COMMUTATIVE ALGEBRA

NEIL STRICKLAND

1. Rings

Definition 1.1. [defn-ring]

A *commutative ring* is a set A equipped with elements $0, 1 \in A$ and operations of addition and multiplication such that the following axioms be satisfied:

(a) For all $a, b \in A$ we have $a + b, ab \in A$.
(b) For all $a \in A$ we have $0 + a = a$ and $1a = a$.
(c) For all $a, b \in A$ we have $a + b = b + a$ and $ab = ba$.
(d) For all $a, b, c \in A$ we have $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$.
(e) For all $a, b, c \in A$ we have $a(b + c) = ab + ac$.
(f) For all $a \in A$ there is an element $-a \in A$ with $a + (-a) = 0$.

In other words, addition and multiplication should be commutative and associative with 0 and 1 as neutral elements, and multiplication should distribute over addition.

Remark 1.2. We will not consider noncommutative rings in this course, so we will just use the word "ring" to mean "commutative ring". We will use without comment various standard consequences of the axioms, such as the facts that $-(-a) = a$, $0.a = 0$ and $(-1).a = -a$.

Remark 1.3. If we have two rings A and B and we need to distinguish between the additive identity elements in A and B, then we will call them 0_A and 0_B rather than just 0. Similarly, we may write 1_A and 1_B for the multiplicative identity elements.

Example 1.4. [eg-numbers]

The sets $\mathbb{Z}$ (of integers), $\mathbb{Q}$ (of rational numbers), $\mathbb{R}$ (of real numbers) and $\mathbb{C}$ (of complex numbers) are all rings. Here of course we are using the standard definitions of addition and multiplication, and of the elements 0 and 1. The set $\mathbb{N}$ (of nonnegative integers) satisfies all the axioms except for axiom (f).

Example 1.5. [eg-two-local]

There are also various other rings of numbers that are slightly less obvious. For example, let $\mathbb{Z}_{(2)}$ denote the set of rational numbers of the form a/b, where a and b are integers and b is odd. Using the equations $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$ and $\frac{a}{b}\,\frac{c}{d} = \frac{ac}{bd}$ we see that $\mathbb{Z}_{(2)}$ is closed under addition and multiplication. It also contains $0 = 0/1$ and $1 = 1/1$, so it is a ring. For another example, consider the set $\mathbb{Z}[i]$ of complex numbers of the form $a + ib$, with $a, b \in \mathbb{Z}$. It is not hard to check that this is also a ring.

In the above example, we define addition and multiplication on $\mathbb{Z}_{(2)}$ by restricting the corresponding operations on $\mathbb{Q}$, and we define addition and multiplication on $\mathbb{Z}[i]$ by restricting the corresponding operations on $\mathbb{C}$. It will be convenient to consider this construction more generally:

Definition 1.6. [defn-subring]

Let A be a ring. A *subring* of A is a subset $B \subseteq A$ such that

(a) $0_A, 1_A \in B$
(b) Whenever $b \in B$ we have $-b \in B$
(c) Whenever $b, c \in B$ we have $b + c \in B$ and $bc \in B$.

It is clear that any subring of A can be considered as a ring in its own right, using the restricted operations.

Example 1.7. [eg-subrings]

$\mathbb{Z}$ and $\mathbb{Z}_{(2)}$ are subrings of $\mathbb{Q}$, and $\mathbb{Q}$ is a subring of $\mathbb{R}$, and $\mathbb{R}$ and $\mathbb{Z}[i]$ are subrings of $\mathbb{C}$.

Example 1.8. [eg-modular]

Now consider an integer $n > 0$. We will define a ring $\mathbb{Z}/n$ to serve as a home for modular arithmetic. The most conceptual way to do this is to use the framework of quotient rings, which we will introduce in Section 5. For the moment we take a more pedestrian approach. We first define $\mathbb{Z}/n = \{0, 1, \dots, n-1\}$. Any integer $a \in \mathbb{Z}$ then has a unique representation $a = nq + r$ with $q \in \mathbb{Z}$ and $r \in \mathbb{Z}/n$; we define $\pi(a)$ to be r, giving a function $\pi\colon \mathbb{Z} \to \mathbb{Z}/n$. Given $a, b \in \mathbb{Z}/n$ we define $a \oplus b = \pi(a+b)$ and $a \otimes b = \pi(ab)$; this defines binary operations $\oplus$ and $\otimes$ on $\mathbb{Z}/n$. We also define a unary operation $\ominus a = \pi(-a)$. Standard ideas about modular arithmetic show that these operations make $\mathbb{Z}/n$ into a commutative ring. We will usually just write $a + b$ and ab and $-a$ instead of $a \oplus b$ and $a \otimes b$ and $\ominus a$, relying on the context to distinguish between operations in $\mathbb{Z}/n$ and operations in $\mathbb{Z}$.

Example 1.9. [eg-trivial-ring]

Consider a set T with one element, say $T = \{t\}$. We can make this into a ring by defining $0_T = t$ and $1_T = t$ and $t + t = t$ and $tt = t$. A ring of this form is called *trivial*. Note here that $1_T = 0_T$. Conversely, if A is any ring in which $1_A = 0_A$, then for any element $a \in A$ we have $a = 1_A a = 0_A a = 0_A$, so $A = \{0_A\}$, so A is trivial.

Example 1.10. [eg-square-matrices]

Consider the set $A = M_n(\mathbb{Z})$ of $n \times n$ matrices with integer entries. Let 0_A denote the zero matrix, and let 1_A denote the identity matrix. With these elements and the standard definition of matrix multiplication, A satisfies all axioms except that multiplication is not commutative (provided that $n \geq 2$). The same applies to $M_n(B)$ for any commutative ring B.

Example 1.11. [eg-F-four]

Let $\mathbb{F}_4$ denote the following set of matrices over $\mathbb{Z}/2$:

$$\mathbb{F}_4 = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\}.$$

We will allow ourselves to write 0 for the zero matrix $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and 1 for the identity matrix $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. We also write $\alpha = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right]$. Note that $\alpha^2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right]$, which is the same as $\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ because we are working with matrices over $\mathbb{Z}/2$. We thus have $\mathbb{F}_4 = \{0, 1, \alpha, \alpha^2\}$. One can check that $\alpha^3 = 1$ and thus $\alpha^4 = \alpha$, and also that $1 + \alpha + \alpha^2 = 0$. From this it follows that $\mathbb{F}_4$ is closed under the operations of addition and multiplication, which can be tabulated as follows:

$+$	0	1	α	α^2
0	0	1	α	α^2
1	1	0	α^2	α
α	α	α^2	0	1
α^2	α^2	α	1	0

$\cdot$	0	1	α	α^2
0	0	0	0	0
1	0	1	α	α^2
α	0	α	α^2	1
α^2	0	α^2	1	α

We see that in this context matrix multiplication is commutative, so we have a commutative ring with just four elements.

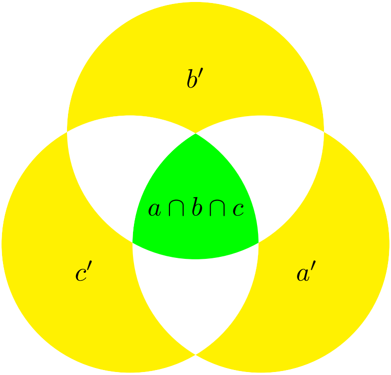
Example 1.12. [eg-boolean]

Let S be any set, and let $\mathrm{Sub}(S)$ be the set of all subsets of S. Put $0_{\mathrm{Sub}(S)} = \emptyset$ and $1_{\mathrm{Sub}(S)} = S$. Given $a, b \in \mathrm{Sub}(S)$ (so $a \subseteq S$ and $b \subseteq S$) we put

$$\begin{aligned} a + b &= (a \cup b) \setminus (a \cap b) = (a \setminus b) \cup (b \setminus a) \\ ab &= a \cap b. \end{aligned}$$

One can check that this gives a ring, in which $-a = a$ for all a. All the axioms are straightforward except for the associativity of addition. For that, consider three elements $a, b, c \in \mathrm{Sub}(S)$. Put

$$\begin{aligned}
a' &= a \setminus (b \cup c) = \{s \mid s \text{ lies in } a \text{ but not } b \text{ or } c\} \\
b' &= b \setminus (a \cup c) = \{s \mid s \text{ lies in } b \text{ but not } a \text{ or } c\} \\
c' &= c \setminus (a \cup b) = \{s \mid s \text{ lies in } c \text{ but not } a \text{ or } b\} \\
u &= a' \cup b' \cup c' \cup (a \cap b \cap c).
\end{aligned}$$



By a check of cases, we find that

$$a + (b + c) = u = (a + b) + c$$

as required.

Definition 1.13. [defn-binary-product]
Let A and B be commutative rings. As usual, we write $A \times B$ for the cartesian product, so the elements of $A \times B$ are pairs (a, b) with $a \in A$ and $b \in B$. We define

$$\begin{aligned}
0_{A\times B} &= (0_A, 0_B) \\
1_{A\times B} &= (1_A, 1_B) \\
(a, b) + (a', b') &= (a + a', b + b') \\
(a, b)(a', b') &= (aa', bb') \\
-(a, b) &= (-a, -b).
\end{aligned}$$

It is easy to see that this makes $A \times B$ into a commutative ring.

Remark 1.14. [rem-axis-not-subring]
The set $A' = \{(a, 0_B) \mid a \in A\} \subseteq A \times B$ is naturally identified with the ring A, but it is not a subring of $A \times B$ because it does not contain the element $1_{A\times B} = (1_A, 1_B)$ (unless B is trivial). Similarly, the set $B' = \{(0_A, b) \mid b \in B\}$ is not a subring unless A is trivial.

Remark 1.15. [rem-infinite-product]
If we have rings $A_1, \dots, A_n$, we can make the product $A_1 \times \cdots \times A_n$ into a ring by an obvious generalisation of the above definition. We can even define the product of infinitely many factors, but we choose to postpone this until we have discussed rings of functions.

Definition 1.16. [defn-map]
For any sets S and T, we write $\mathrm{Map}(S, T)$ for the set of all functions from S to T.

Definition 1.17. [defn-map-ring]
Now suppose we have a set S and a ring A, and we put $M = \mathrm{Map}(S, A)$.

(a) We let 0_M denote the constant function $S \to A$ with value 0_A, so $0_M(s) = 0_A$ for all $s \in S$.
(b) Similarly, we define $1_M : S \to A$ by $1_M(s) = 1_A$ for all $s \in S$.
(c) Given elements $a, b \in M$ (so $a : S \to A$ and $b : S \to A$) we define $a + b \in M$ by $(a + b)(s) = a(s) + b(s)$ for all $s \in S$.

(d) Similarly, we define $ab \in M$ by $(ab)(s) = a(s)\, b(s)$ for all $s \in S$.

(e) We also define $(-a)(s) = -a(s)$.

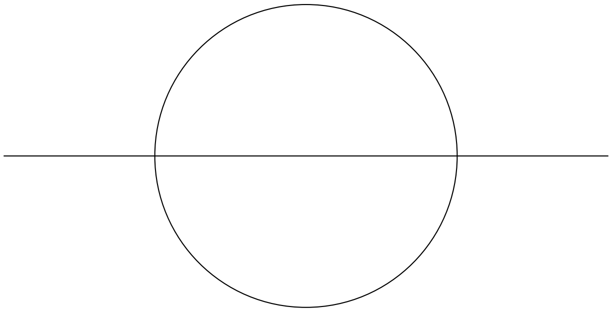
It is clear that these operations make M into a ring.

Remark 1.18. [`rem-function-rings`]

We rarely want to work with the full ring $\mathrm{Map}(S, A)$; instead, we consider various subrings. For example, the ring $\mathrm{Map}(\mathbb{R}, \mathbb{R})$ is too unstructured to be interesting, but it is useful to study the subring of continuous functions, or the subring of smooth (= infinitely differentiable) functions, or the subring of polynomial functions. Similarly, if X is any compact hausdorff space X then we can consider the ring $C(X)$ of continuous functions from X to $\mathbb{R}$. It can be shown that the topology of X is very closely related to the ring structure of $C(X)$. This is just the first of many different contexts where we can study spaces via suitable rings of functions.

Example 1.19. [eg-sunset]
We often want to consider subsets of $\mathbb{R}^n$ defined by polynomial equations, such as the set

$$X = \{(x, y) \in \mathbb{R}^2 \mid (x^2 + y^2 - 1)y = 0\}.$$



We can let A denote the ring of polynomial functions on X. If we have a system of coefficients $c_{ij} \in \mathbb{R}$ for $0 \leq i < N$ and $0 \leq j < M$, we get an element $f \in A$ given by

$$f(x, y) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} c_{ij} x^i y^j.$$

The defining equation for X can be written as $y^3 = y(1 - x^2)$. This in turn gives $y^4 = y^2(1 - x^2)$ and $y^5 = y^3(1 - x^2) = y(1 - x^2)^2$ and so on. This shows that we do not really need powers of y beyond y^2: every element $f \in A$ can be expressed in the form $f(x, y) = \sum_{i=0}^{L-1} \sum_{j=0}^{2} d_{ij} x^i y^j$ for some system of coefficients d_{ij}.

There is an extensive theory of the geometry of sets defined by polynomial equations, and its relationship with the structure of the corresponding rings of polynomial functions. This is called *algebraic geometry.*

Note that an element of $\mathrm{Map}(S, A)$ can be thought of as a family of elements $a(s) \in A$ parametrised by the elements $s \in S$. Sometimes it is more natural to use the notation a_s rather than $a(s)$. Moreover, we sometimes want to assume that a_s is an element of a ring A_s that depends on s, rather than having all the elements a_s lie in the same ring A. This leads us to the following construction:

Definition 1.20. [defn-general-product]
Suppose we have a set S and a ring A_s for each element $s \in S$. We define a new ring $P = \prod_{s \in S} A_s$ as follows. An element of P is a family of elements $(a_s)_{s \in S}$ with $a_s \in A_s$ for all s. The zero element is the family $0_P = (0_{A_s})_{s \in S}$, and similarly $1_P = (1_{A_s})_{s \in S}$. Given elements $a, b \in P$ we put $(a + b)_s = a_s + b_s$ and $(ab)_s = a_s b_s$ and $(-a)_s = -a_s$, which defines elements $a + b, ab, -a \in P$. It is clear that these operations make P into a commutative ring.

Remark 1.21. [rem-product-subring]
It is easiest to understand this definition in the case where there is a single ring A^* such that A_s is a subring of A^* for all s. It is then easy to identify $\prod_{s \in S} A_s$ with the ring

$$P' = \{a \in \mathrm{Map}(S, A^*) \mid a(s) \in A_s \text{ for all } s\},$$

which is a subring of $\mathrm{Map}(S, A^*)$. In particular, if $A_s = A^*$ for all s then we just have $\prod_{s \in S} A_s = \prod_{s \in S} A^* = \mathrm{Map}(S, A^*)$.

Example 1.22. [eg-padic]
Fix a prime number p, and consider the ring $P = \prod_{k=1}^{\infty} \mathbb{Z}/p^k$. We have defined $\mathbb{Z}/p^k$ to be a subset (but not a subring) of $\mathbb{N}$, so P can be regarded as a subset (but not a subring) of $\mathrm{Map}(\mathbb{N} \setminus 0, \mathbb{N})$. Specifically, P is the set of sequences $a = (a_1, a_2, \dots)$ of integers with $0 \leq a_k < p^k$ for all k. Now consider the subset

$$\mathbb{Z}_p = \{a \in P \mid a_k = a_{k+1} \pmod{p^k} \text{ for all } k\}.$$

One can check that this is a subring of P; it is called the ring of p-adic integers, and is important in algebraic number theory. For example, the sequences

$$\begin{aligned} a &= (1,\ 1+p,\ 1+p+p^2,\ 1+p+p^2+p^3,\ \dots) \\ b &= (p-1,\ p-1,\ p-1,\ p-1,\ \dots) \end{aligned}$$

are elements of $\mathbb{Z}_p$ with $ab+1=0$.

We now turn to polynomial rings and formal power series rings. We will first define them using notation that is rigorous but somewhat cumbersome, then we will introduce the more traditional notation.

Definition 1.23. [`defn-poly-initial`]

Let A be a ring. We write $F(A)$ for the set of sequences $a=(a_0,a_1,a_2,\dots)\in\operatorname{Map}(\mathbb{N},A)$, considered as a ring with the following operations:

$$\begin{aligned} 0_{F(A)} &= (0,0,0,0,\dots) \\ 1_{F(A)} &= (1,0,0,0,\dots) \\ (a+b)_i &= a_i+b_i \\ (ab)_i &= \sum_{j=0}^{i} a_j b_{i-j}. \end{aligned}$$

(Verification of the ring axioms is left to the reader.) We then put

$$\begin{aligned} P_{\leq d}(A) &= \{a\in F(A) \mid a_i=0 \text{ for } i>d\} \\ P(A) &= \bigcup_{d\geq 0} P_{\leq d}(A) = \{a\in F(A) \mid a_i = 0 \text{ for } i\gg 0\}. \end{aligned}$$

This is easily seen to be a subring of $F(A)$. We call $F(A)$ a *formal power series ring*, and $P(A)$ a *polynomial ring*. For $a\in P(A)$, we define the *degree* $\deg(a)$ to be the largest d such that $a_d\neq 0$, or $\deg(a)=-\infty$ if $a_i=0$ for all i.

The subset $P_{\leq 0}(A)\subseteq P(A)$ consists of sequences of the form $(a,0,0,0,\dots)$; it is a subring, which can be identified with A itself.

We now introduce a symbol, say x, for the sequence $(0,1,0,0,\dots)\in P(A)$. It is easy to see by induction that x^k is the sequence with 1 in position k and 0 elsewhere. More generally, if $u\in A=P_{\leq 0}(A)$ then ux^k has u in position k and 0 elsewhere. This means that any element

$$a=(a_0,a_1,\dots,a_d,0,0,0,\dots)\in P_{\leq d}(A)$$

can be expressed as $a=\sum_{i=0}^{d} a_i x^i$. More generally, for any $a\in F(A)$ it is natural to write $a=\sum_{i=0}^{\infty} a_i x^i$, with the understanding that this is just a notational convention, because we do not have any independent definition for sums of infinitely many terms.

The traditional notation is to write $A[x]$ for $P(A)$ if we want to use the symbol x for the sequence $(0,1,0,0,\dots)$, or $A[t]$ if we want to use the symbol t, and so on. We can then use the notation $A[x,y]$ for $A[x][y]=P(P(A))$. We find that any element of $A[x,y]$ can be expressed as $\sum_{i=0}^{d}\sum_{j=0}^{d} a_{ij}x^i y^j$ for some $d\in\mathbb{N}$ and some system of coefficients $a_{ij}\in A$. More generally, we can define multivariable polynomial rings $A[x_1,\dots,x_n]$ in essentially the same way. For formal power series rings we use double brackets like $A[\![t]\!]$ or $A[\![u,v,w]\!]$. We also write $A[x]_{\leq d}$ for $P_{\leq d}(A)$.

We conclude by discussing division of polynomials, which is central to many of the special properties of polynomial rings.

Definition 1.24. [`defn-monic`]

A polynomial $f(t)\in A[t]$ is *monic* if it has the form $f(t)=\sum_{i=0}^{d} a_i t^i$ for some $d\in\mathbb{N}$ and some sequence of coefficients $a_0,\dots,a_d\in A$ with $a_d=1$ (so f has degree d).

Proposition 1.25. [`prop-poly-div`]

Let $f\in A[t]$ be a monic polynomial of degree d over A, and define $\mu\colon A[t]\times A[t]_{<d}\to A[t]$ by

$$\mu(q,r)=fq+r.$$

Then μ is a bijection. (In other words, for any $g \in A[t]$ there is a unique pair of polynomials q and r such that $g = qf + r$ and r has degree less than d.)

Proof. First, our hypothesis is that

$$f = \sum_{i=0}^{d} a_i x^i,$$

where $a_d = 1$. It follows that fq has degree $d+\deg(q)$, and in particular, it has degree at least d unless $q = 0$. Now, if $\mu(q,r) = 0$ then $\deg(fq) = \deg(-r) < d$ so $q = 0$, and substituting this in the relation $\mu(q,r) = 0$ gives $r = 0$ as well. More generally, if $\mu(q_0, r_0) = \mu(q_1, r_1)$ then $\mu(q_0 - q_1, r_0 - r_1) = 0$, and we conclude that $(q_0, r_0) = (q_1, r_1)$. Thus, μ is injective.

Now consider a polynomial $g \in A[t]_{\leq k}$. We will show by induction on k that g lies in the image of μ. This is clear if $k < d$, because we then have $g = \mu(0, g)$. Suppose instead that $k \geq d$. Let u be the coefficient of t^k in g, and put $q_0 = u_k t^{k-d}$ and $g_1 = g - q_0 f$. Then $g_1 \in A[t]_{<k}$, so by induction we can find q_1, r with $g_1 = \mu(q_1, r) = q_1 f + r$. We now put $q = q_0 + q_1$ and observe that $g = qf + r$ as required. □

2. Ring homomorphisms

Definition 2.1. [defn-ring-hom]
Let A and B be rings. A *homomorphism* from A to B is a function $\phi\colon A \to B$ such that

(a) For all $a, a' \in A$ we have $\phi(a + a') = \phi(a) + \phi(a')$, and $\phi(aa') = \phi(a)\phi(a')$.
(b) $\phi(1_A) = 1_B$.

An *isomorphism* is a ring homomorphism that is also a bijection.

Remark 2.2. [rem-inv-hom]
If $\phi\colon A \to B$ is an isomorphism, it is straightforward to check that the inverse map $\phi^{-1}\colon B \to A$ is also a ring homomorphism (and therefore an isomorphism).

Remark 2.3. [rem-ring-hom]
As a special case of (a) we have $\phi(0_A) = \phi(0_A) + \phi(0_A)$, and we can add $-\phi(0_A)$ to both sides to get $\phi(0_A) = 0_B$. However, axiom (b) does not follow from (a) in the same way, as we see by considering the function $n \mapsto (n, 0)$ from $\mathbb{Z}$ to $\mathbb{Z} \times \mathbb{Z}$. (The same line of argument gives $\phi(1_A)^2 = \phi(1_A)$, which would be enough if we knew that $\phi(1_A)$ had a multiplicative inverse, but that might not be the case.)

Another special case of (a) gives $\phi(a) + \phi(-a) = \phi(a - a) = \phi(0_A) = 0_B$, so $\phi(-a) = -\phi(a)$.

Example 2.4. [eg-inc-hom]
The obvious inclusion maps $\mathbb{Z} \to \mathbb{Q} \to \mathbb{R} \to \mathbb{C}$ are ring homomorphisms.

Example 2.5. [eg-mod-hom]
The function $\pi\colon \mathbb{Z} \to \mathbb{Z}/n$ (as in Example 1.8) is a ring homomorphism.

Example 2.6. [eg-boolean-iso]
Let S be a set, so we have a ring $A = \mathrm{Sub}(S)$ as in Example 1.12 and a ring $B = \mathrm{Map}(S, \mathbb{Z}/2)$ as in Definition 1.17. We would like to define homomorphisms $\phi\colon A \to B$ and $\psi\colon B \to A$. Suppose that $a \in A$, so a is a subset of S. Then $\phi(a)$ should be a function from S to $\mathbb{Z}/2 = \{0, 1\}$, so for each $s \in S$ we should have an element $\phi(a)(s) \in \{0, 1\}$. We define

$$\phi(a)(s) = \begin{cases} 1 & \text{if } s \in a \\ 0 & \text{if } s \notin a. \end{cases}$$

In the opposite direction, suppose that $b \in B$, so $b\colon S \to \mathbb{Z}/2$, and $\psi(b)$ should be a subset of S. We put

$$\psi(b) = \{s \in S \mid b(s) = 1\}.$$

We leave it to the reader to check that ϕ and ψ are ring homomorphisms. It is also easy to see that $\psi(\phi(a)) = a$ and $\phi(\psi(b)) = b$, so ϕ and ψ are inverse to each other, so they are isomorphisms.

Example 2.7. [eg-Z-initial]

Let A be any ring. For $n \in \mathbb{N}$ we define $\eta(n) \in A$ inductively by the rules $\eta(0) = 0_A$ and $\eta(n{+}1) = \eta(n){+}1_A$ (so $\eta(4) = 1_A{+}1_A{+}1_A{+}1_A$, for example). We then define $\eta(-n) = -\eta(n)$, which gives a function $\eta\colon \mathbb{Z} \to A$. One can check that this is a homomorphism, and that it is the only possible homomorphism from $\mathbb{Z}$ to A. (In the language of category theory, this means that $\mathbb{Z}$ is an initial object in the category of rings.)

Example 2.8. [eg-eval]

Let A be any ring, and let u be an element of A. We can define a homomorphism $\epsilon_u\colon A[x] \to A$ by

$$\epsilon_u\left(\sum_{i=0}^{d} a_i x^i\right) = \sum_{i=0}^{d} a_i u^i,$$

or more briefly $\epsilon_u(f) = f(u)$. This is called an *evaluation homomorphism.* Note in particular that $\epsilon_u(x) = u$.

More generally, given a vector $u = (u_1, \dots, u_d) \in A^d$, we can define a homomorphism

$$\epsilon_u\colon A[x_1, \dots, x_d] \to A$$

by

$$\epsilon_u\left(\sum_{i_1,\dots,i_d \geq 0} a_{i_1,\dots,i_d} x_1^{i_1} \cdots x_d^{i_d}\right) = \sum_{i_1,\dots,i_d \geq 0} a_{i_1,\dots,i_d} u_1^{i_1} \cdots u_d^{i_d},$$

or more briefly $\epsilon_u(f) = f(u_1, \dots, u_d)$.

Example 2.9. [eg-gelfand]

Let X and Y be topological spaces, and let $f\colon X \to Y$ be a continuous map. Let $C(X)$ be the ring of continuous real-valued functions on X, and similarly for $C(Y)$. Note that if $u\colon Y \to \mathbb{R}$ is continuous, then so is the composite $u \circ f\colon X \to \mathbb{R}$. We can thus define a function $f^*\colon C(Y) \to C(X)$ by $f^*(u) = u \circ f$. If $v\colon Y \to \mathbb{R}$ is another continuous function, it is clear that $(u{+}v){\circ}f = (u{\circ}f){+}(v{\circ}f)$, and $(uv){\circ}f = (u{\circ}f)(v{\circ}f)$. Using this, we see that f^* is a ring homomorphism.

If X and Y are compact hausdorff spaces, it can be shown that every ring homomorphism $C(Y) \to C(X)$ arises in this way from a unique continuous map $X \to Y$.

For our final example, we need the following well-known congruence:

Lemma 2.10. [lem-frobenius]

Let p be a prime number, and suppose that $0 < k < p$. Then the binomial coefficient $\begin{pmatrix} p \\ k \end{pmatrix}$ is divisible by p.

Proof. Let X be the set of subsets $K \subset \mathbb{Z}/p$ with $|K| = k$, so $|X| = \begin{pmatrix} p \\ k \end{pmatrix}$. We can let $\mathbb{Z}/p$ act on X by the rule

$$a + \{u_1, \dots, u_k\} = \{a + u_1, \dots, a + u_k\}.$$

We claim that if $a \neq 0$ then $a + K$ can never be equal to K. Indeed, if $a + K = K$ then $ab + K = K$ for all b, but we can choose b so that $ab = 1 \pmod{p}$, so $1 + K = K$. Thus, if $u \in K$ then $u + 1 \in K$, then $u + 2 \in K$ and so on, so $K = \mathbb{Z}/p$, which contradicts the fact that $|K| = k$. Thus all orbits for the action of $\mathbb{Z}/p$ on X are free, so $|X|$ is p times the number of orbits, as required. □

Example 2.11. [eg-frobenius]

Let A be a ring in which $1_A + 1_A = 0_A$ (or more briefly $2 = 0$). For example, A could be the ring $\mathbb{F}_4$ from Example 1.11, or the polynomial ring $\mathbb{Z}/2[x]$. Define $\phi\colon A \to A$ by $\phi(u) = u^2$. This clearly sends 1 to 1 and preserves multiplication. Less obviously, it preserves addition, because $\phi(u + v) - \phi(u) - \phi(v) = 2uv = 0$. Thus, ϕ is a ring homomorphism, called the *frobenius map.*

More generally, if p is a prime number and B is a ring with $p = 0$ one can check using Lemma 2.10 that $(u + v)^p = u^p + v^p$ in B, so we have a frobenius map $\phi\colon B \to B$ given by $\phi(u) = u^p$.

3. Properties of elements

Definition 3.1. [defn-el-props]

Let a be an element in a ring A.

(a) We say that a is *invertible* if there is an element b such that $ab = 1$. Such an element is called an *inverse* for a, and we write a^{-1} for b. We write $A^{\times}$ for the set of invertible elements in A.
(b) We say that a is a *zero-divisor* if there is an element $x \neq 0$ such that $ax = 0$. Otherwise, we say that a is *regular*.
(c) We say that a is *nilpotent* if $a^n = 0$ for some $n \in \mathbb{N}$.
(d) We say that a is *idempotent* if $a^2 = a$, or equivalently $a(1-a) = 0$.

The following result shows that the notation a^{-1} is unambiguous.

Proposition 3.2. [prop-inv-unique]

If an element $a \in A$ is invertible, then it has a unique inverse.

Proof. Suppose that b and c are both inverses for a. Then

$$b = b1 = b(ac) = (ba)c = 1c = c.$$

□

Example 3.3. [eg-C-el-props]

In $\mathbb{C}$, every nonzero element is regular and invertible, 0 is the only nilpotent element, and 0 and 1 are the only idempotent elements.

Proposition 3.4. [prop-Zn-el-props]

Consider an element $a \in \mathbb{Z}/n = \{0, 1, \dots, n-1\}$.

(a) *a is invertible iff it is regular iff a and n are coprime, or equivalently a is not divisible by any prime that divides n.*
(b) *a is nilpotent iff it is divisible by every prime that divides n.*

Proof. The proof will rely on various facts from elementary number theory, which can also be recovered as a special case of the results that will will be discussed in Section 16.

(a) a is invertible iff there is another integer b such that $ab = 1 \pmod{n}$, or equivalently there exist integers b and m such that $ab + nm = 1$, or equivalently a and n are coprime. Moreover, we can define a map $\mu\colon \mathbb{Z}/n \to \mathbb{Z}/n$ by $\mu(x) = ax$. It is clear that μ is injective iff a is regular, and μ is bijective iff as a is invertible. However, if μ is injective then $|\mu(\mathbb{Z}/n)| = |\mathbb{Z}/n| = n$ so $\mu(\mathbb{Z}/n)$ is all of $\mathbb{Z}/n$ so μ is automatically bijective. Thus a is invertible iff it is regular.
(b) Let the prime factorisation of n be $p_1^{v_1} \cdots p_r^{v_r}$, with $p_1 < \cdots < p_r$. Put $v = \max(v_1, \dots, v_r)$. If a is divisible by $p_1, \dots, p_r$, then a^v will be divisible by n in $\mathbb{Z}$, so $a^v = 0$ in $\mathbb{Z}/n$, so a is nilpotent in $\mathbb{Z}/n$. Conversely, if a is not divisible by p_i for some i, then a^k will never be divisible by p_i and so will never be divisible by n, so a^k will always be nonzero in $\mathbb{Z}/n$, so a will not be nilpotent in $\mathbb{Z}/n$.

□

Proposition 3.5. [prop-inv-prod]

For any two elements $a, b \in A$, the product ab is invertible iff a and b are both invertible.

Proof. Put $c = ab$. If a and b are both invertible, then $a^{-1}b^{-1}$ is an inverse for c. Conversely, if c is invertible then bc^{-1} is an inverse for a, and ac^{-1} is an inverse for b. □

Proposition 3.6. [prop-regular-prod]

For any two elements $a, b \in A$, the product ab is regular iff a and b are both regular. Moreover, every invertible element is regular.

Proof. Put $c = ab$. Suppose that a and b are both regular. Consider an element x such that $cx = abx = 0$. As a is regular we must have $bx = 0$, and as b is also regular we see that $x = 0$. Thus c is regular.

Conversely, suppose that c is regular. Consider an element x with $bx = 0$. It follows that $cx = abx = 0$, but c is regular so $x = 0$. This proves that b is regular, and essentially the same argument also shows that a is regular.

Finally, suppose that a is invertible. If $ax = 0$ we can multiply by a^{-1} to get $x = 0$; so a is regular. □

Proposition 3.7. [prop-finite-regular]
Let A be a ring with only finitely many elements. Then every regular element of A is invertible.

Proof. This is a straightforward generalisation of Proposition 3.4(a). We can enumerate the distinct elements as $a_1, \dots, a_n$ say. Suppose that $u \in A$ is a regular element, so all the elements $u(a_i - a_j)$ are nonzero, so the elements $ua_1, \dots, ua_n$ are distinct. As there are n elements in this list, every element of A must appear. In particular, we have $ua_k = 1$ for some k, so a_k is an inverse for u. □

Proposition 3.8. [prop-nilpotent-sum]
If a and b are nilpotent, then so is $a + b$. More precisely, if $a^{n+1} = b^{m+1} = 0$ then $(a + b)^{n+m+1} = 0$.

Proof. If $a^i b^j \neq 0$ then we must have $i \leq n$ and $j \leq m$, so $i + j \leq n + m$. Thus, if $i + j = n + m + 1$ we must have $a^i b^j = 0$. In other words, all terms in the binomial expansion of $(a + b)^{n+m+1}$ are zero, so $(a + b)^{n+m+1} = 0$ as claimed. □

Proposition 3.9. [prop-nilp-inv]
If a is nilpotent then $1 + a$ is invertible.

Proof. For some n we have $a^{n+1} = 0$. Put $u = \sum_{j=0}^{n} (-a)^j$; we then find that $(1 + a)u = 1 - a^{n+1} = 1$, so u is the required inverse for $1 + a$. □

Proposition 3.10. [cor-nilp-inv]
If u is invertible and a is nilpotent then $u + a$ is invertible.

Proof. We can write $u + a$ as $u(1 + au^{-1})$. Here au^{-1} is nilpotent so $1 + au^{-1}$ is invertible so $u + a$ is invertible. □

Proposition 3.11. [prop-idempotent-ops]
The elements 0 and 1 are idempotent. Moreover, if a and b are idempotent then so are the elements $1 - a$ and $1 - b$ and $a + b - ab$.

Proof. Straightforward, especially if we note that the condition $a^2 = a$ is equivalent to $a(1 - a) = 0$ and that $a + b - ab = 1 - (1 - a)(1 - b)$. □

Proposition 3.12. [prop-root-one]
If e is idempotent then the element $u = 1 - 2e$ has $u^2 = 1$ and so is invertible.

Proof. By expanding everything out and recalling that $e(1 - e) = 0$ we see that $u^2 = 1 - 4e(1 - e) = 1$. □

Proposition 3.13. [prop-lifting]

(a) *If e and e' are idempotent and $e' - e$ is nilpotent then $e' = e$.*
(b) *Let e be an element of A such that the element $x = e(1 - e)$ is nilpotent. Then there is an element $a \in A$ such that $e + ax$ is idempotent. Moreover, this is the unique idempotent e' such that $e' - e$ is nilpotent.*

Proof.

(a) Put $x = e' - e$ and $u = 1 - 2e$. If we expand out $x(1 - xu)$ using $e^2 = e$ and $(e')^2 = e'$ repeatedly, we get zero. As x is nilpotent we see that $1 - xu$ is invertible, so we can multiply by the inverse to get $x = 0$.
(b) Suppose that $x^n = 0$, and consider the element

$$y = 1 - e^n - (1 - e)^n.$$

It is clear from the binomial expansion that $(e + f)^n - e^n - f^n$ is always divisible by ef. Taking $f = 1 - e$, we see that y is divisible by x, say $y = vx$ for some v. It follows that the element $u = e^n + (1 - e)^n$ can be written as $1 - vx$, and so is invertible. We put $e' = e^n u^{-1}$, so $1 - e' =$

$(1-e)^n u^{-1}$, so $e'(1-e') = x^n u^{-2} = 0$, so e' is idempotent. Note also that if we put $b = \sum_{k=0}^{n-2} e^k$ we have $e - e^n = bx$ and $u^{-1} - 1 = u^{-1}vx$ so

$$e' - e = e^n u^{-1} - e = e^n(u^{-1} - 1) - (e - e^n) = (e^n u^{-1} v - b)x.$$

Thus, if we put $a = e^n u^{-1} v - b$ we have $e' = e + ax$ as required. Uniqueness follows from part (a) together with Proposition 3.8.

□

Remark 3.14. [rem-lifting]

One can give an interesting alternative formula for the idempotent in part (b). Put $x = e(1-e)$ and $b = \sum_{k=0}^{\infty} \left(\begin{array}{c} 2k+1 \\ k \end{array} \right) x^k$ (noting that this is really only a finite sum, because x is nilpotent). Using some standard combinatorics one can check that $(1-4x)(b^2x - b) = 1$. Now put $a = (2e-1)b$ and $e' = e + ax$. We find that $(1-4x)e'(1-e') = e(1-e) - x = 0$ but $1 - 4x$ is invertible so e' is idempotent.

The following result is a nice illustration of all the above concepts.

Proposition 3.15. *Let A be a ring with only finitely many elements, and suppose that the only idempotents in A are 0 and 1. Then every element of A is either nilpotent or invertible.*

Proof. Let a be an element of A. As A is finite, the powers of a cannot all be different. It follows that there exist integers $p, q > 0$ with $a^{p+q} = a^p$. We then see by induction that $a^{p+kq} = a^p$ for all $k \geq 0$. In particular, we have $a^{p+pq} = a^p$. Multiplying both sides by $a^{(q-1)p}$ gives $a^{2pq} = a^{pq}$, so a^{pq} is idempotent, so either $a^{pq} = 0$ or $a^{pq} = 1$. In the first case, a is nilpotent. In the second case, a^{pq-1} is an inverse for a, so a is invertible. □

Proposition 3.16. [prop-poly-el-props]

Let A be a ring, and let $f = \sum_k a_k x^k \in A[x]$ be a polynomial over A.

(a) *If a_0 is invertible in A, and a_i is nilpotent in A for all $i > 0$, then f is invertible in $A[x]$. Conversely, if f is invertible in $A[x]$ then a_0 is invertible in A.*
(b) *If the first nonzero coefficient in f is regular in A, then f is regular in $A[x]$. Similarly, if the last nonzero coefficient in f is regular in A, then f is regular in $A[x]$.*
(c) *f is nilpotent in $A[x]$ iff all coefficients a_i are nilpotent in A.*
(d) *f is idempotent in $A[x]$ iff a_0 is idempotent in A and $a_i = 0$ for $i > 0$.*

Remark 3.17. In fact, claim (a) is fully reversible: if f is invertible in $A[x]$, then the coefficients a_i are automatically nilpotent for $i > 0$. However, we will defer the proof, as it will become much easier when we have more theory available.

Proof. (c) If all the coefficients a_i are nilpotent, then all the individual terms $a_i x^i$ are nilpotent, so f is nilpotent by Proposition 3.8. Conversely, suppose that f is nilpotent, say $f^n = 0$. If $f = a_d x^d + \text{ lower terms }$ then $f^n = a_d^n x^{nd} + \text{ lower terms }$, so we must have $a_d^n = 0$, so a_d is nilpotent. It follows using Proposition 3.8 again that the polynomial

$$g = \sum_{i=0}^{d-1} a_i x^i = f + (-a_d x^n)$$

is again nilpotent in $A[x]$, so by induction on d we can conclude that the coefficients $a_0, \ldots, a_{d-1}$ are also nilpotent in A.

(a) First suppose that a_0 is invertible and that a_i is nilpotent for $i > 0$. Then the polynomial $g = \sum_{i>0} a_i x^i$ is nilpotent by (c), so the polynomial $f = a_0 + g$ is invertible by Proposition 3.9. Conversely, if f is invertible with inverse $g = \sum_i b_i x^i$, we find that $a_0 b_0 = 1$, so a_0 is invertible.
(b) Suppose that f has lowest term $a_n x^n$ and highest term $a_m x^m$, whereas g has lowest term $b_p x^p$ and highest term $b_q x^q$. Then

$$fg = a_n b_p x^{n+p} + \text{ intermediate terms } + a_m b_q x^{m+q},$$

so fg can only be zero if $a_n b_p = 0$ and $a_m b_q = 0$. (All this is still valid even if $n = m$ or $p = q$.) The claim follows easily.

(d) If a_0 is idempotent and $a_i = 0$ for $i > 0$ then it is clear that f is idempotent. Conversely, suppose that f is idempotent, so $f^2 = f$. By looking at the constant term, we see that a_0 is idempotent. It follows that the poynomial $g = (1 - a_0)f$ is also idempotent, and is divisible by x. Now for all $n > 0$ we have $g = g^n$ and so g is divisible by x^n; this makes it clear that $g = 0$. It follows that $f = a_0 f$, so $a_i = a_0 a_i$ for $i > 0$. Suppose that $a_i = 0$ for $0 < i < m$; then the coefficient of x^m in $f^2 - f$ is $2a_0 a_m - a_m = a_m$, so $a_m = 0$ as well. We deduce by induction that $a_i = 0$ for all $i > 0$, as claimed. □

If $A = B \times C$ then the element $e = (1_B, 0)$ is clearly idempotent, with $1_A - e = (0, 1_C)$. The following proposition shows that all idempotents arise in essentially this way.

Proposition 3.18. [prop-idempotent-splitting]
Let A be a ring, and let $e \in A$ be idempotent. Put

$$B = \{b \in A \mid be = b\} \qquad\qquad C = \{c \in A \mid c(1 - e) = c\}.$$

Then B can be regarded as a ring with $1_B = e$, and C can be regarded as a ring with $1_C = 1_A - e$. Moreover, there is a ring isomorphism $\phi\colon A \to B \times C$ given by $\phi(a) = (ae, a(1 - e))$ with $\phi^{-1}(b, c) = b + c$.

Proof. It is clear that B contains 0 and is closed under addition, subtraction and multiplication. If we define $1_B = e$ then we have $1_B b = b$ by the definition of B. All other ring axioms for B follow immediately from the corresponding axioms for A. We can thus regard B as a ring, and the same argument works for C.

Next, we have $e^2 = e$, so for any $a \in A$ we have $(ae)e = ae$, so $ae \in B$. Similarly $a(1 - e) \in C$, so the formula $\phi(a) = (ae, a(1 - e))$ defines a function $A \to B \times C$. This clearly respects addition and sends 0 to 0. We also have $\phi(1_A) = (e, 1 - e) = (1_B, 1_C) = 1_{B\times C}$, and using $e^2 = e$ and $(1 - e)^2 = 1 - e$ we see that $\phi(aa') = \phi(a)\phi(a')$. Thus, ϕ is a ring homomorphism. In the opposite direction, we define $\psi(b, c) = b + c$, which clearly respects addition and sends 0 to 0. Note that if $b \in B$ and $c \in C$ we have $bc = bec(1 - e) = bc(e - e^2) = 0$. Using this, we find that $\psi(bb', cc') = \psi(b, c)\psi(b', c')$. We also have $\psi(1_{B\times C}) = \psi(e, 1 - e) = 1_A$, so ψ is a ring homomorphism. It is also straightforward to check that $\phi\psi\colon A \to A$ and $\psi\phi\colon B \times C \to B \times C$ are identity maps, so ϕ and ψ are isomorphisms and are inverse to each other. □

4. Properties of rings

Definition 4.1. [defn-ring-props]
Let A be a nontrivial ring.

(a) We say that A is a *field* if every nonzero element is invertible.
(b) We say that A is *local* if for every element $a \in A$, either a is invertible or $1 - a$ is invertible.
(c) We say that A is a *domain* if all nonzero elements are regular, or equivalently, the product of any two nonzero elements is nonzero.
(d) We say that A is *irreducible* if whenever a and b are nonzero elements of A, there are elements x and y with $ax = by \neq 0$.
(e) We say that A is a *predomain* if all non-nilpotent elements are regular.
(f) We say that A is *reduced* if the only nilpotent element in A is 0.

By definition, the trivial ring is considered to be reduced, but not to have any of the other properties listed above.

Remark 4.2. [rem-ring-props]
It is clear that every field is a local domain, and that every domain is a reduced predomain. Moreover, every domain is also irreducible, because we can take $x = b$ and $y = a$ in the definition. Also, every irreducible noetherian ring is a predomain; the noetherian condition will be introduced in Section 18, where we will show that it is satisfied by many of the most commonly studied rings.

Proposition 4.3. [prop-Zn-props]
Consider a ring $A = \mathbb{Z}/n$ with $n > 0$.

(a) *A is a field iff it is a domain iff n is prime.*
(b) *A is local iff $n = p^k$ for some prime number p and some $k > 0$.*

(c) *A is reduced iff there is no prime p such that n is divisible by p^2.*

Proof. Recall from Proposition 3.4 that $a \in \mathbb{Z}/n$ is regular iff invertible iff a and n are coprime, and that a is nilpotent iff it is divisible by every prime that divides n. Claims (a) and (c) are clear from this. For claim (b), we first suppose that $n = p^k$ for some prime p and some $k > 0$. If $a \in \mathbb{Z}/n$ then a and $1-a$ cannot both be divisible by p, so one of them is coprime to n, so one of them is invertible in $\mathbb{Z}/n$; this proves that $\mathbb{Z}/n$ is local. Convesely, suppose that n is not of the form p^k. If $n = 1$ then $\mathbb{Z}/n$ is the trivial ring, which by definition is not local. If $n > 1$ then we can write $n = uv$, where u and v are coprime integers that are both larger than one. As they are coprime, we have $ux + vy = 1$ for some integers x, y. Take $a = \pi(ux) \in \mathbb{Z}/n$, so $1 - a = \pi(vy)$. As $\gcd(ux, n) = u > 1$ and $\gcd(vy, n) = v > 1$ we see that neither a nor $1-a$ is invertible in $\mathbb{Z}/n$, so $\mathbb{Z}/n$ is not local. □

Example 4.4. [`eg-Zpl-local`]
Let p be a prime number, and let $\mathbb{Z}_{(p)}$ denote the set of rational numbers of the form u/v with $u \in \mathbb{Z}$ and $v \in \mathbb{Z} \setminus p\mathbb{Z}$. This is easily seen to be a subring of $\mathbb{Q}$ (and thus a domain). It is not a field, because p is a nonzero element of $\mathbb{Z}_{(p)}$ that has no inverse in $\mathbb{Z}_{(p)}$. However, we claim that it is a local ring. To see this, consider an element $a \in \mathbb{Z}_{(p)}$, say $a = u/v$ in lowest terms, so v is not divisible by p. It follows that u and $v - u$ cannot both be divisible by p. If u is not divisible by p then v/u is an inverse for a in $\mathbb{Z}_{(p)}$, and if $v - u$ is not divisible by p then $v/(v-u)$ is an inverse for $1 - a$ in $\mathbb{Z}_{(p)}$.

Proposition 4.5. *Let A be a nontrivial ring.*

(a) *$A[x]$ is never a field or a local ring.*
(b) *$A[x]$ is a domain iff A is a domain.*
(c) *$A[x]$ is reduced iff A is reduced.*

Proof. Recall from Proposition 3.16 that a polynomial $f \in A[x]$ is nilpotent iff it has nilpotent coefficients, and that f is regular if either the lowest or the highest nonzero coefficient is regular. Claims (b) and (c) follow easily from this. Next note that if $f = ax^d + \text{ lower terms }$ then $xf = ax^{d+1} + \text{ lower terms }$ and $(1-x)f = -ax^{d+1} + \text{ lower terms }$, so neither xf nor $(1-x)f$ can be equal to 1. It follows that neither x nor $1 - x$ is invertible, so $A[x]$ is not a field or a local ring. □

5. Ideals

Definition 5.1. An *ideal* in a ring A is a subset $I \subseteq A$ such that

- $0 \in I$
- If $a, b \in I$ then $a + b \in I$
- If $a \in A$ and $b \in I$ then $ab \in I$.

Example 5.2. [`eg-degenerate-ideals`]
The sets $\{0\}$ and A are ideals in A. We will usually write 0 rather than $\{0\}$.

Example 5.3. [`defn-principal-ideal`]
For any element $x \in A$ the set $Ax = \{ax \mid a \in A\}$ is an ideal. Ideals of this type are called *principal* ideals.

Example 5.4. [`eg-annihilator`]
For any subset $S \subseteq A$ we put

$$\operatorname{ann}_A(S) = \{a \in A \mid as = 0 \text{ for all } s \in S\}.$$

This is called the *annihilator* of S in A; it is easily seen to be an ideal. Important special cases are where S consists of a single element (in which case we write $\operatorname{ann}(s)$ rather than $\operatorname{ann}(\{s\})$) or where S itself is also an ideal.

Definition 5.5. [`defn-lin-comb`]
Consider again a subset $S \subseteq A$. We say that an element $a \in A$ is an *A-linear combination* of S if there exists a finite list $s_1, \dots, s_n$ of elements of S and a finite list $c_1, \dots, c_n$ of elements of A such that $a = \sum_i c_i s_i$. We write $\operatorname{span}_A(S)$ for the set of all linear combinations of S. It is easy to see that this is an ideal, and that any ideal containing S also contains $\operatorname{span}_A(S)$

Proposition 5.6. [prop-ker-ideal]

Let $\phi\colon A \to B$ be a ring homomorphism. Then the kernel $\ker(\phi) = \{a \in A \mid \phi(a) = 0\}$ *is an ideal in A. Moreover, we have* $\ker(\phi) = 0$ *if and only if ϕ is injective.*

Proof. It is clear that $0 \in \ker(\phi)$. Suppose that $a, b \in \ker(\phi)$ and $r \in A$, so $\phi(a) = \phi(b) = 0$. It follows that $\phi(a + b) = \phi(a) + \phi(b) = 0 + 0 = 0$ and $\phi(ra) = \phi(r)\phi(a) = \phi(r)0 = 0$, so $a + b, ra \in \ker(\phi)$.

Now suppose that $\ker(\phi) = 0$. If $a, b \in A$ with $\phi(a) = \phi(b)$ then we have $\phi(a - b) = \phi(a) - \phi(b) = 0$ so $a - b \in \ker(\phi) = 0$ so $a = b$; this proves that ϕ is injective. Conversely, suppose that ϕ is injective. For $a \in \ker(\phi)$ we have $\phi(a) = 0 = \phi(0)$, so injectivity means that $a = 0$; thus $\ker(\phi) = 0$. □

Proposition 5.7. [prop-eval-ker]

Consider a vector $u \in A^d$ and the evaluation homomorphism

$$\epsilon_u\colon A[x_1, \dots, x_d] \to A$$

as in Example 2.8. Then

$$\ker(\epsilon_u) = \operatorname{span}_{A[x_1,\dots,x_d]}(x_1 - u_1, \dots, x_d - u_d).$$

Proof. Put $B = A[x_1, \dots, x_d]$ and $K_u = \ker(\epsilon_u\colon B \to A)$ and $L_u = \operatorname{span}_B(x_i - u_i \mid 1 \leq i \leq d)$. By definition we have $\epsilon_u(x_i - u_i) = u_i - u_i = 0$, so $x_i - u_i \in K$ for all i, so $L_u \leq K_u$. For the reverse inclusion, we first consider the special case where $u_i = 0$ for all i, so ϵ_u just sends every polynomial to its constant term. It follows that if $\epsilon_0(f) = 0$, then f is an A-linear combination of monomials $x_1^{i_1} \cdots x_d^{i_d}$ where at least one exponent has $i_k > 0$, so the monomial is a multiple of x_k. From this it is clear that $K_0 = L_0$. For the general case, note that we can define homomorphisms

$$B \xrightarrow{\alpha} B \xrightarrow{\beta} B$$

by

$$\begin{aligned} \alpha(f(x_1, \dots, x_d)) &= f(x_1 + u_1, \dots, x_d + u_d) \\ \beta(f(x_1, \dots, x_d)) &= f(x_1 - u_1, \dots, x_d - u_d). \end{aligned}$$

These satisfy $\alpha\beta = \beta\alpha = 1$, so they are isomorphisms. We have $\epsilon_u = \epsilon_0 \circ \alpha$, so

$$K_u = \ker(\epsilon_0 \circ \alpha) = \alpha^{-1}(K_0) = \beta(K_0) = \beta(L_0),$$

but it is clear that $\beta(L_0) = L_u$, so $K_u = L_u$ as claimed. □

Remark 5.8. [rem-congruence]

Ideals do not usually contain 1 and so are usually not subrings. However, they can be related to subrings as follows. Given an ideal $I \subseteq A$, we put $E_I = \{(a, b) \in A \times A \mid a - b \in I\}$. As with any subset of $A \times A$, this can be regarded as a relation on the set A, with a related to b iff $(a, b) \in E_I$. We find that E_I is both a subring of $A \times A$ and an equivalence relation. Conversely, if F is a subring of $A \times A$ that is an equivalence relation, then there is a unique ideal I such that $F = E_I$, namely $I = \{a \in A \mid (a, 0) \in F\}$. We will not need this, so we leave the proof as an exercise.

Definition 5.9. [defn-ideal-ops]

Let I and J be ideals in A.

- We write $I \cap J$ for the intersection, so $a \in I \cap J$ iff $a \in I$ and $a \in J$.
- We write $I + J$ for the set of all elements $a \in A$ that can be expressed in the form $a = b + c$ with $b \in I$ and $c \in J$.
- We write IJ for the set of elements $a \in A$ that can be expressed in the form $a = \sum_{i=1}^n b_i c_i$, with $b_i \in I$ and $c_i \in J$.
- We write $(I : J)$ for the set of elements $a \in A$ such that $aJ \subseteq I$.

Remark 5.10. [rem-ideal-ops]

It is easy to see that all the above sets are ideals. Moreover, we have $IJ \subseteq I \cap J \subseteq I \subseteq I + J$ and $IJ \subseteq I \cap J \subseteq J \subseteq I + J$.

We can apply the operations in Definition 5.9 repeatedly to define $I_1 \cap \dots \cap I_t$ and $I_1 + \dots + I_t$ and $I_1 \cdots I_t$ for any finite list of ideals $I_1, \dots, I_t$. There are also versions of the first two operations for infinite families:

Definition 5.11. [defn-infinite-ideal-ops]

Suppose we have a family of ideals I_s for s in some index set S. We put

$$\bigcap_s I_s = \{a \in A \mid a \in I_s \text{ for all } s \in S\}$$

$$\sum_s I_s = \operatorname{span}_A\left(\bigcup_s I_s\right).$$

These are again easily seen to be ideals.

The union of a family of ideals is not generally an ideal, but there are some important special cases where it is an ideal.

Proposition 5.12. [prop-chain-union]

Suppose we have a family of ideals I_n for $n \in \mathbb{N}$ such that $I_n \subseteq I_{n+1}$ for all n. Then the set $I_\infty = \bigcup_n I_n$ is also an ideal.

Proof. Suppose that $a, b \in I_\infty$, so $a \in I_n$ and $b \in I_m$ for some $n, m \in \mathbb{N}$. Put $p = \max(n, m)$, so both a and b lie in the ideal I_p, so $a + b$ lies in $I_p \subseteq I_\infty$. Thus I_∞ is closed under addition, and it is clear that it also contains zero and is closed under multiplication by elements of R. □

Example 5.13. [eg-torsion]

Fix an element $a \in A$, and put

$$I = \{x \in A \mid a^n x = 0 \text{ for some } n \in \mathbb{N}\}.$$

We find that I is the union of the chain

$$\operatorname{ann}_A(a) \subseteq \operatorname{ann}_A(a^2) \subseteq \operatorname{ann}_A(a^3) \subseteq \operatorname{ann}_A(a^3) \subseteq \cdots,$$

and so I is an ideal.

Lemma 5.14. [lem-unit-ideal]

Let I be an ideal in a ring A. Then $I = A$ iff $1 \in I$ iff I contains an invertible element.

Proof. If I contains an invertible element u then it contains $u^{-1}u = 1$, and so for every element $a \in A$ it contains $a.1 = a$, so $I = A$. Conversely, if $I = A$ then I contains the element 1 which is invertible. □

Corollary 5.15. [cor-field-ideals]

If A is a field then the only ideals in A are 0 and A.

Proof. Any nonzero ideal contains an invertible element and so is all of A. □

Definition 5.16. [defn-radical]

For any ideal I in A we put

$$\sqrt{I} = \{a \in A \mid a^n \in I \text{ for some } n \geq 0\},$$

and we call this the *radical* of I. We also use the notation $\operatorname{Nil}(A)$ for $\sqrt{0} = \{$ nilpotent elements in $A\}$, and call this the *nilradical* of A. We say that A is *reduced* if $\operatorname{Nil}(A) = 0$.

Proposition 5.17. [prop-radical]

If I is an ideal in A then $\sqrt{I}$ is also an ideal and $I \subseteq \sqrt{I}$. In particular, $\operatorname{Nil}(A)$ *is an ideal in A.*

Proof. Suppose that $a, b \in \sqrt{I}$, so $a^{n+1}, b^{m+1} \in I$ for some $n, m \in \mathbb{N}$. We then have $a^i b^j \in I$ whenever $i > n$ or $j > m$, so $(a+b)^{n+m+1} \in I$ by the same logic as in Proposition 3.8, so $a + b \in \sqrt{I}$. The other two axioms are easy, so $\sqrt{I}$ is an ideal as claimed. It is also clear that $I \subseteq \sqrt{I}$. □

Definition 5.18. [defn-jacobson]

We put

$$\operatorname{Rad}(A) = \{a \in A \mid 1 + ax \text{ is invertible for all } x \in A\},$$

and call this the *Jacobson radical* of A.

Proposition 5.19. [prop-nil-rad]
$\text{Rad}(A)$ *is an ideal in* A*, and* $\text{Nil}(A) \subseteq \text{Rad}(A)$.

Proof. Suppose that $a, b \in \text{Rad}(A)$. For any $x \in A$ we see that $1 + bx$ is invertible, and then that $1 + ax(1 + bx)^{-1}$ is invertible, so the product $(1 + ax(1 + bx)^{-1})(1 + bx)$ is invertible, but that product is $1 + (a + b)x$. Using this we see that $\text{Rad}(A)$ is closed under addition. The other two axioms are easy, so $\text{Rad}(A)$ is an ideal. If a is nilpotent then so is ax for all x, so $1 + ax$ is invertible by Proposition 3.9. This proves that $\text{Nil}(A) \subseteq \text{Rad}(A)$. □

Proposition 5.20. [prop-local-max]
Let A *be a local ring, and let* I *be the set of elements that are not invertible. Then* I *is an ideal in* A*, and in fact is equal to* $\text{Rad}(A)$*. Moreover, if* J *is any other ideal then either* $J = A$ *or* $J \subseteq I$.

Proof. Suppose that $a+b$ is invertible. We claim that either a or b is invertible. To see this, put $x = a(a+b)^{-1}$, so $1 - x = b(a + b)^{-1}$. As A is local, either x or $1 - x$ must be invertible. If x is invertible then $a = x(a + b)$ is invertible, and if $1 - x$ is invertible then $b = (1 - x)(a + b)$ is invertible, which proves the claim. By the contrapositive, if a and b are non-invertible then $a + b$ is non-invertible, so I is closed under addition. The other two axioms are easy, so I is an ideal. Any ideal not contained in I must contain an invertible element and so must be equal to A. Thus every ideal is either contained in I or is equal to A, as claimed.

As A is local we have $1 \neq 0$, so 0 is not invertible, so $1 \notin \text{Rad}(A)$. We must therefore have $\text{Rad}(A) \subseteq I$. On the other hand, if $a \in I$ then $-ax$ is non-invertible for all x (by Proposition 3.5), so $1 + ax$ must be invertible by the locality condition. This shows that $I \subseteq \text{Rad}(A)$, so $I = \text{Rad}(A)$ as claimed. □

Definition 5.21. [defn-quotient-ring]
Let I be an ideal in a ring A. A *coset* of I in A is a subset $u \subseteq A$ of the form $u = a + I$ for some $a \in A$. We write A/I for the set of all cosets, and we define $\pi \colon A \to A/I$ by $\pi(a) = a + I$. Given cosets $u, v \in A/I$ we put

$$\begin{aligned} u + v &= \{a + b \mid a \in u,\ b \in v\} \subseteq A \\ uv &= \{ab + x \mid a \in u,\ b \in v,\ x \in I\}. \end{aligned}$$

Proposition 5.22. [prop-quotient-ring]
In the above context, the sets $u + v$ *and* uv *are cosets. More specifically, if* $u = \pi(a)$ *and* $v = \pi(b)$ *then* $u + v = \pi(a + b)$ *and* $uv = \pi(ab)$*. Moreover, with the above definition of addition and multiplication, the set* A/I *becomes a ring, with* $0_{A/I} = \pi(0) = I$ *and* $1_{A/I} = \pi(1) = 1 + I$*. The map* $\pi \colon A \to A/I$ *is a ring homomorphism with* $\ker(\pi) = I$.

Proof. First, if $x \in \pi(a) + \pi(b)$ then $x = (a + u) + (b + v)$ for some $u, v \in I$. This can be rewritten as $x = (a + b) + (u + v)$ with $u + v \in I$, so we see that $\pi(a) + \pi(b) \subseteq \pi(a + b)$. Conversely, if $x \in \pi(a + b)$ then $x = a + b + v$ for some $v \in I$, and this is the sum of elements $a \in \pi(a)$ and $b + v \in \pi(b)$, so we see that $\pi(a + b) \subseteq \pi(a) + \pi(b)$, so $\pi(a) + \pi(b) = \pi(a + b)$ as claimed. In particular, the sum of any two cosets is again a coset.

Now suppose instead that $x \in \pi(a)\pi(b)$, so $x = (a + u)(b + v) + w$ for some $u, v, w \in I$. This can be rewritten as $x = ab + (av + ub + uv + w)$, with $av + ub + uv + w \in I$, so we see that $\pi(a)\pi(b) \subseteq \pi(ab)$. The reverse inclusion is clear, so we have $\pi(a)\pi(b) = \pi(ab)$; in particular, the product of any two cosets is again a coset.

The set A/I now has well-defined operations of addition and multiplication. We claim that for all cosets u, v and w we have $u(v+w) = uv+uw$. Equivalently, for all elements $a, b, c \in A$ we claim that $\pi(a)(\pi(b)+\pi(c)) = \pi(a)\pi(b) + \pi(a)\pi(c)$. Indeed, we have

$$\pi(a)(\pi(b) + \pi(c)) = \pi(a)\pi(b + c) = \pi(a(b + c)) = \pi(ab + ac) = \pi(ab) + \pi(ac) = \pi(a)\pi(b) + \pi(a)\pi(c)$$

as claimed. The other ring axioms follow in a similar way. The identities $\pi(a + b) = \pi(a) + \pi(b)$ and $\pi(ab) = \pi(a)\pi(b)$ and $\pi(1_A) = 1_{A/I}$ show that π is a ring homomorphism.

If $a \in \ker(\pi)$ then $\pi(a) = 0_{A/I}$, or in other words $a + I = I$, so in particular $a \in I$. Conversely, if $a \in I$ then any other element $b \in I$ can be expressed as $b = a + (b - a)$ with $b - a \in I$, so $\pi(a) = I = 0_{A/I}$. This proves that $\ker(\pi) = I$. □

Example 5.23. [eg-modular-quotient]
Let n be a positive integer. The set $n\mathbb{Z}$ is then an ideal in $\mathbb{Z}$, and we have $a + n\mathbb{Z} = b + n\mathbb{Z}$ if and only if $a = b \pmod{n}$. Using this, we can identify the quotient $\mathbb{Z}/n\mathbb{Z}$ with the ring $\mathbb{Z}/n$ as defined in Example 1.8.

In keeping with the above example, we adopt the following convention:

Definition 5.24. [defn-principal-quotient]
Given a principal ideal $I = Aa$, we write A/a for A/I.

Proposition 5.25. [prop-monic-quotient]
Let A be a ring, and let f be a monic polynomial of degree d over A. Put $\overline{x} = x + A[x]f \in A[x]/f$. Then every element $u \in A[x]/f$ can be expressed in a unique way as

$$u = \sum_{i=0}^{d-1} a_i \overline{x}^i$$

with $a_0, \dots, a_{d-1} \in A$.

Proof. By definition, u is a coset $\pi(g)$ say. Proposition 1.25 shows that $g = qf + r$ for some polynomial $r(x) = \sum_{i=0}^{d-1} a_i x^i$, with $a_i \in A$. This gives $u = \pi(qf + r) = \pi(r) = \sum_{i=0}^{d-1} a_i \overline{x}^i$. Uniqueness can be proved similarly. □

Example 5.26. [eg-C-as-quotient]
Consider the quotient $K = \mathbb{R}[x]/(x^2 + 1)$. Every element can be expressed uniquely as $a + b\overline{x}$, with $a, b \in \mathbb{R}$. We also have $\overline{x}^2 + 1 = \pi(x^2 + 1) = 0$. Using this, we can identify K with $\mathbb{C}$.

Proposition 5.27. [prop-quotient-ring-map]
Let $\phi \colon A \to B$ be a ring homomorphism, and let I be an ideal in A such that $\phi(I) = 0$, or equivalently $I \subseteq \ker(\phi)$. Then there is a unique homomorphism $\overline{\phi} \colon A/I \to B$ such that $\overline{\phi} \circ \pi = \phi$. Moreover:

(a) *$\overline{\phi}$ is injective iff $\ker(\phi) = I$.*
(b) *$\overline{\phi}$ is surjective iff ϕ is surjective.*
(c) *$\overline{\phi}$ is an isomorphism iff $\ker(\phi) = I$ and ϕ is surjective.*

Proof. Consider a coset $u \subseteq A$. We claim that the set $\phi(u) = \{\phi(x) \mid x \in u\}$ consists of a single element. Indeed, we can write $u = \pi(a) = \{a + t \mid t \in I\}$ for some a, and then we find that

$$\phi(u) = \{\phi(a) + \phi(t) \mid t \in I\} = \phi(a) + \phi(I) = \phi(a) + \{0\} = \{\phi(a)\}$$

as claimed. We define $\overline{\phi}(u)$ to be the unique element of $\phi(u)$. The above calculation shows that $\overline{\phi}(\pi(a)) = \phi(a)$ for all a, so $\overline{\phi} \circ \pi = \phi$, and it is clear that $\overline{\phi}$ is the only function with this property. In particular, we have $\overline{\phi}(1_{A/I}) = \overline{\phi}(\pi(1_A)) = \phi(1_A) = 1_B$. We next claim that $\overline{\phi}(u + v) = \overline{\phi}(u) + \overline{\phi}(v)$ for all $u, v \in A/I$. Indeed, we can choose elements a, b with $u = \pi(a)$ and $v = \pi(b)$, and we find that

$$\overline{\phi}(u + v) = \overline{\phi}(\pi(a) + \pi(b)) = \overline{\phi}(\pi(a + b)) = \phi(a + b) = \phi(a) + \phi(b) = \overline{\phi}(u) + \overline{\phi}(v)$$

as claimed. The same argument gives $\overline{\phi}(uv) = \overline{\phi}(u)\overline{\phi}(v)$, so $\overline{\phi}$ is a ring homomorphism.

(a) Suppose that $\ker(\phi) = I$. We then have $\overline{\phi}(\pi(a)) = 0$ iff $\phi(a) = 0$ iff $a \in I$ iff $\pi(a) = 0$, so $\ker(\overline{\phi}) = 0$, so $\overline{\phi}$ is injective. The converse is similar and is left to the reader.
(b) Suppose that $\overline{\phi}$ is surjective. Then for each $b \in B$ we can choose $u \in A/I$ with $\overline{\phi}(u) = b$, then we can choose $a \in A$ with $u = \pi(a)$, and we find that $\phi(a) = b$. This proves that ϕ is surjective. The converse is similar and is left to the reader.
(c) This follows from (a) and (b).

□

Definition 5.28. [defn-ideal-props]
Let I be an ideal in a ring A.

(a) We say that I is *maximal* iff A/I is a field.
(b) We say that I is *prime* iff A/I is a domain.
(c) We say that I is *coirreducible* if A/I is irreducible.

(d) We say that I is *primary* iff A/I is a predomain.
(e) We say that I is a *radical ideal* if A/I is reduced.

Remark 5.29. [rem-ideal-props]
It follows from Remark 4.2 that maximal ideals are prime, and that prime ideals are are coirreducible and primary and radical. In the noetherian setting (to be discussed later) we will see that all coirreducible ideals are primary.

Example 5.30. [prop-Z-ideal-props]
Consider an ideal $I = \mathbb{Z}.n$ in $\mathbb{Z}$. If $n = 0$ then $\mathbb{Z}/I = \mathbb{Z}$, which is a domain but not a field; so I is prime but not maximal. If $n = 1$ then $\mathbb{Z}/I$ is the trivial ring, so I is radical but not maximal, prime, coirreducible or primary. Suppose instead that $n > 1$, so $\mathbb{Z}/I = \mathbb{Z}/n$. Using Proposition 3.4, we see that I is maximal iff I is prime iff n is a prime number. We also see that I is radical iff there is no prime p such that p^2 divides n.

Example 5.31. [eg-poly-max]
Let F be a field, and put $B = F[x_1, \dots, x_d]$. For any $u \in F^d$ we have a an evaluation homomorphism $\epsilon_u \colon B \to F$, and we saw in Proposition 5.7 that $\ker(\epsilon_u)$ is the same as the ideal

$$L_u = \operatorname{span}_B(x_1 - u_1, \cdots, x_d - u_d).$$

It is clear that ϵ_u is surjective (because it is just the identity on the subring $F \subseteq B$) so it induces an isomorphism $B/L_u \to F$. This shows that L_u is a maximal ideal. We will see in Section 23 that every maximal ideal in $\mathbb{C}[x_1, \dots, x_d]$ has the form L_u for some $u \in \mathbb{C}^d$. On the other hand, Example 5.26 shows that $\mathbb{R}[x].(x^2 + 1)$ is a maximal ideal in $\mathbb{R}[x]$, and it does not have the form L_u.

Proposition 5.32. [prop-local-max-ii]
Let A be a local ring, and let M be the set of non-invertible elements. Then M is the unique maximal ideal in A.

Proof. We saw in Proposition 5.20 that M is an ideal. It clearly does not contain 1, so A/M is a nontrivial ring. If u is a nontrivial element of A/M then it has the form $u = a + M$ for some $a \in A \setminus M$. This means that a is invertible in A, and $a^{-1} + M$ is an inverse for u, so u is invertible in A/M. This proves that A/M is a field, so M is a maximal ideal.

Now let N be any maximal ideal in A. Then A/N is a field, so it must be nontrivial, so N cannot be all of A, so N cannot contain any invertible element of A, so $N \subseteq M$. Conversely, suppose that $a \notin N$. As A/N is a field, there is another element $b \in A$ with $ab \in 1 + N$. Now $N \subseteq M$ so $1 + N \subseteq 1 + M$, which is disjoint from M and so consists of invertible elements. This means that ab is invertible, so a is invertible, so $a \notin M$. We conclude that $N = M$ as required. □

Proposition 5.33. [prop-max-ideal]
An ideal $I \subseteq A$ is maximal iff $I \neq A$, and the only ideal J with $I < J$ is $J = A$.

Proof. First suppose that I is maximal, so A/I is a field. Then A/I is by definition nontrivial, so $I \neq A$. Consider another ideal J with $I < J$, so we can choose $a \in J \setminus I$. As $a \notin I$ the element $\pi(a)$ is nonzero in the field A/I, so it has an inverse. This means that there exists an element $b \in A$ with $ab + I = 1 + I$, so $1 = ab + c$ for some $c \in I \subseteq J$. As $a, c \in J$ we deduce that $1 \in J$, so $J = A$ as required.

Conversely, suppose that the only ideal J with $I < J$ is $J = A$. Any nontrivial element of A/I has the form $\pi(a) = a + I$ for some $a \notin I$. This means that the ideal $J = Ra + I$ is strictly larger than I and so must be all of A. In particular we see that $1 = ab + c$ for some $b \in A$ and $c \in I$. We can now apply π to see that $\pi(a)$ has an inverse $\pi(b)$, as required. □

Proposition 5.34. [prop-prime-complement]
An ideal $I \subseteq A$ is prime iff the complement $A \setminus I$ contains 1 and is closed under multiplication.

Proof. Suppose that I is prime, so A/I is a domain. In particular, A/I is nontrivial, so $1 \in A \setminus I$. If $a, b \in A \setminus I$ then $\pi(a)$ and $\pi(b)$ are nontrivial elements of the domain A/I, so $\pi(ab) = \pi(a)\pi(b)$ is also nontrivial, as $ab \in A \setminus I$ as required. The converse is essentially the same. □

Corollary 5.35. [cor-prime-for-ideals]

Suppose that I, J and P are ideals and that P is prime and $IJ \leq P$ or $I \cap J \leq P$; then either $I \leq P$ or $J \leq P$.

Proof. If $I \not\leq P$ and $J \not\leq P$ then we can choose $a \in I \setminus P$ and $b \in J \setminus P$. We then have $ab \in IJ \leq I \cap J$ and $ab \notin P$, so $IJ \not\leq P$ and $I \cap J \not\leq P$. The claim follows by taking the contrapositive. □

The next result is called *prime avoidance.*

Proposition 5.36. [prop-prime-avoidance]

Let $I, P_1, \dots, P_n$ be ideals in A such that $I \subseteq \bigcup_i P_i$ and P_i is prime for $i > 2$. Then $I \leq P_i$ for some i.

Proof. We argue by induction on n. The case $n = 0$ is vacuous and the case $n = 1$ is trivial. Now suppose that $n \geq 2$. For $1 \leq k \leq n$ put $S_k = \bigcup_{i \neq k} P_i$. We claim that $I \subseteq S_k$ for some k (so that the desired conclusion follows immediately from the induction hypoothesis). If not, we can choose $a_k \in I \setminus S_k$ for all k. As $I \subseteq \bigcup_i P_i$ we see that $a_k \in P_k$. Now put $b = \prod_{i<n} a_i \in I$ and $c = b + a_n$. As $a_i \notin P_n$ for $i < n$ we see that $b \notin P_n$. (If $n = 2$ then $b = a_1$ so this is trivial, and if $n > 2$ then if follows from the primality of P_n.) As $b \notin P_n$ and $a_n \in P_n$ we have $c \notin P_n$. As $c \in I$ we therefore have $c \in P_j$ for some $j < n$. As $a_j \in P_j$ we also have $b \in P_j$ and so $a_n = c - b \in P_j$, contrary to our choice of a_n. This contradiction shows that we must have $I \subseteq S_k$ for some k after all. □

Proposition 5.37. *If A is a predomain, then $A/\operatorname{Nil}(A)$ is a domain.*

Proof. First, as A is a predomain it must be nontrivial, so $1 \neq 0$, so $1 \notin \operatorname{Nil}(A)$, so $A/\operatorname{Nil}(A)$ is nontrivial.

Now let u and v be nontrivial elements of $A/\operatorname{Nil}(A)$. This means that $u = \pi(a)$ and $v = \pi(b)$ for some elements $a, b \in A$ that are not nilpotent. As A is a predomain, it follows that a and b are regular, and thus ab is also regular. In a nontrivial ring a regular element cannot be nilpotent, so $ab \notin \operatorname{Nil}(A)$, so $uv = \pi(ab) \neq 0$. This proves that $A/\operatorname{Nil}(A)$ is a domain. □

Definition 5.38. [defn-idl-functor]

We write $\operatorname{idl}(A)$ for the set of all ideals in A, and $\operatorname{zar}(A)$ for the subset of prime ideals, and $\max(A)$ for the subset of maximal ideals.

For any ring homomorphism $\phi\colon A \to B$, we define maps $\phi_*\colon \operatorname{idl}(A) \to \operatorname{idl}(B)$ and $\phi^*\colon \operatorname{idl}(B) \to \operatorname{idl}(A)$ by

$$\phi_*(I) = \operatorname{span}_B(\phi(I))$$
$$\phi^*(J) = \{a \in A \mid \phi(a) \in J\} = \ker(A \xrightarrow{\phi} B \xrightarrow{\pi} B/J).$$

Example 5.39. [eg-ideals-in-product]

Consider a product ring $C = A \times B$, and let $\alpha\colon C \to A$ and $\beta\colon C \to B$ be the projections, so we have a map $\phi\colon \operatorname{idl}(C) \to \operatorname{idl}(A) \times \operatorname{idl}(B)$ sending K to $(\alpha_*(K), \beta_*(K))$. On the other hand, we can define $\psi\colon \operatorname{idl}(A) \times \operatorname{idl}(B) \to \operatorname{idl}(C)$ by $\psi(I, J) = I \times J$. It is not hard to see that ϕ and ψ are inverse to each other, so both are bijections.

Remark 5.40. Consider the case where A is a subring of B and ϕ is just the inclusion map. We then have $\phi_*(I) = \operatorname{span}_B(I)$, and we may also write BI for this. We also have $\phi^*(J) = A \cap J$.

Remark 5.41. [rem-idl-triangle]

It is easy to see that for all $I \in \operatorname{idl}(A)$ and $J \in \operatorname{idl}(B)$ we have $\phi_*(I) \subseteq J$ iff $I \subseteq \phi^*(J)$. In particular, we have $I \subseteq \phi^*(\phi_*(I))$ and $\phi_*(\phi^*(J)) \subseteq J$. Later we will see various special cases where $I = \phi^*(\phi_*(I))$ or $\phi_*(\phi^*(J)) = J$, but neither of these holds in general.

Remark 5.42. [rem-idl-functor]

If we have homomorphisms $A \xrightarrow{\phi} B \xrightarrow{\psi} C$, then it is easy to see that

$$(\psi\phi)_* = \psi_*\phi_*\colon \operatorname{idl}(A) \to \operatorname{idl}(C)$$
$$(\psi\phi)^* = \phi^*\psi^*\colon \operatorname{idl}(C) \to \operatorname{idl}(A).$$

In other words, these constructions are functorial.

Proposition 5.43. [prop-zar-functor]

Consider a homomorphism $\phi\colon A \to B$. If J is a prime ideal in B, then $\phi^*(J)$ is a prime ideal in A; so ϕ^* restricts to give a map $\operatorname{zar}(B) \to \operatorname{zar}(A)$.

Proof. Suppose that $x, y \in A \setminus \phi^*(J)$. Then $\phi(x), \phi(y) \in B \setminus J$, but J is prime so the element $\phi(xy) = \phi(x)\phi(y)$ also lies in $B \setminus J$, so $xy \in A \setminus \phi^*(J)$. It follows that $\phi^*(J)$ is prime as claimed.

Alternatively, we can say that ϕ induces an injective homomorphism $A/\phi^*(J) \to B/J$, so $A/\phi^*(J)$ is isomorphic to a subring of the domain B/J, so it is itself a domain. □

Remark 5.44. It is not true in general that $\phi^*(\max(B)) \subseteq \max(A)$ or $\phi_*(\operatorname{zar}(A)) \subseteq \operatorname{zar}(B)$ or $\phi_*(\max(A)) \subseteq \max(B)$. Indeed, none of these things is true when ϕ is the inclusion $\mathbb{Z} \to \mathbb{Q}$.

Proposition 5.45. [prop-quotient-ideals]

Let $\phi\colon A \to B$ be a surjective homomorphism, with kernel K. Then the ideals in B are essentially the same as the ideals in A that contain K. More precisely:

(a) *For all $I \in \operatorname{idl}(A)$, the set $\phi(I) \subseteq B$ is an ideal, and is the same as $\phi_*(I)$.*

(b) *Moreover, we have $\phi^*(\phi_*(I)) = I + K$, so $\phi^*(\phi_*(I)) = I$ if and only if $K \subseteq I$.*

(c) *For all $J \in \operatorname{idl}(B)$ we have $K \subseteq \phi^*(J)$ and $\phi_*(\phi^*(J)) = J$.*

(d) *Thus, we have maps*

$$\operatorname{idl}(B) \xrightarrow{\phi^*} \{I \in \operatorname{idl}(A) \mid K \subseteq I\} \xrightarrow{\phi_*} \operatorname{idl}(B)$$

which are inverse to each other and so are bijections.

(e) *Moreover, if $I \in \operatorname{idl}(A)$ corresponds to $J \in \operatorname{idl}(B)$ under the above bijection, then ϕ induces an isomorphism $A/I \to B/J$. Thus, I is maximal, prime, coirreducible, primary or radical iff J has the same property.*

Proof.

(a) Consider elements $u \in B$ and $v, w \in \phi(I)$. As ϕ is surjective we can choose $r \in A$ with $\phi(r) = u$, and by the definition of $\phi(I)$ we can choose $s, t \in I$ with $\phi(s) = v$ and $\phi(t) = w$. As I is an ideal we have $0, rs, s + t \in I$. This gives $0 = \phi(0) \in \phi(I)$ and $uv = \phi(rs) \in \phi(I)$ and $v + w = \phi(s + t) \in \phi(I)$. It follows that $\phi(I)$ is an ideal as claimed. Now $\phi_*(I)$ is by definition the ideal spanned by $\phi(I)$, and as $\phi(I)$ is already an ideal it follows that $\phi_*(I) = \phi(I)$.

(b) If $r \in I + K$ then $r = s + t$ for some $s \in I$ and $t \in K$. Now $\phi(t) = 0$, so $\phi(r) = \phi(s) \in \phi(I) = \phi_*(I)$, so $r \in \phi^*(\phi_*(I))$. Conversely, suppose that $r \in \phi^*(\phi(I))$, so $\phi(r) \in \phi(I)$, so $\phi(r) = \phi(s)$ for some element $s \in I$. This means that the element $t = r - s$ has $\phi(t) = 0$ and so $t \in K$. We therefore have $r = s + t \in I + K$ as claimed.

(c) First, for $t \in K$ we have $\phi(t) = 0 \in J$, so $t \in \phi^*(J)$; this shows that $K \subseteq \phi^*(J)$. As in Remark 5.41, we have $\phi_*(\phi^*(J)) \subseteq J$ for trivial reasons. Conversely, suppose that $u \in J$. As ϕ is surjective there exists $r \in R$ with $\phi(r) = u$. As $\phi(r) \in J$, we have $r \in \phi^*(J)$. Thus, the equation $\phi(r) = u$ shows that $u \in \phi_*(\phi^*(J))$ as required.

(d) This follows from (a), (b) and (c).

(e) Consider an ideal $J \subseteq B$ and the corresponding ideal $I = \phi^*(J) \subseteq B$. Let ψ be the composite $A \xrightarrow{\phi} B \xrightarrow{\pi_J} B/J$, and note that this is surjective. We have $\psi(a) = 0$ iff $\phi(a) \in J$ iff $a \in \phi^*(J) = I$, so $\ker(\psi) = I$. Proposition 5.27 therefore gives us an isomorphism $\overline{\psi}\colon A/I \to B/J$.

□

Corollary 5.46. [cor-quotient-ideals]

Let K be an ideal in A. Then there is a bijection

$$\{I \in \operatorname{idl}(A) \mid K \subseteq I\} \to \operatorname{idl}(A/K)$$

given by $I \mapsto I/K$.

Proof. Just apply the proposition to the standard quotient homomorphism $\pi\colon A \to A/K$. □

Definition 5.47. A *multiplicative set* in a ring A is a subset $U \subseteq A$ that contains 1 and is closed under multiplication.

Remark 5.48. If $P \subseteq A$ is a prime ideal, then $A \setminus P$ is a multiplicative set by Proposition 5.34.

Proposition 5.49. [`prop-primes-exist`]

Let A be a ring, let I be an ideal in A, and let U be a multiplicative set with $U \cap I = \emptyset$. Let $\mathcal{E}$ be the set of all ideals $J \subseteq A$ such that $I \subseteq J$ and $J \cap U = \emptyset$. Then:

(a) *$\mathcal{E}$ has a maximal element.*
(b) *Every maximal element in $\mathcal{E}$ is a prime ideal.*
(c) *If $U = \{1\}$ then every maximal element in $\mathcal{E}$ is a maximal ideal.*

Proof.

(a) A *chain* in $\mathcal{E}$ means a subset $\mathcal{C} \subseteq \mathcal{E}$ such that for all $K, L \in \mathcal{E}$ we have either $K \subseteq L$ or $L \subseteq K$. Let $\mathcal{C}$ be a nonempty chain, and put

$$J = \bigcup \mathcal{C} = \{a \in A \mid \text{ there exists } K \in \mathcal{C} \text{ with } a \in K\}.$$

We claim that $J \in \mathcal{E}$. Indeed, as $\mathcal{C} \neq \emptyset$ we can choose an ideal K with $K \in \mathcal{C}$, and certainly $0 \in K$, so $0 \in J$. If $u \in A$ and $v, w \in J$ then we can choose ideals $K, L \in \mathcal{C}$ with $v \in K$ and $w \in L$. We then have $uv \in K \subseteq J$. Moreover, as $\mathcal{C}$ is a chain we have either $K \subseteq L$ (so $v + w \in L \subseteq J$) or $L \subseteq K$ (so $u + v \in K \subseteq J$). Either way we have $v + w \in J$, so we see that J is an ideal. As every ideal in $\mathcal{C} \subseteq \mathcal{E}$ is contained in $A \setminus U$ it is clear that J is also contained in $A \setminus U$, so $J \in \mathcal{C}$ as claimed. Note also that $\mathcal{E}$ is nonempty, because $I \in \mathcal{E}$. We have now verified the conditions of the principle known as Zorn's Lemma, which guarantees that $\mathcal{E}$ has a maximal element.

As Zorn's Lemma is often considered to be somewhat mysterious, we will outline a proof (for this particular application).

For each maximal element $K \in \mathcal{E}$ (if any) we define $\phi(K) = K$. For each element $K \in \mathcal{E}$ that is not maximal, we choose an element $K' \in \mathcal{E}$ that strictly contains K, and we define $\phi(K) = K'$. (Note that in general we need the Axiom of Choice to make all these choices simultaneously, although for particular rings it my be possible to specify a choice explicitly. For example, if we are given a surjective function $f\colon \mathbb{N} \to A$, we can let n be the smallest integer such that $f(n) \notin K$ and $K + Af(n) \in \mathcal{E}$, and then put $\phi(K) = K + Af(n)$.) It will also be convenient to put $\phi(K) = A$ for any subset $K \subseteq A$ that is not an element of $\mathcal{E}$.

Now define K_α recursively for $\alpha \in \mathbb{N}$ by $K_0 = I$ and $K_{\alpha+1} = \phi(K_\alpha)$. It may be that K_α is maximal for some α, in which case $K_\beta = K_\alpha$ for all $\beta \geq \alpha$ and we can take $J = K_\alpha$. If not, we put $K_\omega = \bigcup_{\alpha \in \mathbb{N}} K_\alpha$, and note that this is again an element of $\mathcal{E}$. We then put $K_{\omega+1} = \phi(K_\omega)$, and $K_{\omega+2} = \phi(K_{\omega+1})$ and so on, and then $K_{2\omega} = \bigcup_{\alpha \in \mathbb{N}} K_{\omega+\alpha}$. To organise this, we need some theory of the "numbers" that we are using a subscripts. These are called *ordinals*, and the relevant theory can be found in any text on axiomatic set theory. In particular, it is possible to make inductive definitions and arguments, as one does for the integers. Using this, we can define subsets $K_\alpha \subseteq A$ for all ordinals α, with $K_{\beta+1} = \phi(K_\beta)$ for all β, and $K_\lambda = \bigcup_{\alpha<\lambda} K_\alpha$ whenever λ does not have the form $\beta + 1$ for any β. One can then check using the chain condition that $K_\alpha \in \mathcal{E}$ for all α, and that if K_α is not maximal then all the ideals K_β with $\beta < \alpha$ are distinct. Some further theory of ordinals provides an ordinal α that is so large that this last condition is impossible, so K_α must be maximal, as required. (There are also proofs of Zorn's Lemma that avoid the use of ordinals, but they are less easy to explain.)

(b) Let P be a maximal element in $\mathcal{E}$, and let a and b be elements of $A \setminus P$. As $P + Aa$ is strictly larger than P, and P is maximal in $\mathcal{E}$, we see that $P + Aa \notin \mathcal{E}$, so $(P + Aa) \cap U \neq \emptyset$. Thus, there are elements $p \in P$ and $x \in A$ and $u \in U$ with $p + ax = u$. Similarly, there are elements $q \in P$ and $y \in A$ and $v \in U$ with $q + by = v$. This gives $uv = (pq + pby + qax) + abxy \in P + Aab$, so $P + Aab$ meets U and thus cannot be equal to P, so $ab \in A \setminus P$. Thus, P is prime.

(c) Now consider the case where $U = \{1\}$. Lemma 5.14 tells us that $\mathcal{E}$ is just the set of proper ideals in A. Thus, Proposition 5.33 tells us that the maximal elements in $\mathcal{E}$ are precisely the maximal ideals.

□

Proposition 5.50. [`prop-radical-intersection`]

Let a be an element in a ring A. Then

(a) *a lies in* $\mathrm{Nil}(A)$ *iff a is nilpotent iff a lies in every prime ideal. In other words,* $\mathrm{Nil}(A)$ *is the intersection of all prime ideals.*

(b) *a lies in* $\mathrm{Rad}(A)$ *iff* $1 + aR \subseteq R^\times$ *iff a lies in every maximal ideal. In other words,* $\mathrm{Rad}(A)$ *is the intersection of all maximal ideals.*

Proof. The first claim, that a lies in $\mathrm{Nil}(A)$ iff a is nilpotent, is simply a reminder of the definition. Similarly, the first part of (b) is simply a reminder of the definition of $\mathrm{Rad}(A)$.

Suppose that there is a prime P such that $a \notin P$. Then the complement $A \setminus P$ is a multiplicative set that contains a but not 0, so no power of a can be 0, so a is not nilpotent. Conversely, if a is not nilpotent then the set $U = \{a^n \mid n \geq 0\}$ is multiplicative and disjoint from the ideal $I = 0$, so we can use Proposition 5.49 to see that there is a prime P with $P \cap U = \emptyset$. In particular, $a \notin P$. This proves (a).

Now suppose that there is a maximal ideal M with $a \notin M$. We then see that a corresponds to an element in the field A/M that is nonzero and therefore invertible. It follows that there is an element $b \in A$ such that $1 - ab \in M$. Recall also that M cannot be equal to A and so cannot contain any invertible elements. Thus $1 - ab$ is not invertible, so a is not in $\mathrm{Rad}(A)$.

Conversely, suppose that a is not in $\mathrm{Rad}(A)$, so there exists $b \in A$ such that $1 - ab$ is not invertible, or equivalently the ideal $I = A(1 - ab)$ is not equal to A. It follows by Proposition 5.49 (with $U = \{1\}$) that there is a maximal ideal M that contains $1 - ab$. If we also had $a \in M$ then we would have $1 \in M$ so $M = A$, which is false. We therefore see that M is a maximal ideal not containing a, as claimed. □

6. Basics of algebraic geometry

Let K be a field. We will typically draw pictures corresponding to the case $K = \mathbb{R}$, but some of the theory will be valid for all fields. Some aspects work better if we assume that K is *algebraically closed*, which means that for every nonconstant polynomial $f(t) \in K[t]$ there is a root $\alpha \in K$ with $f(\alpha) = 0$. For example, it is well known that $\mathbb{C}$ is algebraically closed, but $\mathbb{Q}$ and $\mathbb{R}$ are not (consider $f(t) = t^2 + 1$).

Definition 6.1. [`defn-algebraic-set`]

Consider the ring $P_n = K[x_1, \dots, x_n]$. Note that given $f \in P_n$ and $u \in K^n$ we can evaluate f at u to get $f(u) \in K$.

(a) For any ideal $J \leq P_n$, we put

$$V(J) = \{u \in K^n \mid f(u) = 0 \text{ for all } f \in J\}.$$

(b) For any set $X \subseteq K^n$, we put

$$I(X) = \{f \in P_n \mid f(u) = 0 \text{ for all } u \in X\}.$$

(c) We say that a subset $X \subseteq K^n$ is *algebraic* if $X = V(J)$ for some ideal J.

(d) We say that an ideal $J \leq P_n$ is *geometric* if $J = I(X)$ for some subset $X \subseteq K^n$.

Remark 6.2. [`rem-hilbert`]

Suppose we have polynomials $f_1, \dots, f_r$, and we put

$$X = \{u \in K^n \mid f_1(u) = \cdots = f_r(u) = 0\}.$$

It is easy to see that this is the same as $V(J)$, where $J = Af_1 + \cdots + Af_r$. Thus, X is algebraic. For example, the set

$$X = \{(x, y) \in \mathbb{R}^2 \mid y(x^2 + y^2 - 1) = 0\}$$

(from Example 1.19) is algebraic.

Much later (in Theorem 18.10) we will show that any ideal $J \leq P_n$ can be expressed as $J = P_nf_1 + \cdots + P_nf_r$ for some finite list $f_1, \dots, f_r$, so every algebraic set can be described by a finite system of polynomial equations, as above.

Example 6.3. We have $V(P_n) = \emptyset$ and $V(0) = K^n$ and $I(\emptyset) = P_n$, so the sets $\emptyset$ and K^n are algebraic, and the ideal P_n is geometric. However, if $K = \mathbb{F}_p$ then $u^p = u$ for all $u \in K$, so the elements $x_i^p - x_i$ lie in $I(X)$ for any X. Thus, there is no set X with $I(X) = 0$, and the ideal 0 is not geometric. However, we will show later than whenever K is infinite we have $I(K^n) = 0$, so the ideal 0 is geometric.

Example 6.4. [`eg-affine-line`]

Consider the case where $K = \mathbb{C}$ and $n = 1$. Here it is well-known that every nonzero ideal J can be written as $J = Af$, for some polynomial

$$f(x) = \prod_{i=1}^{r} (x - \lambda_i)^{n_i},$$

where we may assume that the roots λ_i are distinct and the exponents n_i are strictly positive. This gives $V(J) = \{\lambda_1, \dots, \lambda_r\}$. Using this, we see that the algebraic subsets of $\mathbb{C}$ are just the finite subsets, together with the subset $\mathbb{C}$ itself. If X is an infinite subset of $\mathbb{C}$, then $I(X) = 0$, but if $X = \{\lambda_1, \dots, \lambda_r\}$ (with all the λ_i distinct) then $I(X) = P_1 . \prod_i (x - \lambda_i)$; this determines the geometric ideals. (Note here that we allow the case $r = 0$ where $X = \emptyset$; the product of no terms is equal to 1, so $I(\emptyset) = P_1$ as before.

Lemma 6.5.

(a) *If $X_1 \subseteq X_2$ then $I(X_1) \geq I(X_2)$*
(b) *If $J_1 \leq J_2$ then $V(J_1) \supseteq V(J_2)$*
(c) $I(\bigcup_i X_i) = \bigcap_i I(X_i)$
(d) $V(\sum_i J_i) = \bigcap_i V(J_i)$
(e) $I(X_1) + I(X_2) \leq I(X_1 \cap X_2)$
(f) $V(I_1 I_2) = V(I_1 \cap I_2) = V(I_1) \cup V(I_2)$.

Proof. Parts (a) to (e) are immediate from the definitions. For part (f), note that $I_1 I_2 \leq I_1 \cap I_2 \leq I_1$, so $V(I_1 I_2) \supseteq V(I_1 \cap I_2) \supseteq V(I_1)$. Similarly $V(I_1 I_2) \supseteq V(I_1 \cap I_2) \supseteq V(I_2)$, so $V(I_1 I_2) \supseteq V(I_1 \cap I_2) \supseteq V(I_1) \cup V(I_2)$. On the other hand, if $u \notin V(I_1) \cup V(I_2)$, then we can choose $f_1 \in I_1$ and $f_2 \in I_2$ with $f_1(u) \neq 0$ and $f_2(u) \neq 0$ in K. As K is a field, this means that $f_1(u) f_2(u) \neq 0$, so $u \notin V(I_1 I_2)$. The claim follows. □

Corollary 6.6.

(a) *The intersection of any family of algebraic sets is again algebraic.*
(b) *The union of any two algebraic sets is again algebraic.*
(c) *The intersection of any family of geometric ideals is again geometric.* □

Proposition 6.7.

(a) *Given a set $X \subseteq K^n$ and an ideal $J \leq P_n$, we have $X \subseteq V(J)$ iff $J \leq I(X)$.*
(b) *For every ideal $J \leq P_n$, we have $J \leq I(V(J))$.*
(c) *For every subset $X \subseteq K^n$, we have $X \subseteq V(I(X))$.*
(d) *We have $J = I(V(J))$ iff J is geometric.*
(e) *We have $X = V(I(X))$ iff X is algebraic.*

Proof.

(a) Both conditions are equivalent to the condition that $f(u) = 0$ for all $u \in X$ and $f \in J$.
(b) Take $X = V(J)$ in part (a). The condition $V(J) \subseteq V(J)$ is certainly true, so the condition $J \leq I(V(J))$ is also true.
(c) Take $J = I(X)$ in part (a). The condition $I(X) \leq I(X)$ is certainly true, so the condition $X \subseteq V(I(X))$ is also true.
(d) If $J = I(V(J))$ then $J = I(\text{something})$ so J is geometric. Conversely, suppose that J is geometric, so $J = I(X)$ for some set X. This gives $V(J) = V(I(X)) \supseteq X$, but the $I(-)$ operator reverses order, so $I(V(J)) \leq I(X)$. We have $I(X) = J$ by assumption, so $I(V(J)) \leq J$. The reverse inequality is given by (b), so $J = I(V(J))$.
(e) If $X = V(I(X))$, then $X = V(\text{something})$, so X is algebraic. Conversely, suppose that X is algebraic, so $X = V(J)$ for some J. This gives $I(X) = I(V(J)) \geq J$, but the $V(-)$ operator reverses order, so $V(I(X)) \subseteq V(J)$. We have $V(J) = X$ by assumption, so $V(I(X)) \subseteq X$. The reverse inequality is given by (c), so $X = V(I(X))$.

□

Remark 6.8. The above argument is obviously very abstract, and not really specific to the present situation. For a more general version, you can look up the theory of *Galois connections*.

Proposition 6.9. [`prop-IKn`]
If K is infinite, then $I(K^n) = 0$.

Proof. First consider the case $n = 1$. If $f \in K[x] \setminus \{0\}$ and $f(u) = 0$, then $\deg(f) > 0$ and we have $f(x) = (x-u)g(x)$ for some nonzero polynomial g with $\deg(g) = \deg(f)-1$. By a straightforward induction, we see that $|V(P_1 f)| \leq \deg(f) < \infty$. In particular, $V_1(f) \neq K$, so $f \notin I(K)$. This completes the proof for $n = 1$.

In general, if f is a nonzero element of P_n, we can write $f = \sum_{i=0}^{d} c_i x_n^i$, where $c_0, \dots, c_d \in P_{n-1}$ with $c_d \neq 0$. By induction, we can find $u \in K^{n-1}$ with $c_d(u) \neq 0$. Now $f(u, x_n)$ is a nonzero polynomial in $K[x_n]$, so by the $n = 1$ case, we can find $v \in K$ with $f(u,v) \neq 0$, as required. □

7. Product splittings

Definition 7.1. we say that ideals $I, J \leq R$ are *comaximal* if $I + J = R$, or equivalently there exists $a \in I$ with $1 - a \in J$.

The following result (or a result closely related to it) is often called the Chinese Remainder Theorem.

Proposition 7.2. *Suppose that $I_1, \dots, I_n$ are ideals in R such that $I_i + I_j = R$ for all $i \neq j$. Then*

$$I_1 I_2 \cdots I_n = I_1 \cap I_2 \cap \cdots \cap I_n,$$

and there is a natural isomorphism

$$R/\bigcap_i I_i \to \prod_{i=1}^{n} R/I_i.$$

Proof. Put $J = \bigcap_i I_i$ and $K = \prod_i I_i$. It is easy to see that $K \leq I_i$ for all i, and thus that $K \leq J$.

Next, we can define a homomorphism $\pi\colon R \to \prod_i R/I_i$ by $\pi(x) = (x + I_1, \dots, x + I_n)$. Note that $\pi(x)$ is zero iff $x + I_i = 0 + I_i$ for all i iff $x \in I_i$ for all i iff $x \in J$. In other words, we have $\ker(\pi) = J$, so there is an induced homomorphism $\overline{\pi}\colon R/J \to \prod_i R/I_i$ given by $\overline{\pi}(x + J) = \pi(x)$, and this is injective.

Next, for $i \neq j$ we have $I_i + I_j = R$, so we can choose $a_{ij} \in I_i$ with $1 - a_{ij} \in I_j$. Put

$$b_i = \prod_{j \neq i}(1 - a_{ij}) \in \prod_{j \neq i} I_j \leq \bigcap_{j \neq i} I_j.$$

As all the elements a_{ij} lie in I_i we see that $1 - a_{ij} = 1 \pmod{I_i}$ and so $b_i = 1 \pmod{I_i}$.

Suppose we have an element $y = (y_1 + I_1, \dots, y_n + I_n) \in \prod_i R/I_i$. Put $x = \sum_j y_j b_j \in R$. If we fix i then for $j \neq i$ we have $b_j \in I_i$, so $y_j b_j$ does not contribute to $x + I_i$; it follows that $x + I_i = y_i b_i + I_i$. On the other hand, we have $b_i = 1 \pmod{I_i}$ so $y_i b_i + I_i = y_i + I_i$. It follows that $\overline{\pi}(x + J) = \pi(x) = y$; so $\overline{\pi}$ is surjective, and thus an isomorphism.

Now put $c = \prod_i (1 - b_i)$. As $1 - b_i \in I_i$ we have $c \in K$. Now suppose that $x \in J$. For each i we have $x \in I_i$ and $b_i \in \prod_{j \neq i} I_j$ so $x b_i \in K$ so $x = x(1 - b_i) \pmod{K}$. As this holds for all i, we see that $xc = x \pmod{K}$. However, $c \in K$ so $xc = 0 \pmod{K}$ so $x \in K$. This proves that $J = K$. □

The following result is often useful when checking the hypotheses of the Chinese Remainder Theorem.

Proposition 7.3. [`prop-comaximal-powers`]
If I and J are comaximal, then I^n and J^m are also comaximal, for any natural numbers n and m.

Proof. By hypothesis, we can choose $a \in I$ and $b \in J$ such that $a + b = 1$. Now consider the quotient ring $B = A/(I^n + J^m)$, and let $\overline{a}$ and $\overline{b}$ be the images of a and b in B. It is clear that $\overline{a}^n = \overline{b}^m = 0$ in B, so $\overline{a}$ and $\overline{b}$ are nilpotent. By Proposition 3.8, it follows that $(\overline{a} + \overline{b})^{n+m-1} = 0$, but $\overline{a} + \overline{b} = 1$, so $1 = 0$ in B, so B is the trivial ring, so $I^n + J^m = A$ as required. □

8. Rings of fractions

Let A be a ring, and let $U \subseteq A$ be a multiplicative set. We will define (in several stages) a new ring $A[U^{-1}]$, whose elements can be regarded as fractions a/u with $a \in A$ and $u \in U$.

Definition 8.1. [defn-fraction-ops]
We define addition and multiplication operations on the set $A \times U$ by the rules

$$(a,u) + (b,v) = (av + bu, uv)$$
$$(a,u)(b,v) = (ab, uv).$$

We also define $\eta_0 \colon A \to A \times U$ by $\eta_0(a) = (a,1)$, and we write $(a,u) \sim (b,v)$ iff there exists $x \in U$ with $avx = bux$.

Lemma 8.2. [lem-fraction-ops]
The above addition rule is commutative and associative, with $\eta_0(0)$ as an identity element, and $(a,u) + (-a,u) = (0,u^2)$. Similarly, the multiplication rule is commutative and associative, with $\eta_0(1)$ as an identity element. The map η_0 respects addition and multiplication. We also have

$$(a,u).((b,v) + (c,w)) = (abw + acv, uvw)$$
$$(a,u).(b,v) + (a,u).(c,w) = (u(abw + acv), u^2vw).$$

Proof. Straightforward expansion of the definitions. □

Lemma 8.3. [lem-fraction-equiv]
The relation $\sim$ is an equivalence relation. Moreover, if we have elements $p, p', q, q' \in A \times U$ with $p \sim p'$ and $q \sim q'$ then we also have $p + q \sim p' + q'$ and $pq \sim p'q'$.

Proof. The definition of our relation is visibly symmetric, and we can take $x = 1$ to see that $(a,u) \sim (a,u)$, so the relation is also reflexive. Now suppose that $(a,u) \sim (b,v)$ and $(b,v) \sim (c,w)$. We can then choose $x, y \in U$ such that $avx = bux$ and $bwy = cvy$. After multiplying the first of these equations by wy and the second by ux we see that $avwxy = buwxy = cuvxy$, so the element $t = vxy \in U$ satisfies $awt = cut$, so $(a,u) \sim (c,w)$. Thus, the relation is also transitive, and so is an equivalence relation.

Now suppose we have elements $p = (a,u)$ and $p' = (a',u')$ and $q = (b,v)$ and $q' = (b',v')$ such that $p \sim p'$ and $q \sim q'$. This means that there is an element $x \in U$ with $au'x = a'ux$, and an element $y \in U$ with $bv'y = b'vy$. Note that $pq = (ab, uv)$ and $p'q' = (a'b', u'v')$. After multiplying the equation $au'x = a'ux$ by $bv'y$ and multiplying the equation $bv'y = b'vy$ by $a'ux$ we see that

$$abu'v'xy = a'buv'xy = a'b'uvxy,$$

which shows that $pq \sim p'q'$. Similarly, we have $p + q = (av + bu, uv)$ and $p' + q' = (a'v' + b'u', u'v')$. If we add $vv'y$ times the equation $au'x = a'ux$ to $uu'x$ times the equation $bv'y = bvy'$ we get

$$(av + bu)u'v'xy = (a'v' + b'u)uvxy,$$

which proves that $p + q \sim p' + q'$ □

Definition 8.4. We write $A[U^{-1}]$ for the quotient set $(A \times U)/\sim$, and a/u for the equivalence class of the pair (a,u). We also define $\eta(a) = a/1$, which gives a map $\eta \colon A \to A[U^{-1}]$.

Proposition 8.5. *The operations in Definition 8.1 induce well-defined operations on $A[U^{-1}]$, which make $A[U^{-1}]$ into a ring. The map $\eta \colon A \to A[U^{-1}]$ is a ring homomorphism. For any element $u \in U$, the corresponding element $\eta(u) = u/1 \in A[U^{-1}]$ is invertible, with inverse $1/u$.*

Proof. Lemma 8.3 shows that we have well-defined addition and multiplication operations on $A[U^{-1}]$ with $a/u + b/v = (av + bu)/(uv)$ and $(a/u)(b/v) = (ab)/(uv)$. As the operations on $A \times U$ are commutative, associative and unital, the same is true of the induced operations on $A[U^{-1}]$. In Lemma 8.2, it is clear that

$$(abw + acv, uvw) \sim (u(abw + acv), u^2vw);$$

using this, we deduce that $\frac{a}{u}(\frac{b}{v} + \frac{c}{w}) = \frac{ab}{uv} + \frac{ac}{vw}$ in $A[U^{-1}]$. Thus, we have a ring structure on $A[U^{-1}]$. The rest is clear. □

Remark 8.6. [rem-fraction-field]

If A is an integral domain, then the set $U = A \setminus \{0\}$ is multiplicative, so we can form the ring $K = A[U^{-1}]$. This is easily seen to be a field, called the *field of fractions* of A. For example:

- The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$
- The field of fractions of $\mathbb{Z}_{(2)}$ is also $\mathbb{Q}$
- If A is the ring of holomorphic functions on $\mathbb{C}$, then the field of fractions of A is the ring of meromorphic functions (but some nontrivial complex analysis is needed to prove this).
- Any field is its own field of fractions.

Proposition 8.7. [prop-trivial-fractions]

The kernel of the homomorphism $\eta \colon A \to A[U^{-1}]$ is

$$\{a \in A \mid \text{ there exists } u \in U \text{ with } au = 0\} = \bigcup_{u \in U} \operatorname{ann}_A(u).$$

In particular, we have $A[U^{-1}] = 0$ if and only if $1 \in \ker(\eta)$ if and only if $0 \in U$.

Proof. It is clear that $\eta(a) = 0$ if and only if $\eta_0(a) \sim \eta_0(0)$, or in other words $(a, 1) \sim (0, 1)$. From the definition of the equivalence relation, this happens if and only if $au = 0$ for some u, as claimed.

Next, it is true in general that a ring B is trivial if and only if $0 = 1$. Taking $B = A[U^{-1}]$, we see that B is trivial if and only if $1 \in \ker(\eta)$, and by the previous paragraph that can only happen if $0 \in U$. □

Corollary 8.8. [cor-domain-fractions]

If A is a domain and $0 \notin U$ then $A[U^{-1}]$ is also a domain and the map $\eta \colon A \to A[U^{-1}]$ is injective.

Proof. Clear. □

Proposition 8.9. [prop-fractions-universal]

Let $\phi \colon A \to B$ be a ring homomorphism, and let U be a multiplicative subset of A such that $\phi(U) \subseteq B^\times$. Then there is a unique homomorphism $\overline{\phi} \colon A[U^{-1}] \to B$ such that $\overline{\phi} \circ \eta = \phi$.

Proof. We can define $\overline{\phi}_0 \colon A \times U \to B$ by $\overline{\phi}_0(a, u) = \phi(a)\phi(u)^{-1}$. It is straightforward to check that this respects addition and multiplication (as defined in Definition 8.1), and that it satisfies $\overline{\phi}_0 \eta_0 = \phi$.

Now suppose we have pairs (a, u) and (b, v) in $A \times U$ with $(a, u) \sim (b, v)$, so there is an element $x \in U$ with $avx = bux$. Applying ϕ gives $\phi(a)\phi(v)\phi(x) = \phi(b)\phi(u)\phi(x)$, but $\phi(u)$, $\phi(v)$ and $\phi(x)$ are invertible by assumption, so we can multiply by $\phi(u)^{-1}\phi(v)^{-1}\phi(x)^{-1}$ to get $\phi(a)\phi(u)^{-1} = \phi(b)\phi(v)^{-1}$, or in other words $\overline{\phi}_0(a, u) = \overline{\phi}_0(b, v)$. We therefore have a well-defined map $\overline{\phi} \colon A[U^{-1}] \to B$ given by $\overline{\phi}(a/u) = \overline{\phi}_0(a, u) = \phi(a)\phi(u)^{-1}$. It is now easy to see that this is a homomorphism with $\overline{\phi} \circ \eta = \phi$.

On the other hand, if $\psi \colon A[U^{-1}] \to B$ is any homomorphism with $\psi\eta = \phi$, we can apply ψ to the identity $(a/u)\eta(u) = \eta(a)$ to get $\psi(a/u)\phi(u) = \phi(a)$, so $\psi(a/u) = \phi(a)\phi(u)^{-1} = \overline{\phi}(a/u)$. Thus, $\overline{\phi}$ is the unique homomorphism with the stated properties. □

Definition 8.10. [defn-P-loc]

Let A be a ring, and let P be a prime ideal in A, so $A \setminus P$ is a multiplicative set. We write A_P for $A[(A \setminus P)^{-1}]$, and call this the *localisation* of A at P.

Proposition 8.11. [prop-loc-local]

If P is a prime ideal in a ring A, then the localisation A_P is a local ring, with maximal ideal

$$M = P_P = \{a/u \mid a \in P, u \notin P\}.$$

Proof. Write $U = A \setminus P$, so $A_P = A[U^{-1}]$. We have $0 \in P$ so $0 \notin U$ so $A_P \neq 0$. Consider an element $x \in A_P$, so $x = a/u$ for some $a \in A$ and $u \notin P$. Note that $1 - x = (u - a)/u$. As the element $u = a + (u - a)$ is not in P, at least one of the elements a and $u - a$ must be outside P. If a is outside P then x is invertible with inverse u/a, and if $u - a$ is outside P then $1 - x$ is invertible with inverse $u/(u - a)$. Thus A_P is local as claimed. In any local ring the unique maximal ideal is the set of elements that are not invertible, which is easily seen to be the set M described above. □

Now consider a ring A and a multiplicative set $U \subseteq A$. We next discuss the relationship between ideals in A and ideals in $A[U^{-1}]$.

Lemma 8.12. [lem-eta-star]
If I is an ideal in A then

$$\eta_*(I) = \{a/u \mid a \in I,\ u \in U\} \subseteq A[U^{-1}].$$

Proof. Put $I' = \{a/u \mid a \in I,\ u \in U\}$. It is straightforward to check that this is an ideal in $A[U^{-1}]$ containing $\eta(I)$. On the other hand, if I'' is any other ideal that contains $\eta(a) = a/1$ for all $a \in I$, it must also contain the elements $(a/1).(1/u) = a/u$, so $I' \subseteq I''$. Thus, I' is the smallest ideal in $A[U^{-1}]$ containing $\eta(I)$, so it must be equal to $\eta_*(I)$. □

Definition 8.13. [defn-saturated]
For any ideal $I \in \mathrm{idl}(A)$ we put

$$I^\# = \{a \in A \mid \text{ there exists } u \in U \text{ with } au \in I\}.$$

It is clear that $I \subseteq I^\#$, and we say that I is *U-saturated* if $I^\# = I$. We write $\mathrm{sat}_U(A)$ for the set of all U-saturated ideals.

Remark 8.14. [rem-saturated]
For any ideal I we have $I^{\#\#} = I^\#$. Indeed, if $a \in I^{\#\#}$ then $au \in I^\#$ for some $u \in U$, so $auv \in I$ for some $v \in U$, but $uv \in U$ so $a \in I^\#$. Thus, $I^\#$ is always U-saturated.

Proposition 8.15. [prop-saturated]
There is a natural bijection between ideals in $A[U^{-1}]$ and U-saturated ideals in A. In more detail:

(a) *For any ideal $J \subseteq A[U^{-1}]$, the ideal $\eta^*(J) \subseteq A$ is U-saturated. Thus, η^* gives a map $\mathrm{idl}(A[U^{-1}]) \to \mathrm{sat}_U(A)$.*

(b) *Moreover, we have $\eta_*(\eta^*(J)) = J$, so the composite*

$$\mathrm{idl}(A[U^{-1}]) \xrightarrow{\eta^*} \mathrm{sat}_U(A) \xrightarrow{\eta_*} \mathrm{idl}(A[U^{-1}])$$

is the identity.

(c) *For any ideal $I \subseteq A$ we have $\eta^*(\eta_*(I)) = I^\#$. In particular if I is U-saturated then $\eta^*(\eta_*(I)) = I$. Thus, the composite*

$$\mathrm{sat}_U(A) \xrightarrow{\eta_*} \mathrm{idl}(A[U^{-1}]) \xrightarrow{\eta^*} \mathrm{sat}_U(A)$$

is the identity.

Proof.

(a) Suppose that $a \in (\eta^*(J))^\#$, so for some $u \in U$ we have $ua \in \eta^*(J)$, which means that the element $\eta(u)\eta(a) = \eta(ua)$ lies in J. As $\eta(u)$ is invertible we can multiply by the inverse to see that $\eta(a) \in J$, or equivalently $a \in \eta^*(J)$.

(b) Any element $x \in J$ can be written as $x = a/u$ for some $a \in A$ and $u \in U$. It follows that the element $\eta(a) = x\eta(u)$ also lies in J, so $a \in \eta^*(J)$, so $\eta(a) \in \eta_*(\eta^*(J))$, so the element $x = \eta(u)^{-1}\eta(a)$ also lies in $\eta_*(\eta^*(J))$. This proves that $J \subseteq \eta_*(\eta^*(J))$, and we mentioned in Remark 5.41 that the reverse inclusion is automatic.

(c) If $a \in I^\#$ then we can choose $u \in U$ such that $au \in I$, so $\eta(au) \in \eta_*(I)$, so the element $\eta(a) = \eta(au).\eta(u)^{-1}$ also lies in $\eta_*(I)$, so $a \in \eta^*(\eta_*(I))$.

Conversely, if $a \in \eta^*(\eta_*(I))$ then $\eta(a) \in \eta_*(I)$, so $a/1 = b/v$ for some $b \in I$ and $v \in U$. This means that $au = bvu$ for some $u \in U$, but $bvu \in I$, so $a \in I^\#$.

□

Now suppose we have a ring A, a multiplicative set $U \subseteq A$ and an ideal $I \subseteq A$. We have various ways to construct new rings from these. We can form the quotient ring A/I, which has a multiplicative set $\pi(U)$. We can invert this to get a ring $(A/I)[\pi(U)^{-1}]$ (which we will often denote more briefly by $(A/I)[U^{-1}]$). Alternatively, we can form $A[U^{-1}]$ and then the quotient $A[U^{-1}]/\eta_*(I)$ (which we will often denote more briefly by $A[U^{-1}]/I$). It turns out that $A[U^{-1}]/I$ is the same as $(A/I)[U^{-1}]$. A more careful statement is as follows:

Proposition 8.16. [prop-fraction-quotient]
There is a unique natural isomorphism ϕ making the diagram below commute:

$$\begin{tikzcd}
A[U^{-1}] \arrow[d, "\pi"'] & A \arrow[l, "\eta"'] \arrow[r, "\pi"] & A/I \arrow[d, "\eta"] \\
A[U^{-1}]/\eta_*(I) \arrow[rr, "\simeq", "\phi"'] & & (A/I)[\pi(U)^{-1}].
\end{tikzcd}$$

Proof. The basic definition is just that $\phi(a/u + \eta_*(I)) = (a + I)/(u + I)$. To check that this is well-defined we note that $\pi\eta$ sends elements of U to invertible elements of $(A/I)[\pi(U)^{-1}]$, so Proposition 8.9 gives a unique homomorphism $\theta\colon A[U^{-1}] \to (A/I)[\pi(U)^{-1}]$ with $\theta(a/u) = \eta\pi(a)\eta\pi(u)^{-1} = (a + I)/(u + I)$. If $x \in \eta_*(I)$ then we can write $x = a/u$ with $a \in I$ and so $\theta(x) = 0$. Thus, Proposition 5.27 gives us a unique homomorphism $\phi\colon A[U^{-1}]/\eta_*(I) \to (A/I)[\pi(U)^{-1}]$ with $\phi(x + \eta_*(I)) = \theta(x)$, or equivalently $\phi(a/u+\eta_*(I)) = (a+I)/(u+I)$ as before. Similarly, we can use Proposition 5.27 followed by Proposition 8.9 to get a well-defined homomorphism $\psi\colon (A/I)[\pi(U)^{-1}] \to A[U^{-1}]/\eta_I(I)$ with $\psi((a+I)/(u+I)) = a/u+\eta_*(I)$, and then it is clear that ψ is an inverse for ϕ. □

Example 8.17. [eg-residue-field]
Let P be a prime ideal in A, and take $I = P$ and $U = A\setminus P$. We obtain an isomorphism $A_P/P_P = (A/P)_P$. This ring is just the field of fractions of the integral domain A/P. We call it the *residue field* of P and use the notation $K(P)$.

Proposition 8.18. [prop-sat-prime]
Let P be a prime ideal in A.

(a) *If $P \cap U \neq \emptyset$ then $\eta_*(P) = A[U^{-1}]$ and $P^\# = A$. In particular, neither $\eta_*(P)$ nor $P^\#$ is prime.*
(b) *Suppose instead that $P\cap U = \emptyset$. Then an element $x \in A[U^{-1}]$ lies in $\eta_*(P)$ iff for some representation $x = a/u$ we have $a \in P$, iff for every representation $x = a/v$ we have $a \in P$. Moreover, $\eta_*(P)$ is a prime ideal in $A[U^{-1}]$ and $P^\# = P$ (so P is U-saturated).*

Thus, the maps η_* and η^* give a bijection

$$\{ \text{ prime ideals } P \subseteq A \text{ with } P \cap U = \emptyset\} \simeq \{ \text{ prime ideals in } A[U^{-1}]\}.$$

Proof. (a) If $P \cap U \neq \emptyset$ then we can choose $u \in P \cap U$, so the element $1 = u/u$ lies in $\eta_*(P)$, so $\eta_*(P) = A[U^{-1}]$ and $P^\# = \eta^*(\eta_*(P)) = A$.

(b) Suppose instead that $P\cap U = \emptyset$, so U is contained in the set $A\setminus P$, which is closed under multiplication by Proposition 5.34. Consider an element $x = a/u \in A[U^{-1}]$. By Lemma 8.12 we see that $x \in \eta_*(P)$ iff there is a representation $x = a/u$ with $a \in P$. Now suppose we have another representation $x = b/v$, so $avw = buw$ for some $w \in U$. The left hand side lies in P, but on the right hand side u and w are in $A \setminus P$. As $A \setminus P$ is closed under multiplication we must have $b \in P$ as claimed. In particular we have $a/1 \in \eta_*(P)$ iff $a \in P$, so $P^\# = \eta^*(\eta_*(P)) = P$.

We also see from Proposition 8.16 that $A[U^{-1}]/\eta_*(P)$ can be identified with $(A/P)[\pi(U)^{-1}]$. Here A/P is a domain, and $\pi(U)$ does not contain zero, so $(A/P)[\pi(U)^{-1}]$ is a domain, so $A[U^{-1}]/\eta_*(P)$ is a domain, so $\eta_*(P)$ is prime. Alternatively, we can use the previous paragraph to see that $A[U^{-1}] \setminus \eta_*(P)$ contains 1 and is closed under multiplication, which again proves that $\eta_*(P)$ is prime.

Now put

$$\mathcal{P} = \{ \text{ prime ideals } P \subseteq A \text{ with } P \cap U = \emptyset\}$$
$$\mathcal{Q} = \{ \text{ prime ideals in } A[U^{-1}]\} = \operatorname{zar}(A[U^{-1}]).$$

If $Q \in \mathcal{Q}$ then $\eta^*(Q)$ is prime by Proposition 5.43 and saturated by Proposition 8.15, so it lies in $\mathcal{P}$. Thus we have a map $\eta^*\colon \mathcal{Q} \to \mathcal{P}$. On the other hand, points (a) and (b) above show that η_* gives a map $\mathcal{P} \to \mathcal{Q}$. Proposition 8.15 shows that these maps are inverse to each other. □

9. Matrices and determinants

Many things involving vectors, matrices and determinants can be generalised from the traditional context where the entries are real numbers; we can instead allow entries in any ring. Normally we consider matrices M with entries M_{ij} for $0 \leq i < n$ and $0 \leq j < m$ say. However, we will find it convenient to allow the indices i and j to come from an arbitrary finite set. The most basic definitions are as follows.

Definition 9.1. [defn-matrix]
Let A be a ring, and let I and J be finite sets. We write $\text{Free}_I(A)$ for the set of vectors with entries in A indexed by I, so $\text{Free}_I(A) = \prod_{i \in I} A$. We also write $\text{Mat}_{IJ}(A)$ for the set of matrices with entries in A indexed by $I \times J$. For $M \in \text{Mat}_{IJ}(A)$ and $v \in \text{Free}_J(A)$ we define $Mv \in \text{Free}_I(A)$ by $(Mv)_i = \sum_{j \in J} M_{ij} v_j$. Similarly, given $M \in \text{Mat}_{IJ}(A)$ and $N \in \text{Mat}_{JK}(A)$ we define $MN \in \text{Mat}_{IK}(A)$ by $(MN)_{ik} = \sum_{j \in J} M_{ij} N_{jk}$. We write $\text{Mat}_I(A)$ for $\text{Mat}_{II}(A)$, and define $1_I \in \text{Mat}_I(A)$ by

$$(1_I)_{ii'} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise;} \end{cases}$$

this is called the *identity matrix*. Finally, given $M \in \text{Mat}_{IJ}(A)$ we define $M^T \in \text{Mat}_{JI}(A)$ by $(M^T)_{ji} = M_{ij}$, and we call this the *transpose* of A.

Various familiar properties, such as $(MN)P = M(NP)$, $M(u+v) = Mu + Mv$, $(MN)^T = N^T M^T$ and so on, are easily generalised to this new context. We leave details to the reader.

Definition 9.2. [defn-trace]
For a matrix $M \in \text{Mat}_I(A)$ we put $\text{trace}(A) = \sum_{i \in I} M_{ii}$.

Proposition 9.3. [prop-trace]
For $M \in \text{Mat}_{IJ}(A)$ and $N \in \text{Mat}_{JI}(A)$ we have

$$\text{trace}(MN) = \text{trace}(NM) = \sum_{i \in I} \sum_{j \in J} M_{ij} N_{ji}.$$

Proof. Just unwind the definitions. □

Definition 9.4. Let I and J be finite, totally ordered sets with $|I| = |J|$. (The most common case is where $I = J = \{0, \dots, n-1\}$, but it is convenient to allow a little more flexibility.) We write $\Theta(I, J)$ for the set of all maps from I to J, and $\Sigma(I, J)$ for the subset of bijective maps. We also write $\Theta(I) = \Theta(I, I)$ and $\Sigma(I)$ for $\Sigma(I, I)$.

Definition 9.5. [defn-sgn]
Let $P(I)$ be the set of pairs in I, or in other words subsets $p \subseteq I$ with $|p| = 2$. If $\sigma \in \Sigma(I, J)$ and $p = \{i, j\} \in P(I)$ then the set $\sigma_*(p) = \{\sigma(i), \sigma(j)\}$ is an element of $P(J)$. This construction gives a bijection $\sigma_* \colon P(I) \to P(J)$. We also let $L(\sigma)$ denote the set of pairs $p \in P(I)$ for which the map $\sigma \colon p \to \sigma_*(p)$ is order-reversing. Thus, if $p = \{i, j\}$ with $i < j$, then we have $p \in L(\sigma)$ iff $\sigma(i) > \sigma(j)$. We define $\text{sgn}(\sigma) = (-1)^{|L(\sigma)|} \in \{1, -1\}$, and call this the *signature* of σ. We also put $\text{sgn}(\sigma) = 0$ if $\sigma \colon I \to J$ is a map that is not a permutation.

Example 9.6. [eg-transposition]
Suppose that $p, q \in I$ with $p < q$, and let $\tau \colon I \to I$ be the transposition defined by

$$\tau(i) = \begin{cases} q & \text{if } i = p \\ p & \text{if } i = q \\ i & \text{otherwise} \end{cases}$$

Put $J = \{i \mid p < i < q\}$. We find that

$$L(\tau) = \{\{p, j\} \mid j \in J\} \amalg \{\{j, q\} \mid j \in J\} \amalg \{\{p, q\}\},$$

so $|L(\tau)| = 2|J| + 1$ so $\text{sgn}(\tau) = -1$.

Proposition 9.7. [prop-signature]
For all maps $I \xrightarrow{\tau} J \xrightarrow{\sigma} K$ (with $|I| = |J| = |K|$) we have $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$. Moreover, if σ is a bijection then $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$.

Proof. First, it is easy to see that $\sigma\tau$ is a permutation iff both σ and τ are both permutations. We can restrict attention to this case, because in all other cases both $\operatorname{sgn}(\sigma\tau)$ and $\operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$ are zero. Consider a pair $p \in P(I)$. Note that the composite $p \xrightarrow{\tau} \tau_*(p) \xrightarrow{\sigma} (\sigma\tau)_*(p)$ is order-reversing iff precisely one of the maps $p \xrightarrow{\tau} \tau_*(p)$ and $\tau_*(p) \xrightarrow{\sigma} (\sigma\tau)_*(p)$ is order-reversing, so p lies in $L(\sigma\tau)$ iff it lies in precisely one of the two sets $L(\tau)$ and $\tau_*^{-1}L(\sigma)$. It follows that

$$|L(\sigma)| + |L(\tau)| = |\tau_*^{-1}L(\sigma)| + |L(\tau)| \cong |L(\sigma\tau)| \pmod 2,$$

so $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$. It is clear that the signature of any identity map is one, so we can take $\tau = \sigma^{-1}$ to get $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$. □

Definition 9.8. [defn-det]
For any matrix $M \in \operatorname{Mat}_{IJ}(A)$ with $|I| = |J|$ we put

$$\det(M) = \sum_{\sigma\in\Sigma(I,J)} \operatorname{sgn}(\sigma) \prod_{i\in I} M_{i,\sigma(i)} \in A.$$

Proposition 9.9. [prop-det-triangle]
Suppose that $M \in \operatorname{Mat}_I(A)$ and $M_{ij} = 0$ whenever $i < j$. Then $\det(M) = \prod_i M_{ii}$. In particular, we have $\det(1_I) = 1$.

Proof. If $\sigma\colon I \to I$ is not the identity, then let i be the smallest element of i where $\sigma(i) \neq i$. For $j < i$ we have $\sigma(j) = j$ so we cannot have $\sigma(i) = j$; it therefore follows that $i < \sigma(i)$, and so $M_{i,\sigma(i)} = 0$. Thus, the term in $\det(M)$ corresponding to σ is zero. The only remaining term is where σ is the identity, which gives $\prod_i M_{ii}$. □

Proposition 9.10. [prop-det-transpose]
Suppose that $M \in \operatorname{Mat}_{IJ}(A)$ with $|I| = |J|$. Then $\det(A^T) = \det(A)$.

Proof. By unwinding the definitions we have $\det(M^T) = \sum_{\tau\in\Sigma(J,I)} \operatorname{sgn}(\tau) \prod_{j\in J} M_{\tau(j),j}$. We can reindex this in terms of $\sigma = \tau^{-1}$ and $i = \tau(i)$ to get $\det(M^T) = \sum_{\sigma\in\Sigma(I,J)} \operatorname{sgn}(\sigma) \prod_{i\in I} M_{i,\sigma(i)} = \det(M)$. □

Proposition 9.11. [prop-det-repeated]
Suppose that there are indices $p, q \in I$ with $p < q$ such that $M_{pj} = M_{qj}$ for all $j \in J$. Then $\det(M) = 0$.

Proof. Let $\tau \in \Sigma(I)$ be the transposition that exchanges p and q, so $\tau^{-1} = \tau$ and $M_{\tau(i),j} = M_{ij}$ for all i and j. For any $\sigma \in \Sigma(I,J)$ put $m(\sigma) = \operatorname{sgn}(\sigma) \prod_i M_{i,\sigma(i)}$, so $\det(M) = \sum_\sigma m(\sigma)$. In $m(\sigma\tau)$ we can reindex the product in terms of $j = \tau(i)$ to get

$$m(\sigma\tau) = \operatorname{sgn}(\sigma\tau) \prod_i M_{i,\sigma\tau(i)} = -\operatorname{sgn}(\sigma) \prod_j M_{\tau(j),\sigma(j)} = -\operatorname{sgn}(\sigma) \prod_j M_{j,\sigma(j)} = -m(\sigma).$$

Now put $\Sigma_0 = \{\sigma \in \Sigma(I,J) \mid \sigma(p) < \sigma(q)\}$. We find that the maps in $\Sigma \setminus \Sigma_0$ are precisely those of the form $\sigma\tau$ with $\sigma \in \Sigma_0$, so the terms $m(\sigma)$ for $\sigma \in \Sigma_0$ cancel the terms $m(\sigma)$ for $\sigma \notin \Sigma_0$ and we are left with $\det(M) = 0$ as claimed. □

Proposition 9.12. [prop-prod-det]
Suppose we have matrices $M \in \operatorname{Mat}_{IJ}(A)$ and $N \in \operatorname{Mat}_{JK}(A)$ with $|I| = |K| = n$ say. Then if $|J| = n$ we have $\det(MN) = \det(M)\det(N)$, but if $|J| < n$ then $\det(MN) = 0$.

The proof relies on the following observation:

Lemma 9.13. *For any $U \in \operatorname{Mat}_{IJ}(A)$ we have*

$$\prod_{i\in I}\sum_{j\in J} u_{ij} = \sum_{\theta\colon I\to J} \prod_{i\in I} u_{i,\theta(i)}.$$

Proof. This is just a codification of the usual process of expanding a product of sums as a sum of products. □

Proof of Proposition 9.12. From the definitions, we have

$$\det(MN) = \sum_{\sigma\in\Sigma(I,K)} \operatorname{sgn}(\sigma) \prod_{i\in I} \sum_{j\in J} M_{ij} N_{j,\sigma(i)}.$$

Using the lemma, this becomes

$$\det(MN) = \sum_{\sigma\in\Sigma(I,K)} \sum_{\theta\colon I\to J} \operatorname{sgn}(\sigma) \prod_{i\in I} M_{i,\theta(i)} N_{\theta(i),\sigma(i)}. = \sum_{\sigma\in\Sigma(I)} \sum_{\theta\colon I\to J} \operatorname{sgn}(\sigma) \prod_{i\in I} M_{i,\theta(i)} \prod_{i'\in I} N_{\theta(i'),\sigma(i')}.$$

We write Δ for the sum of terms where θ is injective, and Δ' for the sum of all other terms. If $|J| < n$ then clearly $\Delta = 0$. If $|J| = n$ then any injective map $\theta\colon I \to J$ is a bijection, and we can reindex everything in Δ using $\tau = \sigma\theta^{-1} \in \Sigma(J,K)$ and $j = \theta(i')$ (so $\sigma = \tau\theta$ and $i' = \theta^{-1}(j)$). This gives

$$\Delta = \sum_{\theta\in\Sigma(I,J)} \sum_{\tau\in\Sigma(J,K)} \operatorname{sgn}(\tau\theta) \prod_{i\in I} M_{i\,\theta(i)} \prod_{j\in J} N_{j,\tau(j)} = \det(M)\det(N).$$

To complete the proof, it will therefore suffice to show that $\Delta' = 0$.

Consider a function $\theta\colon I \to J$ that is not injective, so $\theta(p) = \theta(q)$ for some $p < q$. Next, for any $\sigma \in \Sigma(I,K)$ we put

$$\Gamma(\theta,\sigma) = \operatorname{sgn}(\sigma) \prod_{i\in I} N_{\theta(i),\sigma(i)}.$$

Let $\tau \in \Sigma(I)$ be the transposition that exchanges p and q, and note that $\theta\tau = \theta$ and $\operatorname{sgn}(\tau) = -1$. Using these facts, we can reindex the above product in terms of $j = \tau(i)$ to get $\Gamma(\theta,\sigma) = -\Gamma(\theta,\sigma\circ\tau)$. Now consider the sum $\Gamma(\theta) = \sum_\sigma \Gamma(\theta,\sigma)$. We can divide the permutations into two groups: those for which $\sigma(p) < \sigma(q)$, and those for which $\sigma(p) > \sigma(q)$. If σ is in the first group then $\sigma\circ\tau$ is in the second group and *vice-versa*. It follows that the terms $\Gamma(\theta,\sigma)$ from the first group cancel the terms $\Gamma(\theta,\sigma)$ from the second group, leaving $\Gamma(\theta) = 0$.

Finally, from our earlier expansion of $\det(MN)$ we have

$$\Delta' = \sum_\theta \left(\prod_{i\in I} M_{i,\theta(i)} \right) \Gamma(\theta),$$

where the sum runs over all functions $\theta\colon I \to J$ that are not permutations. We have seen that $\Gamma(\theta) = 0$, so $\Delta' = 0$ as required. □

Definition 9.14. [`defn-adj`]

Suppose that $M \in \mathrm{Mat}_{IJ}(A)$, where $|I| = |J|$. For $p \in I$ and $q \in J$ we let $\mu_{pq}(M)$ be the evident matrix in $\mathrm{Mat}_{I\setminus\{p\},J\setminus\{q\}}(A)$ obtained by forgetting some of the entries in M. We also define $\rho(p) = |\{i \in I \mid i < p\}|$, and similarly for $\rho(q)$. We define $\operatorname{adj}(M) \in \mathrm{Mat}_{JI}(A)$ by

$$\operatorname{adj}(M)_{qp} = (-1)^{\rho(p)+\rho(q)} \det(\mu_{pq}(M)).$$

Proposition 9.15. [`prop-adjugate`]

$M\operatorname{adj}(M) = \det(M).1_I$ *and* $\operatorname{adj}(M)M = \det(M).1_J$.

Proof. It will be harmless to assume that $I = J = \{0,\dots,n{-}1\}$. Put $N = \operatorname{adj}(M)$, so $N_{ji} = (-1)^{i+j}\det(\mu_{ij}(M))$. Then put $P = MN$, so

$$P_{pp} = \sum_q (-1)^{p+q} M_{pq} \det(\mu_{pq}(M)).$$

For any $p \in I$ and $q \in J$, put $S(p,q) = \Sigma(I\setminus\{p\}, J\setminus\{q\})$. For any $\tau \in S(p,q)$, let τ^+ denote the unique map $I \to J$ extending τ with $\tau^+(p) = q$. If we fix p then this construction gives a bijection $\coprod_q S(p,q) \to \Sigma(I,J)$. We claim that $\operatorname{sgn}(\tau^+) = (-1)^{p+q}\operatorname{sgn}(\tau)$. This is clear when $p = q = 0$, because we then have $L(\tau^+) = L(\tau)$. For the general case, we define $\rho_r \in \Sigma(I)$ by

$$\rho_r(i) = \begin{cases} r & \text{if } i = 0 \\ i-1 & \text{if } 0 < i \le r \\ i & \text{if } i > r. \end{cases}$$

We also write λ_r for the map $I \setminus \{0\} \to I \setminus \{r\}$ obtained by restricting ρ_r. As λ_r is an order-preserving bijection we have $\mathrm{sgn}(\lambda_r) = 1$, but one can also check directly that $\mathrm{sgn}(\rho_r) = (-1)^r$.

If $\tau \in S(p, q)$ we find that the map $\tau_0 = \lambda_q^{-1}\tau\lambda_p$ lies in $S(0, 0)$ and that $\tau^+ = \rho_q\tau_0^+\rho_p^{-1}$. From this it follows that $\mathrm{sgn}(\tau_0) = \mathrm{sgn}(\tau)$ and $\mathrm{sgn}(\tau^+) = (-1)^{p+q}\,\mathrm{sgn}(\tau)$ as claimed.

We can now use the decomposition $\Sigma(I) \simeq \coprod_q S(p, q)$ to get

$$\det(M) = \sum_q \sum_{\tau\in S(p,q)} \mathrm{sgn}(\tau^+) \prod_i M_{i,\tau^+(i)} = (-1)^{p+q} M_{pq} \sum_{q,\tau} \mathrm{sgn}(\tau) \prod_{i\neq p} M_{i,\tau(i)} = P_{pp}.$$

Now consider instead an entry

$$P_{tp} = \sum_q (-1)^{p+q} M_{tq} \det(\mu_{pq}(M)),$$

where $t \neq p$. Define a new matrix M^* by $M^*_{ij} = M_{ij}$ when $i \neq p$, and $M^*_{pj} = M_{tj}$. Put $P^* = M^* \operatorname{adj}(M^*)$. Replacing M by M^* in the previous paragraphs, we get $P^*_{tt} = \det(M^*)$, and this is zero by Proposition 9.11. On the other hand, we have $\mu_{pq}(M^*) = \mu_{pq}(M)$, and using this we see that $P_{tp} = P^*_{tp} = 0$. This completes the proof that $M \operatorname{adj}(M) = P = \det(M).1_I$ as claimed.

We can now replace M by M^T in the above argument to get $M^T \operatorname{adj}(M^T) = \det(M^T).1_J$. We have seen that $\det(M^T) = \det(M)$, and after applying this to the matrices $\mu_{pq}(M)$ we see that $\operatorname{adj}(M^T) = \operatorname{adj}(M^T)$. We can also see from the definitions that $1_J^T = 1_J$ and $(XY)^T = Y^T X^T$. By combining these ingredients, we deduce that $\operatorname{adj}(M)M = \det(M).1_J$. □

Definition 9.16. Consider finite ordered sets I and J, and a matrix $M \in \mathrm{Mat}_{IJ}(A)$. For any $I' \subseteq I$ and $J' \subseteq J$ we form a matrix $M|_{I'\times J'} \in \mathrm{Mat}_{I'J'}(A)$ in the obvious way. If $|I'| = |J'|$ we can then take the determinant to get an element of A. We let $D_k(M)$ denote the ideal generated by all determinants $\det(M|_{I'\times J'})$ with $|I'| = |J'| = k$.

Remark 9.17. The determinant of the empty matrix is taken to be one, so $D_0(M) = A$. It is also clear that $D_1(M)$ is the ideal generated by all the elements M_{ij}. On the other hand, if $|I| = |J| = n$ then $D_n(M) = A.\det(M)$, and if $n > \min(|I|, |J|)$ then $D_n(M) = 0$.

Proposition 9.18. [`prop-inj-mat`]
Consider a matrix $M \in \mathrm{Mat}_{IJ}(A)$ and the corresponding map $\mu\colon \mathrm{Free}_J(A) \to \mathrm{Free}_I(A)$ given by $\mu(u) = Mu$. Then μ is injective iff $\mathrm{ann}_A(D_{|J|}(M)) = 0$. In particular, if $A \neq 0$ and $|J| > |I|$ then μ cannot be injective.

The proof will be given after some preparatory results.

Lemma 9.19. [`lem-minor-ideal`]
For all $k > 0$ we have $D_k(M) \subseteq D_{k-1}(M)$.

Proof. Suppose we have $I' \subseteq I$ and $J' \subseteq I$ with $|I'| = |J'| = k$, and put $N = M|_{I'\times J'}$. It is then clear from the definitions that the entries in $\operatorname{adj}(N)$ lie in $D_{k-1}(M)$, so the identity $\operatorname{adj}(N).N = \det(N).1_{J'}$ shows that $\det(N) \in D_{k-1}(M)$. As $D_k(M)$ is generated by determinants of this form, we have $D_k(M) \subseteq D_{k-1}(M)$. □

Lemma 9.20. [`lem-minor-kernel`]
Suppose we have $k > 0$ and $I' \subseteq I$ and $J' \subseteq J$ with $|I'| = k - 1$ and $|J'| = k$. For $j \in J'$ put $\rho(j) = |\{j' \in J' \mid j' < j\}|$, Define $v \in \mathrm{Free}_J(A)$ by

$$v_j = \begin{cases} (-1)^{\rho(j)} \det(M_{I'\times J'\setminus\{j\}}) & \text{if } j \in J' \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$(Mv)_i = \begin{cases} 0 & \text{if } i \in I' \\ \pm\det(M|_{I'\cup\{i\}\times J'}) & \text{otherwise.} \end{cases}$$

Proof. If $i \notin I'$ then we put $N = M|_{I'\cup\{i\}\times J'}$. We then note that $(Mv)_i = \sum_{j\in J'} M_{ij}v_j$, which is the same up to sign as $(N.\operatorname{adj}(N))_{ii} = \det(N)$.

Similarly, if $i \in I'$ we choose an element $z \notin I$ and define $N \in \mathrm{Mat}_{I' \cup \{z\}, J'}(A)$ by

$$N_{tj} = \begin{cases} M_{tj} & \text{if } t \in I' \\ M_{ij} & \text{if } t = z. \end{cases}$$

We again find that $(Mv)_i = \pm(N.adj(N))_{ii} = \pm\det(N)$, but $N_{ij} = N_{zj}$ for all j so $\det(N) = 0$. □

Proof of Proposition 9.18. Put $n = |J|$.

First suppose that $\mathrm{ann}_A(D_n(M)) = 0$. Consider a vector $v \in \mathrm{Free}_J(A)$ with $Mv = \mu(v) = 0$. For any $I' \subseteq I$ with $|I'| = n$ we then have $(M|_{I' \times J})v = (Mv)|_{I'} = 0$, and we can multiply on the left by $\mathrm{adj}(M|_{I' \times J})$ to get $\det(M|_{I' \times J})v = 0$. This implies that $D_n(M)v = 0$, so each component of v lies in $\mathrm{ann}_A(D_n(M)) = 0$, so $v = 0$. Thus μ is injective as claimed.

Conversely, suppose that μ is injective. Consider an element $a \in \mathrm{ann}_A(D_k(M))$ with $k > 0$. For each $I' \subseteq I$ and $J' \subseteq J$ with $|I'| = k - 1$ and $|J'| = k$, we define $v \in \mathrm{Free}_J(A)$ as in Lemma 9.20. We find that the entries in Mv lie in $D_k(M)$, and thus that $\mu(av) = aMv = 0$. As μ is injective, it follows that $av = 0$. Every generator of $D_{k-1}(M)$ appears (up to sign) as an entry in some such vector v, so it follows that $a \in \mathrm{ann}(D_{k-1}(M))$. Extending this inductively, we see that $\mathrm{ann}(D_n(M)) \subseteq \mathrm{ann}(D_0(M))$, but $D_0(M) = A$ so $\mathrm{ann}(D_n(M)) = 0$.

Note in particular that if $|I| < n$ then $D_n(M) = 0$ so $1 \in \mathrm{ann}(D_n(M))$. Thus, M can only be injective if $1 = 0$, or equivalently A is the trivial ring. □

10. The Cayley-Hamilton Theorem

Definition 10.1. For $M \in \mathrm{Mat}_I(A)$ we define $\chi_M(t) = \det(t.1_I - M) \in A[t]$. We call this the *characteristic polynomial* of M.

Proposition 10.2. *If $|I| = n$ then $\chi_M(t)$ is a monic polynomial of degree n in t.*

Proof. In the definition of $\det(t.1_I - M)$, the identity permutation contributes $\prod_{i \in I}(t - m_{ii})$, and all other permutations contribute terms of degree lower than n. The claim is clear from this. □

The following result is called the Cayley-Hamilton Theorem.

Proposition 10.3. *If $\chi_M(t) = \sum_{i=0}^n a_i t^i$, then the matrix $\chi_M(M) = \sum_{i=0}^n a_i M^i \in \mathrm{Mat}_I(A)$ is zero.*

Proof. Given $P \in \mathrm{Mat}_I(A[t])$ and $Q \in \mathrm{Mat}_I(A)$ we can expand P as $\sum_{k=0}^d P_k t^k$ with $P_k \in \mathrm{Mat}_I(A)$, and we can then define $P * Q = \sum_k P_k Q M^k$. One can check that this satisfies some obvious rules:

$$\begin{aligned} (P + P') * (Q + Q') &= P * Q + P * Q' + P' * Q + P' * Q' \\ (PP') * Q &= P * (P' * Q) \\ 1_I * Q &= Q. \end{aligned}$$

We also have $(t.1_I - M) * 1_I = M - M = 0$. Now consider the characteristic polynomial

$$\chi_M(t) = \det(t.1_I - M) = \sum_{i=0}^n a_i t^i \in A[t].$$

Proposition 9.15 gives

$$\mathrm{adj}(t.1_I - M)(t.1_I - M) = \chi_A(t).1_I = \sum_{i=0}^n a_i t^i.1_I \in \mathrm{Mat}_I(A[t]).$$

It follows that

$$\begin{aligned} \sum_i a_i M^i &= \left(\sum_{i=0}^n a_i t^i.1_I\right) * 1_I \\ &= \mathrm{adj}(t.1_I - M) * ((t.1_I - M) * 1_I) = \mathrm{adj}(t.1_I - M) * 0 = 0, \end{aligned}$$

as claimed. □

One interesting application of the characteristic polynomial is that it gives a way to convert idempotent matrices to idempotent elements.

Definition 10.4. Let E be a matrix in $M_I(A)$ such that $E^2 = E$. Put

$$\begin{aligned}\Phi_E(u) &= 1_I + (u-1)E = (1-u)\left((1-u)^{-1}1_I - E\right) \in M_I(A[u]) \\ \phi_E(u) &= \det(\Phi_E(u)) = (1-u)^n \chi_E((1-u)^{-1}) \in A[u].\end{aligned}$$

Proposition 10.5. *The polynomial $\phi_E(u)$ can be expanded as $\sum_{i=0}^n e_i u^i$ where $e_i^2 = e_i$ (so e_i is idempotent) and $\sum_i e_i = 1$ and $e_i e_j = 0$ when $i \neq j$.*

Proof. Using $E^2 = E$ it is straightforward to check that $\Phi_E(1) = 1_I$ and

$$\Phi_E(u)\Phi_E(v) = 1_I + (uv-1)E = \Phi_E(uv).$$

Taking determinants gives $\phi_E(1) = 1$ and $\phi_E(u)\phi_E(v) = \phi_E(uv)$, or equivalently $\sum_i e_i = 1$ and $\sum_{i,j} e_i e_j u^i u^j = \sum_k e_k u^k v^k$. Comparing coefficients gives $e_i^2 = e_i$, and $e_i e_j = 0$ for $j \neq i$. □

11. Modules

Definition 11.1. [defn-module]

Let A be a ring. An *A-module* is a set M equipped with an element $0 \in M$, an addition operation $M \times M \to M$ and a multiplication operation $A \times M \to M$ such that:

(a) M is an abelian group under addition, with 0 as the identity element.
(b) For all $m \in M$ and $a, b \in A$ we have $1m = m$, and $a(bm) = (ab)m$.
(c) For all $a, b \in A$ and $m, n \in M$ we have $(a+b)m = am + bm$ and $a(m+n) = am + an$.

It is an exercise to check that $0m = 0$ for all m and that $(-1)m = -m$.

Example 11.2. For any finite sets I and J, the sets $\text{Free}_I(A)$ and $\text{Mat}_{IJ}(A)$ are A-modules in an obvious way. In the case $I = \{0, \ldots, n-1\}$ we also write R^n for $\text{Free}_I(R)$.

Definition 11.3. [defn-algebra]

An *A-algebra* is just a ring B equipped with a specified ring homomorphism $\phi\colon A \to B$ (which may be called the *unit map* or the *A-algebra structure map*). For example, if B is any ring and A is a subring of B, then we can use the inclusion map $A \to B$ to regard B as an A-algebra. In particular, $\mathbb{C}$ and $\mathbb{Q}[t]$ can both be regarded as $\mathbb{Q}$-algebras.

Example 11.4. [eg-algebra-as-module]

Any A-algebra can be regarded as an A-module. Indeed, if B is an A-algebra with structure map $\phi\colon A \to B$, then we can use the rule $ab = \phi(a)b$ to define multiplication of elements of B by elements of A, and it is straightforward to check that this satisfies the axioms in Definition 11.1.

Example 11.5. [eg-group-as-module]

If M is any abelian group then we can regard it as a $\mathbb{Z}$-module in an obvious way. More explicitly, for $a \geq 0$ and $m \in M$ we define am recursively by $0m = 0$ and $(a+1)m = am + m$. We then define $(-a)m$ to be the additive inverse of am. It is tedious but essentially straightforward to check all the axioms.

Example 11.6. [eg-module-sum]

If M and N are A-modules, we can define a new A-module $M \oplus N$ as follows. The elements are pairs (m, n) with $m \in M$ and $n \in N$, and the addition and multiplication rules are $(m, n) + (m', n') = (m + m', n + n')$ and $a.(m, n) = (am, an)$. We call $M \oplus N$ the *direct sum* of M and N. This generalises in an obvious way to define $M_0 \oplus \cdots \oplus M_{n-1}$ for any finite list of modules M_i.

Definition 11.7. [defn-module-hom]

If M and N are A-modules, an *A-module homomorphism* (or *A-linear map*) from M to N is a function $\alpha\colon M \to N$ that satisfies $\alpha(m + m') = \alpha(m) + \alpha(m')$ for all $m, m' \in M$, and $\alpha(am) = a\,\alpha(m)$ for all $a \in A$ and $m \in M$. We write $\text{Hom}_A(M, N)$ for the set of all A-module homomorphisms (or just $\text{Hom}(M, N)$ if A is clear from the context).

Remark 11.8. [rem-hom-module]

If M and N are A-modules, then the set $\mathrm{Hom}(M,N)$ is itself an A-module in a natural way. Indeed, if α and β are two elements of $\mathrm{Hom}(M,N)$, then we can define a new map $\alpha+\beta\colon M\to N$ by the obvious rule $(\alpha+\beta)(m)=\alpha(m)+\beta(m)$. This satisfies

$$\begin{aligned}(\alpha+\beta)(am+a'm')&=\alpha(am+a'm')+\beta(am+a'm')\\&=a\alpha(m)+a'\alpha(m')+a\beta(m)+a'\beta(m')\\&=a\,(\alpha+\beta)(m)+a'\,(\alpha+\beta)(m'),\end{aligned}$$

so it is a homomorphism. Similarly, given $\alpha\in\mathrm{Hom}(M,N)$ and $t\in A$ we can define $t\alpha\colon M\to N$ by $(t\alpha)(m)=t\,\alpha(m)$. This is again a homomorphism, so we have addition and scalar multiplication rules for the set $\mathrm{Hom}(M,N)$. To see that this makes $\mathrm{Hom}(M,N)$ into an A-module, we need to check various axioms, for example that $\alpha+\beta=\beta+\alpha$. This is clear because $(\alpha+\beta)(m)=\alpha(m)+\beta(m)$ and $(\beta+\alpha)(m)=\beta(m)+\alpha(n)$, and the right hand sides are the same because addition in N is commutative. The other axioms are equally easy.

Example 11.9. [eg-matrix-hom]

Given finite sets I and J, and a matrix $M\in\mathrm{Mat}_{IJ}(A)$, we can define a homomorphism $\mu_M\colon\mathrm{Free}_J(A)\to\mathrm{Free}_I(A)$ by $\mu_M(u)=Mu$.

Definition 11.10. Let I be any set, and let $\mathrm{Map}(I,A)$ denote the set of all functions from I to A. For $m,n\in\mathrm{Map}(I,A)$ and $a\in A$ we define $m+n,am\in\mathrm{Map}(I,A)$ by $(m+n)(i)=m(i)+n(i)$ and $(am)(i)=a\,m(i)$. It is easy to see that this makes $\mathrm{Map}(I,A)$ into an A-module.

Next, define the *support* of an element $u\in\mathrm{Map}(I,A)$ to be the set $\mathrm{supp}(u)=\{i\in I\mid u(i)\neq 0\}$. We say that u is *finitely supported* if $\mathrm{supp}(u)$ is a finite set. For example, for each $i\in I$ we can define $e_i\colon I\to A$ by

$$e_i(j)=\begin{cases}1 & \text{if } j=i\\ 0 & \text{otherwise.}\end{cases}$$

We then have $\mathrm{supp}(e_i)=\{i\}$. We write $\mathrm{Free}_I(A)$ or $\mathrm{Map}_0(I,A)$ for the set of all finitely supported functions. (Previously we defined $\mathrm{Free}_I(A)=\mathrm{Map}(I,A)$ for finite sets I; our new definition is clearly compatible with that.) It is clear that $\mathrm{supp}(u+v)\subseteq\mathrm{supp}(u)\cup\mathrm{supp}(v)$, and $\mathrm{supp}(au)\,\mathrm{supp}(u)$, so $\mathrm{Free}_I(A)$ is a submodule of $\mathrm{Map}(I,A)$. We say that a module M is *free* if it is isomorphic to $\mathrm{Free}_I(A)$ for some set I.

Definition 11.11. [defn-free-universal]

Suppose we have a ring A, an A-module M and a set I. For any map $m\colon I\to M$ we define $\phi_m\colon\mathrm{Free}_I(A)\to M$ by

$$\phi_m(u)=\sum_{i\in\mathrm{supp}(u)}u(i)\,m(i)=\sum_{i\in I}u(i)\,m(i).$$

Proposition 11.12. [prop-free-universal]

For any map $m\colon I\to M$, the resulting map $\phi_m\colon\mathrm{Free}_I(A)\to M$ is a homomorphism, with $\phi_m(e_i)=m(i)$ for all i. Conversely, if we start with any homomorphism $\psi\colon\mathrm{Free}_I(A)\to M$ and define $m(i)=\psi(e_i)$, then we have $\psi=\phi_m$. Thus, these constructions give a natural bijection $\mathrm{Hom}_A(\mathrm{Free}_I(A),M)\simeq\mathrm{Map}(I,M)$.

Proof. Straightforward. The key point is that any element $u\in\mathrm{Free}_I(A)$ can be expressed as $u=\sum_{i\in\mathrm{supp}(u)}u(i)\,e_i$, so if $\psi\colon\mathrm{Free}_I(A)\to M$ is a homomorphism we have $\psi(u)=\sum_{i\in\mathrm{supp}(u)}u(i)\,\psi(e_i)$. □

Remark 11.13. Definition 11.11 and Proposition 11.12 are formulated in terms of maps $I\to M$. Often we have a list $m_0,\dots,m_{n-1}$ of elements of M and we consider the map from the set $N=\{0,\dots,n-1\}$ to M given by $i\mapsto m_i$; this gives a homomorphism $\phi_m\colon A^n\simeq\mathrm{Free}_N(A)\to M$. It is also common to have a subset $S\subset M$ and to consider the inclusion map $S\to M$, giving a homomorphism $\phi_S\colon\mathrm{Free}_S(A)\to M$. We will need some straightforward translations between these contexts, most of which we leave to the reader.

The following result is essentially a special case of Proposition 11.12, but written in slightly different notation.

Proposition 11.14. [prop-matrix-hom]

Let I and J be finite sets. Then any A-module homomorphism $\mathrm{Free}_J(A) \to \mathrm{Free}_I(A)$ has the form α_M for a unique matrix $M \in \mathrm{Mat}_{IJ}(A)$, so $\mathrm{Hom}_A(\mathrm{Free}_J(A), \mathrm{Free}_I(A))$ can be identified with $\mathrm{Mat}_{IJ}(A)$.

Proof. Let $\phi\colon \mathrm{Free}_J(A) \to \mathrm{Free}_I(A)$ be an A-module homomorphism. For each $j \in J$ we have an element $\phi(e_j) \in \mathrm{Free}_I(A)$. We write M_{ij} for the i'th coefficient of $\phi(e_j)$, so that $\phi(e_j) = \sum_i M_{ij} e_i$. An arbitrary element $v \in \mathrm{Free}_J(A)$ can be expressed as $v = \sum_j v_j e_j$, giving

$$\phi(v) = \sum_j v_j \phi(e_j) = \sum_{i,j} M_{ij} v_j e_i,$$

or in other words $\phi(v)_i = \sum_j M_{ij} v_j$. This shows that $\phi = \alpha_M$. □

Definition 11.15. Consider a map $m\colon I \to M$, and the corresponding homomorphism $\phi_m\colon \mathrm{Free}_I(A) \to M$.

(a) We say that a map m is *A-linearly independent* if ϕ_m is injective.
(b) We say that m *spans* M if ϕ_m is surjective.
(c) We say that m is a *basis* for M if ϕ_m is an isomorphism. (Thus, M is free iff it has a basis.)

Remark 11.16. If $m\colon I \to M$ is A-linearly independent, or spans, or is a basis, then the inclusion map $m(I) \to M$ has the same property. Conversely, if $m(I) \to M$ spans then so does $m\colon I \to M$. However, the corresponding statements for (a) and (c) are not true.

Definition 11.17. We say that a module M is *finitely generated* if there is a surjective homomorphism $A^n \to M$, or equivalently there is a map $I \to M$ that spans M with I finite.

Proposition 11.18. *If $A \neq 0$ and $n > m$ then there is no injective A-algebra homomorphism from $A[x_0, \dots, x_{n-1}]$ to $A[x_0, \dots, x_{m-1}]$.*

Proof. Write $P_n = A[x_0, \dots, x_{n-1}]$ for brevity, and let $F_d P_n$ denote the span of the monomials of total degree at most d. This is a free module over A of rank $\begin{pmatrix} d+n \\ n \end{pmatrix}$, which is a polynomial in d with leading term $d^n/n!$. Let $\phi\colon P_n \to P_m$ be an R-algebra homomorphism, where $m < n$. Choose t such that $\phi(x_i) \in F_t P_m$ for all i, and note that $\phi(F_d P_n) \subseteq F_{td} P_m$. As $m < n$ we will have $\begin{pmatrix} d+n \\ n \end{pmatrix} > \begin{pmatrix} td+m \\ m \end{pmatrix}$ for large d, and it follows from Proposition 9.18 that ϕ cannot be injective. □

Proposition 11.19. [prop-rank-unique]

If A is a nontrivial ring and $n < m$ then there is no surjective homomorphism from A^n to A^m. In particular, A^n and A^m are not isomorphic.

Proof. Suppose we have a surjective homomorphism $\phi\colon A^n \to A^m$. We can then choose $u_i \in A^n$ with $\phi(u_i) = e_i$ for $0 \leq i < m$. Using these, we define $\psi\colon A^m \to A^n$ by $\psi(t) = \sum_{i=0}^{m-1} t_i u_i$; we find that $\phi(\psi(e_i)) = e_i$ for all i, and thus that $\phi\psi$ is the identity.

Next, by Proposition 11.14, there are matrices $M \in \mathrm{Mat}_{m,n}(A)$ and $N \in \mathrm{Mat}_{n,m}(A)$ with $\phi = \alpha_M$ and $\psi = \alpha_N$. As $\phi\psi$ is the identity we see that $MN = 1_m$ and in particular $\det(MN) = 1$. As A is nontrivial we have $\det(MN) \neq 0$, and using Proposition 9.12 it follows that $n \geq m$. □

Corollary 11.20. [cor-rank-unique]

Suppose we have a finite set I and a possibly infinite set J with $|J| > |I|$. Then there is no surjective homomorphism from $\mathrm{Free}_I(A)$ to $\mathrm{Free}_J(A)$.

Proof. We can put $n = |I|$ and choose $K \subseteq J$ with $|K| = n + 1$. We can then define a surjective homomorphism $\pi\colon \mathrm{Free}_J(A) \to \mathrm{Free}_K(A) \simeq R^{n+1}$ by $\pi(e_k) = e_k$ when $k \in K$, and $\pi(e_j) = 0$ for $j \in J \setminus K$. If $\phi\colon \mathrm{Free}_I(A) \to \mathrm{Free}_J(A)$ is surjective, then the same is true of the composite

$$\pi\phi\colon \mathrm{Free}_I(A) \simeq R^n \to \mathrm{Free}_K(A) \simeq R^{n+1},$$

but that is impossible by Proposition 11.19. □

This allows us to make the following definition.

Definition 11.21. If M is isomorphic to $\mathrm{Free}_I(A)$ for some finite set I, then we define the *rank* of M to be the cardinality of I, which is well-defined by Proposition 11.19.

Definition 11.22. Let M be an A-module. A *submodule* of M is a subset $N \subseteq M$ such that

(a) $0 \in N$
(b) For all $n, n' \in N$ we have $n + n' \in N$
(c) For all $a \in A$ and $n \in N$ we have $an \in N$.

Remark 11.23. If N is a submodule of M, then we can regard N itself as an A-module by restricting the addition and multiplication operations on M.

Example 11.24. [eg-ideal-submodule]
We can regard A itself as an A-module; then submodules of A are just the same as ideals in A.

Example 11.25. [eg-ker-img]
If $\alpha\colon M \to N$ is any homomorphism of A-modules, we put

$$\ker(\alpha) = \{m \in M \mid \alpha(m) = 0\}$$
$$\mathrm{image}(\alpha) = \{n \in N \mid n = \alpha(m) \text{ for some } m \in M\}.$$

It is straightforward to check that $\ker(\alpha)$ is a submodule of M and $\mathrm{image}(\alpha)$ is a submodule of N.

Example 11.26. [eg-submodule-ops]
If P and Q are submodules of M, then so are the subsets $P \cap Q$ and $P + Q = \{p + q \mid p \in P,\ q \in Q\}$.

Example 11.27. [eg-ann-submodule]
If I is an ideal in A, then the set

$$\mathrm{ann}_M(I) = \{m \in M \mid am = 0 \text{ for all } a \in I\}$$

is a submodule of M.

We also write IM for the set of elements $m \in M$ that can be expressed in the form $m = a_1m_1 + \cdots + a_km_k$ with $a_i \in I$ and $m_i \in M$. This is again a submodule of M.

Example 11.28. [eg-submodule-span]
Let M be an A-module, and let G be any subset of M. We say that an element $m \in M$ is an *A-linear combination* of G if there exist elements $a_0, \dots, a_{n-1} \in A$ and $g_0, \dots, g_{n-1} \in G$ such that $m = \sum_i a_ig_i$. (The case $n = 0$ is permitted, so 0 is always a linear combination even if $G = \emptyset$.) We write $\mathrm{span}_A(G)$ for the set of all possible linear combinations. This is easily seen to be a submodule of M, and in fact it is the smallest submodule that contains G. We call it the *span of* G, or the *submodule generated by* G.

Definition 11.29. [prop-infinite-product]
Suppose we have a set I (which may be infinite) and a family of modules M_i for $i \in I$. We define $\prod_i M_i$ to be the set of all indexed families $(m_i)_{i\in I}$ where $m_i \in M_i$ for all i. We can define addition of such families termwise, and similarly for multiplication by elements of A; this makes $\prod_i M_i$ into an A-module. For $j \in I$ we define a homomorphism $\pi_j\colon \prod_i M_i \to M_j$ by $\pi_j((m_i)_{i\in I}) = m_j$.

Example 11.30. [eg-map-as-prod]
If all the modules M_i are equal to the same module M, then $\prod_i M_i$ is just the same as $\mathrm{Map}(I, M)$. More generally, if all the modules M_i are submodules of a fixed module P, then we have

$$\prod_i M_i = \{u \in \mathrm{Map}(I, P) \mid u(i) \in M_i \text{ for all } i\}.$$

Example 11.31. [eg-finite-prod]
If $I = \{0, \dots, n-1\}$ then $\prod_i M_i$ is the same as $M_0 \oplus \cdots \oplus M_{n-1}$.

Remark 11.32. [rem-prod-categorical]
If we have another module N and homomorphisms $f_i\colon N \to M_i$ for all i, we can combine them to get a single homomorphism $f\colon N \to \prod_i M_i$ given by $f(n) = (f_i(n))_{i\in I}$. This is the unique homomorphism that satisfies $\pi_i \circ f = f_i$ for all i. It follows that $\prod_i M_i$ is the product of the modules M_i in the sense of category theory.

Definition 11.33. [defn-infinite-sum]

For an element $m \in \prod_i M_i$, we put $\operatorname{supp}(m) = \{i \mid m_i \neq 0\}$. We say that m is *finitely supported* if $\operatorname{supp}(m)$ is a finite set. We write $\bigoplus_i M_i$ for the set of all finitely supported elements. It is not hard to see that $\operatorname{supp}(am + a'm') \subseteq \operatorname{supp}(m) \cup \operatorname{supp}(m')$ and thus that $\bigoplus_i M_i$ is a submodule of $\prod_i M_i$.

We define homomorphisms $\iota_j \colon M_j \to \bigoplus_i M_i$ by

$$\iota_j(m)_i = \begin{cases} m & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Remark 11.34. [rem-infinite-sum-elements]

Note that an arbitrary element $m \in \bigoplus_i M_i$ can be expressed as $m = \sum_{i \in \operatorname{supp}(m)} \iota_i(m_i)$.

Example 11.35. [eg-free-as-sum]

In the case where $M_i = A$ for all i, we have $\bigoplus_i A = \operatorname{Free}_I(A)$.

Example 11.36. [eg-sum-as-product]

If I is finite it is clear that $\bigoplus_i M_i = \prod_i M_i$.

Remark 11.37. [rem-coproduct-categorical]

If we have another module P and homomorphisms $g_i \colon M_i \to P$ for all i, we can combine them to get a single homomorphism $g \colon \bigoplus_i M_i \to P$ defined by

$$g(m) = \sum_{i \in \operatorname{supp}(m)} g_i(m_i) = \sum_{i \in I} g_i(m_i).$$

This is the unique homomorphism such that $g \circ \iota_i = g_i$ for all i. It follows that $\bigoplus_i M_i$ is the coproduct of the modules M_i in the sense of category theory.

Remark 11.38. [rem-submodule-sum]

Suppose we have a module M and a family of submodules $N_i \subseteq M$ for i in some set I. It is then easy to check that the set $\bigcap_i N_i = \{m \in M \mid m \in N_i \text{ for all } i\}$ is again a submodule. Note that a submodule $P \subseteq M$ satisfies $P \subseteq \bigcap_i N_i$ iff we have $P \subseteq N_i$ for all i.

Next, we define $\sigma \colon \bigoplus_i N_i \to M$ by $\sigma(n) = \sum_{i \in \operatorname{supp}(n)} n_i$, and we put $\sum_i N_i = \operatorname{image}(\sigma)$, which is another submodule of M. Given a submodule $P \subseteq M$, it is not hard to check that $P \supseteq \sum_i N_i$ iff we have $P \supseteq N_i$ for all i.

Remark 11.39. [rem-chain-union]

The union of a family of submodules is not generally a submodule. However, if we have a nested chain of submodules

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq M,$$

then $\bigcup_i N_i$ is a submodule of M, by the argument that we gave in Proposition 5.12.

Definition 11.40. [defn-quotient-module]

Let M be an A-module, and let N be a submodule. A *coset* of N is a subset $u \subseteq N$ that can be expressed in the form $u = m + N$ for some $m \in M$. We write M/N for the set of all cosets. Given cosets $u, v \in M/N$ and an element $a \in A$ we put

$$\begin{aligned} u + v &= \{x + y \mid x \in u,\ y \in v\} \subseteq M \\ au &= \{ax + n \mid x \in u, y \in N\} \subseteq M. \end{aligned}$$

One can check that $u + v$ and au are cosets, and that these operations make M/N into another A-module. There is an A-module homomorphism $\pi \colon M \to M/N$ given by $\pi(m) = m + N$.

Proposition 11.41. [prop-quotient-module-map]

Let $\alpha \colon M \to P$ be a homomorphism of A-modules, and let $N \subseteq M$ be a submodule such that $\alpha(N) = 0$, or equivalently $N \subseteq \ker(\alpha)$. Then there is a unique homomorphism $\overline{\alpha} \colon M/N \to P$ such that $\overline{\alpha} \circ \pi = \alpha$, or equivalently $\overline{\alpha}(m + N) = \alpha(m)$ for all $m \in M$. Moreover:

(a) *$\overline{\alpha}$ is injective iff* $\ker(\alpha) = N$.
(b) *$\overline{\alpha}$ is surjective iff α is surjective.*

(c) $\overline{\alpha}$ *is an isomorphism iff* α *is surjective with* $\ker(\alpha) = N$.

Proof. This is very similar to Proposition 5.27. If $u \in M/N$ and $m, m' \in u$ then $u - u' \in N \leq \ker(\alpha)$ so $\alpha(m - m') = 0$ so $\alpha(m) = \alpha(m')$. We define $\overline{\alpha}(u)$ to be the common value of $\alpha(m)$ for all $m \in u$. This gives a well-defined map $\overline{\alpha}\colon M/N \to P$ with $\overline{\alpha}(m + N) = \alpha(m)$ for all m, so $\overline{\alpha} \circ \pi = \alpha$. If $u = m + N$ and $u' = m' + N$ and $a, a' \in A$ then we can choose $m \in u$ and $m' \in u'$ and we find that

$$\overline{\alpha}(au + a'u') = \overline{\alpha}(am + a'm' + N) = \alpha(am + a'm') = a\alpha(m) + a'\alpha(m') = a\overline{al}(u) + a'\overline{al}(u'),$$

so $\overline{\alpha}$ is a module homomorphism.

(a) Suppose that $\ker(\alpha) = N$, and that $\overline{\alpha}(u) = \overline{\alpha}(u')$. We can choose $m \in u$ and $m' \in u'$, and we deduce that $\alpha(m) = \alpha(m')$, so $m - m' \in \ker(\alpha) = N$, so $m + N = m' + N$, or in other words $u = u'$. This shows that $\overline{\alpha}$ is injective as claimed. The converse is similar and is left to the reader.

(b) For any $m \in M$ we have $\alpha(m) = \overline{al}(m + N)$, and for any $u \in M/N$ we can choose $m \in u$ and we have $\overline{\alpha}(u) = \alpha(m)$. This shows that α and $\overline{\alpha}$ have the same image, so in particular α is surjective iff $\overline{\alpha}$ is surjective.

(c) This is clear from (a) and (b).

□

Proposition 11.42. [`prop-quotient-submodules`]

Consider a module M, a submodule N and the quotient map $\pi\colon M \to M/N$.

(a) *For any submodule $P \subseteq M$, the image $\pi(P)$ is the same as $(P+N)/N$ and is a submodule of M/N. If $P \supseteq N$ then $\pi(P)$ is just P/N.*

(b) *For any submodule $Q \subseteq M/N$, the preimage $\pi^{-1}(Q) = \{m \in M \mid \pi(m) \in Q\}$ is a submodule of M that contains N.*

(c) *For any submodule $P \subseteq M$ we have $\pi^{-1}(\pi(P)) = P + N$, which is the same as P iff $P \supseteq N$.*

(d) *For any submodule $Q \subseteq M/N$ we have $\pi(\pi^{-1}(Q)) = Q$.*

(e) *Thus, we have a natural bijection*

$$\{\ \textit{submodules of } M \textit{ containing } N\} \simeq \{\ \textit{submodules of } M/N\}.$$

Proof. (a) For any $p \in P$ and $n \in N$ we have $\pi(p) = \pi(p + n)$. It follows that $\pi(P) = \pi(P + N)$. This is the set of all cosets of the form $x + N$ with $x \in P + N$, and by definition this is $(P + N)/N$, which is clearly a submodule of M/N. If $P \supseteq N$ then $P = P + 0 \subseteq P + N \subseteq P + P = P$, so $P + N = P$, so $\pi(P) = P/N$.

(b) First, as $\pi(N) = 0_{M/N} \in Q$, we see that $N \subseteq \pi^{-1}(Q)$, and in particular $0_M \in \pi^{-1}(Q)$. Now suppose that $u, v \in \pi^{-1}(Q)$ and $a \in A$. Then $\pi(u + v) = \pi(u) + \pi(v) \in Q + Q = Q$, so $u + v \in \pi^{-1}(Q)$. Similarly $\pi(au) = a\pi(u) \in aQ \subseteq Q$, so $au \in \pi^{-1}(Q)$. This shows that $\pi^{-1}(Q)$ is a submodule of M.

(c) Suppose we start with a submodule $P \subseteq M$. If $u \in P + N$ then $\pi(u) \in \pi(P + N) = \pi(P)$, so $u \in \pi^{-1}(\pi(P))$. Conversely, suppose that $u \in \pi^{-1}(\pi(P))$, so $\pi(u) \in \pi(P)$, so there exists $v \in P$ such that $\pi(u) = \pi(v)$. This means that $\pi(u - v) = 0$, so $u - v \in \ker(\pi) = N$, so $u = v + (u - v) \in P + N$. We conclude that $\pi^{-1}(\pi(P)) = P + N$, and we have already seen that this is the same as P if $N \subseteq P$.

(d) Consider a submodule $Q \subseteq M/N$. If $u \in \pi^{-1}(Q)$ then $\pi(u) \in Q$; it follows that $\pi(\pi^{-1}(Q)) \subseteq Q$. Conversely, suppose that $v \in Q$. As $Q \subseteq M/N$, we have $v = u + N = \pi(u)$ for some element $u \in M$. As $\pi(u) = v \in Q$, we actually have $u \in \pi^{-1}(Q)$. This means that $v = \pi(u) \in \pi(\pi^{-1}(Q))$ as required. We conclude that $Q = \pi(\pi^{-1}(Q))$.

(e) Put

$$\begin{aligned} \mathrm{Sub}(M, N) &= \{\ \text{submodules of } M \text{ containing } N\} \\ \mathrm{Sub}(M/N) &= \{\ \text{submodules of } M/N\}. \end{aligned}$$

Part (a) shows that we can define a map $\alpha\colon \mathrm{Sub}(M, N) \to \mathrm{Sub}(M/N)$ by $\alpha(P) = (P+N)/N = \pi(P)$. Part (b) shows that we can define a map $\beta\colon \mathrm{Sub}(M/N) \to \mathrm{Sub}(M, N)$ by $\beta(Q) = \pi^{-1}(Q)$. Part (c) shows that $\beta\alpha$ is the identity, and part (d) shows that $\alpha\beta$ is the identity. Thus, α and β are bijections.

□

The name *Nakayama's Lemma* is often attached to the following result, or to one of several special cases or closely related statements.

Proposition 11.43. [`prop-nakayama`]
Let M be a finitely generated A-module, and let I be an ideal such that $M = IM$. Then there is an element $a \in I$ such that $(1-a)M = 0$.

Proof. By hypothesis, there is a finite list $(m_0, \dots, m_{n-1}) \in M^n$ such that the corresponding map $\phi_m \colon A^n \to M$ is surjective. We also have $M = IM$, so $m_i \in IM = \phi_m(I^n)$, so we can choose elements $u_{ij} \in I$ such that $m_i = \sum_j u_{ij} m_j$ for all i. If we form a matrix U with entries u_{ij} we find that $m = Um$ in A^n, or equivalently $(1_n - U)m = 0$. We can multiply by $\operatorname{adj}(1_n - U)$ to deduce that $\det(1_n - U)m = 0$, but the elements m_i generate M, so $\det(1_n - U)M = 0$. On the other hand, as the entries in U are all elements of I, we find that $\det(1_n - U) = \det(1_n) = 1 \pmod{I}$, so $\det(1_n - U) = 1 - a$ for some $a \in I$. □

Corollary 11.44. [`cor-nakayama`]
Let M be a finitely generated A-module, let N be a submodule, and let I be an ideal such that $M = IM + N$. Then there is an element $a \in I$ such that $(1-a)M \leq N$. Moreover, if $I \leq \operatorname{Rad}(A)$ then $M = N$.

Proof. For the first statement, we note that the quotient $L = M/N$ is finitely generated and has $IL = L$, so there exists $a \in I$ with $(1-a)L = 0$, or equivalently $(1-a)M \leq N$. If $I \leq \operatorname{Rad}(A)$ then $1 - a$ is invertible and so $M \leq N$. □

Definition 11.45. [`defn-fraction-module`]
Let A be a ring and let U be a multiplicative subset of A. For any A-module M, we define $M[U^{-1}] = (M \times U)/\sim$, where $(m, u) \sim (m', u')$ iff there is an element $v \in U$ with $mu'v = m'uv$. We write m/u for the equivalence class of (m, u), and $\eta(m)$ for $m/1$.
If $U = A \backslash P$ for some prime ideal P, we also use the notation M_P for $M[U^{-1}]$ (generalising Definition 8.10).

By essentially the same arguments as those given in Section 8, we see that $M[U^{-1}]$ has a natural structure as a module over $A[U^{-1}]$, with $(a/u).(m/v) = (am)/(uv)$ and $m/v + n/w = (wm + vn)/(vw)$.

Example 11.46. Recall that we have a ring homomorphism $\eta \colon A \to A[U^{-1}]$, and for any ideal $I \subseteq A$ we define $\eta_*(I) = \operatorname{span}_{A[U^{-1}]}(\eta(I))$. We can also regard I as an A-module and thus define $I[U^{-1}]$. Using Lemma 8.12 one can identify $\eta_*(I)$ with $I[U^{-1}]$.

Remark 11.47. [`rem-fraction-functor`]
Given an A-module homomorphism $\phi \colon M \to N$, we can define an $A[U^{-1}]$-module homomorphism $\phi[U^{-1}] \colon M[U^{-1}] \to N[U^{-1}]$ by $\phi[U^{-1}](m/u) = \phi(m)/u$. A little work is required to show that this is well-defined and is a homomorphism, but we leave that to the reader. We will often just write ϕ instead of $\phi[U^{-1}]$.

Proposition 11.48. [`prop-stalks-zero`]

(a) *Let M be an A-module such that $M_P = 0$ for all maximal ideals P; then $M = 0$.*
(b) *Let M be a finitely generated A-module such that $M/PM = 0$ for all maximal ideals P; then $M = 0$.*

Proof. In both cases we assume that we have a nonzero element $m \in M$, and derive a contradiction. As $m \neq 0$ we see that $\operatorname{ann}(m) \neq A$. By Proposition 5.49, there exists a maximal ideal $P \geq \operatorname{ann}(m)$. This means that for $s \in A \setminus P$ we have $s \notin \operatorname{ann}(m)$ and so $sm \neq 0$. It follows that $m/1$ is a nonzero element of M_P, so $M_P \neq 0$. This provides the required contradiction for (a). For (b) we instead use Proposition 11.43, which gives us an element $a \in P$ with $1 - a \in \operatorname{ann}(M)$. As $\operatorname{ann}(M) \leq P$ this gives $1 \in P$, which is again impossible. □

Remark 11.49. To guard against over-optimistic generalisations, we can consider the following example. Take $A = \mathbb{Z}$ and $M = \mathbb{Q}/\mathbb{Z}$ (which is not finitely generated). We find that $M \otimes_A K(P) = 0$ for all prime ideals P, but $M \neq 0$. (Here $K(P)$ is the residue field as in Example 8.17, so $M \otimes_A K(P) = M/PM$ when M is maximal. Thus for $A = \mathbb{Z}$ we have $M \otimes_A K(\mathbb{Z}p) = M/pM$ and $M \otimes_A K(0) = M \otimes_{\mathbb{Z}} \mathbb{Q}$.)

12. Exact sequences

Definition 12.1. [defn-exact]

Consider a sequence of A-module homomorphisms $L \xrightarrow{\phi} M \xrightarrow{\psi} N$. We then have submodules $\text{image}(\phi), \ker(\psi) \subseteq M$, and it is not hard to see that $\text{image}(\phi) \subseteq \ker(\psi)$ if and only if the composite $\psi\phi$ is zero. We say that this sequence is *exact* if in fact $\text{image}(\phi) = \ker(\psi)$.

More generally, we say that a sequence

$$M_0 \xrightarrow{\phi_0} M_1 \xrightarrow{\phi_1} \cdots M_{n-1} \xrightarrow{\phi_{n-1}} M_n$$

is *exact* if each of the sequences $M_i \to M_{i+1} \to M_{i+2}$ is exact. We make the same definition for sequences that extend infinitely to the left or to the right or both.

Remark 12.2. [rem-exact]

Some special cases are as follows.

(a) A sequence of the form $L \xrightarrow{\phi} M \xrightarrow{0} N$ is exact iff ϕ is surjective. In particular, a sequence of the form $L \xrightarrow{\phi} M \to 0$ is exact iff ϕ is surjective.

(b) A sequence of the form $L \xrightarrow{0} M \xrightarrow{\psi} N$ is exact iff ψ is injective. In particular, a sequence of the form $0 \to M \xrightarrow{\psi} N$ is exact iff ψ is surjective.

(c) A sequence of the form $K \xrightarrow{0} L \xrightarrow{\phi} M \xrightarrow{0} N$ is exact iff ϕ is an isomorphism.

(d) A sequence of the form $L \xrightarrow{0} M \xrightarrow{0} N$ is exact iff $M = 0$.

Definition 12.3. [defn-ses]

A *short exact sequence* is an exact sequence of the form

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0.$$

Exactness here means that ϕ is injective, ψ is surjective, and $\text{image}(\phi) = \ker(\psi)$.

Example 12.4. [eg-standard-ses]

Suppose we have a module M and a submodule L. We let $j\colon L \to M$ be the inclusion map, and we let $q\colon M \to M/L$ be the projection, so $j(x) = x$ and $q(y) = y + L$. We then find that the sequence

$$0 \to L \xrightarrow{j} M \xrightarrow{q} M/N \to 0$$

is short exact.

Example 12.5. [eg-sum-ses]

Suppose we have modules L and N, and we define maps

$$0 \to L \xrightarrow{j} L \oplus N \xrightarrow{q} N \to 0$$

by $j(x) = (x, 0)$ and $q(x, z) = z$. It is then clear that the sequence is short exact.

Example 12.6. [eg-nm-ses]

For any integers $n, m > 0$ we can define maps

$$0 \to \mathbb{Z}/n \xrightarrow{\phi} \mathbb{Z}/nm \xrightarrow{\psi} \mathbb{Z}/m \to 0$$

by $\phi(a + n\mathbb{Z}) = ma + nm\mathbb{Z}$ and $\psi(b + nm\mathbb{Z}) = b + m\mathbb{Z}$. One can check that these give a short exact sequence.

In fact, every short exact sequence is essentially the same as one coming from example 12.4. More formally:

Proposition 12.7. [prop-ses]

Let $0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0$ be a short exact sequence. Then there is a commutative diagram as follows, in which the vertical maps are isomorphisms.

$$\begin{tikzcd}
L \arrow[r, "\phi"] \arrow[d, "\alpha"', "\simeq"] & M \arrow[r, "\psi"] \arrow[d, "1"'] & N \arrow[d, "\beta"', "\simeq"] \\
\phi(L) \arrow[r, "j"'] & M \arrow[r, "q"'] & M/\phi(L)
\end{tikzcd}$$

Proof. As the given sequence is short exact, we see that ϕ is injective, so it restricts to give an isomorphism $L \to \phi(L)$ which we call α. By definition we have $j \circ \alpha = \phi$ so the left square commutes. Next, using Proposition 11.41 we see that there is a unique homomorphism

$$\overline{\psi}\colon M/\ker(\psi) = M/\operatorname{image}(\phi) \to N$$

such that $\psi = q \circ \overline{\psi}$. As the original sequence is exact we see that ψ is surjective and thus that $\overline{\psi}$ is an isomorphism. We put $\beta = \overline{\psi}^{-1}$, so the equation $\psi = q \circ \overline{\psi}$ gives $q = \psi \circ \beta$. Thus, the right hand square commutes. □

We next discuss some ideas which help us decide whether an arbitrary short exact sequence is essentially the same as one coming from Example 12.5.

Definition 12.8. [`defn-splitting`]

Suppose we have a short exact sequence

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0.$$

(a) A *left splitting* is a homomorphism $\rho\colon M \to L$ such that $\rho\phi = 1_L$.
(b) A *right splitting* is a homomorphism $\sigma\colon N \to M$ such that $\psi\sigma = 1_N$.
(c) A *full splitting* is a pair of homomorphisms (ρ, σ) such that ρ is a left splitting and σ is a right splitting and $\rho\sigma = 0\colon N \to L$ and $\phi\rho + \sigma\psi = 1_M$.

We say that the sequence is *split* if it admits a full splitting.

Example 12.9. [`eg-nm-split`]

Take $n = m$ in Example 12.6 to get a short exact sequence $\mathbb{Z}/n \to \mathbb{Z}/n^2 \to \mathbb{Z}/n$. If this is split then $\mathbb{Z}/n^2 \simeq \mathbb{Z}/n \oplus \mathbb{Z}/n$, so every element $x \in \mathbb{Z}/n^2$ satisfies $nx = 0$. In particular we can take $x = 1 + n^2\mathbb{Z}$ to see that n^2 divides n, so $n = 1$. Thus, when $n > 1$ the sequence does not split. A similar argument shows more generally that the sequence in Example 12.6 splits iff n and m are coprime.

Proposition 12.10. [`prop-splitting`]

(a) *For every left splitting ρ there is a unique σ such that (ρ, σ) is a full splitting.*
(b) *For every right splitting σ there is a unique ρ such that (ρ, σ) is a full splitting.*
(c) *For every full splitting (ρ, σ) we have an isomorphism $\alpha\colon M \to L \oplus N$ given by $\alpha(y) = (\rho(y), \psi(y))$, with inverse $\alpha^{-1}(x, z) = \phi(x) + \sigma(z)$. Moreover, the diagram*

$$\begin{tikzcd}
L \arrow[r, "\phi"] \arrow[d, "1"'] & M \arrow[r, "\psi"] \arrow[d, "\alpha"', "\simeq"] & N \arrow[d, "1"] \\
L \arrow[r, "j"'] & L \oplus N \arrow[r, "q"'] & N
\end{tikzcd}$$

commutes.

Proof.

(a) Let ρ be a left splitting. For $z \in L$ we can choose $y \in M$ with $\psi(y) = z$. We would like to define $\sigma(z) = y - \phi\rho(y)$. To check that this is well-defined, suppose we have another element $y' \in M$ with $\psi(y') = z$. Then $y' - y \in \ker(\psi) = \operatorname{image}(\phi)$, so $y' = y + \phi(x)$ for some $x \in L$. This gives

$$y' - \phi\rho(y') = y - \phi\rho(y) + \phi(x) - \phi\rho\phi(x),$$

but $\rho\phi$ is the identity so the last two terms cancel and $y' - \phi\rho(y') = y - \phi\rho(y)$. We thus have a well-defined function σ as claimed.

Now consider elements $z_0, z_1 \in N$ and $a_0, a_1 \in A$. Choose y_i with $\psi(y_i) = z_i$, so $\sigma(z_i) = y_i - \phi\rho(y_i)$. The element $y = a_0y_0 + a_1y_1$ is mapped by ψ to the element $z = a_0z_0 + a_1z_1$, so we can use y when calculating $\sigma(z)$, and we find that $\sigma(z) = a_0\sigma(z_0) + a_1\sigma(z_1)$. This shows that σ is an A-module homomorphism.

As the given sequence is exact we have $\psi\phi = 0$, so $\psi\sigma(z) = \psi(y) - \psi\phi\rho(y) = \psi(y) = z$. Thus $\psi\sigma = 1_N$, which means that σ is a right splitting. We also have $\rho\sigma(z) = \rho(y) - \rho\phi\rho(z)$, which is zero because $\rho\phi = 1_L$.

For any $y \in M$ we can use y to calculate $\sigma\psi(y)$, and we get $\sigma\psi(y) = y - \phi\rho(y)$. This shows that $\phi\rho + \sigma\psi = 1_M$, so we have a full splitting as claimed.

(b) Now suppose instead that we have a right splitting σ, so $\psi\sigma = 1_N$. For $y \in M$ the element $y - \sigma\psi(y)$ lies in the kernel of ψ, which is the same as the image of ϕ, so there is an element $x \in L$ with $\phi(x) = y - \sigma\psi(y)$. This element is unique because ϕ is injective, so we can define $\rho\colon M \to L$ by $\rho(y) = x$. Now suppose that $y = a_0y_0 + a_1y_1$ for some $a_0, a_1 \in A$ and $y_0, y_1 \in M$. We find that the element $x = a_0\rho(y_0) + a_1\rho(y_1)$ satisfies $\phi(x) = y - \sigma\psi(y)$, so $x = \rho(y)$; this proves that ρ is a homomorphism.

Now suppose instead that $y = \phi(x)$ for some x. We then have $\psi(y) = 0$ so $\phi(x) = y = y - \sigma\psi(y)$, so $x = \rho(y)$. This proves that $\rho\phi = 1_L$.

Now suppose instead that $y = \sigma(z)$ for some $z \in N$. Using $\psi\sigma = 1_N$ we see that $y - \sigma\psi(y) = 0$ and thus that $\rho(y) = 0$. This proves that $\rho\sigma = 0$.

Finally, for any $y \in M$ it is built into the definition of ρ that $y - \sigma\psi(y) = \phi\rho(y)$, so $\phi\rho + \sigma\psi = 1_M$, so we again have a full splitting.

(c) Now let (ρ, σ) be any full splitting. We can certainly define maps $M \xrightarrow{\alpha} L \oplus N \xrightarrow{\beta} M$ by $\alpha(y) = (\rho(y), \psi(y))$ and $\beta(x, z) = \phi(z) + \sigma(z)$. The axioms for a full splitting, together with the identity $\psi\phi = 0$, give $\beta\alpha(y) = y$ and $\alpha\beta(x, z) = (x, z)$, so α is an isomorphism with inverse β. We also have $\alpha\phi(x) = (\rho\phi(x), \psi\phi(x)) = (x, 0) = j(x)$ and $q\alpha(y) = q(\rho(y), \psi(y)) = \psi(y)$, so the diagram commutes as claimed.

□

Proposition 12.11. [prop-free-projective]

Suppose we have homomorphisms $F \xrightarrow{\alpha} N \xleftarrow{\beta} M$ where β is surjective and F is a free module. Then there is a homomorphism $\gamma\colon F \to M$ with $\beta\gamma = \alpha$.

Proof. By hypothesis F is isomorphic to $\mathrm{Free}_I(A)$ for some set I, and it will be harmless to assume that F is actually equal to $\mathrm{Free}_I(A)$. Proposition 11.12 then tells us that $\alpha = \phi_n$ for some map $n\colon I \to N$. As β is surjective, we can choose an element $m(i) \in M$ with $\beta(m(i)) = n(i)$ for all i. We can then put $\gamma = \phi_m\colon F \to M$, and it is straightforward to check that $\beta\gamma(e_i) = \beta(m(i)) = n(i) = \alpha(e_i)$ for all i, so $\beta\gamma = \alpha$. □

Corollary 12.12. [cor-split-projective]

Suppose we have a short exact sequence

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0$$

in which N is a free module. Then the sequence is split.

Proof. Apply the proposition to the maps $N \xrightarrow{1} N \xleftarrow{\psi} M$. This gives a homomorphism $\sigma\colon N \to M$ with $\psi\sigma = 1_N$, or in other words a right splitting. Using Proposition 12.10 we can extend this to a full splitting, so the sequence is split. □

Proposition 12.13. [prop-fractions-exact]

Let U be a multiplicative set in A, and let $L \xrightarrow{\phi} M \xrightarrow{\psi} N$ be an exact sequence of A-modules. Then the sequence

$$L[U^{-1}] \xrightarrow{\phi[U^{-1}]} M[U^{-1}] \xrightarrow{\psi[U^{-1}]} N[U^{-1}]$$

is also exact.

Proof. For any element $x/u \in L[U^{-1}]$ we have

$$\psi[U^{-1}](\psi[U^{-1}](x/u)) = \psi[U^{-1}](\phi(x)/u) = \psi(\phi(x))/u.$$

As the original sequence is exact we see that $\psi\phi = 0$, so the above element is zero. This proves that $\mathrm{image}(\phi[U^{-1}]) \subseteq \ker(\psi[U^{-1}])$.

Conversely, suppose we have an element $y/v \in \ker(\psi[U^{-1}])$. This means that $\psi(y)/v = 0$ in $N[U^{-1}]$, so $\psi(y)w = 0$ in N for some element $w \in U$. This in turn means that $yw \in \ker(\psi) = \text{image}(\phi)$, so we can choose $x \in L$ with $\phi(x) = yw$. It follows that $y/v = \phi[U^{-1}](x/(vw)) \in \text{image}(\phi[U^{-1}])$ as required. □

Remark 12.14. [`rem-fractions-exact`]

If we have a longer exact sequence

$$M_0 \to M_1 \to \cdots \to M_n$$

we can apply the above to all the subsequences $M_i \to M_{i+1} \to M_{i+2}$ and deduce that the resulting sequence

$$M_0[U^{-1}] \to M_1[U^{-1}] \to \cdots \to M_n[U^{-1}]$$

is also exact.

Remark 12.15. Suppose we have a module M and a submodule L. This gives an exact sequence

$$0 \to L \to M \to M/L \to 0$$

and thus an exact sequence

$$0 \to L[U^{-1}] \to M[U^{-1}] \to (L/M)[U^{-1}] \to 0$$

This means that we can regard $L[U^{-1}]$ as a submodule of $M[U^{-1}]$ and that $M[U^{-1}]/L[U^{-1}] = (M/L)[U^{-1}]$. We can use this to give another proof of Proposition 8.16.

13. Simple modules

Definition 13.1. [`defn-simple`]

A module M is *simple* if it is nontrivial, and the only submodules of M are 0 and M itself.

Proposition 13.2. [`prop-simple-max`]

If M is simple then the ideal $P = \text{ann}_A(M) = \{a \in A \mid aM = 0\}$ is maximal, and M is isomorphic to A/P.

Proof. Choose a nontrivial element $m \in M$, and define $\phi\colon A \to M$ by $\phi(a) = am$. The image $\phi(A) = Am$ is a nontrivial submodule of M, so it must be all of M, so ϕ is surjective. It follows that an element $a \in A$ satisfies $am = 0$ iff $aAm = 0$ iff $aM = 0$ iff $a \in P$, so the kernel of ϕ is P. We therefore have an induced isomorphism $\overline{\phi}\colon A/P \to M$. Next, if $a \in A \setminus P$ then $\phi(Aa)$ is a nontrivial submodule of M so it must be all of M, and so must contain $m = \phi(1)$. It follows that there exists $b \in A$ with $\phi(ab) = \phi(1)$ or $\phi(ab-1) = 0$, so $ab-1 \in P$, so a becomes invertible in A/P. This proves that A/P is a field, or equivalently P is maximal. □

Definition 13.3. Let M be an A-module. A *composition series* for M is a chain of submodules

$$0 = M_0 \subset M_1 \subset \cdots \subset M_n = M$$

for some $n \geq 0$, such that M_i/M_{i-1} is simple for $1 \leq i \leq n$. We call n the *length* of the series. We say that a module M has *finite length* if there exists a composition series.

Lemma 13.4. [`lem-ses-series`]

Suppose we have a short exact sequence

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0,$$

and a composition series $(M_i)_{i=0}^q$ for M. Put $L_i = \phi^{-1}(M_i) \subseteq L$ and $N_i = \psi(M_i) \subseteq N$, so $L_{i-1} \subseteq L_i$ and $N_{i-1} \subseteq N_i$. Then there are natural short exact sequences

$$0 \to \frac{L_i}{L_{i-1}} \xrightarrow{\alpha_i} \frac{M_i}{M_{i-1}} \xrightarrow{\beta_i} \frac{N_i}{N_{i-1}} \to 0.$$

Moreover, for each i we have either

(a) *$L_i = L_{i-1}$ and N_i/N_{i-1} is isomorphic to M_i/M_{i-1}; or*
(b) *$N_i = N_{i-1}$ and L_i/L_{i-1} is isomorphic to M_i/M_{i-1}.*

Thus, after eliminating repetitions the submodules L_i give a composition series for L of some length n, and the submodules N_i give a composition series for N of some length m, where $n + m = p$.

Proof. We define $\alpha_i\colon L_i/L_{i-1} \to M_i/M_{i-1}$ by $\alpha_i(x + L_{i-1}) = \phi(x) + M_{i-1}$. This is zero iff $\phi(x) \in M_{i-1}$ iff $x \in L_{i-1}$ iff $x + L_{i-1} = 0$, so α_i is injective.

Next, we define $\beta_i\colon M_i/M_{i-1} \to N_i/N_{i-1}$ by $\beta_i(y + M_{i-1}) = \psi(y) + N_{i-1}$. As N_i is $\psi(M_i)$ by definition, we see that β_i is surjective. As $\psi\phi = 0$ we have $\beta_i\alpha_i = 0$, so $\text{image}(\alpha_i) \subseteq \ker(\beta_i)$.

Now suppose that $y + M_{i-1} \in \ker(\beta_i)$. This means that $\psi(y) \in N_{i-1} = \psi(M_{i-1})$, so we can find $u \in M_{i-1}$ with $\psi(y - u) = 0$. This in turn means that $y - u = \phi(x)$ for some $x \in L$. More precisely, as $\phi(x) = y - u \in M)i$ we have $x \in L_{i-1}$. We also have $\alpha_i(x + L_{i-1}) = \phi(x) + M_{i-1} = y - u + M_{i-1} = y + M_{i-1}$, so $y + M_{i-1} \in \text{image}(\alpha_i)$ as required. We thus have a short exact sequence as claimed.

Moreover, the set $\text{image}(\alpha_i) = \ker(\beta_i)$ is a submodule of the simple module M_i/M_{i-1}, so it must be zero or the whole of M_{i-1}. If it is zero then $\alpha_i = 0$ and β_i is an isomorphism, but if it is all of M_i/M_{i-1} then α_i is an isomorphism and $\beta_i = 0$. Note also that α_i is injective, so if $\alpha_i = 0$ then $L_i/L_{i-1} = 0$ so $L_i = L_{i-1}$. Similarly, β_i is surjective so if $\beta_i = 0$ then $N_i = N_{i-1}$. □

Proposition 13.5. [`prop-unique-length`]

Any two composition series for M have the same length.

Proof. We will prove by induction on n that if M admits a composition of length n then every composition series has length n. This is clear if $n = 0$ (in which case $M = 0$) or if $n = 1$ (in which case M is simple). Now suppose that M has a composition series of length $n > 1$. This means that there exists a simple submodule $S \leq M$ such that M/S admits a composition series of length $n-1$. Now suppose we have another composition series, of length m say. We can apply the lemma to the short exact sequence $S \to M \to M/S$; this gives composition series for S and M/S of length i and j say, where $i + j = m$. As S is simple we must have $i = 1$, so $j = m - 1$. On the other hand, M/S admits a composition series of length $n - 1$, so by the induction hypothesis we must have $m - 1 = n - 1$, so $m = n$ as required. □

Definition 13.6. If M has finite length, we define $\text{len}(M)$ to be the length of any composition series, which is well-defined by Proposition 13.5.

Proposition 13.7. [`prop-len-additive`]

Suppose we have a short exact sequence

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0.$$

Then M has finite length iff both L and N have finite length, and if so, we have $\text{len}(M) = \text{len}(L) + \text{len}(N)$.

Proof. First, if we have composition series $0 = L_0 \subset \cdots \subset L_n = L$ and $0 = N_0 \subset \cdots \subset N_m = N$ then we can put

$$M_i = \begin{cases} \phi(L_i) & \text{if } 0 < i < n \\ \phi(L_n) = \text{image}(\phi) = \ker(\psi) = \psi^{-1}(N_0) & \text{if } i = n \\ \psi^{-1}(N_{i-n}) & \text{if } n < i \leq n + m. \end{cases}$$

For $0 < i \leq n$ we find that ϕ gives an isomorphism $L_i/L_{i-1} \to M_i/M_{i-1}$, and for $n < i \leq n + m$ we find that ψ gives an isomorphism $M_i/M_{i-1} \to N_{i-n}/N_{i-n-1}$. Thus, all quotients M_i/M_{i-1} are simple, and we have a composition series of length $n + m$ as claimed. The converse follows easily from Lemma 13.4. □

Corollary 13.8. [`cor-len-additive`]

If M has finite length, then every submodule $L \subseteq M$ and every quotient module M/L also have finite length, with $\text{len}(M) = \text{len}(L) + \text{len}(M/L)$. *Moreover, if P and Q have finite length then so does $P \oplus Q$, and* $\text{len}(P \oplus Q) = \text{len}(P) + \text{len}(Q)$ □.

Remark 13.9. For a more refined invariant, we can fix a maximal ideal $P \in \max(A)$ and define $\text{mult}_P(M)$ to be the number of composition factors M_i/M_{i-1} that are isomorphic to A/P. Arguments very similar to Propositions 13.5 and 13.7 show that $\text{mult}_P(M)$ is independent of the choice of composition series, and the $\text{mult}_P(M) = \text{mult}_P(L) + \text{mult}_P(N)$ whenever we have a short exact sequence $L \to M \to N$.

14. Tensor products

The *tensor product* is a construction that combines two A-modules M and N to form a new A-module denoted by $M \otimes_A N$ (or just $M \otimes N$, if there is no danger of confusion). The most basic examples are that $A^m \otimes A^n \simeq A^{mn}$ and $A/I \otimes A/J \simeq A/(I+J)$.

Construction 14.1. [`cons-tensor`]
Let M and N be modules over a ring A. Let $M \square N$ denote the module $\text{Free}_{M\times N}(A) = \text{Map}_0(M \times N, A)$. Given $m \in M$ and $n \in N$ we write $m \square n$ for the basis element $e_{(m,n)} \in M \square N$. Next, let $G(M,N)$ denote the subset of $M \square N$ consisting of all elements of the following forms:

- $(m+m')\square n - m\square n - m'\square n$ (with $m, m' \in M$ and $n \in N$)
- $m\square(n+n') - m\square n - m\square n'$ (with $m \in M$ and $n, n' \in N$)
- $(am)\square n - a(m\square n)$ (with $a \in A$ and $m \in M$ and $n \in N$)
- $m\square(an) - a(m\square n)$ (with $a \in A$ and $m \in M$ and $n \in N$).

Put $M \otimes N = (M\square N)/\operatorname{span}_A(G(M,N))$, and let $m \otimes n$ denote the coset corresponding to $m\square n$. Thus $m \otimes n \in M \otimes N$, and by construction we have

$$\begin{aligned} (m+m')\otimes n &= m\otimes n + m'\otimes n & \qquad m\otimes(n+n') &= m\otimes n + m\otimes n' \\ (am)\otimes n &= a.(m\otimes n) & m\otimes(an) &= a.(m\otimes n). \end{aligned}$$

Definition 14.2. [`defn-bilinear`]
Let M, N and T be A-modules, and let $\phi\colon M\times N \to T$ be a function. We say that ϕ is *bilinear* if we have

$$\begin{aligned} \phi(m+m',n) &= \phi(m,n)+\phi(m',n) & \qquad \phi(m,n+n') &= \phi(m,n)+\phi(m,n') \\ \phi(am,n) &= a\,\phi(m,n) & \phi(m,an) &= a\,\phi(m,n) \end{aligned}$$

for all $m, m' \in M$ and $n, n' \in N$ and $a \in A$.

Note that we have a map $\mu\colon M\times N \to M\otimes N$ given by $\mu(m,n) = m\otimes n$, and this is bilinear by construction. Thus, for any A-module homomorphism $\overline{\phi}\colon M\otimes N \to T$, we have a bilinear map $\phi\colon M\times N \to T$ given by $\phi(m,n) = \overline{\phi}(m\otimes n)$, or equivalently $\phi = \overline{\phi}\circ\mu$. In fact, this construction gives all possible bilinear maps out of $M\times N$:

Proposition 14.3. [`prop-tensor-universal`]
If $\phi\colon M\times N \to Z$ is A-bilinear, then there is a unique A-linear map $\overline{\phi}\colon M\otimes N \to Z$ such that $\phi(m,n) = \overline{\phi}(m\otimes n)$ for all m and n.

Proof. Proposition 11.12 shows that there is a unique homomorphism

$$\psi\colon M\square N = \text{Free}_{M\times N}(A) \to Z$$

with $\psi(m\square n) = \phi(m,n)$ for all m and n. Using the bilinearity of ϕ, we see that

$$\psi((m+m')\square n - m\square n - m'\square n) = \psi(m+m',n) - \psi(m,n) - \psi(m',n) = 0.$$

Similarly, all other elements of $G(M,N)$ lie in the kernel of ψ, so $\operatorname{span}_A(G(M,N)) \subseteq \ker(\psi)$, so Proposition 11.41 gives a unique homomorphism

$$\overline{\phi}\colon M\otimes N = (M\square N)/\operatorname{span}_A(G(M,N)) \to Z$$

with $\overline{\phi}(m\otimes n) = \psi(m\square n) = \phi(m,n)$ as required. □

Example 14.4. [`eg-tensor-universal`]
Define $\phi\colon \mathbb{R}^3\times\mathbb{R}^3 \to \mathbb{R}^3$ by $\phi(u,v) = u\times v$ (the traditional cross product of 3-dimensional vectors). This is clearly bilinear, so Proposition 14.3 guarantees that there is a unique $\mathbb{R}$-module homomorphism $\overline{\phi}\colon \mathbb{R}^3\otimes\mathbb{R}^3 \to \mathbb{R}^3$ satisfying $\overline{\phi}(u\otimes v) = u\times v$ for all u and v. Rather than spelling this out in full detail, we will usually just say that we define $\overline{\phi}\colon \mathbb{R}^3\otimes\mathbb{R}^3 \to \mathbb{R}^3$ by $\overline{\phi}(u\otimes v) = u\times v$, leaving to the reader the definition of ϕ, the check of bilinearity, and the appeal to Proposition 14.3.

Example 14.5. [`eg-tensor-functor`]

Suppose we have modules M, M', N and N', and A-linear maps $\alpha\colon M \to M'$ and $\beta\colon N \to N'$. We then define $\alpha \otimes \beta\colon M \otimes N \to M' \otimes N'$ by

$$(\alpha \otimes \beta)(m \otimes n) = \alpha(m) \otimes \alpha(n).$$

(As discussed above, we are implicitly applying Proposition 14.3 to the bilinear map $(m, n) \mapsto \alpha(m) \otimes \beta(n)$.) It is easy to see that this fits into the following commutative diagram:

$$\begin{tikzcd}
M \otimes N \arrow[r, "\alpha\otimes 1"] \arrow[d, "1\otimes\beta"'] \arrow[dr, "\alpha\otimes\beta"] & M' \otimes N \arrow[d, "1\otimes\beta"] \\
M \otimes N' \arrow[r, "\alpha\otimes 1"'] & M' \otimes N'.
\end{tikzcd}$$

It is convenient to rephrase Proposition 14.3 in terms of the following definition:

Definition 14.6. Let $\nu\colon M \times N \to T$ be a bilinear map. We say that ν is *universal* if for every other bilinear map $\phi\colon M \times N \to Z$, there is a unique A-module homomorphism $\overline{\phi}\colon T \to Z$ with $\phi = \overline{\phi} \circ \nu$.

The proposition says that the bilinear map $\mu\colon M \times N \to M \otimes N$ is universal. In fact, this characterises $M \otimes N$ up to isomorphism:

Corollary 14.7. [`cor-tensor-unique`]

Let $\nu\colon M \times N \to T$ be a bilinear map, giving a homomorphism $\overline{\nu}\colon M \otimes N \to T$ with $\nu = \overline{\nu}\mu$. Then $\overline{\nu}$ is an isomorphism iff ν is universal.

Proof. Suppose that ν is universal. There is then a unique homomorphism $\overline{\mu}\colon T \to M \otimes N$ with $\mu = \overline{\mu}\nu = \overline{\mu}\overline{\nu}\mu$. Now $\overline{\mu}\overline{\nu}$ and $1_{M\otimes N}$ are both homomorphisms out of $M \otimes N$ that become the same when composed with μ. By the uniqueness clause in Proposition 14.3, they must be the same. A similar argument (using the universality of ν) shows that $\overline{\nu}\overline{\mu} = 1_T$, so $\overline{\mu}$ is the required inverse for $\overline{\nu}$. □

Example 14.8. [`eg-free-tensor`]

We claim that $\mathrm{Free}_I(A) \otimes \mathrm{Free}_J(A)$ is naturally isomorphic to $\mathrm{Free}_{I\times J}(A)$. To prove this, define

$$\nu\colon \mathrm{Free}_I(A) \times \mathrm{Free}_J(A) \to \mathrm{Free}_{I\times J}(A)$$

by $\nu(u, v)(i, j) = u(i)v(j)$. This is clearly bilinear. Consider another bilinear map $\phi\colon \mathrm{Free}_I(A) \times \mathrm{Free}_J(A) \to T$. Define $\phi_0\colon I \times J \to T$ by $\phi_0(i, j) = \phi(e_i, e_j)$, then let $\overline{\phi}\colon \mathrm{Free}_{I\times J}(A) \to T$ be the unique homomorphism satisfying $\overline{\phi}(e_{(i,j)}) = \phi_0(i, j) = \phi(e_i, e_j)$. It is straightforward to check that $\overline{\phi}\nu = \phi$, and that $\overline{\phi}$ is uniquely characterised by this property. This means that ν is universal, so the map

$$\overline{\nu}\colon \mathrm{Free}_I(A) \otimes \mathrm{Free}_J(A) \to \mathrm{Free}_{I\times J}(A)$$

is an isomorphism as required.

Remark 14.9. As a special case of the above example, we have $A^n \otimes A^m \simeq A^{mn}$.

Proposition 14.10. *There are natural isomorphisms*

$$\begin{aligned}
A \otimes M &\simeq M \\
M \otimes N &\simeq N \otimes M \\
L \otimes (M \otimes N) &\simeq (L \otimes M) \otimes N \\
L \otimes (M \oplus N) &\simeq (L \otimes M) \oplus (L \otimes N).
\end{aligned}$$

Proof. First, we define $\alpha\colon A \otimes M \to M$ to be the unique A-module map satisfying $\alpha(a \otimes m) = am$ for all $a \in A$ and $m \in M$. We also define $\beta\colon M \to A \otimes M$ by $\beta(m) = 1 \otimes m$. After recalling that

$$a \otimes m = (a.1) \otimes m = a.(1 \otimes m) = 1 \otimes (am)$$

we see that β is also an A-module homomorphism, and that it is inverse to α.

Next, define homomorphisms

$$M \otimes N \xrightarrow{\tau} N \otimes M \xrightarrow{\sigma} M \otimes N$$

by $\tau(m\otimes n)=n\otimes m$ and $\sigma(n\otimes m)=m\otimes n$. (As usual, these definitions implicitly appeal to Proposition 14.3.) It is clear that τ and σ are inverse to each other, so they are isomorphisms.

Next, for any $l\in L$ we can define a bilinear map

$$\lambda_0(l)\colon M\times N\to (L\otimes M)\otimes N$$

by $\lambda_0(l)(m,n)=(l\otimes m)\otimes n$. It follows that there is a unique linear map $\lambda(l)\colon M\otimes N\to (L\otimes M)\otimes N$ satisfying $\lambda(l)(m\otimes n)=(l\otimes m)\otimes n$. We next clam that $\lambda(l+l')=\lambda(l)+\lambda(l')$ in $\mathrm{Hom}(M\otimes N,(L\otimes M)\otimes N)$. Equivalently, we claim that $\lambda(l+l')(x)=\lambda(l)(x)+\lambda(l')(x)$ for all $x\in M\otimes N$. This is clear by construction if x has the form $m\otimes n$ for some m and n, and $M\otimes N$ is generated by terms of that form, so the claim holds in general. Similarly, we have $\lambda(al)=a\lambda(l)$ for all $a\in A$ and $l\in L$. We can now define a map

$$\gamma_0\colon L\times(M\otimes N)\to (L\otimes M)\otimes N$$

by $\gamma_0(l,x)=\lambda(l)(x)$, so $\gamma_0(l,m\otimes n)=(l\otimes m)\otimes n$. Using the above properties of λ, we see that γ_0 is bilinear, so there is a unique homomorphism

$$\gamma\colon L\otimes(M\otimes N)\to (L\otimes M)\otimes N$$

satisfying $\gamma(l\otimes x)=\gamma_0(l,x)$, and in particular $\gamma(l\otimes(m\otimes n))=(l\otimes m)\otimes n$. A similar argument constructs a homomorphism $\delta\colon (L\otimes M)\otimes N\to L\otimes(M\otimes N)$ satisfying $\delta((l\otimes m)\otimes n)=l\otimes(m\otimes n)$. Now $\delta\gamma(x)=x$ whenever x has the form $l\otimes(m\otimes n)$, and it is not hard to see that terms of that form generate $L\otimes(M\otimes N)$, so $\delta\gamma$ is the identity. A similar argument shows that $\gamma\delta$ is also the identity, completing the proof that $L\otimes(M\otimes N)\simeq(L\otimes M)\otimes N$.

Finally, we can define a map

$$\zeta\colon L\otimes(M\oplus N)\to(L\otimes M)\oplus(L\otimes N)$$

by $\zeta(l\otimes(m,n))=(l\otimes m,l\otimes n)$. We can also define maps

$$L\otimes M\xrightarrow{\phi}L\otimes(M\oplus N)\xleftarrow{\psi}L\otimes N$$

by $\phi(l\otimes m)=l\otimes(m,0)$ and $\psi(l\otimes n)=l\otimes(0,n)$. We can then combine these to define

$$\xi\colon(L\otimes M)\oplus(L\otimes N)\to L\otimes(M\oplus N)$$

by $\xi(x,y)=\phi(x)+\psi(y)$. It is straightforward to check that ξ is inverse to ζ. □

The above proof that $L\otimes(M\otimes N)\simeq(L\otimes M)\otimes N$ can be reorganised and generalised as follows. We say that a map

$$\phi\colon L\times M\times N\to Z$$

is *trilinear* if

- For fixed $m\in M$ and $n\in N$, the map $l\mapsto\phi(l,m,n)$ gives an A-module homomorphism $L\to Z$.
- For fixed $l\in L$ and $n\in N$, the map $m\mapsto\phi(l,m,n)$ gives an A-module homomorphism $M\to Z$.
- For fixed $l\in L$ and $m\in M$, the map $n\mapsto\phi(l,m,n)$ gives an A-module homomorphism $N\to Z$.

We can generalise this in an obvious way to define the notion of an n-linear map from $\prod_{i=0}^{n-1}M_i\to Z$, for any list of modules M_i. Suppose we have a trilinear map ϕ as above. If we fix l then we have a bilinear map $\phi_0(l)\colon M\times N\to Z$, defined by $\phi_0(l)(m,n)=\phi(l,m,n)$. It follows that there is a module homomorphism $\phi_1(l)\colon M\otimes N\to Z$ satisfying $\phi_1(l)(m\otimes n)=\phi(l,m,n)$. Using this, we define $\phi_2\colon L\times(M\otimes N)\to Z$ by $\phi_2(l,x)=\phi_1(l)(x)$. One can check that ϕ_2 is bilinear, so it gives rise to a homomorphism $L\otimes(M\otimes N)\to Z$. A slight elaboration shows that this gives a bijection between trilinear maps $L\times M\times N\to Z$, and homomorphisms $L\otimes(M\otimes N)\to Z$. A similar procedure gives a bijection between trilinear maps and homomorphisms $(L\otimes M)\otimes N\to Z$. By considering $Z_0=(L\otimes M)\otimes N$ and $Z_1=L\otimes(M\otimes N)$ we can produce maps $Z_0\to Z_1\to Z_0$ and check that they are inverse to each other. The whole argument can be extended inductively to show that n-linear maps from $M_0\times\ldots\times M_{n-1}$ to Z biject with A-module homomorphisms $M_0\otimes\cdots\otimes M_{n-1}\to Z$, where the tensor product can be bracketed in any way we choose.

Proposition 14.11. *For any ideal I and module M there is a natural isomorphism $A/I\otimes_A M\simeq M/IM$.*

Proof. We define $\nu\colon A/I \times M \to M/IM$ by $\nu(a+I,m) = am + IM$. This is easily seen to be well-defined and bilinear, so it gives rise to a homomorphism $\overline{\nu}\colon A/I \otimes M \to M/IM$. In the opposite direction, we can define $\phi\colon M \to A/I \otimes M$ by $\phi(m) = (1+I) \otimes m$. If $a \in I$ and $m \in M$ we see that

$$\phi(am) = (1+I) \otimes am = a.((1+I) \otimes m) = (a+I) \otimes m = 0 \otimes m = 0.$$

It follows that $\phi(IM) = 0$, so there is an induced homomorphism $\overline{\phi}\colon M/IM \to A/I \otimes M$ with $\overline{\phi}(m+IM) = (1+I) \otimes m$. It is easy to see that $\overline{\phi}$ is inverse to $\overline{\nu}$. □

Remark 14.12. [rem-quot-tensor]
One can check that $I.(A/J) = (I+J)/J$, so as a special case we have $A/I \otimes A/J \simeq A/(I+J)$.

Proposition 14.13. *For any multiplicative set U and any module M there is a natural isomorphism $A[U^{-1}] \otimes_A M \simeq M[U^{-1}]$.*

Proof. We define $\nu\colon A[U^{-1}] \times M \to M[U^{-1}]$ by $\nu(a/u,m) = (am)/u$. This is easily seen to be well-defined and bilinear, so it gives rise to a homomorphism $\overline{\nu}\colon A[U^{-1}] \otimes M \to M[U^{-1}]$. In the opposite direction, we can define $\phi\colon M[U^{-1}] \to A[U^{-1}] \otimes M$ by $\phi(m/u) = 1/u \otimes m$. To see that this is well-defined, suppose that $m/u = m'/u'$, so there exists $v \in U$ with $u'vm = uvm'$. It follows that

$$\frac{1}{u} \otimes m = \frac{u'v}{uu'v} \otimes m = u'v.\left(\frac{1}{uu'v} \otimes m\right) = \frac{1}{uu'v} \otimes u'vm = \frac{1}{uu'v} \otimes uvm' = \frac{1}{u'} \otimes m'$$

as required. It is now easy to check that ϕ is inverse to $\overline{\nu}$. □

Proposition 14.14. [prop-tensor-exact]
Suppose we have a module P and an exact sequence

$$L \xrightarrow{\alpha} M \xrightarrow{\beta} N \to 0.$$

Then the resulting sequence

$$P \otimes L \xrightarrow{1\otimes\alpha} P \otimes M \xrightarrow{1\otimes\beta} P \otimes N \to 0$$

is also exact.

Proof. By hypothesis we have $\beta\alpha = 0$, and it follows easily that $(1 \otimes \beta)(1 \otimes \alpha) = 0$. This proves that $\ker(1 \otimes \beta) \supseteq \operatorname{image}(1 \otimes \alpha)$. Now put

$$Q = (P \otimes M)/\operatorname{image}(1 \otimes \alpha),$$

and observe that there is an induced homomorphism $\phi\colon Q \to P \otimes N$ given by $\phi(x + \operatorname{image}(1\otimes\alpha)) = (1\otimes\beta)(x)$. We claim that ϕ is actually an isomorphism. To see this, suppose that $p \in P$ and $n \in N$. We can choose $m \in M$ with $\beta(m) = n$, and then put

$$\psi_0(p,n) = p \otimes m + \operatorname{image}(1 \otimes \alpha) \in Q.$$

This is not obviously well-defined, because m is not unique. However, as $\ker(\beta) = \operatorname{image}(\alpha)$, any other choice will have the form $m' = m + \alpha(l)$ for some l, giving $p \otimes m' = p \otimes m + (1 \otimes \alpha)(p \otimes l)$, and this shows that ψ_0 is well-defined after all. It is easily seen to be bilinear, so there is an induced map $\psi\colon P \otimes N \to Q$. We find that ψ is inverse to ϕ. From this in turn it follows that $1 \otimes \beta$ is surjective, with kernel equal to the image of $1 \otimes \alpha$. In other words, the sequence is exact as claimed. □

Remark 14.15. If $\alpha\colon L \to M$ is injective, it does not follow that the map $1 \otimes \alpha\colon P \otimes L \to P \otimes M$ is injective. For example, we can take

$$A = \mathbb{Z} \qquad P = \mathbb{Z}/2 \qquad L = \mathbb{Z} \qquad M = \mathbb{Q},$$

and let $\alpha\colon \mathbb{Z} \to \mathbb{Q}$ be the inclusion map. Then $P \otimes L = \mathbb{Z}/2$ and $P \otimes M = 0$, so $1 \otimes \alpha$ cannot be injective.

The following terminology is often used: Proposition 14.14 says that the functor $P \otimes (-)$ is right exact, but the above example shows that it is not left exact, and therefore not exact.

So far we have discussed tensor products of modules; we now consider tensor products of algebras (as in Definition 11.3).

Remark 14.16. [defn-algebra-mult]

Suppose that B is an A-algebra, with structure map $\phi\colon A \to B$. We can use this to consider B as an A-module, and thus form the tensor product $B \otimes_A B$. We can define an A-bilinear map $\mu_0\colon B \times B \to B$ by $\mu_0(b, b') = bb'$. Proposition 14.3 therefore gives us an A-module homomorphism $\mu\colon B \otimes_A B \to B$ satisfying $\mu(b \otimes b') = bb'$. Either μ_0 or μ may be referred to as the *multiplication map* for B.

Proposition 14.17. [prop-algebra-tensor]

If B and C are A-algebras, then $B \otimes_A C$ also has a structure as an A-algebra, such that for all $b, b' \in B$ and $c, c' \in C$ we have

$$(b \otimes c)(b' \otimes c') = (bb') \otimes (cc').$$

Proof. We can define a 4-linear map

$$\phi\colon B \times C \times B \times C \to B \otimes C$$

by $\phi(b, c, b', c') = bb' \otimes cc'$. As discussed above, this gives rise to an A-module homomorphism

$$\phi'\colon (B \otimes C) \otimes (B \otimes C) \to B \otimes C$$

satisfying $\phi'(b \otimes c \otimes b' \otimes c') = bb' \otimes cc'$. For $u, v \in B \otimes C$ we then define $uv = \phi'(u \otimes v)$. Alternatively, we can define $\tau\colon C \otimes B \to B \otimes C$ to be the unique homomorphism with $\tau(c \otimes b) = b \otimes c$, then we can consider the composite

$$B \otimes C \otimes B \otimes C \xrightarrow{1 \otimes \tau \otimes 1} B \otimes B \otimes C \otimes C \xrightarrow{\mu_B \otimes \mu_C} B \otimes C.$$

It is not hard to see that this is the same as ϕ'.

We have now defined a multiplication rule on $B \otimes C$, but we still need to check that it satisfies the axioms. For example, we must show that $(uv)w = u(vw)$ for all $u, v, w \in B \otimes C$. This is clear from the definitions (and the associativity of B and C) in the case where u, v and w all have the form $b \otimes c$ for some $b \in B$ and $c \in C$. Every element of $B \otimes C$ can be expressed as an A-linear combination of terms of that form, so the general case follows easily. □

Remark 14.18. [rem-ring-pushout]

We can define ring maps $B \xrightarrow{\lambda} B \otimes_A C \xleftarrow{\rho} C$ by $\lambda(b) = b \otimes 1$ and $\rho(c) = 1 \otimes c$. If $\beta\colon A \to B$ and $\gamma\colon A \to C$ are the given structure maps, we find that

$$\lambda\beta(a) = \beta(a) \otimes 1 = a.(1 \otimes 1) = 1 \otimes \gamma(a) = \rho\gamma(a),$$

so the following diagram commutes:

$$\begin{tikzcd}
A \arrow[r, "\beta"] \arrow[d, "\gamma"'] & B \arrow[d, "\lambda"] \\
C \arrow[r, "\rho"'] & B \otimes_A C.
\end{tikzcd}$$

Now suppose we have another commutative square of rings:

$$\begin{tikzcd}
A \arrow[r, "\beta"] \arrow[d, "\gamma"'] & B \arrow[d, "\zeta"] \\
C \arrow[r, "\xi"'] & Z.
\end{tikzcd}$$

Using the homomorphism $\zeta\beta = \xi\gamma\colon A \to Z$ we can regard Z as an A-algebra, and thus as an A-module. We can define a module map $\phi\colon B \otimes C \to Z$ by $\phi(b \otimes c) = \zeta(b)\xi(c)$. It is not hard to see that this is actually a ring map, and that it is the unique ring map that satisfies $\phi\lambda = \zeta$ and $\phi\rho = \xi$. In the language of category theory, this means that our first square is actually a pushout in the category of rings.

15. Modules over fields

Let K be a field. Modules over K are traditionally called *vector spaces*. Although we expect that most readers will be familiar with the theory of vector spaces, we will give a rapid treatment here to explain the relationship with our more general theory of modules.

Proposition 15.1. [prop-basis-test]
Let M be a K-module, and let S be a subset of M, giving a homomorphism $\phi_S\colon \mathrm{Free}_S(K) \to M$. Then the following are equivalent:

(a) *S is maximal among linearly independent subsets of M.*
(b) *S is minimal among sets that span M.*
(c) *S is a basis for M.*

Proof.

(a)⇒(c) Suppose that S is maximal among linearly independent sets, and consider an element $x \in M$. If $x \in S$ then $x = \phi_S(e_x) \in \mathrm{image}(\phi_S)$. Suppose instead that $x \notin S$, so the set $T = S \cup \{x\}$ is strictly larger than S and so cannot be linearly independent. This means that there is a nonzero element $m \in \mathrm{Free}_T(K)$ with $\phi_T(m) = 0$. If $m(x)$ were zero we would have $\phi_S(m|_S) = \phi_T(m) = 0$ and so $m|_S = 0$ because S is linearly independent, but that would give $m = 0$, contrary to assumption. We therefore have $m(x) \neq 0$, and it follows that the element $n = -m(x)^{-1}m|_S \in \mathrm{Free}_S(K)$ has $\phi_S(n) = x$, so $x \in \mathrm{image}(\phi_S)$ again. This proves that ϕ_S is surjective as well as injective, so it is a basis.

(b)⇒(c) Suppose that S is minimal among sets that span M. We will assume that we have a nonzero element $m \in \mathrm{Free}_S(M)$ with $\phi_S(m) = 0$, and derive a contradiction. As $m \neq 0$ we can choose $x \in S$ with $m(x) \neq 0$ and put $T = S \setminus \{x\}$. Put $n = -m(x)^{-1}m|_T$, and note that $\phi_S(n) = x$. Thus, for any $u \in \mathrm{Free}_S(K)$ we have $\phi_S(u) = \phi_T(u|T+u(x)n)$, so $\mathrm{image}(\phi_T) = \mathrm{image}(\phi_S) = M$, so T is a spanning set. This contradicts the assumed minimality of S.

(c)⇒(a) Suppose that S is a basis, and let T be a strictly larger set. Choose $x \in T \setminus S$. As S is a basis, there exists $m \in \mathrm{Free}_S(K)$ with $\phi_S(m) = x$. Let $n \in \mathrm{Free}_T(K)$ be given by

$$n(z) = \begin{cases} m(z) & \text{if } z \in S \\ -1 & \text{if } z = x \\ 0 & \text{otherwise.} \end{cases}$$

Then n is a nontrivial element of $\ker(\phi_T)$, so T is linearly dependent. This proves that S is maximal among linearly independent sets.

(c)⇒(b) Suppose again that S is a basis, and let U be a proper subset of S. Choose $x \in S \setminus U$. We claim that $x \notin \mathrm{image}(\phi_U)$. Indeed, if $x = \phi_U(m)$ then we can define $n \in \mathrm{Free}_S(K)$ by

$$n(z) = \begin{cases} m(z) & \text{if } z \in U \\ -1 & \text{if } z = x \\ 0 & \text{otherwise.} \end{cases}$$

We find that n is a nontrivial element in $\ker(\phi_S)$, contradicting the assumption that S is a basis. We conclude that $x \notin \mathrm{image}(\phi_U)$, so ϕ_U is not surjective, so S is minimal among spanning sets.

□

Corollary 15.2. *Every K-module has a basis, and so is free.*

Proof. Let M be a K-module, and let $\mathcal{L}$ be the set of all linearly independent subsets of M. Note that $\emptyset \in \mathcal{L}$, so $\mathcal{L} \neq \emptyset$. We will apply Zorn's Lemma to $\mathcal{L}$. Let $\mathcal{C}$ be a chain in $\mathcal{L}$, or in other words a subset of $\mathcal{L}$ such that for all $C, D \in \mathcal{C}$ we have either $C \subseteq D$ or $D \subseteq C$. Let S be the union of all the sets in $\mathcal{C}$. We claim that S is linearly independent (or equivalently, $S \in \mathcal{L}$). To see this, consider a nonzero $m \in \mathrm{Free}_S(K)$. For each element x in the finite set $\mathrm{supp}(m)$, we have $x \in S$, so we can choose $C_x \in \mathcal{C}$ such that $x \in C_x$. The chain condition implies that the family $\{C_x \mid x \in \mathrm{supp}(m)\}$ is linearly ordered by inclusion, so there is an element $z \in \mathrm{supp}(m)$ such that $C_x \subseteq C_z$ for all x, and therefore $\mathrm{supp}(m) \subseteq C_z$. As $C_z \in \mathcal{L}$ we deduce that the element $\phi_S(m) = \phi_{C_z}(m|_{C_z})$ is nonzero. This implies that S is linearly independent as claimed. This verifies the key condition in Zorn's Lemma, so $\mathcal{L}$ has a maximal element, which is a basis by Proposition 15.1. □

Corollary 15.3. *Every short exact sequence of K-modules is split.*

Proof. This now follows from Corollary 12.12. □

Proposition 15.4. [prop-field-simple]
Every simple K-module is isomorphic to K.

Proof. Proposition 13.2 implies that every simple module is isomorphic to K/P for some maximal ideal P, but as K is a field, the only maximal ideal is 0. □

Proposition 15.5. [prop-fin-dim]
For a K-module M, the following are equivalent.

(a) *M has a finite basis*
(b) *M is finitely generated*
(c) *M has finite length.*

Moreover, if these conditions hold then the rank of M is the same as the length.

Proof. It is clear that (a) implies (b). Moreover, if (b) holds then $M \simeq K^n/L$ for some integer n and some submodule L, but K^n clearly has finite length, so M has finite length by Corollary 13.8.

Now suppose we assume that M has finite length, say $\text{len}(M) = n$. This implies that there is a submodule $L \subset M$ such that $\text{len}(L) = n-1$ and M/L is simple, which means that $M/L \simeq K$. As $M/L \simeq K$ is free the short exact sequence $L \to M \to M/L$ must split, so $M \simeq L \oplus K$. By induction on n we may assume that L is free of rank $n-1$, and it follows that M is free of rank n. All remaining claims are now clear. □

Remark 15.6. In this context it is traditional to say that M is *finite-dimensional* if it satisfies the above conditions, and to define the *dimension* of M to be the rank or length. Note that Corollary 13.8 shows that submodules, quotients and direct sums of finite-dimensional modules are finite-dimensional, with $\dim(M) = \dim(L) + \dim(M/L)$ and $\dim(P \oplus Q) = \dim(P) + \dim(Q)$.

16. Principal ideal domains

Definition 16.1. [defn-pid]
Recall that an ideal $I \subseteq A$ is *principal* if it has the form $I = Aa$ for some $a \in A$. A *principal ideal domain* (or PID) is a domain in which every ideal is principal.

The simplest way to prove that a ring is a PID is to use the following notion:

Definition 16.2. [defn-ev]
Let A be a domain. A *euclidean valuation* on A is a function $\nu\colon A \to \mathbb{N}$ with the following properties:

(a) $\nu(a) = 0$ if and only if $a = 0$
(b) Whenever $a, b \in A$ with $a \neq 0$ there are elements $q, r \in A$ with $b = qa + r$ and $\nu(r) < \nu(a)$

(Some other sources handle the case $a = 0$ differently, taking $\nu(0)$ to be $-\infty$ or leaving it undefined. Our convention is very natural for Examples 16.3, 16.4 and 16.7, but less natural for Examples 16.5 and 16.6.)

Example 16.3. [eg-Z-ev]
The map $\nu(n) = |n|$ is a euclidean valuation on $\mathbb{Z}$.

Example 16.4. [eg-field-ev]
If K is a field, then we can define a euclidean valuation on K by $\nu(0) = 0$ and $\nu(a) = 1$ for all $a \neq 0$. In axiom (b) we simply take $q = b/a$ and $r = 0$.

Example 16.5. [eg-poly-ev]
Let K be a field, and define $\nu\colon K[x] \to \mathbb{N}$ by $\nu(f) = \deg(f) + 1$ when $f \neq 0$, and $\nu(0) = 0$. It follows easily from Proposition 1.25 that this is a euclidean valuation.

Example 16.6. [eg-Zpl-ev]

Let p be a prime number. Recall that $\mathbb{Z}_{(p)}$ is the set of rational numbers of the form $a = x/u$, where $x, u \in \mathbb{Z}$ and u is not divisible by p. If $x = 0$ we define $\nu(x/u) = 0$. Otherwise, we let d be the largest integer such that x is divisible by p^d, and we put $\nu(x/u) = d+1$. We claim that this is a euclidean valuation. Indeed, suppose we have $a, b \in \mathbb{Z}_{(p)}$ with $a \neq 0$. If $\nu(b) < \nu(a)$ then we just take $q = 0$ and $r = b$. On the other hand, if $\nu(b) \geq \nu(a)$ we find that $b/a \in \mathbb{Z}_{(p)}$ and we just take $q = b/a$ and $r = 0$. Either way we have $b = qa + r$ with $\nu(r) < \nu(a)$ as required.

Example 16.7. [eg-C-ev]

Let A be a subring of $\mathbb{C}$ such that

(a) For all $a \in A$ we have $|a|^2 \in \mathbb{N}$
(b) For all $z \in \mathbb{C}$ there exists $q \in A$ with $|z - q| < 1$.

We then claim that the map $\nu(a) = |a|^2$ is a euclidean valuation. Indeed, if $a, b \in A$ with $a \neq 0$ then we can choose $q \in A$ with $|b/a - q| < 1$, then put $r = b - aq = a(b/a - q) \in A$ so $b = aq + r$ and $|r|^2 < |a|^2$ as required.

The most obvious example of this type is the ring $\mathbb{Z}[i] = \{a + ib \mid a, b \in \mathbb{Z}\}$ of gaussian integers. Given $z = x+iy \in \mathbb{C}$ we can choose integers n, m with $|n-x| \leq 1/2$ and $|m-y| \leq 1/2$. The element $q = n+im \in A$ then has $|z - q| \leq 1/\sqrt{2} < 1$ as required. For another example, we can take $\omega = e^{2\pi i/3} = (i\sqrt{3} - 1)/2$, so $\omega^{-1} = \overline{\omega} = \omega^2 = -1 - \omega$, and put $A = \{n + m\omega \mid n, m \in \mathbb{Z}\}$. One can check that $|n + m\omega|^2 = n^2 + m^2 - nm$, so condition (a) holds. We leave (b) to the reader.

Proposition 16.8. [prop-ev-pid]

If A is a domain with a euclidean valuation $\nu \colon A \to \mathbb{N}$, then it is a PID.

Proof. Consider an ideal $I \subseteq A$. If $I = 0$ then $I = A0$ and so I is principal. Suppose instead that $I \neq 0$, and choose an element $a \in I \setminus \{0\}$ for which $\nu(a)$ is as small as possible. Let b be any other element of I. We then have $b = qa + r$ for some $q, r \in A$ with $\nu(r) < \nu(a)$. Now $r = b - qa$ with $a, b \in I$ so $r \in I$. However, a was chosen to have minimal valuation among nontrivial elements of I, so we must have $r = 0$, and thus $b = qa$. This proves that $I = Aa$, so I is principal as required. □

Corollary 16.9. *The rings $\mathbb{Z}$, $\mathbb{Z}_{(p)}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[e^{2\pi i/3}]$ are all PIDs. Moreover, for every field K, both K and $K[x]$ are PIDs.* □

We next discuss some domains that are not principal ideal domains.

Example 16.10. [eg-not-pid]

Consider a maximal ideal P in a domain A. For any A-module M we can regard M/PM as a module over the field A/P, so we have a well-defined dimension $\dim_{A/P}(M/PM)$. In particular, we can take $M = P$ and consider $\dim_{A/P}(P/P^2)$. If $P = Ap$ then P/P^2 is spanned by the coset $p + P^2$, so $\dim_{A/P}(P/P^2) \leq 1$. Thus, if we can find a maximal ideal P with $\dim_{A/P}(P/P^2) > 1$, then A cannot be a principal ideal domain. There are many examples where this is easy. For example, if $A = \mathbb{C}[x, y]$ and $P = Ax + Ay$ then $A/P = \mathbb{C}$ and P/P^2 has basis $\{x + P^2, y + P^2\}$, so P is not principal.

Example 16.11. [eg-not-pid-dedekind]

For a more subtle example, consider the ring

$$A = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}.$$

Note that $|a + b\sqrt{-5}|^2 = a^2 + 5b^2 \in \mathbb{N}$, so the first condition in Example 16.7 is satisfied, but the example $z = \sqrt{-5}/2$ shows that the second condition is not satisfied.

Now define $\phi \colon A \to \mathbb{Z}/3$ by $\phi(a + b\sqrt{-5}) = a - b + 3\mathbb{Z}$. It is straightforward to check that this is a surjective homomorphism, and that the kernel is the ideal M generated by 3 and $1 + \sqrt{-5}$.

We claim that M is not principal. One can check that $\dim_{A/P}(M/PM) = 1$ for all maximal ideals P, so the method in the previous example is not useful here. Instead, suppose that $M = A\pi$ for some element π. We must then have $3 = \alpha\pi$ and $1 + \sqrt{-5} = \beta\pi$ for some elements $\alpha, \beta \in A$. This gives $9 = |\alpha|^2|\pi|^2$ and $6 = |\beta|^2|\pi|^2$, so $|\pi|^2(|\alpha|^2 - |\beta|^2) = 3$. As $|\pi|^2$ and $|\alpha|^2 - |\beta|^2$ are positive integers, we must have $|\pi|^2 \in \{1, 3\}$. However, it is clear that 3 cannot be represented as $a^2 + 5b^2$, and that 1 can only be represented as $a^2 + 5b^2$ if $(a, b) = (\pm 1, 0)$. We must therefore have $\pi = \pm 1$, which is impossible as $\pm 1 \notin M$.

For the rest of this section, we let A be a PID.

Definition 16.12. We put

$$\operatorname{Idl}(A) = \operatorname{idl}(A) \setminus \{0\} = \{\text{nontrivial ideals}\}.$$

We also put $A^\bullet = A \setminus \{0\}$ and $A^\times = \{\text{invertible elements}\}$.

Lemma 16.13. *If $I \in \operatorname{Idl}(A)$ then $I = Aa$ for some $a \in A^\bullet$. Moreover, for $a, b \in A^\bullet$ we have $Aa = Ab$ iff $b = ua$ for some $u \in A^\times$, so $\operatorname{Idl}(A)$ can be identified with $A^\bullet/A^\times$.*

Proof. The first claim is clear from the definitions. If $Aa = Ab$ then $b \in Aa$ and $a \in Ab$, so there are elements $u, v \in A$ with $b = ua$ and $a = vb$. It follows that $a = uva$, so $a(uv - 1) = 0$. If $a \neq 0$ then (as A is a domain) we have $uv = 1$ so $u \in A^\times$. □

We now discuss various structures on $\operatorname{Idl}(A)$. Recall that in a general ring we defined IJ to be the ideal generated by the set $U = \{xy \mid x \in I,\ y \in J\}$. However, if $I = Aa$ and $J = Ab$ then it is straightforward to check that U itself is an ideal and in fact $IJ = U = Aab$. Note that this is contained in $I \cap J$, so if I and J are nonzero then $I \cap J$ is also nonzero. We now see that when $I, J \in \operatorname{Idl}(A)$ we have ideals $I + J$, $I \cap J$ and IJ, all of which again lie in $\operatorname{Idl}(A)$. We also put $(I : J) = \{a \in A \mid aJ \subseteq I\}$, which contains I and so is again in $\operatorname{Idl}(A)$.

We also consider $\operatorname{Idl}(A)$ as an ordered set using the inclusion relation. The order is related to the above operations, because we have $I \subseteq J$ iff $I \cap J = I$ iff $I + J = J$; this follows directly from the definitions. The following is a little less obvious:

Proposition 16.14. *For $I, K \in \operatorname{Idl}(A)$, the following are equivalent:*

(a) *$K \subseteq I$*
(b) *For some $J \in \operatorname{Idl}(A)$ we have $K = IJ$*
(c) *$K = I.(K : I)$*
(d) *For some $a, b \in A^\bullet$ we have $I = Aa$ and $K = Aab$.*

Proof. First, we can choose $a, c \in A$ such that $I = Aa$ and $K = Ac$.

Note that if $K \subseteq I$ then $Ac \subseteq Aa$ so $c \in Aa$ so $c = ab$ for some b. Thus if we put $J = Ab$ we have $K = IJ$. These arguments show that (a)⇒(d)⇒(b).

Now suppose that $K = IJ$ as in (b). From the definition of $(K : I)$ we have $J \subseteq (K : I)$ and $I.(K : I) \subseteq K$. As $J \subseteq (K : I)$ we also have $K = IJ \subseteq I.(K : I)$, so in fact $I.(K : I) = K$. This proves that (b)⇒(c), and it is clear that (c)⇒(a), which closes the loop. □

Definition 16.15. Suppose we have $a, b \in A^\bullet$. A *GCD system* for the pair (a, b) is a list $(\overline{a}, \overline{b}, x, y, d, m) \in A^6$ such that

$$x\overline{a} + y\overline{b} = 1 \qquad d\overline{a}\overline{b} = m$$
$$d\overline{a} = a \qquad d\overline{b} = b.$$

Proposition 16.16. *For any pair (a, b), there is an associated GCD system $(\overline{a}, \overline{b}, x, y, d, m)$. If we put $I = Aa$ and $J = Ab$, then for any such system we have*

$$\begin{aligned} I + J &= Ad \\ I \cap J &= Am \\ IJ &= Aab \\ (I : J) &= A\overline{a} \\ (J : I) &= A\overline{b}. \end{aligned}$$

Proof. Certainly there exists $d \in A^\bullet$ with $I + J = Ad$. Now $a \in I \subseteq I + J = Ad$, so there is a unique element $\overline{a} \in A$ with $a = d\overline{a}$. Similarly, there is a unique element $\overline{b} \in A$ with $b = d\overline{b}$. Next, as $d \in I + J$ there exist elements $x, y \in A$ with $d = xa + yb$. This in turn gives $d = x\overline{a}d + y\overline{b}d$, and A is a domain so we can cancel the factor of d to conclude that $x\overline{a} + y\overline{b} = 1$. We now put $m = d\overline{a}\overline{b}$, and we find that $(\overline{a}, \overline{b}, x, y, d, m)$ is a GCD system.

Suppose instead that we start with a GCD system $(\overline{a},\overline{b},x,y,d,m)$. We then have $d = d(x\overline{a}+y\overline{b}) = xa+yb$, so $d \in I+J$. On the other hand, any element in $I+J$ has the form $c = ua+vb$ for some $u, v \in A$. This can be rewritten as $c = (u\overline{a}+v\overline{b})d$, so $c \in Ad$. It follows that $I+J = Ad$ as claimed.

Next, as $m = d\overline{a}\overline{b} = a\overline{b} = \overline{a}b$ we see that $Am \subseteq I \cap J$. Conversely, suppose that $u \in I \cap J$, so $u = va = wb$ for some v and w. We then have $u = (x\overline{a}+y\overline{b})u = x\overline{a}u + y\overline{b}u$. In the first term, we use $u = wb$ and $\overline{a}b = m$ to get xwm. In the second term, we use $u = va$ and $a\overline{b} = m$ to get yvm. Combining these gives $u = (xw+yv)m \in Am$. It follows that $I \cap J = Am$.

The relation $IJ = Aab$ is clear.

Next, we have

$$A\overline{a}J = A\overline{a}b = A\overline{a}\overline{b}d = Aa\overline{b} \subseteq Aa = I,$$

so $A\overline{a} \subseteq (I : J)$. In the opposite direction, suppose that $t \in (I : J)$, so $tJ \subseteq I$, or equivalently $tb \in Aa$. This means that $tb = ra$ for some r, or equivalently $t\overline{b} = r\overline{a}$. Now

$$t = (x\overline{a}+y\overline{b})t = \overline{a}(tx+ry) \in A\overline{a}.$$

This shows that $(I : J) = A\overline{a}$ as claimed, and a symmetrical argument shows that $(J : I) = A\overline{b}$ as well. □

Definition 16.17. Let A be an integral domain, and let p be a nonzero element of A. We say that p is a *prime element* if Ap is a prime ideal (so whenever p divides ab, it divides a or b). We say that p is *irreducible* if it is not invertible, but whenever $p = ab$, either a or b is invertible. We say that elements p and q are *associates* if there is an invertible element u such that $q = up$.

Lemma 16.18. [lem-prime-irr]

In any integral domain, prime elements are irreducible.

Proof. Let p be prime. If $p = ab$ then certainly p divides ab, so p divides either a or b. We may assume wlog that p divides a, so $a = px$ for some x. Now $p = ab = pxb$, so $p(1-xb) = 0$, and A is assumed to be an integral domain, so $xb = 1$. This shows that b is invertible, as required. □

Proposition 16.19. [prop-irr-prime]

Let A be a PID, and let P be a nontrivial ideal in A. Then the following conditions are equivalent:

(a) *P is a maximal ideal*
(b) *P is a prime ideal*
(c) *Some generator of P is a prime element*
(d) *Every generator of P is a prime element*
(e) *Some generator of P is an irreducible element*
(f) *Every generator of P is an irreducible element.*

Proof. It is clear from the definitions that (c) and (d) are equivalent, and it is straightforward to check that (e) and (f) are also equivalent. Remark 5.29 shows that (a) implies (b), and (b) is clearly equivalent to (c) and (d), which imply (e) and (f) by Lemma 16.18. The real point is to prove that (e) implies (a). Equivalently, given an irreducible element p, we must show that A/Ap is a field. As p is not invertible by definition, we see that $A/Ap \neq 0$. Any nontrivial element of A/Ap has the form $a + Ap$ for some $a \notin Ap$. Let d be a gcd for a and p, so we can write $a = \overline{a}d$ and $p = \overline{p}d$ and $x\overline{a} + y\overline{p} = 1$ for some $\overline{a}, \overline{p}, x, y \in A$. As p is irreducible, either $\overline{p}$ or d must be invertible. If $\overline{p}$ were invertible we would have $a = \overline{ap}^{-1}p$, which is impossible as p is assumed not to divide a. Thus d must be invertible, and we see that the element $a^* = d^{-1}x$ satisfies $a^*a = 1 - d^{-1}yp = 1 \pmod{p}$, so $a^* + Ap$ is the required inverse for $a + Ap$. □

Lemma 16.20. [lem-inf-divis]

Let $(P_k)_{k\geq 0}$ be a sequence of nonzero prime ideals, and put $I_k = \prod_{i<k} P_i$. Then $\bigcap_k I_k = 0$.

Proof. First, we can choose prime elements p_i such that $P_i = Ap_i$ for all i. We then have $I_k = Am_k$, where $m_k = \prod_{i<k} p_i$. Now put $I_\infty = \bigcap_k I_k$. This is an ideal, and all ideals are principal, so $I_\infty = Aa$ for some $a \in I_\infty$; we must show that $a = 0$. Now $a \in m_k A$ for all k, so there is an element b_k such that $a = m_k b_k$. It follows that

$$m_k(b_k - p_k\, b_{k+1}) = m_k b_k - m_{k+1} b_{k+1} = a - a = 0,$$

and A is an integral domain, so $b_k = p_k b_{k+1}$ for all k. This shows that the ideals $J_k = Ab_k$ satisfy $J_k \subseteq J_{k+1}$, so the union $J_\infty = \bigcup_k J_k$ is again an ideal. As A is a PID, there must be an element b_∞ with $J_\infty = Ab_\infty$. Now $b_\infty \in \bigcup_k J_k$, so $b_\infty \in J_k$ for some $k < \infty$. This gives

$$b_{k+1} \in J_{k+1} \subseteq J_\infty \subseteq J_k = Ab_k = Ap_k b_{k+1},$$

so there exists x with $b_{k+1} = xp_k b_{k+1}$, or in other words $(1 - xp_k)b_{k+1} = 0$. Now $Ap_k = P_k$ is assumed to be prime, so in particular it is not all of A, so $xp_k \neq 1$, so $1 - xp_k \neq 0$. As A is an integral domain, we deduce that $b_{k+1} = 0$. It follows that $a = m_{k+1} b_{k+1} = 0$ as required. □

In particular, if P is any nonzero prime ideal we see that $P^0 = A$ but $\bigcap_k P^k = 0$. This validates the following:

Definition 16.21. [defn-vP]

Let P be a prime ideal in A. For any nonzero element $a \in A$, we let $v_P(a)$ denote the largest natural number n such that $a \in P^n$. For any nonzero ideal I, we let $v_P(I)$ denote the largest natural number n such that $I \subseteq P^n$ (so that $v_P(Aa) = v_P(a)$).

Proposition 16.22. [prop-vP]

(a) *For any element $a \in A^\bullet$ we have $v_P(a) = 0$ iff $a \notin P$.*
(b) *For any elements $a, b \in A^\bullet$ we have $v_P(ab) = v_P(a) + v_P(b)$.*
(c) *For any $a \in A^\bullet$, the set $\{P \mid v_P(a) > 0\}$ is finite.*

Proof.

(a) This is clear from the definitions.
(b) Choose a generator p for P. If $v_P(a) = m$ and $v_P(b) = n$ then $a = p^n x$ and $b = p^m y$ for some elements $x, y \in A \setminus P$. This gives $ab = p^{n+m} xy \in P^{n+m}$. If ab were in P^{n+m+1} we would have $ab = p^{n+m+1} z$ for some z, and we could cancel p^{n+m} to get $xy = pz \in P$. However, this is impossible because P is a prime ideal and $x, y \notin P$. This shows that $v_P(ab) = n + m$.
(c) If not, we can choose a sequence of distinct prime ideals P_i such that $a \in \bigcap_i P_i$. Now choose a generator p_i for each ideal P_i, and put $m_k = \prod_{i<k} p_i$. As the ideals P_i are distinct and maximal, we see that $p_i \notin P_k$ for all $i < k$, and so $m_k \notin P_k$. We claim that there are elements b_k with $a = b_k m_k$ for all k. Indeed, we can take $b_0 = a$. Once we have b_k, we can note that the product $b_k m_k = a$ is divisible by the prime element p_k, but m_k is not divisible by p_k, so b_k must be divisible by p_k, say $b_k = b_{k+1} p_k$. This gives $b_{k+1} m_{k+1} = b_{k+1} p_k m_k = b_k m_k = a$ as required. The claim follows by induction, so we see that $a \in \bigcap_k \prod_{i<k} P_i$. Using Lemma 16.20 we deduce that $a = 0$, contrary to our assumption that $a \in A^\bullet$.

□

Part (c) of the proposition validates the following:

Definition 16.23. For any nonzero element a we put $v^*(a) = \sum_P v_P(a)$. We also put $v^*(I) = \sum_P v_P(I)$, so $v^*(Aa) = v^*(a)$.

Proposition 16.24. *We have $v^*(I) = 0$ iff $I = A$. We also have $v^*(IJ) = v^*(I) + v^*(J)$.*

Proof. It is clear that $v^*(A) = \sum_P v_P(1) = 0$. On the other hand, if I is a nontrivial ideal that is different from A, then I is contained in some maximal ideal P, which is clearly also nontrivial. This gives $v^*(I) \geq v_P(I) > 0$. The formula $v^*(IJ) = v^*(I) + v^*(J)$ follows immediately from Proposition 16.22(b). □

Theorem 16.25. [thm-pid-ufd]

There is a bijection $\mu\colon \bigoplus_P \mathbb{N} \to \mathrm{Idl}(A)$ given by $\mu(m) = \prod_P P^{m(P)}$, with inverse $\mu^{-1}(I)(P) = v_P(I)$. In other words, every nontrivial ideal can be written in an essentially unique way as a product of powers of nontrivial prime ideals.

Proof. We can certainly define a map μ as above. Using Proposition 16.22(c), we can also define a map σ in the opposite direction by $\sigma(I)(P) = v_P(I)$. It is clear that $v_P(P) = 1$ and $v_P(Q) = 0$ for all nontrivial prime ideals $Q \neq P$. Using this together with Proposition 16.22(b), we see that $\sigma\mu = 1$, so μ is injective. We next claim that every nontrivial ideal I lies in the image of μ. If $v^*(I) = 0$ then $I = A = \mu(0)$ and the claim is clear. Suppose instead that $v^*(I) > 0$, so we can choose a nontrivial prime ideal P with $I \subseteq P$. We now have $I = Aa$ and $P = Ap$ for some a and p, and the relation $I \subseteq P$ means that $a = pb$ for some b. Thus, if we put $J = Ab$ we find that $I = PJ$. From this we get $v^*(J) = v^*(I) - v^*(P) = v^*(I) - 1$. We may assume by induction that J lies in the image of μ, say $J = \mu(m)$. Now put $n(P) = m(P) + 1$ and $n(Q) = m(Q)$ for all $Q \neq P$. We find that $\mu(n) = P\mu(m) = PJ = I$ as required. This proves that μ is surjective as well as injective, so it is a bijection as claimed. □

It is more traditional to talk about unique factorisation of elements rather than ideals. For this we need some additional discussion.

Definition 16.26. Let A be an integral domain. A *system of irreducibles* for A is a set $\mathcal{P}$ of irreducible elements, such that for every irreducible element $p_0 \in A$ there is a unique element $p \in \mathcal{P}$ that is an associate of p_0.

Example 16.27.

- The set of (positive) prime numbers is a system of irreducibles for $\mathbb{Z}$.
- The set $\{p\}$ is a system of irreducibles for $\mathbb{Z}_{(p)}$.
- The set $\{x - \lambda \mid \lambda \in \mathbb{C}\}$ is a system of irreducibles for $\mathbb{C}[x]$.

We can now state our unique factorisation result for elements:

Theorem 16.28. [`thm-pid-ufd-elts`]
Let $\mathcal{P}$ be a system of irreducibles in a principal ideal domain A. Define a map

$$\mu\colon A^\times \times \bigoplus_{p\in\mathcal{P}} \mathbb{N} \to A^\bullet$$

by $\mu(u, m) = u \prod_{p\in\mathcal{P}} p^{m(p)}$. Then μ is a bijection.

Proof. Consider an element $a \in A^\bullet$. Note that the map $p \mapsto Ap$ gives a bijection from $\mathcal{P}$ to the set of nontrivial prime ideals. Theorem 16.25 therefore tells us that there is a unique system of exponents $m(p) = v_p(Aa) = v_p(a) \in \mathbb{N}$ (almost all zero) such that $Aa = \prod_P (Ap)^{m(p)} = A.\prod_p p^{m(p)}$. It follows that a is a unit multiple of $\prod_p p^{m(p)}$. The claim follows easily. □

17. Modules over principal ideal domains

Throughout this section, A is assumed to be a principal ideal domain. In this section, we will study the structure of finitely generated A-modules. In Section 20 we will give a cruder classification of modules that works for a much larger class of rings. Some features of this section are designed for compatibility with that more general situation.

In order to state the main result, we need some definitions.

Definition 17.1. [`defn-torsion`]
Let M be any A-module.

(a) A *torsion element* is an element $m \in M$ such that $am = 0$ for some $a \in A \setminus \{0\}$. We write $T(M)$ for the set of all torsion elements. Note that if $am = bn = 0$ with $a, b \neq 0$ then $ab \neq 0$ and $ab(m+n) = 0$. Using this, we see that $T(M)$ is a submodule of M.
(b) Now let P be a nontrivial prime ideal. A *P-torsion element* is an element $m \in M$ such that $P^k m = 0$ for some $k \geq 0$. We write $T_P(M)$ for the set of all P-torsion elements, which is again a submodule.
(c) For compatibility with the notation used for more general rings, we also write $E_P(M)$ for $T_P(M)$ (when $P \neq 0$) and $E_0(M) = M/T(M)$.
(d) We say that M is a *torsion module* if all elements are torsion, so $M = T(M)$ and $E_0(M) = 0$. Similarly, we say that M is a *P-torsion module* if $M = T_P(M)$.
(e) We say that M is *torsion-free* if 0 is the only torsion element, so $T(M) = 0$ and $E_0(M) = M$.

Definition 17.2. [`defn-f-p-k`]
Suppose that P is a nontrivial prime ideal and $k \in \mathbb{N}$. For any A-module M, we define

$$F_P^k(M) = \{x \in P^{k-1}M \mid Px = 0\}.$$

This is easily seen to be a submodule of M. Moreover, as $Px = 0$ for all $x \in F_P^k(M)$, we can regard $F_P^k(M)$ as a module over the field A/P. We define

$$f_P^k(M) = \dim_{A/P}(F_P^k(M)).$$

If the dimensions $f_P^k(M)$ are all finite, we also put

$$g_P^k(M) = f_P^k(M) - f_P^{k+1}(M).$$

The main purpose of this section is to prove the following result.

Theorem 17.3. [thm-pid-modules]
Let M be a finitely generated A-module. Then

(a) $M \simeq E_0(M) \oplus \bigoplus_{P \neq 0} E_P(M)$*, and only finitely many of the terms in this sum are nonzero.*
(b) $E_0(M) \simeq A^d$ *for some d.*
(c) $E_P(M)$ *is isomorphic to a direct sum of copies of the modules A/P^k for various k. The number of copies of A/P^k is $g_P^k(M)$, which is zero for sufficiently large k.*

The proof will be given after Corollary 17.22 below.

Proposition 17.4. [prop-hereditary]
Let n be a natural number. Then any submodule of A^n is isomorphic to A^m for some $m \leq n$.

Proof. We will argue by induction on n. The case $n = 0$ is trival. Suppose we have proved the case $n = k$, and that M is a submodule of $A^{n+1} = \bigoplus_{i=0}^n A$. Define $\pi\colon A^{n+1} \to A$ by $\pi(a_0, \dots, a_n) = a_n$, and put $N = M \cap \ker(\pi)$ and $I = \pi(M)$. Here N is a submodule of $\ker(\pi) \simeq A^n$, so by induction we can choose an isomorphism $\phi\colon A^m \to N$ for some $m \leq n$. On the other hand, I is a submodule of A, or in other words an ideal. If $I = 0$ then $M = N$, so $M \simeq A^m$ as required. If $I \neq 0$, then $I = Ax$ for some nonzero element $x \in A$. As I is defined to be $\pi(M)$, we can choose $u \in M$ with $\pi(u) = x$. Define $\psi\colon A^{m+1} \to M$ by

$$\psi(a_0, \dots, a_m) = \phi(a_0, \dots, a_{m-1}) + a_m u.$$

We claim that this is an isomorphism. Indeed, if $\psi(a_0, \dots, a_m) = 0$ then we can apply π to get $a_m x = 0$. As A is an integral domain, we deduce that $a_m = 0$, so

$$0 = \psi(a_0, \dots, a_m) = \phi(a_0, \dots, a_{m-1}).$$

As ϕ is assumed to be an isomorphism, it follows that $a_0 = \dots = a_{m-1} = 0$ as well. This shows that ψ is injective. In the other direction, suppose we have an element $v \in M$. Then $\pi(v) \in \pi(M) = Ax = A\pi(u)$, so we can choose $t \in A$ with $\pi(v) = t\,\pi(u)$. This means that $v - tu \in M \cap \ker(\pi) = N$, so $v - tu = \phi(a_0, \dots, a_{m-1})$ for some elements $a_0, \dots, a_{m-1}$ in A. This in turn gives

$$v = \psi(a_0, \dots, a_{m-1}, t),$$

showing that ψ is also surjective. □

Remark 17.5. [rem-hereditary]
It is true more generally that any submodule of a free module over a PID is always free, even if the modules in question are infinitely generated. However, a proof would require more set theory than we have space to develop here.

Corollary 17.6. [cor-pid-noetherian]
Let M be a finitely generated A-module. Then every submodule $N \subseteq M$ is also finitely generated.

Remark 17.7. This corollary is also valid for many rings that are not PIDs, but there are some rings for which it fails. This will be discussed in more detail in Section 18.

Proof. Choose a generating set $\{e_0, \dots, e_{n-1}\}$ for M. We can then define a surjective homomorphism $\phi\colon A^n \to M$ by $\phi(x) = \sum_i x_i e_i$. Put $L = \{x \in A^n \mid \phi(x) \in N\}$. This is a submodule of A^n, so it is isomorphic to A^m for some $m \leq n$. We can thus choose a basis $u_0, \dots, u_{m-1}$ for L. We claim that the elements $v_j = \phi(u_j) \in N$ generate N. Indeed, as ϕ is surjective, every element $t \in N \subseteq M$ can be written as $\phi(x)$ for some element $x \in A^n$. As $\phi(x) = t \in N$ we see that $x \in L$, so $x = \sum_j a_j u_j$ for some system of coefficients $a_j \in A$. Applying ϕ gives $t = \sum_j a_j v_j$ as required. □

Remark 17.8. If M is finitely generated, we see in particular that the modules $T(M)$, $E_P(M) = T_P(M)$, $E_0(M) = M/T(M)$ and $F_P^k(M)$ are all finitely generated. This in turn implies that $f_P^k(M) < \infty$ for all k, so $g_P^k(M)$ is well defined.

Proposition 17.9. [prop-tors-free-free]
A finitely generated A-module is free if and only if it is torsion-free.

Proof. First suppose that M is a free. Then $M \simeq A^n$ for some n, so we may assume that $M = A^n$. Suppose that $x \in A^n$ is a torsion element. Then there is a nonzero element $a \in A$ such that $ax = 0$, so $ax_i = 0$ for all i. As A is a domain and $a \neq 0$ we must have $x_i = 0$ for all i, so $x = 0$. Thus M is torsion-free.

Conversely, suppose that M is torsion-free. Clearly, a list $e_0, \dots, e_{n-1}$ generates M iff $\sum_i Ae_i = M$. We will say that such a set *almost generates* M if there is a nonzero element $t \in A$ such that $\sum_i Ae_i \geq tM$. By assumption we can choose a finite list of elements that generates M, so certainly we can choose a finite list that almost generates M. Let $e_0, \dots, e_{n-1}$ be such a list which is as short as possible, and fix an element $t \neq 0$ such that $\sum_i Ae_i \geq tM$.

We claim that the elements e_i are independent. If not, we have a relation $\sum_i a_i e_i = 0$ where some coefficient a_k is nonzero. After reordering everything if necessary, we may assume that $a_0 \neq 0$. For any $x \in M$ we know that tx can be written in the form $tx = \sum_{i=0}^{n-1} b_i e_i$. After multiplying by a_0 and using the substitution $a_0 b_0 e_0 = -\sum_{i=1}^{n-1} a_i b_0 e_i$, we see that $a_0 tx$ lies in the span of $e_1, \dots, e_{n-1}$. Thus, $Ae_1 + \dots + Ae_{n-1} \geq a_0 tM$, so the list $e_1, \dots, e_{n-1}$ almost generates M, contradicting our assumption that the list $e_0, \dots, e_n$ was as short as possible. This contradiction shows that $e_0, \dots, e_{n-1}$ must be independent, after all. This implies that the module $N = \sum_i Ae_i$ is free. By assumption, the module $L := aM$ is contained in N, so it is free by Proposition 17.4, with basis $f_0, \dots, f_{m-1}$ say. As $L = aM$ we have $f_i = ag_i$ for some $g_i \in M$. It is now easy to see that the elements g_i give a basis for M, so M is free as claimed. □

Lemma 17.10. [lem-tf-quotient]
The module $E_0(M) = M/T(M)$ is always torsion free.

Proof. Let $\pi \colon M \to E_0(M)$ be the quotient map. Let $q \in E_0(M)$ be a torsion element, so $aq = 0$ for some $a \in A \setminus \{0\}$. We must have $q = \pi(m)$ for some $m \in M$. Now $\pi(am) = a\,\pi(m) = aq = 0$, so $am \in \ker(\pi) = T(M)$, so we must have $bam = 0$ for some $b \in A \setminus \{0\}$. As A is a domain and $a, b \neq 0$ we must have $ba \neq 0$, but $bam = 0$, so $m \in T(M)$. This means that $\pi(m) = 0$, or in other words $q = 0$ as required. □

Proposition 17.11. [prop-free-summand]
If M is a finitely generated module, then $E_0(M) \simeq A^d$ for some d, and $M \simeq E_0(M) \oplus T(M)$.

Remark 17.12. There are some choices involved in constructing an isomorphism $M \simeq E_0(M) \oplus T(M)$, so this is not a natural isomorphism in the sense of category theory.

Proof. Recall that $E_0(M) = M/T(M)$. This is clearly finitely generated, and it is torsion free by Lemma 17.10, so it is isomorphic to A^d for some d. We can therefore choose elements $m_0, \dots, m_{d-1} \in M$ such that the corresponding cosets $q_i = \pi(m_i)$ form a basis for $E_0(M)$. Let F be the submodule of M generated by the elements m_i. As the elements $\pi(m_i)$ form a basis, it is easy to see that the elements m_i are independent, so they form a basis for F, proving that F is a free module. This in turn means that the only torsion element in F is 0, so $T(M) \cap F = 0$. Now let $m \in M$ be an arbitrary element. We can then write $\pi(m)$ as $\sum_i a_i q_i$ for some system of coefficients a_i. Put $f = \sum_i a_i m_i \in F$ and $t = m - f$. We find that $\pi(f) = \pi(m)$, so $\pi(t) = 0$, so $t \in T(M)$, so $m = f + t \in F \oplus T(M)$. It follows that $M = F \oplus T(M) \simeq E_0(M) \oplus T(M)$ as claimed. □

Lemma 17.13. [lem-totally-tors]
If M is a finitely generated torsion module, then there is a nonzero element $a \in A$ such that $aM = 0$. Similarly, if M is a finitely generated P-torsion module, then there exists $k \geq 0$ such that $P^k M = 0$.

Proof. Choose a finite generating set $\{e_0, \dots, e_{n-1}\}$ for M. As M is a torsion module, for each i we can choose $a_i \neq 0$ such that $a_i e_i = 0$. Put $a = \prod_i a_i$, so $ae_i = (\prod_{j \neq i} a_j)(a_i e_i) = 0$ for all i. As the elements e_i generate M, we deduce that $aM = 0$. Similarly, if M is P-torsion then we can choose k_i such that $P^{k_i} e_i = 0$, then we can put $k = \max(k_0, \dots, k_{n-1})$. We find that $P^k e_i = 0$ for all i, and so $P^k M = 0$. □

Lemma 17.14. [lem-module-coprime-split]

Let L be an A-module. Suppose that $b, c \in A$ are coprime and that $bcL = 0$. Put $M = \{y \in L \mid by = 0\}$ and $N = \{z \in L \mid cz = 0\}$. Then $L = M \oplus N$.

Proof. As b and c are coprime there exist elements u, w such that $ub+wc = 1$. If $x \in M \cap N$ then $bx = cx = 0$ so $x = 1.x = ubx + wcx = 0$; thus $M \cap N = 0$. Now let x be an arbitrary element of M. Put $y = wcx$ and $z = ubx$, so $x = y + z$. We have $by = (bc)(wx)$ but $bcL = 0$ so $by = 0$ so $y \in M$. Similarly, $cz = (bc)(ux) = 0$ so $z \in N$. Thus $x = y + z \in M + N$, which shows that $M + N = L$. As $M \cap N = 0$, the sum is direct. □

Proposition 17.15. [prop-tors-split]

Let M be a finitely generated torsion module. Then $M = \bigoplus_P T_P(M)$, and only finitely many of the terms $T_P(M)$ are nonzero.

Proof. Put $I = \{a \in A \mid aM = 0\}$. This is easily seen to be an ideal, and it is nonzero by Lemma 17.13, so it can be factored as $I = \prod_i P_i^{n_i}$ for some finite list of distinct nontrivial prime ideals P_i and exponents $n_i > 0$. Put $M_i = \{m \in M \mid P_i^{n_i} m = 0\}$. The ideals $P_i^{n_i}$ are pairwise coprime, so an evident inductive extension of Lemma 17.14 gives $M = \bigoplus_i M_i$. It follows that $T_P(M) = M_i$ if $P = P_i$, and that $T_P(M) = 0$ if P is not one of the ideals P_i. The claim is clear from this. □

Remark 17.16. [rem-pid-modules]

By combining Propositions 17.11 and 17.15, we see that any finitely generated module M is isomorphic to $E_0(M) \oplus \bigoplus_{P \neq 0} E_P(M)$, and again only finitely many of the terms are nonzero. To complete the proof of Theorem 17.3, we just need to study the structure of finitely generated P-torsion modules, such as $E_P(M)$.

We start with some results about the numbers $f_P^k(M)$ and $g_P^k(M)$.

Remark 17.17. [rem-g-p-k-additive]

It is easy to see that a pair $(x, y) \in M \oplus N$ lies in $F_P^k(M \oplus N)$ if and only if $x \in F_P^k(M)$ and $y \in F_P^k(N)$. It follows that $F_P^k(M \oplus N) = F_P^k(M) \oplus F_P^k(N)$ and thus that $f_P^k(M \oplus N) = f_P^k(M) + f_P^k(N)$ and $g_P^k(M \oplus N) = g_P^k(M) + g_P^k(N)$.

Remark 17.18. [rem-g-p-k-iso]

It is clear that any isomorphism $M \to M'$ restricts to give isomorphisms $F_P^k(M) \to F_P^k(M')$ for all P and k. Thus, if $M \simeq M'$ then $f_P^k(M) = f_P^k(M')$ and $g_P^k(M) = g_P^k(M')$.

Proposition 17.19. [prop-g-p-k-basic]

We have $g_P^k(A) = 0$ for all P and k, and

$$g_P^k(A/Q^j) = \begin{cases} 1 & \textit{if } p = q \textit{ and } k = j \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. It is clear that $F_P^k(A) = 0$ and so $g_P^k(A) = f_P^k(A) = 0$ for all P and all $k > 0$.

For the case of A/Q^j, it will be convenient to choose generators p and q for P and Q. We start by proving that

$$f_P^k(A/Q^j) = \begin{cases} 0 & \text{if } P \neq Q \\ 0 & \text{if } P = Q \text{ and } k > j \\ 1 & \text{if } P = Q \text{ and } k \leq j. \end{cases}$$

First suppose that $p \neq q$. Then p^k and q^j are coprime, so $ap^k + bq^j = 1$ for some $a, b \in A$. If $x \in F_P^k(A/Q^j)$ then $x = p^{k-1}y$ for some y and $px = 0$ so $p^k y = 0$. On the other hand, it is clear from the definition of A/Q^j that $q^j z = 0$ for all $z \in A/Q^j$, so $q^j y = 0$. We thus have $y = 1.y = ap^k y + bq^j y = 0$, and thus $x = p^{k-1}y = 0$. Thus $F_P^k(A/Q^j) = 0$ and so $f_P^k(A/Q^j) = 0$, as required.

Now suppose that $Q = P$ (so we can choose $q = p$) and $j < k$. Then $k - 1 - j \geq 0$ and $p^{k-1}A/P^j = p^{k-1-j}p^j A/P^j = 0$ and $F_P^k(M) \leq p^{k-1}M$ so $F_P^k(A/P^j) = 0$. This means that $f_P^k(A/P^j) = 0$, as required.

Now suppose instead that $q = p$ and $k \leq j$. Let e be the element $1 + p^j A$ in A/P^j, so that $ae = (a + p^j A)$. Put $f = p^{j-1}e$, so that $f \neq 0$ and $pf = 0$. We also have $f = p^{k-1}(p^{j-k}e)$ so $f \in p^{k-1}A/P^j$, so $f \in F_P^k(A/P^j)$. Let u be another element of $F_P^k(A/P^j)$. We can write $u = ae = (a + p^j A)$ for some $a \in A$. As $pu = 0$ we have $pa = 0 \pmod{p^j}$, or in other words $pa = p^j b$ for some b, so $a = p^{j-1}b$ and thus $u = bf$. This shows that $\{f\}$ generates the vector space $F_P^k(A/P^j)$ over A/P, and $f \neq 0$ so the dimension must be exactly one. Thus $f_P^k(A/P^j) = 1$, as required.

It is now easy to deduce our description of $g_P^k(A/Q^j)$. If $Q \neq P$ then $f_P^k(A/Q^j) = 0$ for all k and it follows easily that $g_P^k(A/Q^j) = 0$. Suppose instead that $Q = P$. If $k > j$ then $k+1 > j$ as well so $f_P^k(A/P^j) = f_P^{k+1}(A/P^j) = 0$ so $g_P^k(A/P^j) = 0$ as claimed. If $k < j$ then both k and $k+1$ are less than or equal to j, so $f_P^k(A/P^j) = f_P^{k+1}(A/P^j) = 1$ so $g_P^k(A/P^j) = 0$ as claimed. If $k = j$ then $f_P^k(A/P^j) = 1$ and $f_P^{k+1}(A/P^j) = 0$ so $g_P^k(A/P^j) = 1$ as claimed. □

Corollary 17.20. [cor-g-p-k]

Let M be a finitely generated torsion module. Then if there is any list of basic modules whose direct sum is isomorphic to M, then that list must contain precisely $g_P^k(M)$ copies of A/P^k. □

Lemma 17.21. [lem-one-summand]

Suppose that M has a generating set $e_0, \dots, e_{n-1}$ such that $P^k e_i = 0$ for all i, and $P^{k-1}e_0 \neq 0$. Then $Ae_0 \simeq A/P^k$, and there is a submodule N such that $M = Ae_0 \oplus N$.

Proof. It will again be convenient to choose a generator p for P.

First, it is clear that the map $a \mapsto ae_0$ induces an isomorphism $A/I \to Ae_0$, where $I = \{a \mid ae_0 = 0\}$. Now $I = Au$ for some u, and $p^k \in I$ so u divides p^k, so u must be a unit multiple of p^j for some $j \leq k$. On the other hand, we are given that $p^{k-1} \notin I$, so we must have $j = k$ and $I = P^k$. This means that $Ae_0 \simeq A/P^k = A/P^k$ as claimed.

For the splitting $M = Ae_0 \oplus N$, we will argue by induction on n. If $n = 1$ we can just take $N = 0$. Suppose instead that $n > 1$, and put $M' = \sum_{i=0}^{n-2} Ae_i \subseteq M$. By induction, there is a submodule $N' < M'$ such that $M' = Ae_0 \oplus N'$. Put $J = \{a \mid ae_{n-1} \in M'\}$. This again contains P^k, so we must have $J = P^l$ for some $l \leq k$. We thus have $p^l e_{n-1} \in M' = Ae_0 \oplus N'$, so $p^l e_{n-1} = ve_0 + n'$ for some $v \in A$ and $n' \in N'$. We can multiply this by p^{k-l} to get $p^{k-l}ve_0 + p^{k-l}n' = 0$ in $M' = Ae_0 \oplus N'$, so $p^{k-l}ve_0 = 0$. As $Ae_0 \simeq A/p^k$ we deduce that $p^{k-l}v \in p^k A$ and so $v \in p^l A$, say $v = p^l w$. Put $e^* = e_{n-1} - we_0$ (so $p^l e^* = n'$) and $N = N' + Ae^*$. It is clear that

$$Ae_0 + N = Ae_0 + N' + Ae^* = Ae_0 + N' + Ae_{n-1} = M' + Ae_{n-1} = M.$$

Now suppose we have an element $x \in Ae_0 \cap N$, so

$$x = re_0 = s + te^* = s + te_{n-1} - twe_0$$

for some $r \in A$ and $s \in N'$ and $t \in A$. This gives $te_{n-1} = (r + tw)e_0 - s \in Ae_0 + N' = M'$, so $t \in J = Ap^l$, so $t = p^l t'$ for some $t' \in A$. As $p^l e^* = n'$ we deduce that $re_0 = s + t'n' \in N'$. However, we have $Ae_0 \cap N' = 0$ by assumption, so $re_0 = 0$, or in other words $x = 0$. This proves that $M = Ae_0 \oplus N$, as required. □

Corollary 17.22. [cor-one-summand]

Let M be a nontrivial finitely generated P-torsion module, and let k be the smallest integer such that $P^k M = 0$. Then $M \simeq A/P^k \oplus N$ for some N.

Proof. Choose a finite system of generators e_i for M. These must all satisfy $P^k e_i = 0$, and at least one of them must satisfy $P^k e_i \neq 0$. After renumbering if necessary we can assume that $P^k e_0 \neq 0$, and then we can apply the lemma. □

Proof of Theorem 17.3. Given Remark 17.16 and Corollary 17.20, we need only show that every finitely generated P-torsion module M is isomorphic to a finite direct sum of modules of the form A/P^k. We will argue by induction on the number

$$f_P^1(M) = \dim_{A/P}\{m \in M \mid Pm = 0\}.$$

If $f_P^1(M) = 0$ then multiplication by p gives an injective map $M \to M$, but every element $m \in M$ also satisfies $p^k m = 0$ for large k. The only way that these can be reconciled is if $M = 0$, in which case M is the direct sum of the empty list. Suppose instead that $f_P^1(M) > 0$, so $M \neq 0$. Let k be the smallest integer such that $P^k M = 0$. Corollary 17.22 gives a splitting $M = A/P^k \oplus N$, and we find that $f_P^1(N) = f_P^1(M) - f_P^1(A/P^k) = f_P^1(M) - 1$. We can thus assume by induction that N splits as a sum of modules of the form A/P^j, and it follows that M has a splitting of the same type. □

Example 17.23. [eg-finab]

Abelian groups are just the same as $\mathbb{Z}$-modules, so we can use Theorem 17.3 to classify the finitely generated ones. Every such group is therefore a direct sum of copies of $\mathbb{Z}$, or $\mathbb{Z}/p^k$ for various prime numbers p and integers $k > 0$. The classification of finite abelian groups is the same, except that we cannot have any summands isomorphic to $\mathbb{Z}$.

Example 17.24. [eg-jordan]

Another application of the above theory is to the classification of matrices up to conjugacy. Given a square matrix $C \in \text{Mat}_d(\mathbb{C})$, we can make $\mathbb{C}^d$ into a module over $\mathbb{C}[x]$ by the rule

$$\left(\sum_i a_i x^i\right).u = \sum_i a_i C^i u.$$

We write V_C for $\mathbb{C}^d$ equipped with this module structure. Note that any $\mathbb{C}$-linear map $\alpha\colon \mathbb{C}^d \to \mathbb{C}^d$ has the form $\alpha(v) = Uv$ for some matrix U, and α gives a $\mathbb{C}[x]$-linear map from V_C to V_D if and only if $UC = DU$. Using this, we see that $V_C \simeq V_D$ if and only if C and D are conjugate. Note that V_C cannot have any summands of the form $\mathbb{C}[x]$, because $\dim_{\mathbb{C}}(\mathbb{C}[x]) = \infty$. Theorem 17.3 therefore tells us that V_C is isomorphic to a finite direct sum of modules of the form $\mathbb{C}[x]/(x-\lambda)^k$, for various complex numbers λ and integers $k > 0$. In particular, if C is a Jordan block of size d and eigenvalue λ, it is not hard to write down an isomorphism $V_C \simeq \mathbb{C}[x]/(x-\lambda)^d$. After a small amount of translation, this proves the familiar theorem that every square matrix over $\mathbb{C}$ is conjugate to a block diagonal sum of some Jordan blocks.

Example 17.25. [eg-diffeq]

A third application is to the study of differential equations. Let $C^\infty(\mathbb{R}, \mathbb{C})$ denote the set of smooth complex-valued functions on the real line. We can make this a module over the polynomial ring $\mathbb{C}[D]$ by the rule $D.f = f'$. Given a differential operator $L = \sum_{k=0}^d a_k D^k$, we want to understand the solution space

$$S(L) = \{f \in C^\infty(\mathbb{R}, \mathbb{C}) \mid Lf = 0\}.$$

This is a finitely generated module over $\mathbb{C}[D]$, so we can use the above theory. We can factor L as

$$L = u \prod_{j=1}^r (D - \lambda_j)^{m_j}$$

for some $u \neq 0$ and some distinct complex numbers λ_j and exponents $m_j > 0$. Using Lemma 17.14 we get

$$S(L) = \bigoplus_{j=1}^r S((D - \lambda_j)^{m_j}).$$

It is straightforward to check that $S(D^m)$ has basis $\{1, x, \dots, x^{m-1}\}$, and that multiplication by $e^{\lambda x}$ gives an isomorphism $S(D^m) \to S((D-\lambda)^m)$. It follows that the functions $x^p e^{\lambda_k x}$ (with $1 \leq k \leq r$ and $0 \leq p < m_k$) give a basis for $S(L)$ over $\mathbb{C}$, and that

$$S(L) \simeq \bigoplus_{k=1}^r \mathbb{C}[D]/((D - \lambda_k)^{m_k}).$$

18. Noetherian rings

Definition 18.1. [defn-noetherian]

Let A be a ring, and let M be an A-module. We say that M is *noetherian* if every submodule of M is finitely generated. We say that A is a *noetherian ring* if it is noetherian as a module over itself, or equivalently, every ideal is finitely generated.

We will see that most of the rings that people usually study are noetherian, but it takes some work to prove this. However, there are a few cases that we can handle immediately:

Example 18.2. [eg-basic-noetherian]

In a principal ideal domain, every ideal is generated by a single element and so is certainly finitely generated. Thus, principal ideal domains are noetherian rings. In particular $\mathbb{Z}$ is noetherian, and for any field K both K and $K[x]$ are noetherian.

Example 18.3. [eg-not-noetherian]

Consider the set

$$A = \mathbb{Z} \oplus x\mathbb{Q}[x] = \{f \in \mathbb{Q}[x] \mid f(0) \in \mathbb{Z}\}.$$

One can check that A is a subring of $\mathbb{Q}[x]$, and that the subset $I = x\mathbb{Q}[x]$ is an ideal. We claim that this is not finitely generated, so A is not noetherian. To see this, define $\pi \colon I \to \mathbb{Q}$ by $\pi(\sum_{i>0} a_i x^i) = a_1$, so $\pi(I) = \mathbb{Q}$. Suppose we have a finite list of elements $g_0, \dots, g_{d-1} \in I$, and we let J be the ideal in A that they generate. By clearing denominators we can find $n > 0$ such that $n\pi(g_i) \in \mathbb{Z}$ for all i. Now any element of J has the form $h = \sum_i f_i g_i$ with $f_i \in A$, which gives $n\pi(h) = \sum_i f_i(0).n\pi(g_i) \in \mathbb{Z}$. This shows that $x/(2n) \in I \setminus J$, so $J \neq I$.

Proposition 18.4. [prop-noetherian-ses]

Suppose we have a short exact sequence of A-modules

$$0 \to L \xrightarrow{\phi} M \xrightarrow{\psi} N \to 0.$$

Then M is noetherian iff both L and N are noetherian. In particular:

(a) *If M is noetherian, then every submodule and every quotient module of M is also noetherian.*
(b) *If L and N are noetherian, then so is $L \oplus N$.*

Proof. First suppose that M is noetherian. Any submodule $L_0 \subseteq L$ is isomorphic to $\phi(L_0) \subseteq M$ and so is finitely generated; so L is noetherian. Consider instead a submodule $N_0 \subseteq N$, and put $M_0 = \psi^{-1}(N_0) \subseteq M$. Now M_0 must be generated by some finite list of elements $(y_i)_{i=0}^{d-1}$, and $\psi \colon M_0 \to N_0$ is surjective, so N_0 is generated by $(\psi(y_i))_{i=0}^{d-1}$. This proves that N is also noetherian.

Suppose instead L and N are both noetherian. Consider a submodule $M_0 \subseteq M$, and put $L_0 = \phi^{-1}(L_0) \subseteq L$ and $N_0 = \psi(M_0) \subseteq M$. One can check that ϕ and ψ restrict to give a short exact sequence

$$0 \to L_0 \xrightarrow{\phi_0} M_0 \xrightarrow{\psi_0} N_0 \to 0.$$

As L and N are noetherian, we can choose finite lists $X = (x_i)_{i=0}^{p-1}$ and $Z = (z_k)_{k=0}^{r-1}$ that generate L_0 and N_0 respectively. Put $x'_i = \phi_0(x_i)$, and choose $z'_k \in M_0$ with $\psi_0(z'_k) = z_k$. We claim that the list $Y = (x'_0, \dots, x'_{p-1}, z'_0, \dots, z'_{r-1})$ generates M_0. To see this, consider an element $m \in M_0$. As Z generates N_0, we can express $\psi_0(m)$ as $\sum_k c_k z_k$ for some coefficients $c_k \in A$. Put $m_1 = \sum_k c_k z'_k \in \operatorname{span}_A(Y)$ and $m_2 = m - m_1$. We then have $\psi_0(m_1) = \psi_0(m)$ so $m_2 \in \ker(\psi_0) = \operatorname{image}(\phi_0)$. Thus, for some system of coefficients $a_i \in A$ we have $m_2 = \phi(\sum_i a_i x_i) = \sum_i a_i x'_i \in \operatorname{span}_A(Y)$. It follows that $m = m_1 + m_2 \in \operatorname{span}_A(Y)$ as claimed. This shows that an arbitrary submodule $M_0 \subseteq M$ is finitely generated, so M is noetherian.

Finally, we can recover statements (a) and (b) by considering short exact sequences $L \to M \to M/L$ and $L \to L \oplus N \to N$ as in Examples 12.4 and 12.5. □

Corollary 18.5. [cor-fg-noetherian]

Let A be a noetherian ring. Then an A-module M is noetherian iff it is finitely generated.

Proof. If M is noetherian then by definition it must be finitely generated (because it is a submodule of itself).

Conversely, we can use Proposition 18.4(b) to see by induction that A^n is noetherian for all $n \geq 0$. Any finitely generated module is isomorphis to a quotient of A^n, and so is noetherian by Proposition 18.4(a). □

Proposition 18.6. [prop-simple-noetherian]

Let A be an arbitrary ring; then every simple module is noetherian. More generally, every module of finite length is noetherian.

Proof. First let S be a simple module, so $S \neq 0$ and every nontrivial submodule is all of S. Choose any nontrivial element $s \in S$ and note that As must be all of S; this proves that S is finitely generated. The only other submodule of S is 0, which is also finitely generated, so S is noetherian.

Now let M be a simple module, with composition series $(M_i)_{i=0}^n$ say. We then have short exact sequences $M_{i-1} \to M_i \to M_i/M_{i-1}$ in which M_i/M_{i-1} is simple. We can use these to prove by induction that M_i is noetherian for all i. In particular, as $M_n = M$, we see that M is noetherian. □

Proposition 18.7. [prop-noetherian-ops]
Let A be a noetherian ring. Then A/K is noetherian for every ideal $K \subseteq A$, and $A[U^{-1}]$ is noetherian for every multiplicative subset $U \subseteq A$. Moreover, if B is another noetherian ring then $A \times B$ is also noetherian.

Proof. Corollary 5.46 tells us that every ideal in A/K has the form I/K for some ideal $I \subseteq A$ with $K \subseteq I$. As A is noetherian we can choose a finite set of generators for I, and the images of these elements in A/K will generate I/K. A similar argument based on Proposition 8.15 shows that $A[U^{-1}]$ is noetherian.

Finally, Example 5.39 tells us that every ideal in $A \times B$ has the form $I \times J$ for some ideals $I \subseteq A$ and $J \subseteq B$. Now I will be generated by some finite list $(a_i)_{i=0}^{p-1}$ and J will be generated by some finite list $(b_j)_{j=0}^{q-1}$. It follows easily that $I \times J$ is generated by the elements $(a_i, 0)$ and $(0, b_j)$. □

Proposition 18.8. [prop-acc]
Let A be a ring and let M be an A-module. Let $\text{submod}(M)$ *be the set of all submodules of M. Then the following are equivalent:*

(a) *M is noetherian.*
(b) *Any chain $M_0 \subseteq M_1 \subseteq M_2 \subseteq \cdots$ in* $\text{submod}(M)$ *is eventually constant (so there exists $n_0 \in \mathbb{N}$ such that $M_n = M_{n_0}$ for all $n \geq n_0$).*
(c) *Every nonempty subset of* $\text{submod}(M)$ *has a maximal element.*

The advantage of conditions (b) and (c) here is that they depend only on the structure of $\text{submod}(M)$ as an ordered set.

Proof.

(a)⇒(b) Suppose that M is noetherian, and that we have a nested chain of submodules M_i as in (b). Put $M_\infty = \bigcup_i M_i$, and recall from Remark 11.39 that this is also a submodule of M. As M is noetherian, we can choose a finite list of elements $(m_k)_{k=0}^{d-1}$ that generate M_∞. As $m_k \in M_\infty$ we can choose p_k such that $m_k \in M_{p_k}$. Now put $n_0 = \max(p_0, \ldots, p_{d-1})$, so $m_k \in M_{n_0}$ for all k. It follows that for all $n \geq n_0$ we have $M_{n_0} = M_n = M_\infty$ as required.

(b)⇒(c) Suppose that there is a nonempty subset $\mathcal{C} \subset \text{submod}(M)$ with no maximal element. As $\mathcal{C}$ is nonempty we can choose $M_0 \in \mathcal{C}$. Now M_0 cannot be maximal, so we can choose $M_1 \in \mathcal{C}$ with $M_0 \subset M_1$. Again M_1 cannot be maximal, so we can choose $M_2 \in \mathcal{C}$ with $M_1 \subset M_2$. We can continue this process recursively to get a strictly increasing chain which never becomes constsnt. This shows that the negation of (c) implies the negation of (b), or equivalently that (b) implies (c).

(c)⇒(a) Suppose that (c) holds. Given any submodule $N \subseteq M$, put

$$\mathcal{C} = \{L \in \text{submod}(M) \mid L \subseteq N \text{ and } N/L \text{ is not finitely generated }\}.$$

We claim that $\mathcal{C}$ is empty. If not, there is a maximal element $L \in \mathcal{C}$. As N/N is certainly finitely generated, we see that L must be a proper subset of N, so we can choose $x_0 \in N \setminus L$. Now $L + Ax_0$ is strictly larger than L, so by the maximality of L we see that the quotient $N/(L + Ax_0)$ is generated by some finite list of elements, which will be the images of some elements $x_1, \ldots, x_r \in N$. This means that $N = L + Ax_0 + Ax_1 + \cdots + Ax_r$, so N/L is generated by the finite list $x_0, \ldots, x_r$, contradicting the assumption that $L \in \mathcal{C}$. Thus, $\mathcal{C}$ must be empty after all, so in particular $0 \notin \mathcal{C}$, so N is finitely generated.

□

We call the next result the *principle of noetherian induction.*

Corollary 18.9. [cor-induction]
Let M be a noetherian module, and let $\mathcal{C}$ be a family of submodules of M. Suppose that whenever $N \subseteq M$ and every strictly larger submodule lies in $\mathcal{C}$, we also have $N \in \mathcal{C}$. (In particular, we suppose that $M \in \mathcal{C}$, as the condition is vacuously satisfied when $N = M$.) Then every submodule lies in $\mathcal{C}$.

Proof. The set $\mathcal{D} = \text{submod}(M) \setminus \mathcal{C}$ cannot have a maximal element, so it must be empty. □

The following result is called the *Hilbert basis theorem.*

Theorem 18.10. [thm-hilbert]

Let A be a noetherian ring; then $A[x]$ is also noetherian.

Proof. For any $f \in A[x]$ we let $\pi_n(f)$ denote the coefficient of x^n in f. This defines an A-module map $\pi_n \colon A[x] \to A$.

Now consider an ideal $I \subseteq A[x]$. Let $I_{\leq n}$ denote the set of polynomials in I of degree at most n, which is an A-submodule of I. Put $J_n = \pi_n(I_{\leq n}) \subseteq A$, which is an A-submodule of A and thus an ideal. We have $x.I_{\leq n} \subseteq I_{\leq n+1}$ and $\pi_{n+1}(xf) = \pi_n(f)$ so $J_n \subseteq J_{n+1}$. We thus have an ascending chain of ideals in A, which must eventually be constant as A is noetherian. Choose N such that $J_n = J_N$ for all $n \geq N$. Now note that $I_{\leq N}$ is an A-submodule of the module $A[x]_{\leq N}$, which is isomorphic to A^{N+1} and is therefore noetherian. We can therefore choose a finite set $F \subseteq I_{\leq N}$ that generates $I_{\leq N}$ as an A-module. Let I^* denote the ideal in $A[x]$ generated by F. It is clear that $I_{\leq N} \subseteq I^* \subseteq I$. Suppose we know that $I_{\leq n-1} \subseteq I^*$ for some $n > N$, and that $f \in I_{\leq n}$. We then have $\pi_n(f) \in J_n = J_N = \pi_N(I_{\leq N})$. We can therefore express $\pi_n(f)$ as $\sum_i a_i \pi_N(f_i)$ with $a_i \in A$ and $f_i \in F$. Put $g = x^{n-N} \sum_i a_i f_i$ and $h = f - g$. Then $g \in I^* \subseteq I$ and $\pi_n(h) = 0$ so $h \in I_{\leq n-1} \subseteq I^*$. It follows that the polynomial $f = g + h$ is also in I^*. As $f \in I_{\leq n}$ was arbitrary we deduce that $I_{\leq n} \subseteq I^*$, and after extending this inductively we see that $I^* = I$. Thus, I is generated by the finite set F, as required. □

Corollary 18.11. [cor-hilbert]

If A is a noetherian ring and B is a finitely generated A-algebra then B is also noetherian.

Proof. Every finitely generated A-algebra is a quotient of a polynomial ring $P_n = A[x_0, \dots, x_{n-1}]$ for some n. As $P_n \simeq P_{n-1}[x]$ we can use Theorem 18.10 repeatedly to see that P_n is noetherian for all n, and we have also seen that quotients of noetherian rings are noetherian. □

19. Supports and associated primes

Throughout this section, A is assumed to be a noetherian ring. Moreover, the symbol M will refer to an A-module that is assumed to be finitely generated unless we explicitly say otherwise.

In the Section 17, we considered the case where A is a principal ideal domain, and we studied the structure of M using certain auxiliary modules $E_P(M)$. We now generalise the definition of these modules.

Definition 19.1. Let M be an arbitrary A-module, and let P be a prime ideal in A.

(a) We again say that $m \in M$ is a *P-torsion element* if $P^k m = 0$ for some $k \geq 0$. We write $T_P(M)$ for the set of P-torsion elements, which is a submodule of M.

(b) There is a natural map from $T_P(M)$ to the localisation $T_P(M)_P = T_P(M)[(A \setminus P)^{-1}]$. We define $E_P(M)$ to be the image of this map (which is a quotient of $T_P(M)$).

(c) We say that M is *P-coprimary* if every element of M is P-torsion, but for every element $a \in A \setminus P$, multiplication by a gives an injective map $M \to M$.

Remark 19.2. Suppose that A is a principal ideal domain. We then find that $T_0(M) = M$, and $T_0(M)_0 = M[(A \setminus 0)^{-1}]$. It follows that the kernel of the map $\eta \colon T_0(M) \to T_0(M)_0$ is just $T(M)$, so the image (which we are now calling $E_0(M)$) can be identified with $M/T(M)$ (which was our previous definition of $E_0(M)$). Similarly, for a nonzero prime ideal P and an element $u \in A \setminus P$ we find that u is invertible mod P, so u is invertible mod P^k for all k, so the map $\eta \colon T_P(M) \to T_P(M)_P$ is an isomorphism. It again follows that our new definition of $E_P(M)$ is essentially the same as the old one.

One of our main tasks is to understand which primes P have $E_P(M) \neq 0$. Just as in the case of a principal ideal domain, it will turn out that there are only finitely many of them (under our standing assumption that M is finitely generated). Next, it turns out that we do not have $M \simeq \bigoplus_P E_P(M)$ in general. Nonetheless, we will be able to prove some weaker and more complicated statements along the same lines, involving the notion of a *primary decomposition.* Our other main task is to set up this theory.

Proposition 19.3. [prop-EPM-coprimary]

Let M be an arbitrary A-module. Then $E_P(M)$ is always P-coprimary, and M is P-coprimary if and only if it is isomorphic to $E_P(M)$.

Proof. Because $E_P(M)$ is a quotient of $T_P(M)$, it is easily seen to be a P-torsion module. It is also a submodule of $T_P(M)_P$, and elements of $A \setminus P$ act as isomorphisms on $T_P(M)_P$, so they act injectively on $E_P(M)$. This proves that $E_P(M)$ is P-coprimary.

Conversely, suppose that M is P-coprimary. This firstly means that $T_P(M) = M$, so $E_P(M)$ is the image of the map $M \to M_P$. This is the same as M/N, where

$$N = \{n \in M \mid un = 0 \text{ for some } u \in A \setminus P\}.$$

From the definitions and the P-coprimary condition we see that $N = 0$, so $E_P(M) = M$. □

Definition 19.4. Let M be an arbitrary A-module.

(a) We put $\mathrm{supp}(M) = \{P \in \mathrm{zar}(A) \mid M_P \neq 0\}$, and call this the *support* of M. (Recall here that $\mathrm{zar}(A)$ is the set of all prime ideals in A.)
(b) A *minimal prime for M* is a minimal element of the set $\mathrm{supp}(M)$. We write $\min(M)$ for the set of minimal primes.
(c) An *associated prime* for M is a prime ideal P such that M contains a submodule isomorphic to A/P. We write $\mathrm{ass}(M)$ for the set of associated primes.
(d) We put $\mathrm{reg}(M) = \{a \in A \mid a.1_M \text{ is injective }\}$.
(e) We put $\mathrm{ann}(M) = \{a \in A \mid aM = 0\}$.
(f) We put $\mathcal{A}(M) = \{\mathrm{ann}_A(m) \mid m \in M \setminus \{0\}\}$.

Example 19.5. [eg-pid-ass]

Let A be a principal ideal domain. As we see from Theorem 17.3, any finitely generated A-module M can be written as $M = E_0(M) \oplus \bigoplus_{P \neq 0} E_P(M)$, where $E_0(M)$ is free and $E_P(M)$ is annihilated by some power of P and only finitely many of the summands $E_P(M)$ are nonzero. One can check that

(a) If $E_0(M) \neq 0$ then $\mathrm{supp}(M) = \mathrm{zar}(A) = \{0\} \amalg \{P \mid P \neq 0\}$. However, if $E_0(M) = 0$ then $\mathrm{supp}(M) = \{P \neq 0 \mid E_P(M) \neq 0\}$.
(b) If $E_0(M) \neq 0$ then $\min(M) = \{0\}$, but if $E_0(M) = 0$ (or equivalently $M = T(M)$) then $\min(M) = \mathrm{supp}(M)$.
(c) In all cases $\mathrm{ass}(M) = \{P \in \mathrm{zar}(A) \mid E_P(M) \neq 0\}$. (In fact, we will see that this holds for arbitrary noetherian rings, not just for principal ideal domains.)
(d) $\mathrm{reg}(M)$ is the set of elements $a \in A$ that do not lie in any associated prime ideal.

There is no real difficulty in describing $\mathrm{ann}(M)$ and $\mathcal{A}(M)$ as well, but the notation would be cumbersome.

Example 19.6. [eg-dedekind-ass]

As in Example 16.11, consider the ring $A = \mathbb{Z}[\sqrt{-5}]$ and the ideal M generated by 3 and $1 + \sqrt{-5}$. We will use this repeatedly as a counterexample for various things related to associated primes and primary decompositions. For the moment we just note that A is an integral domain and $A \simeq 3A \leq M \leq A$; this easily implies that $E_0(M) = T_0(M) = M$, whereas $E_P(M) = T_P(M) = 0$ for all $P \neq 0$. We also find that $\mathrm{supp}(M) = \mathrm{zar}(A)$ and $\min(M) = \mathrm{ass}(M) = \{0\}$ and $\mathrm{reg}(M) = A \setminus 0$ and $\mathrm{ann}(M) = 0$ and $\mathcal{A}(M) = \{0\}$.

Example 19.7. [eg-cross-ass]

Consider the ring $A = \mathbb{C}[x, y]$ and the module $M = A/(xy)$. We claim that

$$\begin{aligned}\mathrm{supp}(M) &= \{P \in \mathrm{zar}(A) \mid x \in P \text{ or } y \in P\} \\ &= \{Ax,\ Ay,\ Ax + Ay\} \amalg \{Ax + A(y - \mu) \mid \mu \in \mathbb{C}^\times\} \amalg \{A(x - \lambda) + Ay \mid \lambda \in \mathbb{C}^\times\}.\end{aligned}$$

To see this, put $e = 1 + Axy$, which is the obvious generator of M. We have $M_P = 0$ iff $e/1 = 0$ in M_P, iff there exists $u \in A \setminus P$ with $ue = 0$, iff there exists $u \in A \setminus P$ with $u \in Axy$, iff $xy \notin P$, iff ($x \notin P$ and $y \notin P$). By the contrapositive, we have $P \in \mathrm{supp}(M)$ iff $x \in P$ or $y \in P$. If $x \in P$ then P corresponds to a prime ideal $\overline{P}$ in the quotient ring $A/x = \mathbb{C}[y]$, so $\overline{P}$ must be zero or generated by $y - \mu$ for some $\mu \in \mathbb{C}$. The situation if $y \in P$ is similar, and our more explicit description of $\mathrm{supp}(M)$ follows easily. It follows in turn that $\min(M) = \{Ax, Ay\}$.

Any element $m \in M$ can be written in a unique way as $a + x\,f(x) + y\,g(y)$, where $a \in \mathbb{C}$, and f and g are polynomials. One can check that

- If $a \neq 0$ then $\mathrm{ann}_A(m) = Axy$

- If $a = 0$ but $f \neq 0$ and $g \neq 0$ then again $\text{ann}_A(m) = Axy$
- If $a = 0$ and $f = 0$ but $g \neq 0$ then $\text{ann}_A(m) = Ax$
- If $a = 0$ and $g = 0$ but $f \neq 0$ then $\text{ann}_A(m) = Ay$
- If $a = 0$ and $f = 0$ and $g = 0$ (so $m = 0$) then $\text{ann}_A(m) = A$.

This gives

$$\begin{aligned}\text{ass}(M) &= \{Ax,\ Ay\} \\ \text{reg}(M) &= A \setminus (Ax \cup Ay) \\ \text{ann}(M) &= Axy \\ \mathcal{A}(M) &= \{Ax,\ Ay,\ Axy\}.\end{aligned}$$

(Here we have used the fact that $\text{ass}(M)$ is the set of prime ideals in $\mathcal{A}(M)$; the proof is straightforward, and is given as part of Proposition 19.15 below.)

Example 19.8. [eg-tick-ass]

The previous example involved the associated prime ideals Ax and Ay, neither of which is contained in the other. Some additional phenomena appear in cases where the relevant prime ideals are nested. For example, we can take $A = \mathbb{C}[x, y]$ again, and $M = A/(Axy + Ay^2)$. We first claim that

$$\text{supp}(M) = \{P \in \text{zar}(A) \mid y \in P\} = \{Ay\} \amalg \{A(x - \lambda) + Ay \mid \lambda \in \mathbb{C}\}.$$

To see this, we again let e denote the standard generator of M, so $M_P = 0$ iff $e/1 = 0$ iff there exists $u \in A \setminus P$ with $u \in Axy + Ay^2$. By the contrapositive, we have $P \in \text{supp}(M)$ iff $Axy + Ay^2 \leq P$. If $y \in P$ then it is clear that $Axy + Ay^2 \leq P$. Conversely, if $Axy + Ay^2 \leq P$ then $y^2 \in P$, which means that $y \in P$ as P is prime. It follows easily that $\text{supp}(M)$ is as described, and thus that $\min(M) = \{Ay\}$.

Any element $m \in M$ can be written as $m = f(x) + ay$ for some some polynomial f and some constant $a \in \mathbb{C}$. One can check that

- If $f(0) \neq 0$ then $\text{ann}_A(m) = Axy + Ay^2$
- If $f(0) = 0$ but $f(x) \neq 0$ then $\text{ann}_A(m) = Ay$
- If $f(x) = 0$ but $a \neq 0$ then $\text{ann}_A(m) = Ax + Ay$
- If $f(x) = 0$ and $b = 0$ (so $m = 0$) then $\text{ann}_A(m) = A$.

This gives

$$\begin{aligned}\text{ass}(M) &= \{Ay,\ Ax + Ay\} \\ \text{reg}(M) &= A \setminus (Ax + Ay) \\ \text{ann}(M) &= Axy + Ay^2 \\ \mathcal{A}(M) &= \{Axy + Ay^2,\ Ay,\ Ax + Ay\}.\end{aligned}$$

Proposition 19.9. [prop-ass-supp]

$\text{ass}(M)$ *is always a subset of* $\text{supp}(M)$.

Proof. If $P \in \text{ass}(M)$ then we have an injective homomorphism $\alpha\colon A/P \to M$. Using Proposition 12.13 we deduce that the map $\alpha_P\colon (A/P)_P \to M_P$ is also injective, but $(A/P)_P$ is easily seen to be nonzero, so $M_P \neq 0$, so $P \in \text{supp}(M)$. □

Proposition 19.10. [prop-sum-ass]

For direct sums we have

$$\begin{aligned}\text{supp}(M \oplus N) &= \text{supp}(M) \cup \text{supp}(N) \\ \text{ass}(M \oplus N) &= \text{ass}(M) \cup \text{ass}(N) \\ \text{reg}(M \oplus N) &= \text{reg}(M) \cap \text{reg}(N) \\ \text{ann}(M \oplus N) &= \text{ann}(M) \cap \text{ann}(N).\end{aligned}$$

Proof. Only the claim about $\text{ass}(M \oplus N)$ requires comment. Using the inclusions $M \to M \oplus N$ and $N \to M \oplus N$, it is clear that $\text{ass}(M \oplus N) \supseteq \text{ass}(M) \cup \text{ass}(N)$. In the opposite direction, suppose that $P \in \text{ass}(M \oplus N)$, so there is an injective homomorphism $\alpha\colon A/P \to M \oplus N$. This consists of a pair of homomorphisms $\beta\colon A/P \to M$ and $\gamma\colon A/P \to N$. We claim that at least one of these is injective. If not, we can choose $a, b \in A \setminus P$ with $\beta(a+P) = 0$ and $\gamma(b+P) = 0$. It follows that $\alpha(ab+P) = 0$, but α is injective, so $ab \in P$, which contradicts the fact that P is a prime ideal. This proves the claim, so $P \in \text{ass}(M) \cup \text{ass}(N)$, as required. □

The above can be partially generalised as follows.

Proposition 19.11. [prop-ass-ses]
Suppose that there is a short exact sequence

$$0 \to M \xrightarrow{\phi} U \xrightarrow{\psi} N \to 0.$$

Then

$$\begin{array}{rcl} & \text{supp}(U) & = \text{supp}(M) \cup \text{supp}(N) \\ \text{ass}(M) \subseteq & \text{ass}(U) & \subseteq \text{ass}(M) \cup \text{ass}(N) \\ \text{reg}(M) \cap \text{reg}(N) \subseteq & \text{reg}(U) & \subseteq \text{reg}(M) \\ \text{ann}(M).\,\text{ann}(N) \subseteq & \text{ann}(U) & \subseteq \text{ann}(M) \cap \text{ann}(N). \end{array}$$

Proof. First, Proposition 12.13 gives us a short exact sequence $M_P \to U_P \to N_P$ for each P, and it follows easily that U_P can only be zero if M_P and N_P are both zero. This means that $\text{supp}(U) = \text{supp}(M) \cup \text{supp}(N)$.

Next, if $P \in \text{ass}(M)$ then we have an injective homomorphism $A/P \to M$, which we can compose with ϕ to get an injective homomorphism $A/P \to U$. This shows that $\text{ass}(M) \subseteq \text{ass}(U)$. Suppose instead that we start with an injective homomorphism $\beta\colon A/P \to U$. If $\psi\beta\colon A/P \to N$ is injective, then $P \in \text{ass}(N)$. Otherwise we can choose an element $x \in A \setminus P$ such that $\psi\beta(x+P) = 0$. As $\text{image}(\phi) = \ker(\psi)$ there exists $m \in M$ with $\phi(m) = \beta(x+P)$. As both ϕ and β are injective, we see that $\text{ann}(m) = \text{ann}(\phi(m)) = P$, so there is an injective homomorphism $\alpha\colon A/P \to M$ with $\alpha(a+P) = am$. This means that $P \in \text{ass}(M)$, as required.

We now consider regular elements. Suppose that $a \in \text{reg}(M) \cap \text{reg}(N)$. If $e \in U$ with $ae = 0$, then we also have $a\,\psi(e) = \psi(ae) = 0$, but a is regular on N so $\psi(e) = 0$. As $\text{image}(\phi) = \ker(\psi)$ we see that $e = \phi(m)$ for some m, and $\phi(am) = a\,\phi(m) = ae = 0$, but ϕ is injective so $am = 0$. Moreover, a is regular on M so $m = 0$, so $e = \phi(m) = 0$. This proves that $\text{reg}(M) \cap \text{reg}(N) \subseteq \text{reg}(U)$. In the other direction, if $a \in \text{reg}(U)$ and $am = 0$ for some $m \in M$ then we can apply the map ϕ to see that $a\,\phi(m) = 0$, but a is regular on U so $\phi(m) = 0$, and ϕ is injective so $m = 0$. Thus, $\text{reg}(U) \subseteq \text{reg}(M)$.

Finally, suppose that $a \in \text{ann}(U)$. Any $n \in N$ has $n = \psi(e)$ for some $e \in U$, so $an = \psi(ae) = \psi(0) = 0$; thus $a \in \text{ann}(N)$. Also, for $m \in M$ we have $\phi(m) \in U$ so $\phi(am) = a\,\phi(m) = 0$, but ϕ is injective so $am = 0$; thus $a \in \text{ann}(M)$. In the opposite direction, suppose that $b \in \text{ann}(M)$ and $c \in \text{ann}(N)$. For $e \in U$ we have $\psi(ce) = c\,\psi(e) = 0$, so $ce = \phi(m)$ for some m, so $bce = \phi(bm) = \phi(0) = 0$; thus $bc \in \text{ann}(U)$. □

Proposition 19.12. *If there exist injective homomorphisms*

$$M \xrightarrow{\alpha} N \xrightarrow{\beta} M,$$

then $\text{supp}(M) = \text{supp}(N)$ *and* $\min(M) = \min(N)$ *and* $\text{ass}(M) = \text{ass}(N)$ *and* $\text{reg}(M) = \text{reg}(N)$ *and* $\text{ann}(M) = \text{ann}(N)$ *and* $\mathcal{A}(M) = \mathcal{A}(N)$.

Proof. Just from the existence of an injective homomorphism $M \to N$ we get $\text{supp}(M) \subseteq \text{supp}(N)$ and $\text{ass}(M) \subseteq \text{ass}(N)$ and $\text{reg}(N) \subseteq \text{reg}(M)$ and $\text{ann}(N) \subseteq \text{ann}(M)$ and $\mathcal{A}(M) \subseteq \mathcal{A}(N)$. As we have injective homomorphisms in both directions, we conclude that $\text{supp}(M) = \text{supp}(N)$ and $\text{ass}(M) = \text{ass}(N)$ and $\text{reg}(M) = \text{reg}(N)$ and $\text{ann}(M) = \text{ann}(N)$ and $\mathcal{A}(M) = \mathcal{A}(N)$. From $\text{supp}(M) = \text{supp}(N)$ it is clear that $\min(M) = \min(N)$. □

Proposition 19.13. [prop-ass-chain]
For any finitely generated module M there is a chain

$$0 = M_0 \subseteq M_1 \subseteq \cdots \subseteq M_n = M,$$

and a list $P_1, \dots, P_n$ of prime ideals, such that $M_i/M_{i-1} \simeq A/P_i$ for $1 \leq i \leq n$. Moreover, we have $\text{ass}(M) \subseteq \{P_1, \dots, P_n\}$ *(so* $\text{ass}(M)$ *is finite).*

Proof. Let $\mathcal{C}$ be the family of all submodules of M that have a chain as described. The zero module lies in $\mathcal{C}$ so $\mathcal{C} \neq \emptyset$. Thus, Proposition 18.8 tells us that $\mathcal{C}$ has a maximal element, say N. As $N \in \mathcal{C}$ we can choose a chain

$$0 = M_0 \subseteq M_1 \subseteq \cdots \subseteq M_n = N$$

for some n, and a list $P_1, \dots, P_n$ of prime ideals, such that $M_i/M_{i-1} \simeq A/P_i$ for $1 \leq i \leq n$. If $N \neq M$ then $\operatorname{ass}(M/N) \neq \emptyset$, so we can choose a prime ideal P_{n+1} and an injective homomorphism $\alpha \colon A/P_{i+1} \to M/N$. The image of this homomorphism will have the form M_{n+1}/N for some submodule $M_{n+1} \subseteq M$ containing $N = M_n$. This module M_{n+1} will then lie in $\mathcal{C}$, contradicting the maximality of N. We conclude that $N = M$ after all, so we have a chain of the required type for M itself. Note that when P is prime, the annihilator of any nontrivial element of A/P is just P. It follows that $\operatorname{ass}(M_i/M_{i-1}) = \operatorname{ass}(R/P_i) = \{P_i\}$. Using the short exact sequences $M_{i-1} \to M_i \to M_i/M_{i-1}$ we get $\operatorname{ass}(M_i) \subseteq \operatorname{ass}(M_{i-1}) \cup \{P_i\}$, and so $\operatorname{ass}(M) = \operatorname{ass}(M_n) \subseteq \{P_1, \dots, P_n\}$. □

Remark 19.14. [rem-extra-quotient]

In general, we cannot arrange to have $\operatorname{ass}(M) = \{P_1, \dots, P_n\}$ in the above construction. To see this, consider the pair (A, M) as in Example 16.11, where $\operatorname{ass}(M) = \{0\}$. Suppose we have a chain of submodules $0 = M_0 < \cdots < M_n = M$ with $M_i/M_{i-1} \simeq A/0 = A$. The short exact sequences $M_{i-1} \to M_i \to A$ must split, so we have $M \simeq A^n$, with basis $m_1, \dots, m_n$ say. As M is a non-principal ideal, we must have $n > 1$. As M is just a subset of A, we have a relation $m_2.m_1 + (-m_1).m_2 = 0$, showing that $m_1, \dots, m_n$ cannot be a basis after all. Thus, there can be no chain of the indicated type.

Proposition 19.15. [prop-max-ann]

(a) *The associated primes are precisely the prime ideals that lie in $\mathcal{A}(M)$, so* $\operatorname{ass}(M) = \operatorname{zar}(A) \cap \mathcal{A}(M)$.
(b) *Every element of $\mathcal{A}(M)$ is contained in a maximal element of $\mathcal{A}(M)$.*
(c) *Every maximal element of $\mathcal{A}(M)$ is an associated prime.*
(d) $\operatorname{reg}(M)$ *is the complement of the union of the associated primes.*

In particular, if $M \neq 0$ then $\operatorname{ass}(M) \neq \emptyset$.

Proof.

(a) If P is an associated prime then there is an isomorphism from A/P to some submodule of M, or equivalently there is an injective homomorphism $\alpha \colon A/P \to M$. Put $m = \alpha(1 + P)$, so $\alpha(x + P) = \alpha(x.(1 + P)) = xm$ for all x. As α is injective we see that $\operatorname{ann}_A(m) = P$. As $P \neq A$ we have $m \neq 0$, so $P \in \mathcal{A}(M)$. This shows that $\operatorname{ass}(M) \subseteq \operatorname{zar}(A) \cap \mathcal{A}(M)$, and the reverse inclusion can be proved in essentially the same way.
(b) If $I \in \mathcal{A}(M)$ then $\{J \in \mathcal{A}(M) \mid J \supseteq I\}$ is a nonempty family of ideals in the noetherian ring A, so it has a maximal element by Proposition 18.8. It is clear that any such element will also be maximal in $\mathcal{A}(M)$.
(c) Now let P be a maximal element in $\mathcal{A}(M)$, and choose $m \in M \setminus \{0\}$ such that $P = \operatorname{ann}_A(m)$. As $m \neq 0$ we have $1 \notin P$. Suppose that $a \notin P$, so $am \neq 0$. It follows that $\operatorname{ann}_A(am) \in \mathcal{A}(M)$ and $P \subseteq \operatorname{ann}_A(am)$ so by maximality $\operatorname{ann}_A(am) = P$. In particular, if b is another element with $b \notin P$ then $b \notin \operatorname{ann}_A(am)$ so $abm \neq 0$ so $ab \notin P$. This proves that P is prime, so $P \in \operatorname{ass}(M)$ by (a).
(d) It is clear that the complement of $\operatorname{reg}(M)$ is the union of all the ideals in $\mathcal{A}(M)$. From (b) and (c) we see that this is the same as the union of the associated prime ideals.

□

Corollary 19.16. [cor-exists-regular]

Let $I \leq A$ be an ideal that is not contained in any of the associated primes for M. Then $\operatorname{reg}(M) \cap I \neq \emptyset$.

Proof. Proposition 19.13 tells us that that there are only finitely many associated primes, we can use Proposition 5.36. This tells us that I is not contained in the union of the associated primes. The claim therefore follows from Proposition 19.15(d). □

Proposition 19.17. [prop-fraction-ass]

For any multiplicative set $U \subseteq A$ we have

$$\mathrm{ass}_{A[U^{-1}]}(M[U^{-1}]) = \{P[U^{-1}] \mid P \in \mathrm{ass}_A(M),\ P \cap U = \emptyset\}.$$

In particular, we have $M[U^{-1}] \neq 0$ iff $\mathrm{ass}_{A[U^{-1}]}(M[U^{-1}]) \neq \emptyset$ iff there is a prime $P \in \mathrm{ass}(M)$ with $P \cap U = \emptyset$.

Proof. First, for an element $m/u \in M[U^{-1}]$ we have $m/u = 0$ iff $mw = 0$ for some $w \in U$ iff $\mathrm{ann}_A(m) \cap U = \emptyset$. Similarly, if $(a/v)(m/u) = 0$ then $awm = 0$ for some $w \in U$, and we can rewrite a/v as $(aw)/(vw)$ with $aw \in \mathrm{ann}_A(m)$ and $vw \in U$. This gives

$$\mathrm{ann}_{A[U^{-1}]}(m/u) = \mathrm{ann}_A(m)[U^{-1}].$$

Suppose that $P \in \mathrm{ass}_A(M)$ with $P \cap U = \emptyset$. Then there is an element $m \in M \setminus \{0\}$ with $\mathrm{ann}_A(m) = P$. It follows that $\mathrm{ann}_{A[U^{-1}]}(m/1) = P[U^{-1}]$, and $P[U^{-1}]$ is a prime ideal in $A[U^{-1}]$ by Proposition 8.18, so $P[U^{-1}] \in \mathrm{ass}_{A[U^{-1}]}(M[U^{-1}])$.

Conversely, suppose we have an ideal $P^* \in \mathrm{ass}_{A[U^{-1}]}(M[U^{-1}])$. Proposition 8.18 tells us that the set $P = \{a \in A \mid a/1 \in P^*\}$ is a prime ideal in A with $P \cap U = \emptyset$ and that $P^* = P[U^{-1}]$. As P^* is an associated prime, there is an element $m/u \in M[U^{-1}]$ with $\mathrm{ann}_{A[U^{-1}]}(m/u) = P^*$. Next, as A is noetherian we can choose a finite list $a_0, \dots, a_{d-1}$ that generates P. Now $a_i/1 \in P^*$ so $a_i m/u = 0$ so there exists $v_i \in U$ with $a_i v_i m = 0$. Put $v = \prod_i v_i \in U$, and note that $a_i v m = 0$ for all i, so $P \subseteq \mathrm{ann}_A(vm)$. On the other hand, if $bvm = 0$ then the element $b/1 = (bv)/v$ lies in $\mathrm{ann}_{A[U^{-1}]}(m/u) = P^*$, so $b \in P$. We conclude that $\mathrm{ann}_A(vm) = P$, so $P \in \mathrm{ass}_A(M)$. □

Corollary 19.18. [cor-loc-ass]

Let Q be a prime ideal in A. Then the following are equivalent:

(a) $Q \in \mathrm{supp}(M)$
(b) $Q \supseteq \mathrm{ann}_A(M)$
(c) $Q \supseteq \sqrt{\mathrm{ann}_A(M)}$
(d) *There is an associated prime $P \in \mathrm{ass}(M)$ with $Q \supseteq P$.*

Proof. Take $U = A \setminus P$ in the Proposition to see that (a) and (d) are equivalent. As Q is prime, we see that $a \in Q$ iff $a^k \in Q$ for some $k > 0$. Using this, we see that (b) and (c) are equivalent.

Next, if there is an element $v \in \mathrm{ann}_A(M) \setminus Q$ then for all $m/u \in M_Q$ we have $m/u = (vm)/(vu) = 0$, so $M_Q = 0$. Conversely, suppose that $M_Q = 0$. By assumption there is a finite list $m_0, \dots, m_{d-1}$ that generates M. The elements $m_i/1 \in M_Q$ must be zero, so there are elements $v_i \in A \setminus Q$ with $v_i m_i = 0$. The product $v = \prod_i v_i$ then lies in $\mathrm{ann}_A(M) \setminus Q$. It follows that (a) and (b) are equivalent. □

Corollary 19.19. [cor-min-ass]

We have $\min(M) \subseteq \mathrm{ass}(M) \subseteq \mathrm{supp}(M)$, so $\min(M)$ (which was originally defined as the set of minimal elements in $\mathrm{supp}(M)$), is also the set of minimal elements in $\mathrm{ass}(M)$.

Proof. We saw in Proposition 19.9 that $\mathrm{ass}(M) \subseteq \mathrm{supp}(M)$. Now suppose that $Q \in \min(M)$, so in particular $Q \in \mathrm{supp}(M)$. By the previous corollary, there is an associated prime $P \in \mathrm{ass}(M)$ with $P \subseteq Q$. Now both P and Q lie in $\mathrm{ass}(M)$, but Q is minimal in $\mathrm{supp}(M)$ by hypothesis, so we must have $P = Q$, so $Q \in \mathrm{ass}(M)$. □

Proposition 19.20. [prop-primary-collect]

Submodules and direct sums of P-coprimary modules are again P-coprimary.

Proof. Let M and N be P-coprimary modules, and let L be a submodule of $M \oplus N$. If $a \in P$ then there are integers $m, n \geq 0$ such that $a^m M = 0$ and $a^n N = 0$; it follows that $a^{\max(m,n)} L = 0$. On the other hand, if $a \notin P$ then $a.1_M$ and $a.1_N$ are injective, so the map $a.1_L = (a.1_M \oplus a.1_N)|_L$ is also injective. It follows that L is P-coprimary. The special cases where $L = M \oplus N$ or $N = 0$ give the two statements in the Proposition. □

Proposition 19.21. [prop-coprimary-ass]

Suppose that M is nontrivial and P-coprimary. Then

(a) $\mathrm{reg}(M) = A \setminus P$, *and* $\sqrt{\mathrm{ann}_A(M)} = P$.
(b) $E_P(M) = T_P(M) = M$.
(c) *If $Q \not\leq P$ then $E_Q(M) = T_Q(M) = 0$.*

(d) *If $Q < P$ then $T_Q(M) = M$ but $E_Q(M) = M_Q = 0$.*
(e) *If $Q \neq P$ then $E_Q(M) = 0$.*
(f) $\operatorname{supp}(M) = \{Q \in \operatorname{zar}(A) \mid Q \geq P\}$.
(g) $\min(M) = \operatorname{ass}(M) = \{P\}$.

Proof. It is clear from the definitions that $\operatorname{reg}(M) = A \setminus P$ and thus that $\sqrt{\operatorname{ann}(M)} \leq P$. For the converse, let $m_1, \dots, m_r$ be generators for M. If $a \in P$ then by assumption there are integers $k_i \geq 0$ with $a^{k_i} m_i = 0$ for all i. It follows that the number $k = \max(k_1, \dots, k_r)$ satisfies $a^k m_i = 0$ for all i, so $a^k \in \operatorname{ann}_A(M)$. This proves (a).

Claim (b) is just a reminder of Proposition 19.3.

For claim (c), suppose that $Q \not\leq P$, so we can choose $u \in Q \setminus P$. If $m \in T_Q(M)$ then we have $u^k m = 0$ for some k, but u acts injectively on M by assumption, so $T_Q(M) = 0$. It follows that $T_Q(M)_Q = 0$ and $E_Q(M) = 0$ as claimed.

For claim (d), suppose instead that $Q < P$. For every $m \in M$ we have $P^k m = 0$ for large k, so certainly $Q^k m = 0$. This proves that $T_Q(M) = M$. Now choose $u \in P \setminus Q$. For any element $m/v \in M_Q$ we have $m \in M$ so $u^k m = 0$ for some k, so $m/v = (u^k m)/(u^k v) = 0$. This proves that $T_Q(M)_Q = M_Q = 0$. As $E_Q(M)$ is the image of the natural map $T_Q(M) \to T_Q(M)_Q$, we see that $E_Q(M) = 0$ as well.

Claim (e) is simply a combination of (c) and (d).

Claim (f) follows from (a) together with Corollary 19.18. It follows in turn that $\min(M) = \{P\}$, so Corollary 19.19 gives $P \in \operatorname{ass}(M)$. Conversely, if $Q \in \operatorname{ass}(M)$ then there must exist $m \in M \setminus 0$ with $\operatorname{ann}_A(m) = Q$, and as Q is prime we have $\sqrt{\operatorname{ann}_A(m)} = \sqrt{Q} = Q$. However, the coprimary condition gives $\sqrt{\operatorname{ann}_A(m)} = P$, so we must have $Q = P$. This completes the proof of (g). □

20. Primary decomposition

In this section we again assume that A is a noetherian ring, and M is a finitely generated A-module.

Definition 20.1. [defn-primdec]

A *primary decomposition* of M consists of a family of modules $M(P)$ (for all $P \in \operatorname{zar}(A)$) together with homomorphisms $\pi_P \colon M \to M(P)$ such that

(a) $M(P)$ is P-coprimary for all P.
(b) $M(P) = 0$ for all but finitely many P.
(c) The homomorphisms $\pi_P \colon M \to M(P)$ are all surjective.
(d) The combined map $\pi \colon M \to \bigoplus_P M(P)$ is injective.
(e) Whenever $M(P) \neq 0$, the combined map $M \to \bigoplus_{Q \neq P} M(Q)$ has nontrivial kernel (which we denote by $M[P]$).

We will show that every finitely generated module has a primary decomposition.

Remark 20.2. Primary decompositions are traditionally defined in terms of the submodules $\ker(\pi_P)$ rather than the quotient modules $M(P)$, but that approach obscures the analogy with the theory of modules over PIDs, so we have avoided it.

Proposition 20.3. [prop-primdec-theta]

Suppose we have a primary decomposition as above, and we put $M' = \bigoplus_P M(P)$ and $M'' = \bigoplus_P M[P]$. Put $\theta_P = \pi_P|_{M[P]} \colon M[P] \to M(P)$. Then

(a) *θ_P is injective for all P.*
(b) *The evident maps $M'' \to M \to M'$ are also injective, and their composite is $\bigoplus_P \theta_P$.*
(c) *The modules $M[P]$ and $M(P)$ are both P-coprimary.*

Proof. Let σ be the evident map $M'' \to M$. The composite

$$\bigoplus_Q M[Q] = M'' \xrightarrow{\sigma} M \xrightarrow{\pi} M' = \bigoplus_P M(P)$$

decomposes into homomorphisms $\pi_P|_{M[Q]} \colon M[Q] \to M(P)$. By the definition of $M[Q]$, we have $\pi_P|_{M[Q]} = 0$ unless $P = Q$. Thus, $\pi\sigma$ is the direct sum of the maps θ_P, and the restriction of π to $M[P]$ is essentially

θ_P. As π is injective, we conclude that θ_P is injective. It follows that the map $\pi\sigma = \bigoplus_P \theta_P$ is injective, and thus that σ is injective. As $\theta_P \colon M[P] \to M(P)$ is injective, Proposition 19.20 tells us that $M[P]$ is P-coprimary. □

Proposition 20.4. *In any primary decomposition, we have $M(P) \neq 0$ iff $P \in \operatorname{ass}(M)$.*

Proof. We have injective homomorphisms $M'' \to M \to M'$, where $M'' = \bigoplus_P M[P]$ and $M' = \bigoplus_P M(P)$. It follows that $\operatorname{ass}(M'') \subseteq \operatorname{ass}(M) \subseteq \operatorname{ass}(M')$. Using Propositions 19.10 and 19.21 we see that $\operatorname{ass}(M'') = \{P \mid M[P] \neq 0\}$ and $\operatorname{ass}(M') = \{P \mid M(P) \neq 0\}$. From the definition of a primary decomposition, we have $M[P] \neq 0$ iff $M(P) \neq 0$. The claim is now clear. □

Proposition 20.5. [prop-T-exact]

Suppose that the sequence $0 \to L \xrightarrow{\alpha} M \xrightarrow{\beta} N$ is exact. Then the rows in the following diagram are also exact, except that the third row need not be exact at $E_P M$.

$$\begin{tikzcd}
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \\
0 \arrow[u, equal] \arrow[r] & T_P(L) \arrow[u, tail] \arrow[r] \arrow[d, two heads] & T_P(M) \arrow[u, tail] \arrow[r] \arrow[d, two heads] & T_P(N) \arrow[u, tail] \arrow[d, two heads] \\
0 \arrow[u, equal] \arrow[d, equal] \arrow[r] & E_P(L) \arrow[r] \arrow[d, tail] & E_P(M) \arrow[r] \arrow[d, tail] & E_P(N) \arrow[d, tail] \\
0 \arrow[r] & T_P(L)_P \arrow[r] & T_P(M)_P \arrow[r] & T_P(N)_P.
\end{tikzcd}$$

Proof. The first row is exact by hypothesis. For the second row, injectivity of $L \to M$ clearly implies injectivity of $T_P(L) \to T_P(M)$. Now consider an element $m \in T_P(M)$, say with $P^k m = 0$, and suppose that $\beta(m) = 0$ in $T_P(N)$. By the exactness of the original sequence, we have $m = \alpha(l)$ for some $l \in L$. This satisfies $\alpha(P^k l) = P^k \alpha(l) = P^k m = 0$, but α is injective so $P^k l = 0$. This means that l is an element of $T_P(L)$ with $\phi(l) = m$, completing the proof that the second row is exact. Exactness of the last row therefore follow by Proposition 12.13. By definition, the third row maps injectively to the last row, and it follows easily that the map $E_P(L) \to E_P(M)$ is injective. This completes the proof (as we are making no claim about exactness at $E_P(M)$). □

Proposition 20.6. *Any primary decomposition of M gives a natural diagram as follows:*

$$\begin{tikzcd}
M[P] \arrow[r, tail] \arrow[d, tail] & E_P(M) \arrow[r, tail] \arrow[d, tail] & M(P) \arrow[d, tail] \\
M[P]_P \arrow[r, "\simeq"'] & T_P(M)_P \arrow[r, tail] & M(P)_P.
\end{tikzcd}$$

Proof. First note that $E_P(M[Q]) = E_P(M(Q)) = 0$ for all $Q \neq P$, by Proposition 19.21. Moreover, both $E_P(\cdot)$ and $T_P(\cdot)_P$ preserve monomorphisms, by Proposition 20.5. We can thus apply these functors to the maps

$$\begin{tikzcd}
\bigoplus_Q M[Q] \arrow[r, tail] & M \arrow[r, tail] & \bigoplus_Q M(Q)
\end{tikzcd}$$

to get a diagram as claimed, except that we do not yet know that the map $M[P]_P \to T_P(M)_P$ is surjective. For this, we consider the sequence

$$0 \to M[P] \to M \to \bigoplus_{Q \neq P} M(Q),$$

which is exact by the definition of $M[P]$. Proposition 20.5 tells us that it will remain exact if we apply $T_P(\cdot)_P$. Here $T_P(M(Q))_P = 0$ for all $Q \neq P$, whereas $T_P(M[P])_P = M[P]_P$. We therefore have an exact sequence

$$0 \to M[P]_P \to F_P M \to 0,$$

showing that the map $M[P]_P \to T_P(M)_P$ is an isomorphism. □

Although our main interest is in coprimary modules, it turns out to be useful to work temporarily with modules satisfying a slightly stronger condition which we now introduce.

Definition 20.7. [defn-coirr-submodule]
We say that an A-module M is *coirreducible* if the intersection of any two nontrivial submodules is nontrivial. We say that a submodule $L \leq M$ is *irreducible* if M/L is coirreducible. Equivalently, L is irreducible iff whenever U and V are strictly larger submodules of M, the intersection $U \cap V$ is also strictly larger than L.

We also say that L is *P-primary* in M if M/L is P-coprimary.

Remark 20.8. [rem-primary-intersection]
If L_0 and L_1 are both P-primary, then the evident embedding $M/(L_0 \cap L_1) \to M/L_0 \times M/L_1$ shows that $L_0 \cap L_1$ is also P-primary.

Proposition 20.9. [prop-coirr-submodule]
Let M be a finitely generated A-module. If M is coirreducible then the ideal $P = \sqrt{\text{ann}_A(M)}$ is prime and M is P-coprimary.

Proof. Consider an element $a \in A$. The submodules $\text{ann}_M(a^k) \subseteq M$ form an ascending chain, which must eventually be constant. Thus, for some n we have $\text{ann}_M(a^n) = \text{ann}_M(a^{n+1})$. We claim that $\text{ann}_M(a) \cap a^n M = 0$. To see this, suppose that $x \in \text{ann}_M(a) \cap a^n M$, so $x = a^n y$ for some y. Now $a^{n+1}y = ax = 0$, so $y \in \text{ann}_M(a^{n+1}) = \text{ann}_M(a^n)$, so $a^n y = 0$ or in other words $x = 0$ as claimed. As M is assumed to be coirreducible, we must either have $\text{ann}_M(a) = 0$ or $a^n M = 0$. If $\text{ann}_M(a) = 0$ then $a.1_M$ is injective. On the other hand, if $a^n M = 0$ then $a.1_M$ is nilpotent and $a^n \in \text{ann}_A(M)$ so $a \in \sqrt{\text{ann}_A(M)} = P$. It follows that P is prime and M is P-coprimary. □

Proposition 20.10. [prop-irr-decomp]
Every submodule of M can be written as the intersection of some finite list of irreducible submodules.

Proof. Let $\mathcal{C}$ be the set of submodules that can be written as the intersection of some finite list of irreducible submodules. Clearly every irreducible submodule lies in $\mathcal{C}$, as does M itself (use the empty list). Moreover, if $L, N \in \mathcal{C}$ then it is clear that $L \cap N \in \mathcal{C}$. Now suppose that N is a submodule of M, and that every strictly larger submodule lies in $\mathcal{C}$. If N is irreducible then it lies in $\mathcal{C}$. Otherwise, we have $N = U \cap V$ for some submodules $U, V \subseteq M$ that are strictly larger than N, so U and V lie in $\mathcal{C}$, so $N = U \cap V \in \mathcal{C}$. It follows by noetherian induction that every submodule lies in $\mathcal{C}$, as claimed. □

Theorem 20.11. [thm-primdec]
Every finitely generated module M has a primary decomposition.

Proof. Propositions 20.9 and 20.10 show that there exist lists $L_1, \dots, L_d$ of primary submodules with $\bigcap_i L_i = 0$. Choose such a list which is as short as possible, and let P_i be the prime ideal such that L_i is P_i-primary. If we had $P_i = P_j$ for some $i \neq j$, then we could replace L_i and L_j by $L_i \cap L_j$, giving a shorter list of the required type. This is impossible by assumption, so the ideals P_i must all be different. We define $M\{P\} = L_i$ if $P = P_i$ for some i, and $M\{P\} = M$ otherwise. We then put $M(P) = M/M\{P\}$, and note that this is P-coprimary for all P. We have $\bigcap_P M\{P\} = \bigcap_i L_i = 0$, so the natural map $M \to \bigoplus_P M(P)$ is injective. Next, observe that $M[P_k] = \bigcap_{i \neq k} L_i$, and this is nontrivial by our minimality assumption. We therefore have a primary decomposition as claimed. □

21. Artinian rings

Definition 21.1. [defn-artinian]
We say that a ring A is *artinian* if for every descending chain

$$A \supseteq J_0 \supseteq J_1 \supseteq J_2 \supseteq \cdots$$

of ideals, there exists N such that $J_n = J_N$ for all $n \geq N$.

More generally, let M be a module over an arbitrary ring A. We say that M is *artinian* if for every descending chain

$$M \supseteq M_0 \supseteq M_1 \supseteq M_2 \supseteq \cdots$$

of submodules, there exists N such that $M_n = M_N$ for all $n \geq N$. Thus, an artinian ring is just a ring A that is artinian as an A-module.

Remark 21.2. [`rem-dcc`]

A ring A is artinian if and only if every nonempty family of ideals has a minimal element. The proof is essentially the same as for Proposition 18.8.

Example 21.3. [`eg-finite-artinian`]

Suppose that A has only finitely many elements. Given a descending chain of ideals J_k, the numbers $|J_k|$ form a weakly decreasing sequence of positive integers, so there exists N such that $|J_n| = |J_N|$ for all $n \geq N$. As $J_n \subseteq J_N$ this implies that $J_n = J_N$. We therefore see that A is artinian.

Example 21.4. [`eg-length-artinian`]

Let M be a module of finite length over a ring A. Given a descending chain of submodules M_k, we again have a weakly decreasing sequence of nonnegative integers $\text{len}(M_k)$, which must eventually be constant, and it again follows that M_k is independent of k when k is sufficiently large. Thus, M is artinian.

Example 21.5. [`eg-dim-artinian`]

Let M be a module over a field K. We can then identify M with $\text{Free}_I(K)$ for some set I. If I is infinite we can choose an infinite sequence of distinct elements $(i_n)_{n\in\mathbb{N}}$, and let M_n be the span of $\{e_{i_k} \mid k \geq n\}$; this gives an infinite descending chain of submodules that never stabilises, showing that M is not artinian. Thus, we see that K-modules are artinian if and only if they have finite dimension.

Remark 21.6. [`rem-artinian-constructs`]

Let A be an artinian ring. For any ideal $I \subseteq A$ the partially ordered set $\text{idl}(A/I)$ can be identified with $\{J \in \text{idl}(A) \mid I \subseteq J\}$, and using this we see that A/I is artinian. Similarly, if $U \subseteq A$ is a multiplicative set then $\text{idl}(A[U^{-1}])$ can be identified with $\text{sat}_U(A) \subseteq \text{idl}(A)$, so $A[U^{-1}]$ is artinian.

Note also that if $A = B \times C$ then $\text{idl}(A) \simeq \text{idl}(B) \times \text{idl}(C)$, and it follows easily that A is artinian iff both B and C are artinian.

Theorem 21.7. [`thm-artinian`]

A ring A is artinian if and only if it is noetherian and all prime ideals are maximal. If so, then A is a finite product of local rings in which the maximal ideal is finitely generated and nilpotent.

Proof. Combine Lemmas 21.8 to 21.11 below. □

Lemma 21.8. [`lem-artinian-a`]

Any artinian ring is a finite product of indecomposable artinian rings.

Proof. Let A be artinian, and let E be the set of idempotents. We say that idempotents e_0 and e_1 are *disjoint* if $e_0e_1 = 0$, and we say that an idempotent e is *primitive* if it is nonzero but cannot be expressed as the sum of two disjoint, nonzero idempotents. Note that we always have a splitting $A = Ae \times A(1 - e)$, and the factor Ae is indecomposable iff e is primitive.

Let E_0 be the set of primitive idempotents, let E_1 be the set of idempotents that can be represented as a finite disjoint sum of primitive idempotents, and put $E_2 = E \setminus E_1$. It will suffice to show that $1 \in E_1$. In fact we claim that $E_1 = E$, or equivalently $E_2 = \emptyset$. To see this, put $\mathcal{E}_2 = \{Ae \mid e \in E_2\}$. If $E_2 \neq \emptyset$, then we can choose $e \in E_2$ such that Ae is minimal in $\mathcal{E}_2$ (by the artinian condition). As e lies in E_2 it cannot be zero or primitive, so $e = e_0 + e_1$ for some disjoint nonzero idempotents e_0, e_1. As $e \notin E_1$, the elements e_0 and e_1 cannot both lie in E_1. We may assume without loss of generality that $e_0 \notin E_1$, so $Ae_0 \in \mathcal{E}_2$, but $e \notin Ae_0$ so this contradicts the assumed minimality of Ae. Thus E_2 must be empty after all. □

Lemma 21.9. [`lem-artinian-b`]

Let A be an indecomposable artinian ring, and put $M = \text{Nil}(A)$. Then M is finitely generated and satisfies $M^n = 0$ for some n, and every element of $A \setminus M$ is invertible, so A is a local ring. Moreover, A is noetherian.

Proof. For any element $a \in A$ we see that the ideals Aa^n form a descending chain, which must eventually stabilise. It follows that for some n we have $a^n \in Aa^{n+1}$, so for some $x \in A$ we have $a^n = a^{n+1}x$. It follows inductively that $a^n = a^{n+k}x^k$ for all $k \geq 0$, and in particular $a^n = a^{2n}x^n$. It follows from this that the element $e = a^n x^n$ is idempotent, with $a^n = a^n e$. As A is indecomposable we have $e = 0$ or $e = 1$. If $e = 0$ then the equation $a^n = a^n e$ shows that a is nilpotent. If $e = 1$ then the equation $a^n x^n = e = 1$ shows that a is invertible. We now see that every element of $A \setminus M$ is invertible, so A is local, with M as the unique maximal ideal. The ideals M^k form a descending chain, so for some $n \in \mathbb{N}$ we must have $M^n = M^{n+1}$. We claim that in fact $M^n = 0$. If not, put $\mathcal{J} = \{J \in \text{idl}(A) \mid JM^n \neq 0\}$, and note that this is nonempty because it contains A. By the artinian condition, we can choose a minimal element $J \in \mathcal{J}$. As $JM^n \neq 0$ we can choose $a \in J$ with $aM^n \neq 0$. As $M^{n+1} = M^n$ we see that $(aM)M^n \neq 0$ so $aM \in \mathcal{J}$ and also $aM \subseteq J$. As J is assumed to be minimal, we must have $aM = J$, so in particular $a \in aM$, so $a(1-b) = 0$ for some $b \in M$. However, every element of M is nilpotent, so $1 - b$ is invertible, so $a = 0$, which contradicts the assumption that $aM^n \neq 0$. Thus, we must have $M^n = 0$ after all.

Next, the quotient $K = A/M$ is a field, and M/M^2 can be regarded as a vector space over K. Any descending chain of vector subspaces gives a descending chain of ideals in the artinian ring A/M^2, and so must eventually stabilise. It follows that M/M^2 has finite dimension over K, so we can choose a finite subset $F \subseteq M$ such that the image in M/M^2 is a basis. If we let I be the ideal in A generated by F, we find that $M = I + M^2$. This in turn gives $M^2 = IM + M^3 \subseteq I + M^3$, and by combining this with $M = I + M^2$ we get $M = I + M^3$. An evident inductive extension gives $M = I + M^k$ for all k, and by taking $k = n$ we get $M = I$. Thus, M is finitely generated. It follows that M^k is finitely generated for all k, so M^k/M^{k+1} has finite dimension over K, and thus has finite length as an A-module. Using the short exact sequences $M^k/M^{k+1} \to A/M^{k+1} \to A/M^k$, we deduce by induction on k that A/M^k has finite length. Taking $k = n$, we see that A itself has finite length. It follows that for all ideals $J \subseteq A$ we have $\text{len}(J) \leq \text{len}(A) < \infty$. Thus, if we have an ascending chain of ideals $(J_k)_{k \in \mathbb{N}}$, then the sequence $(\text{len}(J_k))_{k \in \mathbb{N}}$ is nondecreasing and bounded above, so it must eventually be constant. Thus, for large k we have $J_k \subseteq J_{k+1}$ with $\text{len}(J_{k+1}/J_k) = \text{len}(J_{k+1}) - \text{len}(J_k) = 0$, so $J_{k+1} = J_k$. It follows that A is noetherian. □

Lemma 21.10. [`lem-artinian-c`]

Let A be a noetherian local ring in which every element of the maximal ideal is nilpotent. Then A is artinian.

Proof. Let M be the maximal ideal. As A is noetherian, we can choose a finite list of elements $a_0, \dots, a_{d-1}$ that generates M. By assumption, each element a_i is nilpotent, so $a_i^{n_i+1} = 0$ for some integer n_i. We put $n = \sum_i n_i$.

Next, for any sequence $\alpha = (\alpha_0, \dots, \alpha_{d-1}) \in \mathbb{N}^d$ we put $a^\alpha = \prod_i a_i^{\alpha_i}$ and $|\alpha| = \sum_i \alpha_i$. We then find that $\{a^\alpha \mid |\alpha| = m\}$ is a generating set for M^m. We also find that a^α can only be nonzero if $\alpha_i \leq n_i$ for all i, which implies that $|\alpha| \leq n$. It follows from this that $M^{n+1} = 0$.

Next, put $K = A/M$, which is a field. The quotient M^k/M^{k+1} is a finitely generated K-module and so has finite length as an A-module. It follows by induction using the short exact sequences $M^k/M^{k+1} \to A/M^{k+1} \to A/M^k$ that A/M^k has finite length for all k. As $M^{n+1} = 0$ it follows that A itself has finite length as an A-module, and so is artinian by Example 21.4. □

Lemma 21.11. [`lem-artinian-d`]

Suppose that A is noetherian and that all prime ideals in A are maximal. Then A is a finite product of noetherian local rings in which the maximal ideal is nilpotent.

Proof. As A is noetherian, there is a decomposition $Q_1 \cap \dots \cap Q_n = 0$ say, where each Q_i is a primary ideal and the corresponding prime ideals $P_i = \sqrt{Q_i}$ are distinct. The general theory of primary decompositions also tells us that every minimal prime ideal of A occurs in the list $P_1, \dots, P_n$. However, all primes in A are maximal, so there are no inclusions between distinct primes, which means that all primes are minimal, so $P_1, \dots, P_n$ are the only prime ideals. If $i \neq j$ then $P_i + P_j$ is strictly larger than the maximal ideal P_i, so it must be all of A. We can thus choose $a \in P_i$ and $b \in P_j$ with $a + b = 1$. As $P_i = \sqrt{Q_i}$ and $P_j = \sqrt{Q_j}$ we can choose n and m such that $a^{n+1} \in Q_i$ and $b^{m+1} \in Q_j$. Now $(a+b)^{n+m+1} = 1$, and all terms in the expansion of the left hand side lie in Q_j or Q_j, so $Q_i + Q_j = A$. As $\bigcap_i Q_i = 0$, the Chinese Remainder Theorem now gives $A \simeq \prod_i A/Q_i$. As $\sqrt{Q_i} = P_i$ and P_i is maximal we see that A/Q_i is local, with maximal ideal P_i/Q_i, which is nilpotent. □

22. Finite extensions and integral extensions

Throughout this section, A will denote a noetherian ring.

To do:

- Finite etale extensions.
- Balmer approach to degree.

Definition 22.1. Let B be an A-algebra, and let b be an element of B. There is an evaluation homomorphism $\epsilon_b \colon A[t] \to B$ given by $\epsilon_b(f) = f(b)$. We write $A[b]$ for the image of this homomorphism, which is easily seen to be the smallest A-subalgebra of B containing b.

Proposition 22.2. [prop-integral-tfae]

For A, B and b as above, the following are equivalent:

(a) *There is a monic polynomial $f(t)$ over A with $f(b) = 0$.*
(b) *The subalgebra $A[b] \leq B$ is finitely generated as an A-module.*
(c) *There is a subalgebra $C \leq B$ such that $b \in C$ and C is finitely generated as an A-module.*
(d) *There is a finitely generated $A[b]$-module M that is finitely generated as an A-module and satisfies* $\operatorname{ann}_{A[b]}(M) = 0$.

Proof.

(a)⇒(b): Suppose that $b^n + \sum_{i=0}^n a_i b^i = 0$. Let M be the submodule of B generated by $\{b^i \mid 0 \leq i < n\}$. The relation shows that $bM \leq M$, and it follows easily that $M = A[b]$.

(b)⇒(c): Take $C = A[b]$.

(c)⇒(d): Take $M = C$.

(d)⇒(a): Suppose that M is as in (d). Choose generators $m_1, \dots, m_n$ for M as an A-module. We then have $bm_i = \sum_j a_{ij} m_j$ for some coefficients $a_{ij} \in A$. This can be written as an equation $bm = Am$ in M^n for some matrix $A \in M_n(R)$. Let $f(t)$ be the characteristic polynomial of A, which is monic of degree n over R. The equation $xm = Am$ gives $(bI - A)m = 0$, and we can multiply on the left by $\operatorname{adj}(bI - A)$ to get $f(b)m = 0$ in M^n, so $f(b)m_i = 0$ for all i, so $f(b) \in \operatorname{ann}_{A[b]}(M) = 0$, so $f(b) = 0$ as required.

□

Definition 22.3. [defn-integral]

We say that $b \in B$ is *integral* over A if the above conditions are satisfied. We write $\widetilde{A}$ for the set of integral elements, and call this the *integral closure* of A in B.

Remark 22.4. [rem-gaussian-integral]

If B itself is finitely generated as an A-module, then it is clear that all elements are integral and so $\widetilde{A} = B$. For example, this applies when $A = \mathbb{Z}$ and $B = \mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$. Explicitly, any element $z = a + ib$ satisfies $f(z) = 0$, where $f(t) = t^2 - 2at + a^2 + b^2$.

Lemma 22.5. [lem-integral-tensor]

Let B be an A-algebra, let C and D be A-subalgebras of B that are finitely generated as A-modules, and let CD denote the smallest A-subalgebra containing C and D. Then CD is also finitely generated as an A-module. In particular, if $A[c]$ and $A[d]$ are finitely generated then so is $A[c, d]$.

Proof. Let $\{c_i \mid i < n\}$ be a generating set for C, and let $\{d_j \mid j < m\}$ be a generating set for D. Put

$$E = \sum_{i,j} Ac_i d_j = \sum_i Dc_i = \sum_j Cd_j.$$

This is both an C-submodule and a D-submodule of CD and it contains 1 so it is equal to CD. It is clearly finitely generated. □

Lemma 22.6. [lem-integral-chain]
Suppose we have ring maps $A \to B \to C$ such that B is finitely generated as an A-module and C is finitely generated as a B-module. Then C is also finitely generated as an A-module.

Proof. Choose generating sets so that $B = \sum_{i<n} Ab_i$ and $C = \sum_{j<m} Bc_j$. We then have $C = \sum_{i<n} \sum_{j<m} Ab_i c_j$. □

Proposition 22.7. [prop-double-integral]
$\widetilde{A}$ is an A-subalgebra of B, and every element that is integral over $\widetilde{A}$ is also integral over A.

Proof. It is clear that $\widetilde{A}$ contains the image of A. If $c, d \in \widetilde{A}$ then $A[c]$ and $A[d]$ are finitely generated A-modules, so the same is true of $A[c, d]$, so any element of $A[c, d]$ is integral. In particular $c \pm d$ and cd are integral. It follows that $\widetilde{A}$ is a subalgebra, as claimed.

Now suppose that $b \in B$ is integral over $\widetilde{A}$. We then have a monic polynomial $h(t) = t^p + \sum_{k<p} c_k t^k$ such that $c_k \in \widetilde{A}$ for all k, and $h(b) = 0$. Put $C = A[c_0, \ldots, c_{n-1}]$, and note that this is finitely generated as an A-module. Now $C[b]$ is finitely generated as a C-module, and therefore also as an A-module; so b is integral over A. □

Example 22.8. [eg-cyclotomic]
Let U denote the group of roots of unity in $\mathbb{C}$, so

$$U = \{z \in \mathbb{C} \mid z^n = 1 \text{ for some } n > 0\} = \{e^{2\pi i m/n} \mid m, n \in \mathbb{Z},\ n > 0\}.$$

Let A denote the set of $\mathbb{Z}$-linear combinations of elements of U. This is easily seen to be a subring of $\mathbb{C}$, called the *ring of cyclotomic integers.* Each element of U is a root of some monic polynomial $t^n - 1$, so it is integral over $\mathbb{Z}$. It follows that A is an integral extension of $\mathbb{Z}$, but it is clearly not finitely generated as a $\mathbb{Z}$-module.

Definition 22.9. Suppose that the map $A \to B$ is the inclusion of a subring.
We say that B is an *integral extension* of A if every element of B is integral, so $\widetilde{A} = B$. On the other hand, we say that A is *integrally closed* in B if $\widetilde{A} = A$. If A is an integral domain, we say that A is *integrally closed* if it is integrally closed in its field of fractions.

Example 22.10. [eg-invariant-integral]
Let B be an integral domain, and let G be a finite group of automorphisms of B. Put

$$A = B^G = \{a \in B \mid \gamma(a) = a \text{ for all } \gamma \in G\},$$

and note that this is a subring of B. We claim that B is integral over A. To see this, consider an element $b \in B$, and put $\phi_b(t) = \prod_{\gamma \in G}(t - \gamma(b)) \in B[t]$. This is clearly a monic polynomial, of degree $|G|$. Now note that each automorphism $\gamma \in G$ gives rise to an automorphism of $B[t]$ by the rule $\gamma(\sum_i a_i t^i) = \sum_i \gamma(a_i) t^i$, and we have $B[t]^G = A[t]$. It is easy to see that $\gamma(\phi_b(t)) = \phi_b(t)$ for all γ, so $\phi_b(t) \in A[t]$. Moreover, $\phi_b(t)$ has a factor $t - b$ (corresponding to $\gamma = 1$), so $\phi_b(b) = 0$. This proves that b is integral over A, as claimed.

Proposition 22.11. *If A is a unique factorisation domain, then it is integrally closed.*

Proof. Let K be the field of fractions. If $x \in K$ is integral over A then we have a relation $x^n = \sum_{i=0}^{n-1} a_i x^i$ with $a_i \in A$. For each prime p this gives $v_p(x^n) \geq \min(v_p(a_i x^i) \mid i < n)$, so there exists $i < n$ with $v_p(a_i) + i v_p(x) \leq n v_p(x)$. As $a_i \in A$ we have $v_p(a_i) \geq 0$, and it follows that $v_p(x) \geq 0$. As this holds for all p we must have $x \in A$. □

Proposition 22.12. *If U is a multiplicative subset of A, then the integral closure of $A[U^{-1}]$ in $B[U^{-1}]$ is $\widetilde{A}[U^{-1}]$.*

Proof. First suppose that $x \in B$ is integral over A, so there is an A-subalgebra $C \leq B$ that is a finitely generated A-module and contains x. Then $C[U^{-1}]$ is an $A[U^{-1}]$-subalgebra of $B[U^{-1}]$ that is finitely generated over $A[U^{-1}]$ and contains a/u for all $u \in U$; so all such elements a/u are integral.

Conversely, suppose we have an element $y \in B[U^{-1}]$ that is integral over $A[U^{-1}]$, so there is a relation $y^n + \sum_{i=0}^{n-1} b_i y^i = 0$ with $b_i \in A[U^{-1}]$ after multiplying together all the denominators that appear, we find that there are elements $u \in U$ and $x \in B$ and $a_0, \ldots, a_{n-1} \in A$ such that $y = x/u$ and $b_i = a_i/u$. After multiplying the given relation by u^n we find that the element $r = x^n + \sum_i b_i u^{n-i} x^i \in A$ becomes zero in $A[U^{-1}]$. This means that there is an element $v \in U$ with $vr = 0$ in A, which implies that $(vx)^n + \sum_i b_i (uv)^{n-i} (vx)^i = 0$ in A. This shows that $vx \in \widetilde{A}$, so $y = (vx)/(uv) \in \widetilde{A}[U^{-1}]$. □

Corollary 22.13.

(a) *If B is an integral extension of A, then $B[U^{-1}]$ is an integral extension of $A[U^{-1}]$.*
(b) *If A is integrally closed in B, then $A[U^{-1}]$ is integrally closed in $B[U^{-1}]$.*
(c) *Suppose that A is a domain with field of fractions K and A is integrally closed in K. Then $A[U^{-1}]$ is again a domain with field of fractions K that is integrally closed in K.*

Proof. Clear from the Proposition. □

Proposition 22.14. *Suppose that B is an integral domain and A is a subring of B, and that A_M is integrally closed in B_M for all maximal ideals M. Then A is integrally closed in B.*

Proof. Suppose that $b \in B$ and b is integral over A. Put $I = \{a \in A \mid ab \in A\}$; we must show that $I = A$. For any maximal ideal M we see that b is an element of B_M that is integral over A_M, so $b \in A_M$, so there is an element $a \in A \setminus M$ with $ab \in A$, so $I \not\leq M$. As I is not contained in any maximal ideal, we must have $I = A$ as required. □

Proposition 22.15. [`prop-int-field`]
Suppose that B is an integral domain and that A is a subring such that B is integral over A. Then A is a field if and only if B is a field.

Proof. First suppose that A is a field. For any element $x \in B \setminus \{0\}$ choose a monic polynomial $f(t) = \sum_{i=0}^{n} a_i t^i \in A[t]$ of minimal degree such that $f(x) = 0$. Put $g(t) = -\sum_{i=1}^{n} a_i t^{i-1}$, so $a_0 = g(x)x$. By minimality of f we have $g(x) \neq 0$, and by assumption we have $x \neq 0$, and B is a domain so $a_0 \neq 0$. Moreover, $a_0 \in A$ and A is a field so we have an inverse $a_0^{-1} \in A$, and we find that $a_0^{-1} g(x)$ is an inverse for x in B. Thus B is a field. □

Corollary 22.16. [`cor-int-max`]
Suppose that $\phi\colon A \to B$ makes B into an integral A-algebra, and that Q is a prime ideal in B. Then Q is maximal in B if and only if the ideal $\phi^*(Q) = \{a \in A \mid \phi(a) \in Q\}$ is maximal in A.

Proof. It is easy to see that ϕ induces an injective homomorphism from $A/\phi^*(Q)$ to B/Q, which makes B/Q into an integral extansion of $A/\phi^*(Q)$. The proposition tells us that B/Q is a field if and only if $A/\phi^*(Q)$ is a field. □

Proposition 22.17. [`prop-zar-fibre`]
Let $\phi\colon A \to B$ be an integral extansion of integral domains, and consider the resulting map $\phi^*\colon \mathrm{zar}(B) \to \mathrm{zar}(A)$. Given $P \in \mathrm{zar}(B)$ put

$$\begin{aligned} K(P) &= \textit{ field of fractions of } A/P \\ F(P) &= \{Q \in \mathrm{zar}(B) \mid \phi^*(Q) = P\}. \end{aligned}$$

Then

(a) *$F(P)$ is naturally identified with $\max(A_P \otimes_A B)$ and with $\mathrm{zar}(K(P) \otimes_A B)$.*
(b) *$F(P)$ is always nonempty, so ϕ^* is surjective.*
(c) *If $Q_0, Q_1 \in F(P)$ with $Q_0 \subseteq Q_1$ then $Q_0 = Q_1$.*
(d) *If B is a finitely generated A-module then $F(P)$ is finite.*

Proof. It will be harmless to assume that ϕ is just the inclusion of a subring, so $\phi^*(Q) = Q \cap A$ for all Q. Proposition 8.16 allows us to identify $K(P)$ with A_P/P_P. We have a commutative diagram of ring homomorphisms as follows:

$$\begin{tikzcd}
A \arrow[r, rightarrowtail] \arrow[d, rightarrowtail, "\phi"'] & A_P \arrow[r, twoheadrightarrow] \arrow[d, rightarrowtail, "\phi_P"'] & K(P) \arrow[d] \\
B \arrow[r, rightarrowtail] & B_P = A_P \otimes_A B \arrow[r, twoheadrightarrow] & B_P/PB_P = K(P) \otimes_A B
\end{tikzcd}$$

Using Propositions 5.45 and 8.18, we get natural bijections

$$\begin{aligned}
\operatorname{zar}(A_P) &= \{P' \in \operatorname{zar}(A) \mid P' \leq P\} \\
\operatorname{zar}(K(P)) &= \operatorname{zar}(A_P/P_P) = \{P\} \\
\operatorname{zar}(B_P) &= \{Q \in \operatorname{zar}(B) \mid Q \cap (A \setminus P) = \emptyset\} = \{Q \in \operatorname{zar}(B) \mid Q \cap A \leq P\} \\
\operatorname{zar}(B_P/BP_P) &= \{Q \in \operatorname{zar}(B) \mid Q \cap A \leq P \text{ and } Q \geq BP\} \\
&= \{Q \in \operatorname{zar}(B) \mid Q \cap A = P\} = F(P).
\end{aligned}$$

This proves claim (a).

Next, note that P_P is the unique maximal ideal in the local ring A_P. We can also identify $F(P)$ with $\{Q' \in \operatorname{zar}(B_P) \mid Q' \cap A_P = P_P\}$, or in other words with $\{Q' \in \operatorname{zar}(B_P) \mid Q' \cap A_P \text{ is maximal }\}$. However, Corollary 22.16 tells us that $Q' \cap A_P$ is maximal if and only if Q' is maximal, so $F(P)$ bijects with $\max(B_P)$. The map $A \to B$ is injective, so $A_P \to B_P$ is injective, so $B_P \neq 0$, so $\max(B_P) \neq \emptyset$, so $F(P) \neq \emptyset$, which proves claim (b). Moreover, if $M_0, M_1 \in \max(B_P)$ with $M_0 \leq M_1$ then maximality clearly implies that $M_0 = M_1$; this proves claim (c). Finally, if B is finitely generated as an A-module then $K(P) \otimes_A B$ is finite-dimensional over the field $K(P)$, so it is an artinian ring, so the set $F(P) = \operatorname{zar}(K(P) \otimes_A B)$ is finite by Theorem 21.7. □

Example 22.18. Consider the case where $A = \mathbb{Z}$ and $B = \mathbb{Z}[i]$. Standard arguments from number theory, which we will not explain here, give the following.

(a) $F(0) = \{0\}$

(b) $F(A.2) = B.(1+i) = B.(1-i)$

(c) If p is a prime congruent to 1 mod 4 then there are integers a, b with $p = a^2 + b^2$, and $F(A.p) = \{B.(a+ib), B.(a-ib)\}$.

(d) If p is a prime congruent to 3 mod 4 then $F(A.p) = \{B.p\}$.

In the context of Example 22.10, we can be more precise about the relationship between $\operatorname{zar}(A)$ and $\operatorname{zar}(B)$.

Proposition 22.19. *Let B be an integral domain with a finite group G of automorphisms, and put $A = \{a \in B \mid \gamma(a) = a \text{ for all } \gamma \in G\}$. Then the inclusion $\phi \colon A \to B$ induces a bijection $\operatorname{zar}(B)/G \to \operatorname{zar}(A)$.*

Proof. We saw in Example 22.10 that ϕ is an integral extension, so $\phi^* \colon \operatorname{zar}(B) \to \operatorname{zar}(A)$ is surjective by Proposition 22.17.

For $\gamma \in G$ we have $\gamma\phi = \phi$, so $\phi^*\gamma^* = \phi^* \colon \operatorname{zar}(B) \to \operatorname{zar}(A)$. It follows that ϕ^* induces a map $\operatorname{zar}(B)/G \to \operatorname{zar}(A)$, which must again be surjective.

Now suppose we have two prime ideals $Q, Q' \in \operatorname{zar}(B)$ with $\phi^*(Q) = \phi^*(Q') = P$ say. Put $I = \bigcap_{\gamma \in G} \gamma(Q) \leq B$. Consider an element $b \in I$, and put $\phi_b(t) = \prod_{\gamma \in G}(t - \gamma(b))$ as before. All the elements $\gamma(b)$ lie in Q, so $\phi_b(t) - t^n \in Q[t]$. On the other hand, we know that $\phi_b(t) \in B[t]^G = A[t]$, and $Q \cap A = P$, so $\phi_b(t) - t^n \in P[t] \subseteq Q'[t]$. We can now substitute $t = b$ and recall that $\phi_b(b) = 0$ to get $b^n \in Q'$. As Q' is prime, it follows that $b \in Q'$. We now conclude that the ideal $I = \bigcap_\gamma \gamma^*(Q)$ is contained in Q'. Using Corollary 5.35, we deduce that one of the ideals $\gamma^*(Q)$ must be contained in Q'. Now both Q and $\gamma^*(Q')$ lie in $F(P)$, so part (c) of Proposition 22.17 tells us that $\gamma^*(Q') = Q$. It follows that the map $\operatorname{zar}(B)/G \to \operatorname{zar}(A)$ is bijective as claimed. □

The significance of the following result will become clearer when we define the Krull dimension of commutative rings

Proposition 22.20. [prop-going-up]

Let $\phi \colon A \to B$ be an integral extension. Suppose we have a chain of prime ideals $P_0 < \cdots < P_n$ in A, and another chain of prime ideals $Q_0 < \cdots < Q_m$ in B, where $m \leq n$, and $Q_i \cap A = P_i$ for $0 \leq i \leq m$. Then we can choose further prime ideals $Q_{m+1}, \ldots, Q_n$ in B such that $Q_0 < \cdots < Q_n$, and $Q_i \cap A = P_i$ for all i.

Proof. First consider the special case where $n = 1$ and $m = 0$. As $Q_0 \cap A = P_0$, we have an integral extension $A/P_0 \to B/Q_0$ of integral domains. We can apply Proposition 22.17 to this extension; we learn that there is a prime ideal $\overline{Q}_1 \in \operatorname{zar}(B/Q_0)$ with $\overline{Q}_1 \cap A/P_0 = P_1/P_0$. This must have the form $\overline{Q}_1 = Q_1/Q_0$ for some

$Q_1 \in \operatorname{zar}(B)$ with $Q_1 \cap A = P_1$. This completes the proof of the special case, and we can use that as the induction step in an an obvious inductive proof of the general case. □

23. Noether normalisation and the Nullstellensatz

Fix a field K, and write P_n for the polynomial ring $K[x_0, \dots, x_{n-1}]$.

Theorem 23.1. [thm-normalisation]
Let A be a nontrivial finitely generated algebra over K. Then there is a subalgebra $P \subseteq A$ such that P is isomorphic to P_d for some d, and A is finitely generated as a P-module.

We pause to explain the geometric meaning of this result. Suppose that $K = \mathbb{C}$ and $A = P_m/I(X)$ for some algebraic subset $X \subseteq \mathbb{C}^m$. The theorem gives an inclusion $P_d \to A$, corresponding to a map $f \colon X \to \mathbb{C}^d$. The fact that A is integral over P_d means that f is surjective with finite fibres, which indicates that X is d-dimensional over $\mathbb{C}$.

The proof depends on the following result.

Definition 23.2. [defn-ess-monic]
We say that a polynomial $f \in A[t]$ is *essentially monic in t* if it has the form $f = \sum_{i=0}^{d} a_i t^i$ for some d, where a_d is invertible.

Lemma 23.3. *For any nonzero element $f \in P_n$ there is an automorphism α of P_n such that $\alpha(f)$ is essentially monic as a polynomial in x_0.*

Proof. We can write f as a K-linear combination of monomials $\prod_{t=0}^{n-1} x_t^{i_t}$. Let b be an integer that is strictly larger than any of the exponents i_t that occur in this representation.

Now define $\alpha \colon P_n \to P_n$ by $\alpha(x_0) = x_0$, and $\alpha(x_i) = x_i + x_0^{b^i}$ for $0 < i < n$. We also define $\beta \colon P_n \to P_n$ by $\beta(x_0) = x_0$, and $\beta(x_i) = x_i - x_0^{b^i}$ for $0 < i < n$; this is an inverse for α, proving that α is an automorphism.

Next, for $0 \leq k < b^n$ there is a unique sequence $(i_0, \dots, i_{n-1})$ such that $0 \leq i_t < b$ for all t, and $\sum_t i_t b^t = k$ (this is essentially the base b representation of k). We put $m_k = \prod_t x_t^{i_t}$, so the list $m_0, \dots, m_{p^n-1}$ contains all monomials that have degree less than b in each of the variables $x_0, \dots, x_{n-1}$. We therefore have $f = \sum_{k=0}^{b^n-1} c_k m_k$ for some sequence of coefficients $c_k \in K$ that are not all zero. It is easy to see that

$$\alpha(m_k) = x_0^k + \text{ terms of degree less than } k \text{ in } x_0 \ .$$

Thus, if we let m be the largest integer with $c_m \neq 0$, we have

$$\alpha(f) = c_m x_0^m + \text{ terms of degree less than } m \text{ in } x_0 \ ,$$

so $\alpha(f)$ is essentially monic in x_0. □

Proof of Theorem 23.1. Suppose that A can be generated as a K-algebra by n elements, so we have a surjective homomorphism $\phi \colon P_n \to A$ say. We may assume inductively that theorem holds for any K-algebra that can be generated by $n-1$ elements. If ϕ is injective then it is an isomorphism so we can just take $P = A$. Suppose instead that ϕ is not injective, so we can choose a nonzero polynomial $f \in P_n$ with $\phi(f) = 0$. After replacing f by $\alpha(f)$ and ϕ by $\phi \circ \alpha^{-1}$ for some automorphism ϕ, we may assume that f is essentially monic in x_0, of degree m say. Now put $R = K[x_1, \dots, x_{n-1}]$ and $Q = P_n/f$, so $\{x_0^i \mid 0 \leq i < m\}$ is a finite basis for Q as an R-module. As ϕ is surjective with $\phi(f)$, it induces a surjective homomorphism $\overline{\phi} \colon Q \to A$. It follows that A is finitely generated as a module over the subring $B = \overline{\phi}(R)$. As $R \simeq P_{n-1}$, our induction hypothesis gives a subalgebra $P \subseteq B$ that is isomorphic to P_d for some d, such that B is finitely generated as a P-module. It follows that A is also finitely generated as a P-module, which proves the theorem. □

We next want to prove that the integer d occuring in Theorem 23.1 is independent of the choice of P. For this it is convenient to introduce the following notation:

Definition 23.4. For any monomial $m = \prod_{t=0}^{d-1} x_t^{i_t}$, the *total degree* is $|m| = \sum_t i_t$. We write $B_r P_d$ for the set of all monomials of total degree at most r, and $F_r P_d$ for the K-linear span of $B_r P_d$.

Lemma 23.5. *Suppose that A is a finitely generated module over P_d, and that $\phi\colon P_m \to A$ is a homomorphism of K-algebras. Then there is a polynomial $f(t)$ of degree d such that the leading coefficient is positive, and $\dim_K(\phi(F_rP_m)) \leq f(r)$ for all r.*

Proof. Choose a finite set X that contains 1 and generates A as a module over P_d. Let F_rA denote the K-linear span of $F_rP_d.X$, so $A = \bigcup_r F_rA$. Choose an integer p large enough that $X.X \subseteq F_pA$, and choose q such that $\phi(x_i) \in F_qA$ for all i. Note that

$$\begin{aligned} F_qA.F_{n(p+q)}A &= \operatorname{span}_K(B_qP_d.X.B_{np+nq}P_d.X) = \operatorname{span}_K(B_{np+(n+1)q}P_d.X.X) \\ &\subseteq \operatorname{span}_K(B_{(n+1)(p+q)}P_d.X) = F_{(n+1)(p+q)}A. \end{aligned}$$

Using this, we can prove by induction that $\phi(F_rP_m) \leq F_{(p+q)r}A$. It is standard that $|B_rP_d| = \left(\begin{array}{c} r+d \\ d \end{array}\right)$, and it follows that

$$\dim(\phi(F_rP_m)) \leq \left(\begin{array}{c} (p+q)r+d \\ d \end{array}\right)|X|,$$

which is polynomial of degree d in r. □

Proposition 23.6. [prop-noether-dim]

Let A be a K-algebra, and suppose that A has subalgebras $P \simeq P_d$ and $P' \simeq P_{d'}$ such that A is finitely generated as a module over each of these subalgebras. Then $d = d'$.

Proof. Let $\phi\colon P \to A$ and $\phi'\colon P' \to A$ be the inclusions. The Lemma tells us that $\dim(\phi'(F_rP_{d'})) \leq f(r)$ for some polynomial f of degree d whose leading coefficient is positive. However, ϕ' is injective, so $\dim(\phi'(F_rP_{d'})) = \dim(F_rP_{d'}) = \left(\begin{array}{c} d'+r \\ d' \end{array}\right)$, which is polynomial of degree d', again with positive leading coefficient. This is only consistent if $d' \leq d$. By exchanging the roles of P and P', we deduce that $d = d'$. □

Definition 23.7. [defn-noether-dim]

The number d occuring in Theorem 23.1 will be called the *noether dimension* of A. Proposition 23.6 tells us that this is well defined.

Proposition 23.8. [prop-nsatz-a]

Let K be a field, and let L be a finitely generated K-algebra that is also a field. Then L is finitely generated as a K-module. In particular, if K is algebraically closed, then the map $K \to L$ is an isomorphism.

Proof. By Theorem 23.1, we can choose an integer $d \geq 0$ and a subalgebra $P \leq L$ such that $P \simeq P_d$ and L is a finitely generated P-module. This means that L is integral over P, and L is a field, so P is a field by Proposition 22.15. This can only happen if $d = 0$, so $P = K$ and L is a finitely generated K-module. □

Corollary 23.9. [cor-nsatz-b]

Let K be a field, let A be a finitely generated K-algebra, and let M be a maximal ideal in A. Then A/M is finite-dimensional over K. In particular, if K is algebraically closed then the natural map $K \to A/M$ is an isomorphism.

Proof. Just apply the Proposition to the field $L = A/M$. □

Proposition 23.10. *Suppose that A is a finitely generated algebra over a field K. Then* $\operatorname{Rad}(A) = \operatorname{Nil}(A)$.

Proof. By Proposition 5.19, we always have $\operatorname{Nil}(A) \subseteq \operatorname{Rad}(A)$, for any commutative ring A. Conversely, suppose that $a \in A$ but $a \notin \operatorname{Nil}(A)$. Put $A' = A[a^{-1}] \simeq A[b]/(ab-1)$, and note that this is nontrivial, and finitely generated as a K-algebra. We can therefore choose a maximal ideal $M' \in \max(A')$, and we find that A'/M' is finite-dimensional over K. Now put $M = M' \cap A$, so we have natural injective maps $K \to A/M \to A'/M'$. This implies that A/M is an integral domain that is finite-dimensional over K, so it is a field, so M is maximal. It is clear that $a \notin M$, so $a \notin \operatorname{Rad}(A)$ by Proposition 5.50. □