1j}$ for all $j > 1$ and also that a_{11} divides a_{uv} for all u, v.* To see this, suppose that a_{11} does not divide a_{uv}. Then add the u^{th} row to the first row. Then this means that the $(1, v)$ entry is now a_{uv}. So, go back to Step III.A. At every step we will reduce $\chi(a_{11})$ and so the process must eventually stop.

Step III.E. *Completion of the proof of Step III.* We have reduced A to a matrix of the form $A = \begin{pmatrix} a_{11} & 0 \\ 0 & B \end{pmatrix}$ where $B = (b_{ij})$ is a $(n-1) \times (m-1)$ matrix such that $a_1 | b_{ij}$ for all i, j. By induction on n we can apply the propositon to B to reduce B to the diagonal matrix $B = \text{diag}(b_2, \ldots, b_r)$ where $b_j | b_{j+1}$ for all j. Now in the process we are only doing elementary operations and, as a_{11} divides each b_{ij} at the beginning, we find that a_{11} divides each b_{ij} after each of these operations. In other words, after we have finished we find that a_{11} does divide each b_j. □

Step IV. This claims that, if $A = \text{diag}(\alpha_1, \cdots \alpha_t, 0, 0, \ldots 0)$, where $\alpha_i | \alpha_{i+1}$ for all i and t is chosen such that $\alpha_t \neq 0$, then

$$M \;\cong\; R/R\alpha_1 \oplus R/R\alpha_2 \cdots R/R\alpha_t \oplus R^{(n-t)}.$$

We remark, here, that $R^{(u)}$ is defined to be zero if $u = 0$.

Proof. Look at $R/R\alpha_1 \oplus R/R\alpha_2 \oplus \cdots \oplus R/R\alpha_t \oplus R^{(n-t)}$. This is obviously generated by the n elements $\overline{e}_1 = (1, 0, \ldots, 0), ..., \overline{e}_n = (0, \ldots, 0, 1)$ and so we can write it as $R^{(n)}/K$ for the appropriate submodule K Clearly the annihilator of $\overline{e}_j$ is $R\alpha_j$ for $j \leq t$ but $ann_R(\overline{e}_\ell) = 0$ for $\ell > t$. Thus

$$M \cong R^{(n)}/K \qquad \text{where} \;\; K = (R\alpha_1, R\alpha_2, \ldots, R\alpha_t, 0, \ldots, 0).$$

But on the other hand recall that $M = R^{(n)}/Im(\theta)$ where θ is given by right multiplication by A. Since $Im(\theta)$ is just the module generated by the rows of A, we find that $Im(\theta)$ is precisely K, as required. □

This competes the proof of the existence part of Theorem 6.2. We will now do some examples to see how this algorithm works in practice and only after that will we come back and prove the uniqueness part of Theorem 6.2.

Example 6.4. *Consider*

$$M = \frac{\mathbb{Z} \oplus \mathbb{Z}}{\mathbb{Z}(66,\, 30) + \mathbb{Z}(12,\, 4)}.$$

Clearly here $M = \mathbb{Z}^{(2)}/N$ *where* N *is generated by the rows of* $A = \begin{pmatrix} 66 & 30 \\ 12 & 4 \end{pmatrix}$ *(see Equation 6.3). So, we follow Step III in the proof of Theorem 6.2 and do elementary operations to the matrix* A*. This gives*

$$A \;=\; \begin{pmatrix} 66 & 30 \\ 12 & 4 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 12 & 4 \\ 66 & 30 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 4 & 12 \\ 30 & 66 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 4 & 12 \\ 2 & -18 \end{pmatrix}.$$

Here we swopped rows, then swopped columns, then subtracted 7 times row 1 from row 2.

Next, if we again swop rows and then subtract the appropriate multiple of the first row from the second and then repeat for the columns we get:

$$\begin{pmatrix} 4 & 12 \\ 2 & -18 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 2 & -18 \\ 4 & 12 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 2 & -18 \\ 0 & 48 \end{pmatrix} \;\rightsquigarrow\; \begin{pmatrix} 2 & 0 \\ 0 & 48 \end{pmatrix}.$$

Thus,

$$M \cong \frac{\mathbb{Z}}{2\mathbb{Z}} \oplus \frac{\mathbb{Z}}{48\mathbb{Z}}.$$

Remark 6.5. It would not matter if you wrote the generators of N as the columns of A. The reason is that you would get the transpose of A, which would then given you the transpose of our Smith normal form. But the transpose of a diagonal matrix is itself!

Example 6.6. *To make it a little more complicated, take*

$$M = \frac{\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}}{\mathbb{Z}(4,\, -2,\, 3) + \mathbb{Z}(2,\, 2,\, 0) + \mathbb{Z}(-6,\, 12,\, -9)}.$$

Now, we take $A = \begin{pmatrix} 4 & -2 & 3 \\ 2 & 2 & 0 \\ -6 & 12 & -9 \end{pmatrix}$ *and we get*

$$A \rightsquigarrow \begin{pmatrix} 1 & -2 & 3 \\ 2 & 2 & 0 \\ 3 & 12 & -9 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 2 & 6 & -6 \\ 3 & 18 & -18 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & 18 & -18 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & -6 \\ 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The first operation subtracted column 3 from column 1 and the other operations should be clear. Thus

$$M \;\cong\; \frac{\mathbb{Z}}{\mathbb{Z}} \oplus \frac{\mathbb{Z}}{6\mathbb{Z}} \oplus \frac{\mathbb{Z}}{0} \;\cong\; \frac{\mathbb{Z}}{6\mathbb{Z}} \oplus \mathbb{Z}.$$

Notice that, as happens in this example, we do allow $\alpha_1 = 1$ in the algorithm in which case the first factor of M is just $\mathbb{Z}/\mathbb{Z} = 0$ so it can be ignored. It is only when one worries about uniqueness that one has to delete such modules.

Example 6.7. *We can also do exactly the same game with $\mathbb{Z}$ replaced by $k[x]$ for a field k. However here we do want to be able to factorise polynomials so for simplicity I will always work with $k = \mathbb{C}$. So, let $R = \mathbb{C}[x]$ and consider*

$$M = \frac{R \oplus R \oplus R}{R(x+5,\, 2,\, -3) \,+\, R(-1,\, x,\, 1) \,+\, R(6,\, 2,\, x-4)}.$$

Now, we take $A = \begin{pmatrix} x+5 & 2 & -3 \\ -1 & x & 1 \\ 6 & 2 & x-4 \end{pmatrix}$ *and we get*

$$A \rightsquigarrow \begin{pmatrix} -1 & x & 1 \\ x+5 & 2 & -3 \\ 6 & 2 & x-4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & x & 1 \\ 0 & 2+(x^2+5x) & -3+(x+5) \\ 0 & 2+6x & x+2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & x & 1 \\ 0 & x^2+5x+2 & x+2 \\ 0 & 6x+2 & x+2 \end{pmatrix}.$$

(Here we first swopped the first two rows, and then subtracted the appropriate multiples of the first row from the others and then cleaned things up a bit.)

In the next equation after doing some obvious column operations, we want to subtract the appropriate multiple of the $(x+2)$ from the $(2,2)$ entry. For this we factorise $(x^2+5x+2) = (x+2)(x+3) - 4$. So we get

$$\rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & (x+2)(x+3)-4 & x+2 \\ 0 & 6x+2 & x+2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & -4 & x+2 \\ 0 & * & x+2 \end{pmatrix}$$

Here, the $*$ equals $(6x+2) - (x+2)(x+3) = -x^2 + x - 4$. Next multiply the last column by 4 (which is allowed for $R = \mathbb{C}[x]$) and then add (x+2) times column 2 to column 3. This gives

$$\rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & -4 & 4(x+2) \\ 0 & (-x^2+x-4) & 4(x+2) \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & (-x^2+x-4) & \dagger \end{pmatrix}$$

where

$$\dagger = 4(x+2) + (x+2)(-x^2+x-4) = -(x+2)(x^2-x) = -x(x+2)(x-1)$$

After adding the appropriate multiple of row 2 to row 3 and then multiplying each row by the appropriate scalar we get

$$A \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -x(x+2)(x-1) \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & x(x+2)(x-1) \end{pmatrix}.$$

Thus—finally—we see that

$$M \cong \frac{\mathbb{C}[x]}{\mathbb{C}[x]x(x+2)(x-1)}.$$

What we see from this example that the computations do get messier with $\mathbb{C}[x]$ in place of $\mathbb{Z}$ simply because factorisation is harder. However, the principle is the same.

Example 6.8. *It is important to note that the Fundamental Theorem 6.2 does not work for infinitely generated modules. Indeed, as a $\mathbb{Z}$-module, $\mathbb{Q}$ cannot be written as a direct sum of cyclic modules.*

Proof. Indeed, as $\mathbb{Z}a \cap \mathbb{Z}b$ for any two non-zero rational numbers a, b, if $\mathbb{Q}$ is a direct sum of cyclic left $\mathbb{Z}$-modules, then it must be the direct sum of just one. In other words one would have $\mathbb{Q} = \mathbb{Z}q$ for some rational q. But you cannot write $q/2$ as an integer multiple of q, giving the required contradiction. □

There is a second version of the Fundamental Theorem which can be illustrated for the simplest case of $\mathbb{Z}_6$. One can also write this as $Z_6 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_3$. For this we use the Chinese Remainder Theorem 5.20.

Theorem 6.9. *Let M be a finitely generated module over a Euclidean domain R. Then M can be uniquely written as $M \cong \bigoplus_{n_i \geq 0} R/p_i^{n_i} \oplus R^m$ for appropriate primes p_i. In more detail,*

$$M \;\cong\; R/(p_1^{n_1}) \oplus \cdots \oplus R/(p_t^{n_t}) \oplus R^{(m)}.$$

Here the p_j are primes and the $n(i_j)$ and t are positive integers. This decomposition is unique up to permutation of terms and replacing the p_i by associate primes.

Proof. Recall from Theorem 6.2 that M can be uniquely written as $M \cong \bigoplus R/(\alpha_i) \oplus R^{(m)}$ for some α_i and m. Each α_i can be uniquely written as $\alpha_i = p_1^{u_1} \cdots p_r^{u_r}$ for the appropriate primes p_i and integers u_i (possibly with repetitions). By the Chinese Remainder Theorem 5.20 we can then write $R/(\alpha_i) = R/(p_1^{u_1}) \oplus \cdots \oplus R/(p_r^{u_r})$. Now collect terms. □

Example 6.10. *For example, $\mathbb{Z}_{792} \cong \mathbb{Z}_8 \oplus \mathbb{Z}_9 \oplus \mathbb{Z}_{11}$. This is true as either rings or $\mathbb{Z}$-modules. More generally, the different possible ways of writing $\mathbb{Z}_{792}$ as a sum of cyclic modules is the same as the different possible ways of writing 792 as products of coprime numbers. Thus*

$$792 = 8 \cdot 9 \cdot 11 = 72 \cdot 11 = 88 \cdot 9 = 99 \cdot 8.$$

Thus,

$$\mathbb{Z}_{792} \;\cong\; \mathbb{Z}_8 \oplus \mathbb{Z}_9 \oplus \mathbb{Z}_{11} \;\cong\; \mathbb{Z}_{72} \oplus \mathbb{Z}_{11} \;\cong\; \mathbb{Z}_{88} \oplus \mathbb{Z}_9 \;\cong\; \mathbb{Z}_{99} \oplus \mathbb{Z}_8.$$

One also has the complementary problem of finding all the nonisomorphic groups of a given order:

Example 6.11. *The nonisomorphic $\mathbb{Z}$-modules (or equivalently, the nonisomorphic abelian groups) of order 88 are:*

$$\mathbb{Z}_8 \oplus \mathbb{Z}_{11}, \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_{11} \qquad \textit{and} \qquad \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_{11}.$$

Of course each of these can be written in more than one way; for example $\mathbb{Z}_2 \oplus \mathbb{Z}_4 \oplus \mathbb{Z}_{11} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_{44} \cong \mathbb{Z}_{22} \oplus \mathbb{Z}_4$. *I will let you write down the nonisomorphic* $\mathbb{Z}$*-modules of order 792. There are 6 of them.*

You can do exactly the same thing for $k[x]$-modules, except that it is harder to factorise.

Example 6.12. *Find the* $\mathbb{C}[x]$*-modules of complex dimension 3 that are killed by* $(x-1)^3$*. Answer:*

$$\frac{\mathbb{C}[x]}{((x-1)^3)} \qquad \textit{and} \qquad \frac{\mathbb{C}[x]}{((x-1)^2)} \oplus \frac{\mathbb{C}[x]}{(x-1)} \qquad \textit{and} \qquad \frac{\mathbb{C}[x]}{(x-1)} \oplus \frac{\mathbb{C}[x]}{(x-1)} \oplus \frac{\mathbb{C}[x]}{(x-1)},$$

where I have written $((x-1)^3)$ *for the ideal generated by* $(1-x)^3$.

Note that all but the first one of these modules are actually killed by $(x-1)^2$*, since in* $\mathbb{C}[x]/(1-x)^3$*, the coset* $[1]$ *obviously has annihilator* $((1-x)^3)$*. Thus the answer to the question "*Find the $\mathbb{C}[x]$-modules of complex dimension 3 that are killed by $(x-1)^2$*" would be*

$$\frac{\mathbb{C}[x]}{((x-1)^2)} \oplus \frac{\mathbb{C}[x]}{(x-1)} \qquad \textit{and} \qquad \frac{\mathbb{C}[x]}{(x-1)} \oplus \frac{\mathbb{C}[x]}{(x-1)} \oplus \frac{\mathbb{C}[x]}{(x-1)}.$$

The following material on Jordan canonical forms, and the proof of uniqueness of decomposition in the Fundamental Decomposition Theorem, are not examinable.

The next application of the Fundamental Theorem we want to give is to the Jordan Canonical Form of matrices. This follows from the following amazing idea:

Proposition 6.13. *Given a field* k *then the following are equivalent:*

(1) $k[x]$*-modules* V*.*

(2) k*-vector spaces with a linear transformation* $\theta : V \to V$*.*

Proof. This is really a tautology in the sense that the two definitions coincide. In more detail, given a $k[x]$-modules V then certainly one has a map $V \to V$ given by $v \mapsto xv$. So, check that this satisfies the axioms of a linear transformation. Conversely, if $\theta : V \to V$ is a linear transformation, then we define a module structure on V by $x \cdot v = \theta(v)$. □

Example 6.14. *Think of Example 6.12. First, if one takes* $\mathbb{C}[x]/(1-x^3)$ *then this is a* 3*-dimensional complex vector space and a particularly nice basis is* $e_1 = 1, e_2 = (x-1)$ *and* $e_3 = (x-1)^2$*. Then the action of* x *on this basis (acting by left multiplication) sends*

$$e_1 \mapsto (x-1) + 1 = e_1 + e_2, \qquad e_2 \mapsto x(x-1) = (x-1)^2 + (x-1) = e_2 + e_3,$$

and

$$e_3 \mapsto x(x-1)^2 = (x-1)^3 + (x-1)^2 = (x-1)^2 = e_3.$$

Thus $\theta : v \mapsto xv$ is the linear transformation with matrix

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

The same sort of computation for the module $k[x]/(x-1)^2 \oplus k[x]/(x-1)$ will lead to the matrix

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

Theorem 6.15. Jordan Canonical Form. *Suppose that k is an algebraically closed field and that $\theta : V \to V$ is a linear transformation of an n-dimensional k-vector space V. Then there exists a basis of V with respect to which θ has matrix*

$$\begin{pmatrix} J_1 & 0 & 0 \cdots & \\ 0 & J_2 & 0 \cdots & \\ 0 & 0 & \ddots & \\ 0 & \cdots & 0 & J_t \end{pmatrix},$$

where each J_i is an $r_i \times r_i$ matrix of the form

$$\begin{pmatrix} \lambda_i & 0 & 0 & \cdots & \\ 1 & \lambda_i & 0 & \cdots & \\ 0 & 1 & \lambda_i & 0 & \cdots \\ 0 & & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & \lambda_i \end{pmatrix}$$

for some integers r_i with $\sum r_i = n$ and some scalars λ_i.

This is also unique up to reordering the blocks.

Proof. Using Proposition 6.13 we translate the problem into a module problem. So, think of $\{V, \theta\}$ as a $k[x]$-module $M = V$ of k-dimension n on which θ is just multiplication by x. By the Fundamental Theorem in the form of Theorem 6.9 M can be uniquely written as $M \cong R/(p_1^{r_1}) \oplus \cdots \oplus R/(p_t^{r_t})$ for some primes $p_i \in k[x]$. (Note that there is no term of the form $k[x]^{(m)}$ as M is finite dimensional.) Since k is algebraically closed, each $p_i = (x - \lambda_i)$ for some scalar λ_i.

Thus to prove the theorem, if suffices to consider the case when M just has one summand; say $M = k[x]/(x-\lambda)^u$. But now, exactly as in Example 6.14, if we chose the basis $e_1 = 1, e_2 = (x-\lambda), \cdots e_u =$

$(x-\lambda)^{u-1}$ of M then multiplication by x (which, after all, is the transformation θ) has matrix

$$\begin{pmatrix} \lambda & 0 & 0 & \cdots & \\ 1 & \lambda & 0 & \cdots & \\ 0 & 1 & \lambda & 0 & \cdots \\ 0 & & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & \lambda \end{pmatrix},$$

as required. □

Example 6.16. *Find the Jordan Canonical Form of the matrix* $A = \begin{pmatrix} -5 & 1 & -6 \\ -2 & 0 & -2 \\ 3 & -1 & 4 \end{pmatrix}$.

Answer: In fact in this case the field does not matter but let's call it $k = \mathbb{C}$. So, we translate the problem into module theory. Thus, we take the 3-dimensional $\mathbb{C}$-vector space thought of as a $\mathbb{C}[x]$-module M on which x acts by multiplication by x. Here we should regard M as generated by the 3 basis elements $e_1 = (1,0,0)^T$, etc and we use this to write $M = \mathbb{C}[x]^{(3)}/L$ for a submodule L we have to find. But $x \cdot e_1 = -5e_1 - 2e_2 + 3e_3$ and so $(x+5)e_1 + 2e_2 - 3e_3 = 0$. This says that $L \ni (x+5, 2, -3)$. Repeat this for the other two basis elements and you see that

$$L \supseteq L' = \mathbb{C}[x](x+5,\ 2,\ -3) \ +\ \mathbb{C}[x](-1,\ x,\ 1) \ +\ \mathbb{C}[x](6,\ 2,\ x-4).$$

Counting vector space dimensions we see that $\dim_{\mathbb{C}} C[x]/L = \dim_{\mathbb{C}} M = 3 = \dim_{\mathbb{C}} \mathbb{C}[x]/L'$ and so $L = L'$. Thus

$$M = \frac{R \oplus R \oplus R}{R(x+5,\ 2,\ -3) \ +\ R(-1,\ x,\ 1) \ +\ R(6,\ 2,\ x-4)}\ .$$

But this is the module we saw back in Example 6.7. So, we know that this can be written out as

$$M \cong \mathbb{C}[x]/\mathbb{C}[x]x(x-1)(x+2) \cong \mathbb{C}[x]/(x) \oplus \mathbb{C}[x]/(x-1) \oplus \mathbb{C}[x]/(x+2).$$

Converting back to our linear transformation we find that A has Jordan Canonical Form equal to the diagonal matrix

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

One slightly annoying thing with the way I have set this all up is that the matrix A I start with is not exactly the matrix I computed with in Example 6.7—they are each other's transpose! This this just arises from technical reasons (essentially that one writes the module L above in terms of rows not columns). Fortunately this is irrelevant since *a matrix A and its transpose A^T have the same Jordan Canonical Form. Thus, if you wish you can do the computation corresponding to Example 6.7 for A rather than A^T.* It is rather easy to see why this is true, by the way. For, when we reduce a matrix to its Smith Normal Form,

we are allowed to do row or column matrices. Thus if we start with a matrix A^T then we will end up with the transpose of the normal form of A. But a diagonal matrix is equal to its own transpose!.

With this in mind, lets try:

Example 6.17. *Find the Jordan Canonical Form of the matrix* $A = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

Answer: So, bearing in mind the above observation, we take the matrix

$$X = xI_3 - A = \begin{pmatrix} x-1 & 1 & 1 \\ -1 & x-3 & -1 \\ 0 & 0 & x-2 \end{pmatrix}$$

and then play the same games as before:

$$X \rightsquigarrow \begin{pmatrix} -1 & x-3 & -1 \\ x-1 & 1 & 1 \\ 0 & 0 & x-2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & x-3 & -1 \\ 0 & 1+(x-3)(x-1) & 1-(x-1) \\ 0 & 0 & x-2 \end{pmatrix} = \begin{pmatrix} -1 & x-3 & -1 \\ 0 & (x-2)^2 & -(x-2) \\ 0 & 0 & x-2 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & (x-2)^2 & -(x-2) \\ 0 & 0 & x-2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & (x-2)^2 & 0 \\ 0 & 0 & x-2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -1 & 0 & 0 \\ 0 & (x-2) & 0 \\ 0 & 0 & (x-2)^2 \end{pmatrix}$$

(where the final step actually consisted of first swopping rows and then columns)! In other words, then module

$$M = \frac{\mathbb{C}[x] \oplus \mathbb{C}[x] \oplus \mathbb{C}[x]}{\mathbb{C}[x](x-1,1,1) + \mathbb{C}[x](-1,x-3,-1) + (0,0,x-2)},$$

is isomorphic to

$$\frac{\mathbb{C}[x]}{(x-2)} + \frac{\mathbb{C}[x]}{(x-2)^2}.$$

Also the matrix A has Jordan canonical form

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix} \quad \text{or, if you prefer,} \quad \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{or, if you prefer,} \quad \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

There is a second situation where one wants to find the generators of a module over a Euclidean domain R. This is when one is given a submodule M of a free module $R^{(n)}$. Here I will lead you through the relevant algorithm in one example and then leave you to do a second example by yourself.

Example 6.18. *(i) Prove that if $M \subseteq R^{(n)}$ over a Euclidean domain R, then M is itself a free module; say $\phi : R^{(r)} \xrightarrow{\sim} K$. In this case the elements $\{x_1 = \phi(1, 0, ..., 0), ..., x_r = \phi(0, ..., 0, 1)\}$ are called a basis of K.*

(ii) Show that elements $\{y_j : 1 \leq j \leq s\}$ in the R-module M are a basis if and only if every element $m \in M$ can be uniquely written $m = \sum r_i y_i$ for some $r_i \in R$. (The proofs form part of Example Sheet 8.)

Example 6.19. *Let K be the submodule of $\mathbb{Z}^{(3)}$ generated by $a = (2, -2, -2)$, $b = (5, -2, -3)$, and $c = (1, -4, -3)$. Find a basis for K.*

Here is one way of doing this:

(1) Set $L = (\mathbb{Z}, \mathbb{Z}, 0) \subseteq \mathbb{Z}^{(3)}$. Then $K + L/L$ is generated by the cosets $[d + L]$ of the elements of K whose third coordinate is the GCD of all the third coordinates – in this case we could take $d = b - 2a = (1, 2, 1)$. Add this to your set of generators.

(2) Now take the generators of K and subtract the appropriate multiple of d so that one obtains an element of $K \cap L$; that is an element whose third coordinate is zero. These elements then generate $K \cap L$. In our case this means taking $a + 2d = (4, 2, 0)$, and $c + 3d = (4, 2, 0)$ and $b + 3d = (8, 4, 0)$.

Now induct; this means regard $L \cong \mathbb{Z}^{(2)}$ and find the element of $L \cap K$ whose second coordinate is the GCD of the second coordinates of elements of $K \cap L$. Then subtract multiples of this from the other generators, etc, etc.

Of course in this particular example, this means take $e = a + 2d$. Then the process stops and $\{d = (1, 2, 1), e = (4, 2, 0)\}$ is our basis.

Example 6.20. *If instead we started with the module K' that was generated by $a = (2, -2, -2)$, $b = (5, -2, -3)$, $c = (1, -4, -3)$ and $f = (1, 2, 0)$, then we would again get new generators $d = b - 2a = (1, 2, 1)$, $e = a + 2d = (4, 2, 0)$ and $f = (1, 2, 0)$ after the first 2 steps. Here, I am ignoring the superfluous ones $c + 3d$ and $b + 3d$.*

Now when apply the inductive step I can use $e = (4, 2, 0)$ and $f - e = (-3, 0, 0)$ as my generators of $L \cap K'$. So, now K' has basis $\{d = (1, 2, 1), e = (4, 2, 0), (-3, 0, 0)\}$.

The reason why this works is that:

(I) We have always applied elementary operations to our generating set and so the new elements are indeed a generating set and

(II) the final set of generators are automatically linearly independent (say in $\mathbb{Q}^{(3)}$) and this ensures the uniqueness property of a basis.

Uniqueness of Smith Normal Forms. Finally (no, I had not forgotten, though I may not reach it in the lectures), we will prove the uniqueness part of Theorem 6.2.

Definition-Lemma 6.21. *Let M be a finitely generated module over a commutative domain R. Then*

(1) *the* torsion submodule $T(M)$ of M *is the subset*

$$T(M) \;=\; \{m \in M | rm = 0 \text{ for some non-zero } r \in R\}.$$

It is a submodule.

(2) *The module $M/T(M)$ is* torsion-free *in the sense that $T(M/T(M)) = 0$.*

(3) *If M a PID and $M = R/\alpha_1 R \oplus \cdots \oplus R/\alpha_n R \oplus R^{(m)}$ for some nonzero α_i and some integer m, then $T(M) = R/\alpha_1 R \oplus \cdots \oplus R/\alpha_n R$.* □

The elementary proof of the lemma is left to the reader. There are no good analogues to the theorem for non-domains or non-commutative rings.

Since the torsion submodule of a module is unique, this lemma reduces the proof of uniqueness to two special cases for a module M over a PID R:

(1) If $R^{(n)} \cong R^{(m)}$ then $n = m$.

(2) If M is torsion then M can be uniquely written as $M = \bigoplus_{i=1}^{u} R/\alpha_i R$ where $\alpha_i | \alpha_{i+1}$ for all i and the α_i are neither zero nor units.

Part (1) has already appeared on Example Sheet 3, Question 5, so we need only prove part (2). Note, here that we exclude the case when some α_j is a unit, as this would add a superfluous term of 0 to the sum, and we cannot allow $\alpha_j = 0$ as that would introduce a torsion-free summand. In fact, it is slightly easier to prove uniqueness for the alternate form of Theorem 6.9; thus

(2′) If M is torsion, finitely generated module over a PID R then M can be uniquely written as $M = \bigoplus_{i=1}^{u} R/(p_i^{n_i})R$ for some primes p_j and positive integers n_i. Here the p_i are only unique up to associates.

We leave it to the reader to check, by using the Chinese Remainder Theorem, that (2) and (2′) are equivalent.

To prove (2′) we first separate the non-associate primes using the same trick as we did for the torsion part of M. So, given a prime p in a PID R then define the *p-torsion submodule of M* to be

$$T_p(M) \;=\; \{m \in M : p^r m = 0 \text{ for some } r > 0\}.$$

As before, it is easy to prove that $T_p(M)$ is a submodule. Moreover, we have:

Lemma 6.22. *Suppose that M is a module over a PID R such that $M = (\bigoplus_i R/p_i^{n_i}) \oplus \left(\bigoplus_i R/q_j^{m_j}\right)$, for primes q_j that are not associates of p and primes p_i that are associates of p. Then $T_p(M) = \left(\bigoplus_i R/p_i^{n_i}\right)$.*

Proof. Clearly p^n kills $\bigoplus_i R/p^{n_i}$ provided that $n \geq \max\{n_i\}$. Moreover if $p_i = up$ for a unit u then p^n also kills R/p_i^n. Putting these two observations together shows that $T_p(M) \supseteq \left(\bigoplus_i R/p_i^{n_i}\right)$.

So, we need to prove that nothing else is in $T_p(M)$. To do this, note that, rather like in the last paragraph, the element $x = \prod_j q_j^{m_j}$ does kill $\left(\bigoplus_i R/q_j^{m_j}\right)$. Also, we claim that $GCD(p^u, x) = 1$ for any $u \geq 1$; indeed if not then $Rx + Rp^u = Rv$ for some non-unit v. But now, z is any prime that divides v then it must divide p and also divide some q_j. This is of course impossible and proves the claim.

By the Chinese Remainder Theorem we can therefore find integers a, b such that $1 = pa + xb$. Now, suppose that some $n \in N = \left(\bigoplus_i R/q_j^{m_j}\right)$ is in $T_p(N)$; say with $p^c n = 0$. By the Chinese Remainder Theorem we can therefore find integers a, b such that $1 = p^c a + xb$. Since $xn = 0$ this says that $n = n \cdot 1 = np^c + nx = 0 + 0 = 0$. □

This means that in proving the uniqueness statement (2') we can assume that $M = R/Rp^{n_1} \oplus R/Rp^{n_2} \oplus \cdots \oplus R/Rp^{n_r}$, and we need to prove that (up to reordering) the n_i are unique. Collecting terms and assume that

$$M = (R/Rp)^{(s_1)} \oplus \left(R/Rp^2\right)^{(s_2)} \oplus \cdots \oplus \left(R/Rp^n\right)^{(s_n)}.$$

Here I do now have to allow some $s_j = 0$ and, as for free modules, the notation $(R/Rp)^{(s_1)}$ means the direct sum of s_1 copies of the module R/Rp. So, we have to prove that the s_j are uniquely determined by M.

Now consider $p^{n-1}M$. Clearly $p^{n-1}(R/(p^u) = 0$ whenever $u \leq n-1$, and so p^{n-1} kills all but the final term. So what about the final term? But it is also clear that $p^{n-1}(R/(p^n)) = p^{n-1}R/p^nR$. Moreover, as this module is generated by p^{n-1} and this generator is killed by p, Lemma 2.17 shows that $p^{n-1}R/p^nR \cong R/pR$. Adding together s_n copies of this gives $p^{n-1} \cdot M = p^{n-1} \cdot (R/p^nR)^{(s_n)} \cong (R/pR)^{(s_n)}$. But this implies that s_n is uniquely determined—just use the second Coursework (which shows that we can think of this as an isomorphism of R/p-modules) and then apply Example Sheet 3, Question 5.

So, at least we have proved that the final term is uniquely determined by M. Fortunately the rest follows by induction. Consider $M/p^{n-1}M$. Only the final term is affected and, by the observation of the last paragraph, in $Z = (R/Rp^n)^{s_n}$, the third isomorphism theorem implies that

$$Z/p^{n-1}Z = \left(\frac{R/Rp^n}{p^{n-1}R/Rp^n}\right)^{s_n} \cong \left(R/Rp^{n-1}\right)^{s_n}.$$

Adding terms together we see that

$$M/p^{n-1}M \cong (R/Rp)^{(s_1)} \oplus \left(R/Rp^2\right)^{(s_2)} \oplus \cdots \oplus \left(R/Rp^{n-2}\right)^{(s_{n-2})} \oplus \left(R/Rp^{n-1}\right)^{(s_{n-1}+s_n)}.$$

So, by induction on n we can assert that the numbers $s_1, \ldots, s_{n-2}, s_{n-1} + s_n$ are uniquely determined by M. As s_n was also uniquely determined we see that each of $s_1, \ldots, s_n$ is uniquely determined by M.

This completes the proof of all the uniqueness assertions in this chapter. □

Let's end with a few, somewhat more abstract, applications of the Fundamental Theorem.

Corollary 6.23. *Let M be a submodule of a finitely generated free module $R^{(n)}$ over a PID R. Then $M \cong R^{(s)}$ for some unique $s \leq n$. In particular, M can be generated by at most n elements.*

Proof. By Theorem 6.2, and Lemma 6.21, $M \cong T(M) \oplus R^{(s)}$ for some unique s. However, $R^{(n)}$ has a zero torsion submodule, the same is true for its submodule M. Hence $M \cong R^{(s)}$ for some unique s. It remains to show that $s \leq n$. We use induction on n. Let L denote the direct sum of the first $n-1$) copies of R inside $R^{(n)}$. Clearly L is the kernel of the projection of $R^{(n)}$ onto the final copy of R in the direct sum. Thus, $R^{(n)}/L \cong R$.

Now intersect with M. Clearly $M/(M \cap L) \cong M + L/L$ is a submodule of $R^{(n)}/L \cong R$. Therefore, since R is a PID, $M/(M \cap L)$ is cyclic, say by $x_1 \in M$. On the other hand, $M \cap L \subseteq L \cong R^{(n-1)}$ and so by induction on n can be generate by $n-1$ elements $x_2, ..., x_n$. It follows easily that M can be generated by the n elements $x_1, \ldots, x_n$.

Finally, we need to prove that this implies that $s \leq n$. This is similar to the proof of Example Sheet 3, Question 5 and we leave the proof to the interested reader. □

Appendix: Basic facts about Euclidean domains.

Euclidean Domains. So, lets begin by discussing Euclidean domains. Remember that over the integers one has *Euclid's Algorithm* for finding the GCD or Greatest Common Divisor of two numbers. For example, in order to find $(100, 330)$ one computes:

$$\begin{aligned} 330 &= 100 \cdot 3 + 30 \\ 100 &= 30 \cdot 3 + 10 \\ 30 &= 3 \cdot 10 + 0. \end{aligned} \tag{6.9}$$

Hence $(100, 330) = 10$. Moreover we can write the GCD out as a sum of multiples of 100 and 330 by working "upwards":

$$\begin{aligned} 10 &= 100 - 30 \cdot 3 \\ &= 100 - (330 - 100 \cdot 3) \cdot 3 \\ &= 10 \cdot 100 - 3 \cdot 330. \end{aligned} \tag{6.10}$$

A Euclidean domain is any commutative domain where one has such an algorithm. Formally:

Definition 6.24. *A* Euclidean domain *is a commutative domain with a function $\chi : R \to \mathbb{N} = \mathbb{Z}_{\geq 0}$ such that*

(1) *For all $a, b \in R$ with $b \neq 0$ there exists $q, r \in R$ such that $a = qb + r$ where $\chi(r) < \chi(b)$;*
(2) *For all $a, b \in R$ with $a, b \neq 0$ one has $\chi(ab) \geq \chi(a)$.*
(3) *$\chi(0) = 0$ and hence $\chi(r) > 0 \iff r \neq 0$.*

Some books do not require condition (3), but since you can always reduce to that case by replacing χ by $\chi = \chi(0)$, so it seems simplest to demand it.

Examples 6.25. (1) $R = \mathbb{Z}$ with $\chi(a) = |a|$.

(2) $R = k[x]$ for any field k. Here one can essentially take $\chi(f) = \deg f$. However to fit precisely with our definition one needs to define $\chi(f) = 1 + \deg f$ if $f \neq 0$ and $\chi(0) = 0$. The proof of this is left as an exercise.

(3) After this the examples get harder to define. For example (see pages 130-133 of Cohn's book) the so-called ring of integers inside $\mathbb{Q}(\sqrt{d})$ (this is essentially $\mathbb{Z}(\sqrt{d})$ is a Euclidean domain for $d = 2, 3, 5, -1, -2, -3, -7, -11$. But it is not know in general for which values of d it works!

Many of the standard properties of GCD's for integers work for any Euclidean domain and the proofs are essentially the same. Thus, for example, we have:

Properties/Definitions 6.26. *Let R be a Euclidean domain and let $a, b \in R$.*

(1) *We say $a|b$ or a divides b if $b = ca$ for some $c \in R$. If $a = ub$ for a unit u then we write $a \sim b$ and call a, b* associates.

(2) *The GCD of a and b, written $GCD(a,b)$ or simply (a,b) is the number $c \in R$ such that $c|a$ and $c|b$ and it is maximal in the sense that, if $d|a$ and $d|b$ then $d|c$.*

Note that this definition is not unique—even for $R = \mathbb{Z}$ one has $(5,7) = \pm 1$. However, if $(a,b) = c$ and $(a,b) = c'$ then $c = uc'$ and $c' = vc$ which, forces $c = uvc$. Since R is a domain this implies that $uv = 1$. Thus u is a unit and GCD's are *unique up to taking associates*.

(3) In special cases we do get uniqueness for GCD's by demanding more—for $R = \mathbb{Z}$ we assume that $(a,b) > 0$ while if $R = k[x]$ we use the convention that (a,b) is the unique *monic* polynomial that is the GCD of aq and b.

(4) *a and b are* coprime *if $(a,b) = 1$ (or more formally if (a,b) is a unit).*

Euclid's Algorithm. Given nonzero elements a, b in a Euclidean domain R, we can follow the algorithm for the integers and write

$$\begin{array}{rcll} a & = & bq_0 + r_1 & \text{with } \chi(r_1) < \chi(b) \\ b & = & r_1 q_1 + r_2 & \text{with } \chi(r_2) < \chi(r_1) \\ r_1 & = & r_2 q_2 + r_3 & \text{with } \chi(r_3) < \chi(r_2) \\ & & \cdots & \\ r_{n-2} & = & r_{n-1} q_{n-1} + r_n & \text{with } \chi(r_n) < \chi(r_{n-1}) \text{ and } r_n \neq 0 \\ r_{n-1} & = & r_n q_n + 0 & \end{array} \tag{6.11}$$

Similarly you can unravel it with

$$\begin{array}{rcl} r_n & = & -r_{n-1} q_{n-1} + r_{n-2} \\ & = & -(r_{n-3} - r_{n-2} q_{n-2}) q_{n-1} + r_{n-2} \\ & = & \cdots \\ & = & xa + yb \end{array} \tag{6.12}$$

for some $x, y \in R$.

Proposition 6.27. *Assume that R is a Euclidean domain.*

(1) *In the above equations $r_n = GCD(a,b)$. It is unique up to multiplication by a scalar.*

(2) *Moreover, $r_n = (a,b)$ is the unique element of R (up to multiplication by a scalar as usual) that can be written as $r = xa + yb$ for some $x, y \in R$ and has $\chi(r_n)$ as small as possible with this property.*

Proof. (1) The proof is the same as for the integers. Moving up the equations in (6.11) gives

$$r_n | r_{n-1} \;\Rightarrow\; r_n | r_{n-2} \;\Rightarrow\; \cdots \;\Rightarrow\; r_n | b \;\Rightarrow\; r_n | a.$$

So, r_n divides both a and b. Conversely, if $d|a$ and $d|b$ then $d|(xa + yb) = r_n$. In other words, r_n is a $GCD(a,b)$.

(2) Exercise. □

Exercise 6.28. Given nonzero elements a, b in a Euclidean domain R, define the Least Common Multiple $LCM(a, b)$ and prove that $(a, b)LCM(a, b) = abR$.

Recall that a PID is a commutative ring in which every ideal is cyclic.

Theorem 6.29. *If R is a Euclidean domain then R is a PID.*

Proof. Let I be a nonzero ideal of R and pick $x \in I \setminus \{0\}$ with $\chi(x)$ as small as possible. We claim that $I = Rx$. Indeed, let $y \in I$. Then we can write $y = qx + r$ with $\chi(r) < \chi(x)$. Of course, $r = y - qx \in I$. Thus, by the minimality of $\chi(x)$, this means that $\chi(r) = 0$ and $r = 0$. □

Definition 6.30. *Let R be a commutative domain.*

(1) *An element $z \in R$ is called* prime *if a is not a unit but whenever $a|bc$ then either $a|b$ or $a|c$. Of course this is just the same definition as primality in the integers.*

(2) *A non-unit $x \in R$ is* irreducible *if it has the following property: x is not a unit but if $x = yz$ for some $y, z \in R$ then either y or z is a unit.*

The use of the word "prime" is suppose to make you think of prime numbers and just as for the integers we have:

Theorem 6.31. *Let R be a Euclidean domain. Then:*

(1) *Each irreducible element is prime and vice versa.*

(2) *Each non-unit $x \in R$ can be written $x = r_1 r_2 \cdots r_n$ for some prime elements r_i. This is unique up reordering and taking associates.*

Remark 6.32. Really, this proof shows that a PID is also a *Unique Factorisation Domain*, though talking about the latter objects will take us a bit far afield. You can find them discussed in each of the recommended text; for example Chapter 3.4 of Cohn's book "Introduction to Ring Theory."

Proof. This is just the same as the proof you saw for the integers.

(1) Obviously prime elements are irreducible. Conversely, suppose that a is irreducible and that $a|bc$, where a does not divide b. Then (a, b) divides a and so as a is irreducible, we must have $1 = ax + by$ for some x, y. But now $c = a(xc) + (bc)y$ is divisible by a.

(2) **Existence.** Any PID is Noetherian, so if the result is false we can take the biggest ideal xR such that x cannot be so written. As x is not irreducible, it can be written as $x = ab$ for some non-units a, b. It follows that $xR \subseteq aR$ and $xR \subsetneq bR$ (why?). By the inductive hypothesis both a and b can be written as products of irreducibles and thus so can x.

Uniqueness. Suppose that $z = \mu \prod p_i = \prod q_j$ where the p_i and q_j are primes and μ is some unit. Once again we can assume that zR is the biggest ideal such that z can be written in two distinct ways like this. Thus q_1 divides $z = \prod p_i$ and hence divides some p_ℓ. Thus, $q_1 = \lambda p_\ell$ for some j and unit λ. So, we can

cancel q_1 and get $w = q_1^{-1}z = (\mu\lambda)\prod_{i\neq\ell} p_i = \prod_{j\geq 2} q_j$. Now $wR \supsetneqq zR$ and so we are done by the same Noetherian induction as before. $\square$

Example 6.33. *Consider $R = \mathbb{Z}[\sqrt{-5}]$. Then $(1+\sqrt{-5})(1-\sqrt{-5}) = 6 = 2\cdot 3$, yet all four elements are irreducible. However, they are not associates, and so cannot be prime elements. Thus R is not a Euclidean domain (or a PID or a UFD).*