Algebra III: Rings and Modules
Lecture Notes, Autumn Term 2022-23

John Nicholson

Contents

Practical information

Prerequisites: This course will build on the Year 2 course 'Algebra II: Groups and Rings'. We will assume familiarity with groups but will develop the basic theory of rings from scratch. Whilst it is not essential to have covered rings before but we will go quite quickly over the basics of rings so be prepared to work hard in the first two weeks if you haven't seen them before.

Purpose of the Course: The main goal of this course is to develop mathematical maturity in algebra and expose you to topics which come up in various areas of pure mathematics. Specific courses which this course will serve as good preparation for are: Algebra IV, Algebraic Number Theory, Group Representation Theory, Galois Theory, Algebraic Topology, Algebraic Geometry and Commutative Algebra.

Assessments: This course will be assessed by an end-of-year exam (90%) and two in-class tests (each 5%). Each in-class test will last one hour. The first test is on Friday 11th November 3-4pm (i.e. Week 6) and will be based on material from Weeks 1-4. The second test is on Friday 9th December 3-4pm (i.e. Week 10) and will be based on material from Weeks 5-8.

Office Hour: My office hour will be on Thursdays 2-3pm. You are all welcome to come along as often as you like. I would be happy to discuss any aspects of the course you find confusing and discuss any problems you have been working on.

Problem Sheets: There will be four problem sheets based on Weeks 1-2, Weeks 3-4, Weeks 5-6 and Weeks 7-8. You are encouraged to work on these problems in groups. The problems are intended to be challenging and many will be more difficult than what you will be expected to solve under exam conditions. Whilst this work will not be assessed, I strongly advise writing up your solutions carefully (and by yourself) and bringing them with you to the problems classes. This should serve an excellent preparation for the in-class tests.

Challenge Problems: Each problem sheet ends with a problem marked '+'. These problems are intended to be extremely difficult and are for students who have completed the other questions and are looking for something challenging to work on. Anyone working on these problems is welcome to attend my office hour to discuss any partial progress. However, note that solutions and (major) hints will not be provided. If you have a solution to any one of the problems, at any time during the course, then please let me know and I will happily mark your work.

Problem Classes: Each problem sheet will have a corresponding problems class which will last roughly one hour. They will take place during the Friday lectures in Weeks 3, 5, 8 and 9.

Lecture Notes: Lecture notes will be posted on Blackboard each week on Friday evening. If you encounter any typos in these notes or something you find confusing, please let me know.

Mastery Material: For fourth-years and masters students there will be an additional 'mastery' question on the exam. This question will be a question that requires you to make more sophisticated use of some of the core concepts in the course. I do not currently plan for there to be any specific 'mastery material' that you are required to learn but this might change towards the end of the course if I have a lecture or two worth of material leftover when the course ends.

Textbooks: This course will not follow any textbook in particular. However a useful reference which will cover the first half of the course is Michael Artin's *Algebra*, second edition. (Particularly Chapters 11, 12, 14, and 15.)

Acknowledgements: These notes are based on previous lecture notes from the same course due to David Helm and Travis Schedler as well as lecture notes for the course Groups, Rings and Modules at the University of Cambridge (particularly Dexter Chua's lecture notes for the course given by Oscar Randal-Williams).

1 Rings

1.1 Basic definitions and examples

There are multiple definitions of rings in the literature. Our definition seems, to us, to be the most standard: it requires a multiplicative identity (an element 1 such that $1 \cdot x = x = x \cdot 1$ for all x) and does not require the ring to be commutative (in a commutative ring, we would also assume $x \cdot y = y \cdot x$).

First, we need to define a monoid, which is essentially a group except that we do not assume the existence of inverses:

Definition 1.1. A *monoid* $(M, \cdot)$ is a set M together with a binary operation $\cdot : M \times M \to M$ and an element $1_M \in M$ (called the multiplicative identity) satisfying the axioms:

- $m \cdot 1_M = m = 1_M \cdot m$ for all $m \in M$
- The operation is associative: $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ (for all $x, y, z \in M$).

Example 1.2. The natural numbers $\mathbb{N} = \{1, 2, 3, \cdots\}$ is a monoid under usual multiplication with identity $1_{\mathbb{N}} := 1$.

Example 1.3. The set $\mathbb{N} \cup \{\infty\}$ is a monoid under usual multiplication on $\mathbb{N}$ as well as $n \cdot \infty := \infty$ and $\infty \cdot n := \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$.

Definition 1.4. A *ring* is a set R together with functions $+ : R \times R \to R$ and $\cdot : R \times R \to R$, and given elements $0_R, 1_R \in R$, such that the following holds:

- $(R, +)$ is an abelian group with identity 0_R
- $(R, \cdot)$ is a monoid with identity 1_R
- The distributivity property holds: $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$.

We refer to $+$ as *addition* and write $x + y$ to denote $+(x, y)$. Similarly we refer to $\cdot$ as *multiplication* and write $x \cdot y$ to denote $\cdot(x, y)$.

Notation 1.5. If R is a ring and $r \in R$, we write $-r$ for the inverse to r in $(R, +)$. This satisfies $r + (-r) = 0_R$. We write $r - s$ to mean $r + (-s)$ etc. Since we can add and multiply two elements, by induction, we can add and multiply any finite number of elements.

We will often write 1_R as 1 and 0_R as 0 when it is clear from the context which elements we are referring to.

Definition 1.6. We say a ring R is *commutative* if $a \cdot b = b \cdot a$ for all $a, b \in R$.

Definition 1.7. Let R be a ring, and $S \subseteq R$ be a subset. We say S is a *subring* of R if $0_R, 1_R \in S$, and the operations $+, \cdot$ make S into a ring with identities 0_R and 1_R. In this case we write $S \leq R$.

Example 1.8. The usual number systems are all rings $\mathbb{Z} \leq \mathbb{Q} \leq \mathbb{R} \leq \mathbb{C}$ under the usual $0, 1, +, \cdot$.

Example 1.9. The set $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\} \leq \mathbb{C}$ is the *Gaussian integers*, which is a ring.

We also have the ring $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\} \leq \mathbb{R}$.

Example 1.10. The *trivial ring* is ring $R = \{0\}$ with $0 \cdot 0 := 0$ and $0 + 0 := 0$. It is the unique ring with a single element.

Example 1.11. The natural numbers $\mathbb{N}$ is not a ring (no additive inverses).

Proposition 1.12. Let R be a ring. Then $1_R = 0_R$ if and only if $R = \{0\}$ is the trivial ring.

Proof. If $R = \{0\}$, then $0_R, 1_R \in R$ implies that $1_R = 0_R$.

Conversely, suppose that $1_R = 0_R$. If $r \in R$, then

$$r = r \cdot 1_R = r \cdot 0_R.$$

It now suffices to show that $r \cdot 0_R = 0_R$ for all $r \in R$. If so, then we can conclude that $r = 0_R$ for all $r \in R$ and so $R = \{0_R\}$ is the trivial ring.

Note that $0_R + 0_R = 0_R$, since this is true in the group $(R, +)$. Then for any $r \in R$, we have:

$$r \cdot (0_R + 0_R) = r \cdot 0_R.$$

Since multiplication distributes over addition, we have:

$$r \cdot 0_R + r \cdot 0_R = r \cdot 0_R.$$

Adding $(-r \cdot 0_R)$ to both sides gives that $r \cdot 0_R = 0_R$, as required. □

Definition 1.13. An element $u \in R$ is a *unit* if there is another element $v \in R$ such that $u \cdot v = v \cdot u = 1_R$. We will let $R^\times \subseteq R$ denote the set of units in R.

Definition 1.14. A *division ring* is a non-trivial ring where every $u \neq 0_R \in R$ is a unit, i.e. $R^\times = R \setminus \{0\}$. A *field* is a commutative division ring.

Example 1.15. $\mathbb{Z}$ is not a field, but $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all fields. We have $\mathbb{Z}^\times = \{\pm 1\}$.

Similarly, $\mathbb{Z}[i]$ is not a field, while $\mathbb{Q}[\sqrt{2}]$ is. We have $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$.

Example 1.16. Define the quaternions $\mathbb{H}$ to be the abelian group $\mathbb{R}^4$. Let $1 = (1,0,0,0)$, $i = (0,1,0,0)$, $j = (0,0,1,0)$ and $k = (0,0,0,1)$, i.e. the standard basis for $\mathbb{R}^4$. Let $\mathbb{R} := \mathbb{R} \cdot 1 \subseteq \mathbb{H}$. Then $\mathbb{H}$ is a ring with multiplication determined by asking that, if $r \in R \subseteq \mathbb{H}$, then $r \cdot (a,b,c,d) = (r \cdot_{\mathbb{R}} a, r \cdot_{\mathbb{R}} b, r \cdot_{\mathbb{R}} c, r \cdot_{\mathbb{R}} d)$ and furthermore that the basis elements are related by $1 \cdot i = i \cdot 1 = i$, $1 \cdot j = j \cdot 1 = j$, $1 \cdot k = k \cdot 1 = k$, $ij = l = -ji$ and $i^2 = j^2 = -1$ (this determines all other products of basis elements such as $jk = i$ and $k^2 = -1$). The identities are $1 = (1,0,0,0)$ and $0 = (0,0,0,0)$. We write (a,b,c,d) as $a + bi + cj + dk$.

This is not commutative since $ij \neq ji$. However it is a division ring since

$$(a + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2.$$

In particular, if $(a,b,c,d) \neq 0$, then $a + bi + cj + dk \in \mathbb{H}$ has inverse

$$\frac{1}{a^2 + b^2 + c^2 + d^2} \cdot (a - bi - cj - dk) \in \mathbb{H}.$$

Proposition 1.17. The subset $R^\times \subseteq R$ is a group under multiplication.

Proof. First we have to show that $R^\times$ is actually closed under multiplication. If $a, b \in R^\times$, let $c, d \in R^\times$ be such that $ac = ca = 1 = bd = db$. Then $(ab)(dc) = a(bd)c = a(1)c = ac = 1$, and similarly $(dc)(ab) = d(ca)b = d(1)b = db = 1$. Therefore $ab \in R^\times$. Next, we show that $R^\times$ is a group: the element $1 \in R$ is a unit since $1 \cdot 1 = 1$, and every element of $R^\times$ has a multiplicative inverse by definition. □

1.2 Constructions of rings

Example 1.18. Let R, S be rings. Then the *product* $R \times S$ is a ring via

$$(r, s) + (r', s') = (r + r', s + s'), \quad (r, s) \cdot (r', s') = (r \cdot r', s \cdot s').$$

The zero is $(0_R, 0_S)$ and the one is $(1_R, 1_S)$. We can check that this is a ring.

Example 1.19. Let R be a ring. Then a *polynomial* with coefficients in R is an infinite sequence $f = (a_0, a_1, a_2, \cdots)$ in R which is eventually zero, i.e. $a_i = 0$ for all i sufficiently large. The *degree* of a polynomial $f = (a_0, a_1, a_2, \cdots)$ is defined as the maximal n for which $a_n \neq 0_R$, or $-\infty$ if $a_i = 0$ for all $i \geq 1$. We will write this as $\deg(f)$.

If $a_i = 0$ for all $i \geq n + 1$, then we will often write a polynomial $(a_0, a_1, a_2, \cdots)$ using the notation:

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

Note that, with this notation, f is the same polynomial as $a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n + 0_R X^{n+1}$.

Example 1.20. Let f have degree $n \geq 0$. If $a_n = 1$, then f is called *monic*.

Example 1.21. If R is a ring, then the *polynomial ring* $R[X]$ consists of the set of all polynomials with coefficients in R. For $f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$ and $g = b_0 + b_1 X + \cdots + b_k X^k \in R[X]$, the operations are defined by:

$$f + g := \sum_{r=0}^{\max\{n,k\}} (a_i + b_i) X^i,$$

and

$$f \cdot g := \sum_{i=0}^{n+k} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) X^i,$$

We identify R with the constant polynomials, i.e. polynomials $\sum a_i X^i$ with $a_i = 0$ for $i > 0$. In particular, $0_R \in R$ and $1_R \in R$ are the zero and one of $R[X]$.

We can check that this is a ring. This fits out notation before since the element $a_0 + a_1 X \in R[X]$ coincides with the element $a_0 +_{R[X]} a_1 X \in R[X]$ formed by adding $a_0, a_1 X \in R[X]$.

Remark 1.22. Note that a polynomial is just a sequence of numbers, interpreted as the coefficients of some formal symbols. While it does indeed induce a function in the obvious way, we shall not identify the polynomial with the function given by it, since different polynomials can give rise to the same function.

For example, in $\mathbb{Z}/2\mathbb{Z}[X]$, $f = X^2 + X$ is not the zero polynomial, since its coefficients are not zero. However, $f(0) = 0$ and $f(1) = 0$. As a function, this is identically zero. So $f \neq 0$ as a polynomial but $f = 0$ as a function.

Example 1.23. The *Laurent polynomials* on R is the set $R[X, X^{-1}]$, i.e. each element is of the form

$$f = \sum_{i \in \mathbb{Z}} a_i X^i$$

where $a_i \in R$ and only finitely many a_i are non-zero. The operations are defined similarly to $R[X]$.

In this case, the set of monomials $\{X^i : i \in \mathbb{Z}\}$ forms a group under multiplication. It turns out that Laurent polynomials are a special case of a much more general construction:

Example 1.24. Let G be a group, and let R be a ring. The *group ring* $R[G]$ is the set of expressions of the form:

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

where λ_g is an element of R for all $g \in G$ and $\{g \in G : a_g \neq 0\}$ is finite. Define addition and multiplication by the formulas:

$$\left(\sum_{g \in G} a_g\right) + \left(\sum_{g \in G} b_g g\right) = \sum_{g \in G} (a_g +_R b_g) g$$

$$\left(\sum_{g \in G} a_g g\right) \cdot \left(\sum_{g \in G} b_g g\right) = \sum_{g \in G} \left(\sum_{h \in G} a_h \cdot_R b_{h^{-1}g}\right) g$$

We can check that $R[G]$ is a ring.

We have $R[X, X^{-1}] \cong R[C_\infty]$ where $C_\infty = (\mathbb{Z}, +)$ denotes the infinite cyclic group.

If R is a commutative ring, then $R[G]$ is commutative if and only if G is abelian.

Example 1.25. If R is a ring and $n \geq 1$, then the set of $n \times n$ matrices $M_n(R)$ forms a ring under the usual addition and multiplication.

Note that $M_n(R)$ is not commutative for all $n \geq 2$ whenever R is non-trivial. In particular, for $n = 2$, we have $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. By placing these matrices in the top left corner of an $n \times n$ matrix for $n \geq 2$ (with all other entries being zero), these elements also show that $M_n(R)$ is non-commutative.

Example 1.26. For an abelian group A, let $\mathrm{End}(A) = \{f : A \to A \mid f \text{ is a group homomorphism}\}$. This is a ring with addition $f +_{\mathrm{End}(A)} g$ for $f, g \in \mathrm{End}(A)$ defined by $(f +_{\mathrm{End}(A)} g)(x) := f(x) +_A g(x)$ for $x \in A$, and multiplication given by composition of functions $f \cdot_{\mathrm{End}(A)} g := f \circ g$. We can check that this is a ring.

A group homomorphism $f : A \to A$ from an abelian group to itself is an *endomorphism*, and $\mathrm{End}(A)$ is the *endomorphism ring* of an abelian group A.

The group of units of $\mathrm{End}(A)$ is the *automorphism group* of A and is denote by $\mathrm{Aut}(A)$.

1.3 Homomorphisms, ideals and quotients

Definition 1.27. Let R, S be rings. A function $\varphi : R \to S$ is a *ring homomorphism* if:

1. $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$,
2. $\varphi(0_R) = 0_S$,
3. $\varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2)$,
4. $\varphi(1_R) = 1_S$.

Definition 1.28. If a homomorphism $\varphi : R \to S$ is a bijection, we call it an *isomorphism*.

Definition 1.29. The *kernel* of a homomorphism $\varphi : R \to S$ is

$$\ker(\varphi) = \{r \in R : \varphi(r) = 0_S\}.$$

Definition 1.30. The *image* of $\varphi : R \to S$ is

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

Lemma 1.31. A homomorphism $\varphi : R \to S$ is injective if and only if $\ker \varphi = \{0_R\}$.

Proof. A ring homomorphism is in particular a group homomorphism $\varphi : (R, +, 0_R) \to (S, +, 0_S)$ of abelian groups. So this follows from the case of groups. □

Definition 1.32. A left ideal $I \subseteq R$ is an abelian subgroup satisfying the property that, for every $i \in I$ and $r \in R$, $ri \in I$. Similarly, a right ideal is an abelian subgroup $I \subseteq R$ such that, for every $i \in I$ and $r \in R$, $ir \in I$. A two-sided ideal (or bi-ideal) is a subset which is both a left and right ideal.

For each type of ideal, the property that $ri \in I$ (and/or $ir \in I$) is often referred to as the strong closure property.

We will use the word *ideal* to denote a left ideal and, when we write $I \subseteq R$, this is what we are referring to. However note that, in a commutative ring, all the types of ideals are the same and so "ideal" means any type of ideal defined above.

Note that the multiplicative closure is stronger than what we require for subrings: for subrings, it has to be closed under multiplication by its own elements; for ideals, it has to be closed under multiplication by everything in R. This is similar to how normal subgroups not only have to be closed under internal multiplication, but also conjugation by external elements.

Lemma 1.33. If $\varphi : R \to S$ is a homomorphism, then $\ker(\varphi) \subseteq R$ is a two-sided ideal.

Proof. Since $\varphi : (R, +, 0_R) \to (S, +, 0_S)$ is a group homomorphism, the kernel is a subgroup of $(R, +, 0_R)$.

For the second part, let $a \in \ker(\varphi)$, $r \in R$. We have

$$\varphi(a \cdot r) = \varphi(a) \cdot \varphi(r) = 0 \cdot \varphi(r) = 0.$$

So $a \cdot r \in \ker(\varphi)$. Similarly, we have $r \cdot a \in \ker(\varphi)$. □

Example 1.34. Suppose $I \subseteq R$ is an ideal, and $1_R \in I$. Then for any $r \in R$, the axioms entail $1_R \cdot r \in I$. But $1_R \cdot r = r$. So if $1_R \in I$, then $I = R$.

Definition 1.35. A proper ideal of a ring R is an ideal which is not equal to R.

In other words, every proper ideal does not contain 1. In particular, every proper ideal is not a subring, since a subring must contain 1.

We are starting to diverge from groups. In groups, a normal subgroup is a subgroup, but here an ideal is not a subring.

Example 1.36. We can generalise the above. Suppose $I \subseteq R$ and $u \in I$ is a unit, i.e. there is some $v \in R$ such that $u \cdot v = 1_R$. Then $1_R = u \cdot v \in I$. So $I = R$.

Hence proper ideals are not allowed to contain any unit at all, not just 1_R. That is, if $I \neq R$, then $I \subseteq R \setminus R^\times$.

Example 1.37. Consider the ring $\mathbb{Z}$ of integers. We claim that every ideal of $\mathbb{Z}$ is of the form

$$n\mathbb{Z} = \{\cdots, -2n, -n, 0, n, 2n, \cdots\} \subseteq \mathbb{Z}.$$

It is easy to see this is indeed an ideal.

To show these are all the ideals, let $I \subseteq \mathbb{Z}$. If $I = \{0\}$, then $I = 0\mathbb{Z}$. Otherwise, let $n \in \mathbb{N}$ be the smallest positive element of I. We want to show in fact $I = n\mathbb{Z}$. Certainly $n\mathbb{Z} \subseteq I$ by strong closure.

Now let $m \in I$. By the Euclidean algorithm, we can write

$$m = q \cdot n + r$$

with $0 \leq r < n$. Now $n, m \in I$. So by strong closure, $m, q \cdot n \in I$. So $r = m - q \cdot n \in I$. As n is the smallest positive element of I, and $r < n$, we must have $r = 0$. So $m = q \cdot n \in n\mathbb{Z}$. So $I \subseteq n\mathbb{Z}$. So $I = n\mathbb{Z}$.

The key to proving this was that we can perform the Euclidean algorithm on $\mathbb{Z}$. Thus, for any ring R in which we can run a "Euclidean algorithm", every ideal is of the form $aR = \{a \cdot r : r \in R\}$ for some $a \in R$. We will make this notion precise later.

Definition 1.38. For an element $a \in R$, we write

$$(a) = Ra = \{r \cdot a \mid r \in R\} \subseteq R.$$

This is the *ideal generated by* a.

In general, let $a_1, a_2, \cdots, a_k \in R$, we write

$$(a_1, a_2, \cdots, a_k) = \{r_1 a_1 + \cdots + r_k a_k \mid r_1, \cdots, r_k \in R\}.$$

This is the *ideal generated by* $a_1, \cdots, a_k$.

We can also have ideals generated by infinitely many objects, but we have to be careful, since we cannot have infinite sums.

Definition 1.39. For $A \subseteq R$ a subset, the *ideal generated by* A is

$$(A) = R \cdot A = \left\{ \sum_{a \in A} r_a \cdot a \mid r_a \in R, \text{ only finitely-many non-zero} \right\}.$$

These ideals are rather nice ideals, since they are easy to describe, and often have some nice properties.

Definition 1.40. An ideal I is a *principal ideal* if $I = (a)$ for some $a \in R$.

So what we have just shown for $\mathbb{Z}$ is that all ideals are principal. Not all rings are like this. These are special types of rings, which we will study more in depth later.

Example 1.41. Consider the following subset:

$$\{f \in \mathbb{R}[X] : \text{ the constant coefficient of } f \text{ is } 0\}.$$

This is an ideal, as we can check manually (alternatively, it is the kernel of the "evaluate at 0" homomorphism). It turns out this is a principal ideal. In fact, it is (X).

We have said two-sided ideals are like normal subgroups. In particular, we have a notion of quotienting by a two-sided ideal:

Definition 1.42. Let $I \subseteq R$ be a two-sided ideal. The *quotient ring* R/I consists of the (additive) cosets $r + I$ with the zero and one as $0_R + I$ and $1_R + I$, and operations

$$(r_1 + I) + (r_2 + I) = (r_1 + r_2) + I$$
$$(r_1 + I) \cdot (r_2 + I) = r_1 r_2 + I.$$

Proposition 1.43. The quotient ring is a ring, and the function

$$\begin{aligned} R &\to R/I \\ r &\mapsto r + I \end{aligned}$$

is a ring homomorphism.

This is true, because we defined ideals to be those things that can be quotiented by. So we just have to check we made the right definition.

Just as we could have come up with the definition of a normal subgroup by requiring operations on the cosets to be well-defined, we could have come up with the definition of an ideal by requiring the multiplication of cosets to be well-defined, and we would end up with the strong closure property.

Proof. We know the group $(R/I, +, 0_{R/I})$ is well-defined, since I is a (normal) subgroup of R. So we only have to check multiplication is well-defined.

Suppose $r_1 + I = r_1' + I$ and $r_2 + I = r_2' + I$. Then $r_1' - r_1 = a_1 \in I$ and $r_2' - r_2 = a_2 \in I$. So

$$r_1' r_2' = (r_1 + a_1)(r_2 + a_2) = r_1 r_2 + r_1 a_2 + a_1 r_2 + a_1 a_2.$$

By the strong closure property, the last three objects are in I. So $r_1' r_2' + I = r_1 r_2 + I$.

It is easy to check that $0_R + I$ and $1_R + I$ are indeed the zero and one, and the function given is clearly a homomorphism. □

Example 1.44. We have the ideals $n\mathbb{Z} \subseteq \mathbb{Z}$. So we have the quotient rings $\mathbb{Z}/n\mathbb{Z}$. The elements are of the form $m + n\mathbb{Z}$, so they are just

$$0 + n\mathbb{Z}, 1 + n\mathbb{Z}, 2 + n\mathbb{Z}, \cdots, (n-1) + n\mathbb{Z}.$$

Addition and multiplication are just what we are used to — addition and multiplication modulo n.

Note that it is easier to come up with ideals than normal subgroups — we can just pick up random elements, and then take the ideal generated by them.

Let's now consider a more explicit link to normal subgroups and group quotients via the group ring.

Example 1.45. Let R be a ring and let $f : G \to H$ be a surjective group homomorphism. Then f induces a surjective ring homomorphism $f_* : R[G] \to R[H]$, $\sum_{g \in G} a_g g \mapsto \sum_{g \in G} a_g f(g)$.

What is $I = \ker(f)$? We know that $I \subseteq R[G]$ is a two-sided ideal. Let $N = \ker(f)$. This is a normal subgroup of G such that $G/N \cong H$. Let $N - 1 = \{n - 1 : n \in N\} \subseteq R[G]$. Since $f_*(n-1) = f_*(n) - 1 = f(n) - 1 = 1 - 1 = 0$, we must have $N - 1 \subseteq I$. Since I is an ideal, we also have $(N-1) = R[G] \cdot N \subseteq I$.

In fact, we can show that $(N-1)$ is a two-sided ideal. If $g \in G$ and $n \in N$, then $g^{-1} n g = n'$ for some $n' \in N$ since $N \subseteq G$ is a normal subgroup. In particular, $(n-1) \cdot g = g \cdot (n'-1) \in (N-1)$. It follows from this that, if $r \in R[G]$ and $\lambda \in (N-1)$, then $\lambda \cdot r \in (N-1)$. Hence $(N-1)$ is a two-sided ideal.

It is an exercise on Problem Sheet 1 to show that $\ker(f_*) = (N-1)$. In particular, we have that $R[G]/(N-1) \cong R[H]$ is an isomorphism of rings.

Example 1.46. Consider $(X) \subseteq \mathbb{C}[X]$. What is $\mathbb{C}[X]/(X)$? Elements are represented by

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n + (X).$$

But everything but the first term is in (X). So every such thing is equivalent to $a_0 + (X)$. It is not hard to convince yourself that this representation is unique. So in fact $\mathbb{C}[X]/(X) \cong \mathbb{C}$, with the bijection $a_0 + (X) \leftrightarrow a_0$.

If we want to prove things like this, we have to convince ourselves this representation is unique. We can do that by hand here, but in general, we want to be able to do this properly.

Proposition 1.47 (Euclidean algorithm for polynomials)**.** Let F be a field and $f, g \in F[X]$. Then there is some $r, q \in F[X]$ such that

$$f = gq + r,$$

with $\deg r < \deg g$.

This is like the usual Euclidean algorithm, except that instead of the absolute value, we use the degree to measure how "big" the polynomial is.

Proof. See Problem Sheet 1. □

Now that we have a Euclidean algorithm for polynomials, we should be able to show that every ideal of $F[X]$ is generated by one polynomial. We will not prove it specifically here, but later show that in *general*, in every ring where the Euclidean algorithm is possible, all ideals are principal.

We now look at some applications of the Euclidean algorithm.

Example 1.48. Consider $\mathbb{R}[X]$, and consider the principal ideal $(X^2 + 1) \subseteq \mathbb{R}[X]$. We let $R = \mathbb{R}[X]/(X^2 + 1)$.

Elements of R are polynomials

$$\underbrace{a_0 + a_1X + a_2X^2 + \cdots + a_nX^n}_{f} + (X^2 + 1).$$

By the Euclidean algorithm, we have

$$f = q(X^2 + 1) + r,$$

with $\deg(r) < 2$, i.e. $r = b_0 + b_1X$. Thus $f + (X^2 + 1) = r + (X^2 + 1)$. So every element of $\mathbb{R}[X]/(X^2 + 1)$ is representable as $a + bX$ for some $a, b \in \mathbb{R}$.

Is this representation unique? If $a + bX + (X^2 + 1) = a' + b'X + (X^2 + 1)$, then the difference $(a - a') + (b - b')X \in (X^2 + 1)$. So it is $(X^2 + 1)q$ for some q. This is possible only if $q = 0$, since for non-zero q, we know $(X^2 + 1)q$ has degree at least 2. So we must have $(a - a') + (b - b')X = 0$. So $a + bX = a' + b'X$. So the representation is unique.

What we've got is that every element in R is of the form $a + bX$, and $X^2 + 1 = 0$, i.e. $X^2 = -1$. This sounds like the complex numbers, just that we are calling it X instead of i.

To show this formally, we define the function

$$\begin{aligned} \varphi : \mathbb{R}[X]/(X^2 + 1) &\to \mathbb{C} \\ a + bX + (X^2 + 1) &\mapsto a + bi. \end{aligned}$$

This is well-defined and a bijection. It is also clearly additive. So to prove this is an isomorphism,

we have to show it is multiplicative. We check this manually. We have

$$\begin{aligned}
&\varphi((a+bX+(X^2+1))(c+dX+(X^2+1))) \\
&= \varphi(ac+(ad+bc)X+bdX^2+(X^2+1)) \\
&= \varphi((ac-bd)+(ad+bc)X+(X^2+1)) \\
&= (ac-bd)+(ad+bc)i \\
&= (a+bi)(c+di) \\
&= \varphi(a+bX+(X^2+1))\varphi(c+dX+(X^2+1)).
\end{aligned}$$

So this is indeed an isomorphism.

This is pretty tedious. Fortunately, we have some helpful results we can use, namely the isomorphism theorems. These are exactly analogous to those for groups.

Theorem 1.49 (First isomorphism theorem)**.** Let $\varphi : R \to S$ be a ring homomorphism. Then $\ker(\varphi) \subseteq R$ is a two-sided ideal and

$$\frac{R}{\ker(\varphi)} \cong \mathrm{im}(\varphi) \leq S.$$

Proof. We have already seen $\ker(\varphi) \subseteq R$. Now define

$$\begin{aligned}
\varphi : R/\ker(\varphi) &\to \mathrm{im}(\varphi) \\
r + \ker(\varphi) &\mapsto \varphi(r).
\end{aligned}$$

This well-defined, since if $r + \ker(\varphi) = r' + \ker(\varphi)$, then $r - r' \in \ker(\varphi)$. So $\varphi(r - r') = 0$. So $\varphi(r) = \varphi(r')$.

We don't have to check this is bijective and additive, since that comes for free from the (proof of the) isomorphism theorem of groups. So we just have to check it is multiplicative. To show φ is multiplicative, we have

$$\begin{aligned}
\varphi((r+\ker(\varphi))(t+\ker(\varphi))) &= \varphi(rt+\ker(\varphi)) \\
&= \varphi(rt) \\
&= \varphi(r)\varphi(t) \\
&= \varphi(r+\ker(\varphi))\varphi(t+\ker(\varphi)).
\end{aligned}$$

□

This is more-or-less the same proof as the one for groups, just that we had a few more things to check. In the example above, showing that $\mathbb{R}[X]/(X^2+1) \cong \mathbb{C}$ now just requires showing that the ring homomorphism $\varphi : \mathbb{R}[X]/(X^2+1) \to \mathbb{C}$ is surjective with $\ker(\varphi) = (X^2+1)$.

Theorem 1.50 (Second isomorphism theorem)**.** Let $R \leq S$ be subrings and let $J \subseteq S$ be a two-sided ideal. Then

(i) $R + J = \{r + j : r \in R, j \in J\} \leq S$ is a subring.

(ii) $J \subseteq R + J$ and $J \cap R \subseteq R$ are each two-sided ideals.

(iii) $\dfrac{R+J}{J} = \{r + J : r \in R\} \leq \dfrac{S}{J}$ is a subring, and $\dfrac{R}{R \cap J} \cong \dfrac{R+J}{J}$.

Proof. Define the function

$$\begin{aligned}\varphi : R &\to S/J \\ r &\mapsto r + J.\end{aligned}$$

Since this is the quotient map, it is a ring homomorphism. The kernel is

$$\ker(\varphi) = \{r \in R : r + J = 0, \text{ i.e. } r \in J\} = R \cap J.$$

Then the image is

$$\mathrm{im}(\varphi) = \{r + J : r \in R\} = \frac{R + J}{J}.$$

Then by the first isomorphism theorem, we know $R \cap J \subseteq R$, and $\frac{R+J}{J} \leq S$, and

$$\frac{R}{R \cap J} \cong \frac{R + J}{J}.$$

□

Theorem 1.51 (Third isomorphism theorem)**.** Let R be a ring and let $I, J \subseteq R$ be two-sided ideals such that $I \subseteq J$. Then $J/I \subseteq R/I$ is a two-sided ideal and

$$\left(\frac{R}{I}\right) / \left(\frac{J}{I}\right) \cong \frac{R}{J}.$$

Proof. We define the map

$$\begin{aligned}\varphi : R/I &\to R/J \\ r + I &\mapsto r + J.\end{aligned}$$

This is well-defined and surjective by the groups case. Also it is a ring homomorphism since multiplication in R/I and R/J are "the same". The kernel is

$$\ker(\varphi) = \{r + I : r + J = 0, \text{ i.e. } r \in J\} = \frac{J}{I}.$$

So the result follows from the first isomorphism theorem. □

2 Integral domains

From now on, we will assume that all rings are commutative (unless we explicitly say 'let R be a ring').

2.1 Integral domains, maximal and prime ideals

Many rings can be nothing like $\mathbb{Z}$. For example, in $\mathbb{Z}$, we know that if $a, b \neq 0$, then $ab \neq 0$. However, in, say, $\mathbb{Z}/6\mathbb{Z}$, we get $2, 3 \neq 0$, but $2 \cdot 3 = 0$. Also, $\mathbb{Z}$ has some nice properties such as every ideal is principal, and every integer has an (essentially) unique factorisation. We will now classify rings according to which properties they have.

Definition 2.1. Let R be a commutative ring. An element $x \in R$ is a *zero divisor* if $x \neq 0$ and there is a $y \neq 0$ such that $x \cdot y = 0 \in R$.

Definition 2.2. An *integral domain* (ID) is a non-trivial commutative ring with no zero divisors, i.e. a ring in which $ab = 0$ implies $a = 0$ or $b = 0$.

Example 2.3. All fields are integral domains, since if $a \cdot b = 0$, and $b \neq 0$, then $a = a \cdot (b \cdot b^{-1}) = 0$. Similarly, if $a \neq 0$, then $b = 0$.

Example 2.4. A subring of an integral domain is an integral domain, since a zero divisor in the small ring would also be a zero divisor in the big ring.

Example 2.5. Immediately, we know $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are integral domains, since $\mathbb{C}$ is a field, and the others are subrings of it. Also, $\mathbb{Z}[i] \leq \mathbb{C}$ is also an integral domain.

These are the nice rings we like in number theory, since there we can sensibly talk about things like factorisation.

It turns out there are no interesting finite integral domains.

Lemma 2.6. Let R be a finite ring which is an integral domain. Then R is a field.

Proof. See Problem Sheet 1. □

So far, we know fields are integral domains, and subrings of integral domains are integral domains. We have another good source of integral domain as follows:

Lemma 2.7. Let R be an integral domain. Then $R[X]$ is also an integral domain.

Proof. We need to show that the product of two non-zero elements is non-zero. Let $f, g \in R[X]$ be non-zero, say

$$f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$$
$$g = b_0 + b_1 X + \cdots + b_m X^m \in R[X],$$

with $a_n, b_m \neq 0$. Then the coefficient of X^{n+m} in fg is $a_n b_m$. This is non-zero since R is an integral domain. So fg is non-zero. So $R[X]$ is an integral domain. □

So, for instance, $\mathbb{Z}[X]$ is an integral domain.

Notation 2.8. Write $R[X, Y]$ for $(R[X])[Y]$, the polynomial ring of R in two variables. In general, write $R[X_1, \cdots, X_n] = (\cdots((R[X_1])[X_2])\cdots)[X_n]$.

Then if R is an integral domain, so is $R[X_1, \cdots, X_n]$.

To some people, it is a shame to think of rings as having elements. Instead, we should think of a ring as a god-like object, and the only things we should ever mention are its ideals. We should also not think of the ideals as containing elements, but just some abstract objects, and all we know is how ideals relate to one another, e.g. if one contains the other.

Under this philosophy, we can think of a field as follows:

Lemma 2.9. A non-trivial commutative ring R is a field if and only if its only ideals are $\{0\}$ and R.

Note that we don't need elements to define the ideals $\{0\}$ and R. $\{0\}$ can be defined as the ideal that all other ideals contain, and R is the ideal that contains all other ideals. Alternatively, we can reword this as "R is a field if and only if it has only two ideals" to avoid mentioning explicit ideals.

Proof. ($\Rightarrow$) Let $I \subseteq R$ and R be a field. Suppose $x \neq 0 \in I$. Then as x is a unit, $I = R$.

($\Leftarrow$) Suppose $x \neq 0 \in R$. Then (x) is an ideal of R. It is not $\{0\}$ since it contains x. So $(x) = R$. In other words $1_R \in (x)$. But (x) is defined to be $\{x \cdot y : y \in R\}$. So there is some $u \in R$ such that $x \cdot u = 1_R$. So x is a unit. Since x was arbitrary, R is a field. □

This is another reason why fields are special. They have the simplest possible ideal structure.

This motivates the following definition:

Definition 2.10. An ideal I of a ring R is *maximal* if $I \neq R$ and for any ideal J with $I \leq J \leq R$, either $J = I$ or $J = R$.

The relation with what we've done above is quite simple. There is an easy way to recognize if an ideal is maximal.

Lemma 2.11. Let R be a commutative ring. An ideal $I \subseteq R$ is maximal if and only if R/I is a field.

Proof. Note that R/I is a field if and only if $\{0\}$ and R/I are the only ideals of R/I.

($\Leftarrow$) If there exists $I \subsetneq J \subsetneq R$, then consider $J/I = \{r + I : r \in J\} \subseteq R/I$. This is an ideal and satisfies $\{0\} \subsetneq J/I \subsetneq R/I$.

($\Rightarrow$) If there exists an ideal $\{0\} \subseteq S \subseteq R/I$, then define $J = \{r \in R : r + I \in S\}$. We can check that this is an ideal and that $I \subsetneq J \subseteq R$. □

Remark 2.12. The proof actually shows something else, namely that we have a one-to-one correspondence:

$$\begin{aligned} \{\text{ideals of } R/I\} &\longleftrightarrow \{\text{ideals of } R \text{ which contain } I\} \\ S \leq \frac{R}{I} &\longrightarrow F = \{x \in R : x + I \in S\} \\ \frac{J}{I} \leq \frac{R}{I} &\longleftarrow J \subseteq S \leq R. \end{aligned}$$

This is a nice result. This makes a correspondence between properties of ideals I and properties of the quotient R/I. Here is another one:

Definition 2.13. An ideal I of a ring R is *prime* if $I \neq R$ and whenever $a, b \in R$ are such that $a \cdot b \in I$, then $a \in I$ or $b \in I$.

Example 2.14. A non-zero ideal $n\mathbb{Z} \subseteq \mathbb{Z}$ is prime if and only if n is a prime.

To show this, first suppose $n = p$ is a prime, and $a \cdot b \in p\mathbb{Z}$. So $p \mid a \cdot b$. So $p \mid a$ or $p \mid b$, i.e. $a \in p\mathbb{Z}$ or $b \in p\mathbb{Z}$.

For the other direction, suppose $n = pq$ is a composite number ($p, q \neq 1$). Then $n \in n\mathbb{Z}$ but $p \notin n\mathbb{Z}$ and $q \notin n\mathbb{Z}$, since $0 < p, q < n$.

So instead of talking about prime numbers, we can talk about prime ideals instead, because ideals are better than elements.

Example 2.15. Let R be an integral domain. Then $(X) \subseteq F[X]$ is a prime ideal.

Note that the trivial ideal $\{0\} \subseteq R$ is prime if and only if R is an integral domain. This generalises to the following:

Lemma 2.16. Let R be a commutative ring. An ideal $I \subseteq R$ is prime if and only if R/I is an integral domain.

Proof. See Problem Sheet 1. □

Prime ideals and maximal ideals are the main types of ideals we care about. Note that every field is an integral domain. So we immediately have the following result:

Corollary 2.17. Let R be a commutative ring. Then every maximal ideal is a prime ideal.

Proof. $I \subseteq R$ is maximal implies R/I is a field implies R/I is an integral domain implies I is prime. □

The converse is not true. For example, $\{0\} \subseteq \mathbb{Z}$ is prime but not maximal. Less stupidly, $(X) \in \mathbb{Z}[X]$ is prime but not maximal (since $\mathbb{Z}[X]/(X) \cong \mathbb{Z}$ and $\mathbb{Z}$ is an integral domain but not a field).

We can provide a more explicit proof of this, which is essentially the same.

Alternative proof. Let I be a maximal ideal, and suppose $a, b \notin I$ but $ab \in I$. Then by maximality, $I + (a) = I + (b) = R = (1)$. So we can find some $p, q \in R$ and $n, m \in I$ such that $n + ap = m + bq = 1$. Then

$$1 = (n + ap)(m + bq) = nm + apm + bqn + abpq \in I,$$

since $n, m, ab \in I$. This is a contradiction. □

Note that for any ring R, there is a unique ring homomorphism $\mathbb{Z} \to R$, given by

$$\begin{aligned} \iota : \mathbb{Z} &\to R \\ n \geq 0 &\mapsto \underbrace{1_R + 1_R + \cdots + 1_R}_{n \text{ times}} \\ n \leq 0 &\mapsto -(\underbrace{1_R + 1_R + \cdots + 1_R}_{-n \text{ times}}) \end{aligned}$$

Any homomorphism $\mathbb{Z} \to R$ must be given by this formula, since it must send the unit to the unit, and we can show this is indeed a homomorphism by distributivity. So the ring homomorphism is unique. We then know $\ker(\iota) \subseteq \mathbb{Z}$. Thus $\ker(\iota) = n\mathbb{Z}$ for some n.

Definition 2.18. Let R be a ring, and $\iota : \mathbb{Z} \to R$ be the unique such map. The *characteristic* of R is the unique non-negative n such that $\ker(\iota) = n\mathbb{Z}$.

Example 2.19. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ all have characteristic 0. The ring $\mathbb{Z}/n\mathbb{Z}$ has characteristic n. In particular, all natural numbers can be characteristics.

Lemma 2.20. Let R be an integral domain. Then its characteristic is either 0 or a prime number.

Proof. Consider the unique map $\varphi : \mathbb{Z} \to R$, and $\ker(\varphi) = n\mathbb{Z}$. Then n is the characteristic of R by definition.

By the first isomorphism theorem, $\mathbb{Z}/n\mathbb{Z} = \mathrm{im}(\varphi) \leq R$. So $\mathbb{Z}/n\mathbb{Z}$ is an integral domain. So $n\mathbb{Z} \subseteq \mathbb{Z}$ is a prime. So $n = 0$ or a prime number. □

2.2 Factorisation in integral domains

We now move on to tackle the problem of factorisation in rings. For sanity, we suppose throughout this section that R is an integral domain. We start with some key definitions.

Definition 2.21. Let R be a ring. For elements $a, b \in R$, we say a *divides* b, written $a \mid b$, if there is a $c \in R$ such that $b = ac$. Equivalently, if $(b) \subseteq (a)$.

Definition 2.22. Let R be a ring. We say $a, b \in R$ are *associates* if $a = bc$ for some unit c. Equivalently, if $(a) = (b)$. Equivalently, if $a \mid b$ and $b \mid a$.

In the integers, this can only happen if a and b differ by a sign but, in other rings, more interesting things can happen.

When considering division in rings, we often consider two associates to be "the same". For example, in $\mathbb{Z}$, we can factorize 6 as

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

but this does not violate unique factorisation, since 2 and -2 are associates (and so are 3 and -3), and we consider these two factorisations to be "the same".

Definition 2.23. Let R be a ring. We say $a \in R$ is *irreducible* if $a \neq 0$, a is not a unit, and if $a = xy$, then x or y is a unit.

For integers, being irreducible is the same as being a prime number. However, "prime" means something different in general rings.

Definition 2.24. Let R be a ring. We say $a \in R$ is *prime* if $a \neq 0$, a is not a unit, and whenever $a \mid xy$, either $a \mid x$ or $a \mid y$.

It is important to note all these properties depend on the ring, not just the element itself.

Example 2.25. $2 \in \mathbb{Z}$ is a prime, but $2 \in \mathbb{Q}$ is not (since it is a unit).

Similarly, the polynomial $2X \in \mathbb{Q}[X]$ is irreducible (since 2 is a unit), but $2X \in \mathbb{Z}[X]$ not irreducible.

Lemma 2.26. A principal ideal (r) is a prime ideal in R if and only if $r = 0$ or r is prime.

Proof. ($\Rightarrow$) Let (r) be a prime ideal. If $r = 0$, then done. Otherwise, as prime ideals are proper, i.e. not the whole ring, r is not a unit. Now suppose $r \mid a \cdot b$. Then $a \cdot b \in (r)$. But (r) is prime. So $a \in (r)$ or $b \in (r)$. So $r \mid a$ or $r \mid b$. So r is prime.

($\Leftarrow$) If $r = 0$, then $(0) = \{0\} \subseteq R$, which is prime since R is an integral domain. Otherwise, let $r \neq 0$ be prime. Suppose $a \cdot b \in (r)$. This means $r \mid a \cdot b$. So $r \mid a$ or $r \mid b$. So $a \in (r)$ and $b \in (r)$. So (r) is prime. □

Lemma 2.27. If $r \in R$ is prime, then it is irreducible.

Proof. Let $r \in R$ be prime, and suppose $r = ab$. Since $r \mid r = ab$, and r is prime, we must have $r \mid a$ or $r \mid b$. wlog, $r \mid a$. So $a = rc$ for some $c \in R$. So $r = ab = rcb$. Since we are in an integral domain, we must have $1 = cb$. So b is a unit. □

In $\mathbb{Z}$, all irreducibles are prime. We now give an example to show that this fails in general.

Example 2.28. Let

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

By definition, it is a subring of a field. So it is an integral domain. What are the units of the ring? There is a nice trick we can use, when things are lying inside $\mathbb{C}$. Consider the function

$$N : R \to \mathbb{Z}_{\geq 0}$$

given by

$$N(a + b\sqrt{-5}) \mapsto a^2 + 5b^2.$$

It is convenient to think of this as $z \mapsto z\bar{z} = |z|^2$. This satisfies $N(z \cdot w) = N(z)N(w)$. This is a desirable thing to have for a ring, since it immediately implies all units have norm 1 — if $r \cdot s = 1$, then $1 = N(1) = N(rs) = N(r)N(s)$. So $N(r) = N(s) = 1$.

So to find the units, we need to solve $a^2 + 5b^2 = 1$, for a and b units. The only solutions are ± 1. So only $\pm 1 \in R$ can be units, and these obviously are units. Hence $R^\times = \{\pm 1\}$.

Next, we claim $2 \in R$ is irreducible. We again use the norm. Suppose $2 = ab$. Then $4 = N(2) = N(a)N(b)$. Now note that nothing has norm 2. $a^2 + 5b^2$ can never be 2 for integers $a, b \in \mathbb{Z}$. So we must have, wlog, $N(a) = 4, N(b) = 1$. So b must be a unit. Similarly, we see that $3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are irreducible (since there is also no element of norm 3).

We have four irreducible elements in this ring. We claim that they are not prime. Note that

$$(1 + \sqrt{-5})(1 - \sqrt{-5}) = 6 = 2 \cdot 3.$$

We now claim 2 does not divide $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$. So 2 is not prime.

To show this, suppose $2 \mid 1 + \sqrt{-5}$. Then $N(2) \mid N(1 + \sqrt{-5})$. But $N(2) = 4$ and $N(1 + \sqrt{-5}) = 6$, and $4 \nmid 6$. Similarly, $N(1 - \sqrt{-5}) = 6$ as well. So $2 \nmid 1 \pm \sqrt{-5}$.

There are several things to be learnt here. First is that primes and irreducibles are not the same thing in general. The second is that factorisation into irreducibles is not necessarily unique, since $2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ are two factorisations into irreducibles.

However, there is one situation when unique factorisations holds. This is when we have a Euclidean algorithm available.

Definition 2.29 (Euclidean domain)**.** An integral domain R is a *Euclidean domain* (ED) if there is a *Euclidean function* $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ such that

1. $\phi(a \cdot b) \geq \phi(b)$ for all $a, b \neq 0$
2. If $a, b \in R$, with $b \neq 0$, then there are $q, r \in R$ such that

$$a = b \cdot q + r,$$

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

Every time we said we have a "Euclidean algorithm", we have an example of a Euclidean domain.

Example 2.30. $\mathbb{Z}$ is a Euclidean domain with $\phi(n) = |n|$.

Example 2.31. For any field F, $F[X]$ is a Euclidean domain with

$$\phi(f) = \deg(f).$$

Division is always possible in fields, even with no remainders:

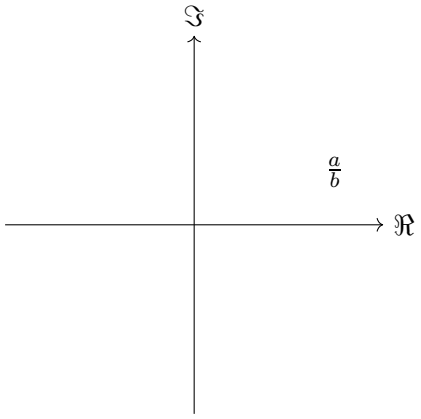
Example 2.32. If F is a field, then F is a Euclidean domain with $\phi(x) = 0$ for all $x \in F^\times = F \setminus \{0\}$. Note that we can always take $r = 0$ in condition (2).

Example 2.33. The Gaussian integers $R = \mathbb{Z}[i] \leq \mathbb{C}$ is a Euclidean domain with $\phi(z) = N(z) = |z|^2$. We now check this:

1. We have $\phi(zw) = \phi(z)\phi(w) \geq \phi(z)$, since $\phi(w)$ is a positive integer.
2. Given $a, b \in \mathbb{Z}[i]$, $b \neq 0$. We consider the complex number

   $$\frac{a}{b} \in \mathbb{C}.$$

   Consider the following complex plane, where the red dots are points in $\mathbb{Z}[i]$.

   By looking at the picture, we know that there is some $q \in \mathbb{Z}[i]$ such that $\left|\frac{a}{b} - q\right| < 1$. So we can write

   $$\frac{a}{b} = q + c$$

   with $|c| < 1$. Then we have

   $$a = b \cdot q + \underbrace{b \cdot c}_{r}.$$

   We know $r = a - bq \in \mathbb{Z}[i]$, and $\phi(r) = N(bc) = N(b)N(c) < N(b) = \phi(b)$. So done.

This is not just true for the Gaussian integers. All we really needed was that $R \leq \mathbb{C}$, and for any $x \in \mathbb{C}$, there is some point in R that is not more than 1 away from x. If we draw some more pictures, we will see this is not true for $\mathbb{Z}[\sqrt{-5}]$.

Before we move on to prove unique factorisation, we first derive something we've previously mentioned. Recall we showed that every ideal in $\mathbb{Z}$ is principal, and we proved this by the Euclidean algorithm. So we might expect this to be true in an arbitrary Euclidean domain.

Definition 2.34 (Principal ideal domain)**.** A ring R is a *principal ideal domain* (PID) if it is an integral domain, and every ideal is a principal ideal, i.e. for all $I \subseteq R$, there is some a such that $I = (a)$.

Example 2.35. $\mathbb{Z}$ is a principal ideal domain.

Proposition 2.36. Let R be a Euclidean domain. Then R is a principal ideal domain.

We have already proved this, just that we did it for a particular Euclidean domain $\mathbb{Z}$. Nonetheless, we shall do it again.

Proof. Let R have a Euclidean function $\phi : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$. We let $I \subseteq R$ be a non-zero ideal, and let $b \in I \setminus \{0\}$ be an element with $\phi(b)$ minimal. Then for any $a \in I$, we write

$$a = bq + r,$$

with $r = 0$ or $\phi(r) < \phi(b)$. However, any such r must be in I since $r = a - bq \in I$. So we cannot have $\phi(r) < \phi(b)$. So we must have $r = 0$. So $a = bq$. So $a \in (b)$. Since this is true for all $a \in I$, we must have $I \subseteq (b)$. On the other hand, since $b \in I$, we must have $(b) \subseteq I$. So we must have $I = (b)$. □

This is exactly, word by word, the same proof as we gave for the integers, except we replaced the absolute value with ϕ.

Example 2.37. $\mathbb{Z}$ is a Euclidean domain, and hence a principal ideal domain. Also, for any field $\mathbb{F}$, $\mathbb{F}[X]$ is a Euclidean domain, hence a principal ideal domain.

Also, $\mathbb{Z}[i]$ is a Euclidean domain, and hence a principal ideal domain.

Example 2.38. Is $\mathbb{Z}[X]$ a Euclidean domain? If so, then it must be a principal ideal domain. This is easier to work with, so let's consider this instead.

We claim that the ideal $(2, X) \subseteq \mathbb{Z}[X]$ is not a principal ideal. Suppose it were. Then $(2, X) = (f)$. Since $2 \in (2, X) = (f)$, we know $2 \in (f)$, i.e. $2 = f \cdot g$ for some g. So f has degree zero, and hence constant. So $f = \pm 1$ or ± 2.

If $f = \pm 1$, since ± 1 are units, then $(f) = \mathbb{Z}[X]$. But $(2, X) \neq \mathbb{Z}[X]$, since, say, $1 \notin (2, X)$. If $f = \pm 2$, then since $X \in (2, X) = (f)$, we must have $\pm 2 \mid X$, but this is clearly false. So $(2, X)$ cannot be a principal ideal. Hence $\mathbb{Z}[X]$ is not principal ideal domain or a Euclidean domain.

Remark 2.39 (Tangential)**.** At first sight, you might expect a proof that $\mathbb{Z}[X]$ is a Euclidean domain to start by saying "Suppose ϕ is a Euclidean function..." and then proceed to a contradiction. Here we instead use that, if $\mathbb{Z}[X]$ is a Euclidean domain, then it is also a principal ideal domain, which is a property that is easier to work with. In fact, it is quite difficult to come up with a ring which is a Euclidean domain but not a principal ideal domain.

This is a typical style of proof in algebra and is worth remembering. For example, in algebraic topology, we might ask whether two spaces X and Y are homeomorphic ('topologically equivalent'). We almost never argue by saying "Let $f : X \to Y$ be a homeomorphism between them...". If X and Y are homeomorphic, then two associated rings $H^*(X)$ and $H^*(Y)$ must be isomorphic as rings. If we can show that these rings are not isomorphic, then we deduce that X and Y are not homeomorphic. This is the approach of *invariants*.

Example 2.40. Let $A \in M_{n \times n}(F)$ be an $n \times n$ matrix over a field F. We consider the following set

$$I = \{f \in F[X] : f(A) = 0\}.$$

This is an ideal — if $f, g \in I$, then $(f+g)(A) = f(A) + g(A) = 0$. Similarly, if $f \in I$ and $h \in F[X]$, then $(fg)(A) = f(A)g(A) = 0$.

But we know $F[X]$ is a principal ideal domain. So there must be some $m \in F[X]$ such that $I = (m)$ for some m.

Suppose $f \in F[X]$ such that $f(A) = 0$, i.e. $f \in I$. Then $m \mid f$. So m is a polynomial that divides all polynomials that kill A, i.e. m is the *minimal polynomial* of A.

We have just proved that all matrices have minimal polynomials, and that the minimal polynomial divides all other polynomials that kill A. Also, the minimal polynomial is unique up to multiplication of units.

For a general ring, we cannot factorize things into irreducibles uniquely. We say this in the example $\mathbb{Z}[\sqrt{-5}]$. However, in some rings, this is possible.

Definition 2.41 (Unique factorisation domain)**.** An integral domain R is a *unique factorisation domain* (UFD) if

1. Every non-zero non-unit may be written as a product of irreducibles (Existence).
2. If $p_1 \cdots p_n = q_1 \cdots q_m$ with p_i, q_j irreducibles, then $n = m$, and they can be reordered such that p_i is an associate of q_i (Uniqueness).

This is a really nice property, and here we can do things we are familiar with in number theory. So how do we know if something is a unique factorisation domain? Our goal is to show the following:

Theorem 2.42 (PID $\Rightarrow$ UFD)**.** If R be a principal ideal domain, then R is a unique factorisation domain.

Sketch proof. Uniqueness: In $\mathbb{Z}$, irreducibles are primes so suppose this is true here. If $p_1 \cdots p_n = q_1 \cdots q_m$ for p_i, q_i irreducible (hence prime). Then $p_n \mid q_1 \cdots q_m$ implies $p_n \mid q_i$ for some i. Wlog $i = m$. So $q_m = p_n r$ for some $r \in R$. Since q_m is irreducible, r is a unit hence q_m and p_n are associates. Since R is an integral domain, this implies that $p_1 \cdots p_{n-1} = q_1 \cdots q'_{m-1}$ where $q'_{m-1} = q_{m-1} r$ which is still irreducible. This process eventually terminates and implies $n = m$.

Existence: Let r be a non-zero non-unit. If r is irreducible, then done. If not then we can write $r = r_1 s_1$ for some $r_1, s_1 \in R$ non-units. If r_1, s_1 are irreducible then done so assume not. Wlog assume that r_1 is not irreducible. Then we can write $r_1 = r_2 s_2$. If this process never terminates, then we have a sequence $\{r_i\}_{i \geq 1}$ in R such that $r_{i+1} \mid r_i$ but r_i and r_{i+1} are not associates. If we take $I_i = (r_i)$, this implies we have a chain of ideals $I_1 \subsetneq I_2 \subsetneq \cdots$ which never terminates. This never happens in $\mathbb{Z}$ so assume this never happens in R. If so, then this process terminates. Hence it terminates for all 'paths' through the r_i, s_i (e.g. picking s_1 instead of r_1 at the first step). This gives the required factorisation. □

The aim of the rest of this section is to fill in the details of the sketch proof and, along the way, to establish basic properties of principal ideal domains.

Firstly recall that, for a commutative ring R, a principal ideal (r) is prime if and only if $r \in R$ is prime or $r = 0$. In the special situation of principal ideal domains, we also have the following relation between maximal ideals and irreducible elements:

Lemma 2.43. Let R be a principal ideal domain. Then a principal ideal (r) is maximal if and only if r is irreducible or, if R is a field, $r = 0$.

Proof. ($\Rightarrow$) If r is not irreducible, then $r = xy$ for $x, y \in R$ non-units. Then $(r) \subseteq (x) \subseteq R$. Since x is not a unit, we have $(x) \neq R$. Since (r) is maximal, $(r) = (x)$. This implies that $x = rz$ for some non-unit z. So $r = xy = ryz$ and $r(1 - yz) = 0$. Since R is an integral domain, we have $yz = 1$ or $r = 0$. The former cannot hold since y is a non-unit, so we deduce that $r = 0$. This is a contradiction unless R is a field since (0) maximal would imply that $R \cong R/(0)$ is a field.

($\Leftarrow$) Note that, if R is a field, then (0) is maximal. Now suppose $r \in R$ is irreducible. If (r) is non-maximal, then there exists an ideal I such that $(r) \subsetneq I \subsetneq R$. Since R is a principal ideal domain, we have $I = (x)$ for some $x \in R$. Since $(x) \neq R$, x is a non-unit. Since $(r) \subseteq (x)$, we have $r = xy$ for y a non-unit. This is a contradiction. Hence (r) is maximal. □

Proposition 2.44. Let R be a principal ideal domain. If $r \in R$ is irreducible, then it is prime.

We already know that, if $r \in R$ is prime, then it is irreducible. Hence it follows that, in a principal ideal domain, irreducible elements and prime elements coincide.

Proof. If $r \in R$ is irreducible, then the above implies that (r) is maximal ideal. Every maximal ideal is prime so (r) is a prime ideal. Since (r) is prime and $r \neq 0$, it follows from a previous lemma that r is prime. □

Alternative proof. Let $p \in R$ be irreducible, and suppose $p \mid a \cdot b$. Also, suppose $p \nmid a$. We need to show $p \mid b$.

Consider the ideal $(p, a) \subseteq R$. Since R is a principal ideal domain, there is some $d \in R$ such that $(p, a) = (d)$. So $d \mid p$ and $d \mid a$.

Since $d \mid p$, there is some q_1 such that $p = q_1 d$. As p is irreducible, either q_1 or d is a unit.

If q_1 is a unit, then $d = q_1^{-1} p$, and this divides a. So $a = q_1^{-1} px$ for some x. This is a contradiction, since $p \nmid a$.

Therefore d is a unit. So $(p, a) = (d) = R$. In particular, $1_R \in (p, a)$. So suppose $1_R = rp + sa$, for some $r, s \in R$. We now take the whole thing and multiply by b. Then we get

$$b = rpb + sab.$$

We observe that ab is divisible by p, and so is p. So b is divisible by p. So done. □

Corollary 2.45. Let R be a principal ideal domain. Then every non-zero prime ideal is maximal.

We already know that every maximal ideal is prime. Hence it follows that, prime ideals and maximal ideals coincide with the exception of $\{0\}$ which is alway prime and which is maximal if and only if R is a field.

Proof. If I is a non-zero prime ideal, then $I = (r)$ for $r \in R$ a prime by a previous lemma. In general, primes are irreducible so r is irreducible. The proposition above then shows that (r) is maximal. □

We will now consider another nice property of principal ideal domains.

Definition 2.46. A commutative ring satisfies the *ascending chain condition* (ACC) if there is no infinite strictly increasing chain of ideals. That is, if $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ be a chain of ideals, then there is some $N \in \mathbb{N}$ such that $I_n = I_{n+1}$ for some $n \geq N$.

Definition 2.47. A commutative ring that satisfies the ascending chain condition is known as a *Noetherian ring*.

Proposition 2.48. If R is a principal ideal domain, then R is Noetherian.

Proof. The obvious thing to do when we have an infinite chain of ideals is to take the union of them. We let

$$I = \bigcup_{n\geq 1}^{\infty} I_n,$$

which is again an ideal. Since R is a principal ideal domain, $I = (a)$ for some $a \in R$. We know $a \in I = \bigcup_{n=0}^{\infty} I_n$. So $a \in I_N$ for some N. Then we have

$$(a) \subseteq I_N \subseteq I = (a)$$

So we must have $I_N = I$. So $I_n = I_N = I$ for all $n \geq N$. □

Notice it is not important that I is generated by one element. If, for some reason, we know I is generated by finitely many elements, then the same argument works. So if every ideal is finitely generated, then the ring must be Noetherian. In fact, a ring is Noetherian if and only if every ideal is finitely generated.

Finally, we have done the setup, and we can prove the theorem.

Proof of "PID ⇒ UFD". Uniqueness: Let $p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$, with p_i, q_i irreducible. So in particular $p_1 \mid q_1 \cdots q_m$. Since p_1 is irreducible, it is prime. So p_1 divides some q_i. We reorder and suppose $p_1 \mid q_1$. So $q_1 = p_1 \cdot a$ for some a. But since q_1 is irreducible, a must be a unit. So p_1, q_1 are associates. Since R is a principal ideal domain, hence integral domain, we can cancel p_1 to obtain

$$p_2 p_3 \cdots p_n = (a q_2) q_3 \cdots q_m.$$

We now rename aq_2 as q_2, so that we in fact have

$$p_2 p_3 \cdots p_n = q_2 q_3 \cdots q_m.$$

We can then continue to show that p_i and q_i are associates for all i. This also shows that $n = m$, or else if $n = m + k$, saw, then $p_{k+1} \cdots p_n = 1$, which is a contradiction.

Existence: Suppose $r \in R$ is a non-unit which cannot be factored as a product of irreducibles. Then it is certainly not irreducible. So we can write $r = r_1 s_1$, with r_1, s_1 both non-units. Since r cannot be factored as a product of irreducibles, wlog r_1 cannot be factored as a product of irreducibles (if both can, then r would be a product of irreducibles). So we can write $r_1 = r_2 s_2$, with r_2, s_2 not units. Again, wlog r_2 cannot be factored as a product of irreducibles. We continue this way.

By assumption, the process does not end, and then we have the following chain of ideals:

$$(r) \subseteq (r_1) \subseteq (r_2) \subseteq \cdots \subseteq (r_n) \subseteq \cdots$$

But then we have an ascending chain of ideals. By the ascending chain condition, these are all eventually equal, i.e. there is some n such that $(r_n) = (r_{n+1}) = (r_{n+2}) = \cdots$. In particular, since $(r_n) = (r_{n+1})$, and $r_n = r_{n+1} s_{n+1}$, then s_{n+1} is a unit. But this is a contradiction, since s_{n+1} is not a unit. So r must be a product of irreducibles. □

Example 2.49. Since $\mathbb{Z}[i]$ is a Euclidean domain, it must be a principal ideal domain and the above implies that it is a unique factorisation domain. This was proven by Gauss in 1832 and provided a lot of motivation for the theory that was later developed.

We can now use this to define other familiar notions from number theory.

Definition 2.50. In a ring R, d is a *greatest common divisor* (gcd) of $a_1, a_2, \cdots, a_n$ if $d \mid a_i$ for all i, and if any other d' satisfies $d' \mid a_i$ for all i, then $d' \mid d$.

Note that the gcd of a set of numbers, if it exists, is not unique. It is only well-defined up to multiplication by units.

This is a definition that says what it means to be a greatest common divisor. However, it does not always have to exist.

Lemma 2.51. Let R be a unique factorisation domain. Then greatest common divisors exists, and is unique up to associates, i.e. if d and d' are greatest common divisors of $a_1, a_2, \cdots, a_n$, then d and d' are associates.

Proof. Let $p_1, p_2, \cdots, p_m$ be a list of all irreducible factors of a_i such that no two of these are associates of each other. We now write

$$a_i = u_i \prod_{j=1}^{m} p_j^{n_{ij}},$$

where $n_{ij} \in \mathbb{N}$ and u_i are units. We let

$$m_j = \min_i \{n_{ij}\},$$

and choose

$$d = \prod_{j=1}^{m} p_j^{m_j}.$$

As, by definition, $m_j \leq n_{ij}$ for all i, we know $d \mid a_i$ for all i.

Finally, if $d' \mid a_i$ for all i, then we let

$$d' = v \prod_{j=1}^{m} p_j^{t_j}.$$

Then we must have $t_j \leq n_{ij}$ for all i, j. So we must have $t_j \leq m_j$ for all j. So $d' \mid d$.

Uniqueness is immediate since any two greatest common divisors have to divide each other. □

Note that, if d and d' are associates, then they generate the same ideal $(d) = (d')$. Therefore the greatest common divisor is, in some sense, better thought of as an ideal in a ring rather than an actual element.

We have now completed the first major goal of this course, namely to establish the following chain of implications:

$$(\mathbb{Z}) \Rightarrow \text{ED} \Rightarrow \text{PID} \Rightarrow \text{UFD} \Rightarrow \text{ID} \Rightarrow \text{Commutative Ring} \Rightarrow \text{Ring}$$

where $(\mathbb{Z})$ just denotes the property of being isomorphic to $\mathbb{Z}$. For each ring R, we can classify how similar it is to $\mathbb{Z}$ by seeing how many properties it satisfies, i.e. how far left it sits in the chain of implications. To show that these all make sense as separate definitions, we also need to find examples showing that each implication cannot be reversed:

$$(\mathbb{Z}) \underbrace{\not\Leftarrow}_{\mathbb{Q},\, \mathbb{Z}[i]} \text{ED} \underbrace{\not\Leftarrow}_{\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]} \text{PID} \underbrace{\not\Leftarrow}_{\mathbb{Z}[X]} \text{UFD} \underbrace{\not\Leftarrow}_{\mathbb{Z}[\sqrt{-5}]} \text{ID} \underbrace{\not\Leftarrow}_{\mathbb{Z}/6\mathbb{Z}} \text{Commutative Ring} \underbrace{\not\Leftarrow}_{M_2(\mathbb{Z})} \text{Ring}.$$

We have not seen two of these examples yet: $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ and $\mathbb{Z}[X]$. The first is the challenge problem on Problem Sheet 1.

We already know that $\mathbb{Z}[X]$ is not principal ideal domain, e.g. consider $(2, X)$. We will see later that $\mathbb{Z}[X]$ is a unique factorisation domain.

Remark 2.52 (Tangential)**.** The above discussion is an example of a 'classification hierarchy'; something which is central to many classification problems in pure mathematics. Take a simple mathematical object (e.g. $\mathbb{Z}$) and a wider class of mathematical object of which this is an example (e.g. Rings). Then find general properties which are satisfied by the object in question (e.g. ED, PID) and relate them to each other (e.g. ED $\Rightarrow$ PID, PID $\not\Rightarrow$ ED). The end result is a broad understanding of our simple mathematical object and what characterises it.

In topology, a simple object is an n-dimensional sphere S^n and a class of objects is n-dimensional manifolds. There are many equivalence relations on manifolds: diffeomorphism ($\cong_{\text{Diff}}$), which implies homeomorphism ($\cong$), which implies homotopy equivalent ($\simeq$). So we have:

$$(M \cong_{\text{Diff}} S^n) \Rightarrow (M \cong S^n) \Rightarrow (M \simeq S^n).$$

Which of these implications can be reversed? This is a separate question for each n. This program is known as the Generalised Poincaré Conjecture. Milnor won a Fields medal for finding an example to show that $(M \cong S^n) \not\Rightarrow (M \cong_{\text{Diff}} S^n)$ in the case $n = 7$ (an 'exotic sphere'). Three more Fields Medals were awarded to Smale, Freedman and Perelman for showing that $(M \cong S^n) \Leftrightarrow (M \simeq S^n)$ for all n. Currently just one case remains open and is in four dimensions: does $M \cong S^4$ implies $M \cong_{\text{Diff}} S^4$?

Remark 2.53. A (not necessarily commutative) ring with no zero divisors is a *domain* and many of the nice properties of integral domains also hold for domains. The quaternions $\mathbb{H}$ is a domain, as is any division ring. An important open problem is: if G is a torsion-free group (i.e. every non-zero element has infinite order) and F is a field, then is $F[G]$ a domain? This is Kaplansky's zero divisor problem and dates back to at least the 1940s.

2.3 Localisation

Let R be a commutative ring. Our goal is to define a ring of fractions "$\frac{a}{s}$" for all $a \in R$ and only certain $s \in R$ (not necessarily all).

Here are the prototypical examples:

- R is an integral domain and we consider all fractions $\frac{a}{b}$ where $b \neq 0$: this is a field, called the *field of fractions of* R (we will define this carefully below).

  Examples: the field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$, the field of fractions of $\mathbb{Z}[\sqrt{d}]$ is $\mathbb{Q}(\sqrt{d})$, the field of fractions of $F[X]$ (for a field F) is $F(X)$ (which means all fractions $\frac{f}{g}$ with $f, g \in F[X]$ and $g \neq 0$).

- R is an integral domain and we invert elements with certain prime factors only. For example, $\mathbb{Z}[\frac{1}{p}]$ is the ring consisting of elements $\frac{a}{p^m}$ for $m \geq 0$ and $a \in \mathbb{Z}$.

Definition 2.54. Let R be an integral domain and $S \subseteq (R, \cdot)$ a submonoid not containing 0. The localisation $S^{-1}R$ is defined as the set of equivalence classes of pairs (r, s) with $r \in R$ and $s \in S$, under the equivalence relation $(r, s) \sim (r', s')$ if $rs' = r's$. The pair (r, s) is denoted $\frac{r}{s}$.

This is a ring with operations $(r, s) \cdot (r', s') := (rr', ss')$ and $(r, s) + (r', s') := (rs' + r's, ss')$.

Check this is an equivalence relation: reflexivity and symmetry are immediate; for transitivity if $(r, s) \sim (r', s')$ and $(r', s') \sim (r'', s'')$, then $rs' = sr'$ and $r's'' = s'r''$, so we get

$rs's'' = sr's'' = ss'r''$, so $s'(rs'' - sr'') = 0$. In an integral domain this implies that $rs'' = sr''$ or else $s' = 0$, but the latter we assumed was not the case.

We can also check that this is a ring with the given operations.

Definition 2.55. Let $R = \mathbb{Z}$ and let $S = R \setminus \{0\}$. Then the *rational numbers* $\mathbb{Q}$ is defined as the ring $S^{-1}R$.

Example 2.56. Let $R = \mathbb{Z}$ and let $S = \{p^i : i \geq 0\}$ for a prime p. Then $S^{-1}R \subseteq \mathbb{Q}$ is the subset of rational numbers $\{\frac{a}{p^m} : m \geq 0\}$. We often denote this by $\mathbb{Z}[\frac{1}{p}]$.

Note that we could take $S = \langle p \rangle$ to be the multiplicative submonoid generated by p, i.e. the smallest multiplicative submonoid of $\mathbb{Z}$ which contains p.

Observe there is an obvious map $\iota : R \to S^{-1}R$ sending a to $(a, 1)$.

Proposition 2.57. Let R be an integral domain and let S be a multiplicative submonoid such that $0 \notin S$. Then the map $\iota : R \to S^{-1}R$ is injective.

Proof. We need to show that $(r, 1) \sim (r', 1)$ implies $r = r'$. This is immediate from the definition of the equivalence relation. □

Remark 2.58. If R and S are rings and $f : R \to S$ is injective ring homomorphism, then R is isomorphic to a subring of S, namely the subring $\mathrm{Im}(f) \leq S$. In general, we view a ring as an equivalence class of rings up to ring isomorphism, i.e. we say that two rings are the same if they are isomorphic. We therefore often say that R *is* a subring of S.

In the above proposition, we could therefore say that R is a subring $S^{-1}R$.

For general commutative rings the definition looks almost the same:

Definition 2.59. For R a commutative ring and $S \subseteq R$ a submonoid, the localisation $S^{-1}R$ is defined as the equivalence classes of pairs $(r, s), r \in R, s \in S$ subject to the relation $(r, s) \sim (r', s')$ if there exists $t \in S$ such that $t(rs' - r's) = 0$. As before, (r, s) can be denoted $\frac{r}{s}$.

This is a ring with operations $(r, s) \cdot (r', s') := (rr', ss')$ and $(r, s) + (r', s') := (rs' + r's, ss')$.

The presence of the t is the difference between the two definitions. In an integral domain, we assume $0 \notin S$ and so $t \neq 0$ and $t(rs' - r's) = 0$ implies $rs' - r's = 0$.

Exercise 2.60. Show that $\sim$ is an equivalence relation in the definition.

Note that in the definition of localisation for an integral domain we require that the submonoid $S \subseteq R$ does not include zero. Indeed, if we apply the equivalence relation for integral domains and $0 \in S$, then we get $(r, s) \sim (0, 0)$ for all r, s and hence $S^{-1}R = 0$. Applying the more general equivalence relation $((r, s) \sim (r', s')$ if there exists $t \in S$ such that $t(rs' - r's) = 0)$ also yields that $(r, s) \sim (0, 0)$ for all r, s when $0 \in S$, so we get:

Example 2.61. If S contains zero, then $S^{-1}R \cong \{0\}$ is the zero ring.

There is still a map $\iota : R \to S^{-1}R$ in general, but the following shows that this is not injective when $0 \in S$ unless $R = \{0\}$.

Remark 2.62. This is why we assumed that, for R an integral domain, the multiplicative submonoid S does not contain 0. We needed this to be true to get that the map $\iota : R \to S^{-1}R$ is injective, i.e. that R can be viewed as a subring of $S^{-1}R$.

In fact, this is still often the case when $0 \notin S$:

Example 2.63. Consider $\{1,2,4,5\}^{-1}\mathbb{Z}/6$. In here $(0,2) \sim (3,1)$ so $3 = 0$. We get therefore a map $\{2,4\}^{-1}\mathbb{Z}/6 \to \mathbb{Z}/3$ sending (a,b) to $b^{-1} \cdot a$ (as 2 and 4 are invertible in $\mathbb{Z}/3$). This map is clearly surjective and the kernel consists only of (a,b) for $a \in \{0,3\}$ and $b \in \{1,2,4,5\}$. All of these elements are equivalent to $(0,1)$. Thus the map is injective and we get $\{1,2,4,5\}^{1}\mathbb{Z}/6 \cong \mathbb{Z}/3$. So here the induced map $\mathbb{Z}/6 \to \mathbb{Z}/3$ is not injective.

Definition 2.64. If R is an integral domain and $S = R\setminus\{0\}$, then $S^{-1}R$ is a field: $\frac{a}{b}\frac{b}{a} = \frac{ab}{ab} = 1$. We refer to $S^{-1}R$ as the *field of fractions* of R. We will denote this by $\mathrm{Frac}(R)$.

Remark 2.65. The integral domain hypothesis is needed here, otherwise $R \setminus \{0\}$ is not a multiplicative submonoid. Indeed for some nonzero a, b, we have $ab = 0$.

Example 2.66. The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$; of $\mathbb{Z}[\sqrt{d}]$ it is $\mathbb{Q}(\sqrt{d})$; of $F[x_1,\ldots,x_n]$ it is $F(x_1,\ldots,x_n)$.

Note that $S \subseteq (S^{-1}R)^\times$ is a submonoid of the unit group of $S^{-1}R$. The following says that $S^{-1}R$ is "the smallest thing obtained from R by ensuring that everything in S becomes a unit".

Proposition 2.67 (Universal property of localisation)**.** If A is any commutative ring and $\varphi : R \to A$ a ring homomorphism such that $\varphi(S) \subseteq A^\times$, then φ factors through the homomorphism $\iota : R \to S^{-1}R$: i.e., there exists a unique $\widetilde{\varphi} : S^{-1}R \to A$ such that $\varphi = \widetilde{\varphi} \circ \iota$.

Proof. Let $\widetilde{\varphi}(\frac{a}{b}) = \varphi(a)\varphi(b)^{-1}$ for $a \in R$ and $b \in S$. One must check this map is well-defined: if $\frac{a}{b} = \frac{c}{d}$, then $\varphi(a)\varphi(b)^{-1} = \varphi(c)\varphi(d)^{-1}$. The latter equality is equivalent to $\varphi(a)\varphi(d) = \varphi(b)\varphi(c)$. Now by assumption, there exists $t \in S$ such that $t(ad - bc) = 0$. Therefore $\varphi(t)\varphi(ad-bc) = 0$. Multiplying by $\varphi(t)^{-1}$, we obtain $\varphi(ad-bc) = 0$, hence $\varphi(a)\varphi(d) = \varphi(b)\varphi(c)$ as desired.

It is straightforward to check that, since φ is a ring homomorphism, so is $\widetilde{\varphi}$ (one just checks that fractions have to add and multiply in the usual way in A). Finally, for uniqueness, note that $\widetilde{\varphi}(\frac{1}{b})$ must be an inverse of $\varphi(b)$, and inverses are unique. By the homomorphism property, we get $\widetilde{\varphi}(\frac{a}{b}) = \varphi(a)\varphi(b)^{-1}$. □

We can use this to give the following alternate, much more abstract, definition of localisation:

Definition 2.68. Let R be a commutative ring and let $S \subseteq R$ be a multiplicative submonoid. Then the *localisation*, written $S^{-1}R$, is the unique ring R' such that there exists a map $\iota : R \to R'$ with the following two properties:

(i) $\iota(S) \subseteq (R')^\times$, i.e. everything in S gets mapped to a unit in R'.

(ii) For all commutative rings A and maps $\varphi : R \to A$ with $\varphi(S) \subseteq A^\times$, there exists a unique $\widetilde{\varphi} : R' \to A$ such that $\varphi = \widetilde{\varphi} \circ \iota$.

Exercise 2.69. Prove that this definition is well defined, i.e. that a ring R' satisfying those properties exists and is unique. Deduce that it is equivalent to the previous definition of localisation.

Note that the existence of such a ring follows from the proposition above, i.e. take $R' := S^{-1}R$ and $\iota : R \to S^{-1}R$ using the old definition. Uniqueness is more difficult.

In the case R is an integral domain so that $R \subseteq S^{-1}R$, the proposition says that every $\varphi : R \to A$ sending S to $A^\times$ must extend to $\widetilde{\varphi} : S^{-1}R \to A$.

Corollary 2.70. If R is an integral domain, F a field, and $\varphi : R \to F$ an injective ring homomorphism, then φ must factor through the map from R to its field of fractions: $\varphi = \iota \circ \widetilde{\varphi}$ for $\iota : R \hookrightarrow \mathrm{Frac}(R)$ the canonical map. Moreover, the resulting $\widetilde{\varphi} : \mathrm{Frac}(R) \to F$ is injective.

Proof. Just apply the proposition with $S = R\setminus\{0\}$: Since φ is injective, $\varphi(S) \subseteq \mathrm{Frac}(R)\setminus\{0\} = \mathrm{Frac}(R)^\times$, so the proposition applies. Then the map $\widetilde{\varphi} : \mathrm{Frac}(R) \to F$ is a homomorphism from a field to a nonzero ring, hence injective (the kernel of $\widetilde{\varphi}$ cannot be $\mathrm{Frac}(R)$ since $1 \neq 0$ in F, hence it must be $\{0\}$ as $\mathrm{Frac}(R)$ is a field). □

Corollary 2.71. If F is a field of characteristic zero, it contains a subfield isomorphic to $\mathbb{Q}$. If F is a field of characteristic p, it contains a subfield isomorphic to $\mathbb{F}_p$.

Proof. The characteristic of a field F is zero if and only if the unique ring homomorphism $\mathbb{Z} \to F$ is injective. But then it factors through $\mathbb{Z} \hookrightarrow \mathbb{Q}$, i.e., $\mathbb{Q} \subseteq F$.

If the characteristic of a field is $p > 0$, it contains $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (indeed, this holds for any ring by definition of characteristic). □

For this reason $\mathbb{F}_p$ and $\mathbb{Q}$ are called *the prime fields*.

Lemma 2.72. If F is a field and F is a subring of R, then R is a vector space over F.

Proof. Just observe that R is an abelian group equipped with the scalar multiplication by F coming from the multiplication in R. Check the axioms! □

Corollary 2.73. Every field is a vector space over $\mathbb{F}_p$ or $\mathbb{Q}$ (for the field having characteristic $p > 0$ or 0 respectively).

There is a nice multiplicative submonoid associated to any integral domain R, namely $R\setminus\{0\}$. This leads to the field of fractions $\mathrm{Frac}(R)$. Are there other nice multiplicative submonoids which can be defined for all integral domains, or even all commutative rings?

Example 2.74. Let R be a commutative ring. If $I \subseteq R$ is a prime ideal, then $S = R \setminus I$ is also a multiplicative submonoid. Then the localisation $S^{-1}R$ is also denoted R_I.

Observe that, for an integral domain R, $\{0\}$ is a prime ideal and the field of fractions $\mathrm{Frac}(R)$ is just the localisation $R_{\{0\}}$. Then it is a field, but in general R_I is not a field:

Example 2.75. If $R = \mathbb{Z}$ and $I = (p)$, then $\mathbb{Z}_{(p)}$ is the subring of $\mathbb{Q}$ of fractions whose denominator is not a multiple of p, i.e. $\{\frac{a}{b} : p \nmid b\} \subseteq \mathbb{Q}$.

Example 2.76. Let R be an integral domain. Then, as before, $(X) \subseteq R[X]$ is a prime ideal. $R[X]_{(X)}$ = the collection of fractions whose denominator has nonzero constant term.

It is worth emphasising that $\mathbb{Z}_{(p)}$ is not the same as $\mathbb{Z}[\frac{1}{p}] = \{\frac{a}{p^m} : m \geq 0\} \subseteq \mathbb{Q}$. In $\mathbb{Z}_{(p)}$ we invert infinitely many primes (everything except p), whereas in $\mathbb{Z}[\frac{1}{p}]$ we invert a single prime (only p). For $p \neq q$ primes, we have $\mathbb{Z} \leq \mathbb{Z}[\frac{1}{q}] \leq \mathbb{Z}_{(p)} \leq \mathbb{Q}$.

Proposition 2.77. Let R be a commutative ring and let $I \subseteq R$ be a prime ideal. Then R_I has a unique maximal ideal given by $\bar{I} = \{(r, s) : r \in I, s \in R \setminus I\}$.

Proof. First of all, if $a \notin \bar{I}$, then $a = (r, s)$ for $r \notin I$ and so is invertible in R_I. Hence the ideal generated by a in R_I is all of R_I. Therefore $\bar{I}$ contains every proper (= not all of R) ideal of R_I. Thus it is maximal and the unique maximal ideal. □

Definition 2.78. A *local ring* is a ring which has a unique maximal ideal.

Remark 2.79. If R is not an integral domain then $\{0\}$ is not a prime ideal, so one cannot take $R_{\{0\}}$ (i.e., $R \setminus \{0\}$ is not multiplicative as in Remark 2.65). One can instead take a different prime ideal (one could replace $\{0\}$ by a minimal prime ideal). For example, in $\mathbb{Z}/6\mathbb{Z}$ there are two minimal prime ideals, $3\mathbb{Z}/6\mathbb{Z}$ and $2\mathbb{Z}/6\mathbb{Z}$, and the corresponding localisations are $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$, respectively (cf. Example 2.63).

Let R be a commutative ring, $I \subseteq R$ an ideal, and $S \subseteq R$ a multiplicative submonoid. What are the ideals in $S^{-1}R$?

Definition 2.80. The set $S^{-1}I := \{\frac{i}{s} \mid s \in S, i \in I\}$ is an ideal in $S^{-1}R$. The ideal $S^{-1}I$ is called the *image* of I under the localisation.

Proposition 2.81. Every ideal $I \subseteq S^{-1}R$ is of the form $S^{-1}J$ for some ideal $J \subseteq R$.

Proof. For $r \in R$ and $s \in S$, note that $\frac{r}{s} \in I$ if and only if $\frac{r}{1} \in I$ (i.e. $\frac{1}{s} \in (S^{-1}R)^\times$). Let $J = \{r \in R : \frac{r}{1} \in I\}$. Then $J \subseteq R$ is an ideal and $S^{-1}J = I$ as ideals inside R. □

Remark 2.82. Note that $S^{-1}I$ is not the image of I under the map $R \to S^{-1}R$. Remember that the image of an ideal in this sense is not an ideal in general. In this case it won't be (unless $S \subseteq R^\times$, i.e., $R \to S^{-1}R$ is an isomorphism by an exercise on Problem Sheet 2).

Later, we will learn about modules over a ring R, or R-modules. An ideal in R is an example of an R-module. We can localise ideals and similarly we can localise R-modules.

3 Polynomial rings

3.1 Factorisation in polynomial rings and Gauss' lemma

We say previously that, if R is an integral domain, then $R[X]$ is an integral domain. Such polynomial rings form an important class of integral domains which enjoy many special properties. This section will be devoted to the study of these properties.

Recall that for F a field, we know $F[X]$ is a Euclidean domain, hence a principal ideal domain, hence a unique factorisation domain. Therefore we know

1. If $I \subseteq F[X]$, then $I = (f)$ for some $f \in F[X]$ and I is maximal if and only if I is prime.
2. If $f \in F[X]$, then f is irreducible if and only if f is prime.

In particular, the following four things are equivalent: f is irreducible, f is prime, $F[X]/(f)$ is a field and $F[X]/(f)$ is an integral domain. The most interesting point to take away from all this, and the one which we shall use the most, is that f irreducible implies that $F[X]/(f)$ is a field. We can use this to construct many interesting fields.

So we want to understand reducibility, i.e. we want to know whether we can factorize a polynomial f. Firstly, we want to get rid of the trivial case where we just factor out a scalar, e.g. $2X^2 + 2 = 2(X^2 + 1) \in \mathbb{Z}[X]$.

Definition 3.1. Let R be a UFD and $f = a_0 + a_1X + \cdots + a_nX^n \in R[X]$. The *content* $c(f)$ of f is

$$c(f) = \gcd(a_0, a_1, \cdots, a_n) \in R.$$

Again, since the gcd is only defined up to a unit, so is the content. We can equivalently define $c(f)$ as the ideal $(\gcd(a_0, \cdots, a_n))$.

Definition 3.2. A polynomial is *primitive* if $c(f)$ is a unit, i.e. the a_i are coprime. In other words, viewing $c(f)$ as an ideal, we have $c(f) = R[X]$.

We have the following basic properties.

Lemma 3.3. Let R be a UFD. If $f \in R[X]$, then $f = c(f) \cdot f_1$ for some $f_1 \in R[X]$ primitive.

Proof. Let $f = a_0 + a_1X + \cdots + a_nX^n$. Let $d = c(f) = \gcd(a_0, \cdots, a_n)$. By the definition of greatest common divisor we have that, for all i, $a_i = b_id$ for some $b_i \in R$ such that $\gcd(b_0, \cdots, b_n) = 1$. Then $f = d \cdot f_1$ for $f_1 = b_0 + b_1X + \cdots + b_nX^n$ with f_1 primitive. □

Lemma 3.4. Let R be a UFD. If $f, g \in R[X]$ are primitive, then so is fg.

Proof. We let

$$f = a_0 + a_1X + \cdots + a_nX^n,$$
$$g = b_0 + b_1X + \cdots + b_mX^m,$$

where $a_n, b_m \neq 0$, and f, g are primitive. We want to show that the content of fg is a unit.

Now suppose fg is not primitive. Then $c(fg)$ is not a unit. Since R is a UFD, we can find an irreducible p which divides $c(fg)$.

By assumption, $c(f)$ and $c(g)$ are units. So $p \nmid c(f)$ and $p \nmid c(g)$. So suppose $p \mid a_0$, $p \mid a_1$, $\ldots$, $p \mid a_{k-1}$ but $p \nmid a_k$. Note it is possible that $k = 0$. Similarly, suppose $p \mid b_0, p \mid b_1, \cdots, p \mid b_{\ell-1}, p \nmid b_\ell$.

We look at the coefficient of $X^{k+\ell}$ in fg. It is given by

$$\sum_{i+j=k+\ell} a_i b_j = a_{k+\ell}b_0 + \cdots + a_{k+1}b_{\ell-1} + a_k b_\ell + a_{k-1}b_{\ell+1} + \cdots + a_0 b_{\ell+k}.$$

By assumption, this is divisible by p. So

$$p \mid \sum_{i+j=k+\ell} a_i b_j.$$

However, the terms $a_{k+\ell}b_0 + \cdots + a_{k+1}b_{\ell-1}$, is divisible by p, as $p \mid b_j$ for $j < \ell$. Similarly, $a_{k-1}b_{\ell+1} + \cdots + a_0 b_{\ell+k}$ is divisible by p. So we must have $p \mid a_k b_\ell$. As p is irreducible, and hence prime, we must have $p \mid a_k$ or $p \mid b_\ell$. This is a contradiction. So $c(fg)$ must be a unit. □

Corollary 3.5. Let R be a UFD. Then for $f, g \in R[X]$, we have that $c(fg)$ is an associate of $c(f)c(g)$.

Again, we cannot say they are equal, since content is only well-defined up to a unit.

Proof. We can write $f = c(f)f_1$ and $g = c(g)g_1$, with f_1 and g_1 primitive. Then

$$fg = c(f)c(g)f_1 g_1.$$

Since $f_1 g_1$ is primitive, so $c(f)c(g)$ is a gcd of the coefficients of fg, and so is $c(fg)$, by definition. So they are associates. □

We now want to prove the following important lemma:

Lemma 3.6 (Gauss' lemma)**.** Let R be a UFD, and $f \in R[X]$ be a primitive polynomial. Then f is irreducible in $R[X]$ if and only if f is irreducible $F[X]$, where $F = \mathrm{Frac}(R)$ is the field of fractions of R.

One direction is straightforward: if f is reducible in $R[X]$, then clearly f is reducible in $F[X]$. To see this, let $f = gh$ be a product in $R[X]$ with g, h not units. As f is primitive, so are g and h. So both have degree > 0. So g, h are not units in $F[X]$. So f is reducible in $F[X]$.

The converse is more difficult: if f is irreducible in $R[X]$, why must it be irreducible in $F[X]$? This is useful since characterising irreducibles in $F[X]$ is a priori harder than characterising irreducibles in $R[X]$. We can see this through the following example.

Example 3.7. Consider $X^3 + X + 1 \in \mathbb{Z}[X]$. This has content 1 so is primitive.

Suppose f is reducible in $\mathbb{Q}[X]$. Then by Gauss' lemma, this is reducible in $\mathbb{Z}[X]$. So we can write

$$X^3 + X + 1 = gh,$$

for some polynomials $g, h \in \mathbb{Z}[X]$, with g, h not units. But if g and h are not units, then they cannot be constant, since the coefficients of $X^3 + X + 1$ are all 1 or 0. So they have degree at least 1. Since the degrees add up to 3, we wlog suppose g has degree 1 and h has degree 2. So let

$$g = b_0 + b_1 X, \quad h = c_0 + c_1 X + c_2 X^2.$$

Multiplying out and equating coefficients, we get

$$b_0 c_0 = 1$$
$$c_2 b_1 = 1$$

Since we are working over $\mathbb{Z}$, b_0 and b_1 must be ± 1. This implies that g is either $1+X, 1-X, -1+X$ or $-1-X$, and hence has ± 1 as a root. But this is a contradiction, since ± 1 is not a root of X^3+X+1. Hence f is irreducible in $\mathbb{Q}[X]$. In particular, f has no root in $\mathbb{Q}$.

From this, we can see the utility of using Gauss' lemma: if we worked in $\mathbb{Q}$ instead, we could have gotten to the step $b_0 c_0 = 1$, and then we can do nothing, since there are many solutions for b_0 and c_0.

Proof of Gauss' lemma. We will show that a primitive $f \in R[X]$ is reducible in $R[X]$ if and only if f is reducible in $F[X]$.

One direction is straightforward. Let $f = gh$ be a product in $R[X]$ with g, h not units. As f is primitive, so are g and h. So both have degree > 0. So g, h are not units in $F[X]$. So f is reducible in $F[X]$.

We will now prove the converse. Let $f = gh$ in $F[X]$, with g, h not units. So g and h have degree > 0, since F is a field. So we can clear denominators by finding $a, b \in R \setminus \{0\}$ such that $(ag), (bh) \in R[X]$ (e.g. let a be the product of denominators of coefficients of g). Then we get

$$abf = g'h',$$

where $g' = ag$, $h' = bh$ and this is a factorisation in $R[X]$. Note that $g' = ag$ is not a factorisation in $R[X]$ since we only know $g \in F[X]$. Now write

$$\begin{aligned} g' &= c(g')g_1, \\ h' &= c(h')h_1, \end{aligned}$$

where g_1, h_1 are primitive. Since f is primitive, we have

$$ab = c(abf) = c(g'h') = u \cdot c(g')c(h'),$$

where $u \in R$ is a unit, by the previous corollary. But also we have

$$abf = g'h' = c(g')c(h')g_1h_1 = u^{-1}abg_1h_1.$$

Since R is an integral domain $ab \neq 0$ and, since $R[X]$ is an integral domain, we can therefore cancel ab to get:

$$f = u^{-1}g_1h_1 \in R[X].$$

Hence f is reducible in $R[X]$. □

This might seem like a difficult proof. A useful exercise is to trace through the argument explicitly in the case where $R = \mathbb{Z}$ and $F = \mathbb{Q}$.

From this we can get ourselves a large class of UFDs.

Theorem 3.8 (Polynomial rings over UFDs)**.** If R is a UFD, then $R[X]$ is a UFD.

In particular, if R is a UFD, then $R[X_1, \cdots, X_n]$ is also a UFD.

Proof. We know $R[X]$ has a notion of degree. So we will combine this with the fact that R is a UFD.

Existence: Let $f \in R[X]$. We can write $f = c(f)f_1$, with f_1 primitive. Firstly, as R is a UFD, we may factor

$$c(f) = p_1 p_2 \cdots p_n,$$

for $p_i \in R$ irreducible (and also irreducible in $R[X]$). Now we want to deal with f_1.

Assume for contradiction that f_1 is not the product of irreducibles. Then f_1 is not irreducible, so we can write

$$f_1 = f_2 g_2,$$

with f_2, g_2 both not units. Since f_1 is primitive, f_2, g_2 also cannot be constants. So we must have $\deg f_2, \deg g_2 > 0$. Also, since $\deg f_2 + \deg g_2 = \deg f_1$, we must have $\deg f_2, \deg g_2 < \deg f_1$. If f_2, f_3 are both products of irreducibles, we have a contradiction since then f_1 would be. Wlog we may assume that f_1 is not the product of irreducibles. Continuing like this gives a sequence $f_1, f_2, \cdots$ where $\deg(f_1) > \deg(f_2) > \cdots$. This is a contradiction since the f_i are non-constant and so $\deg(f_i) > 0$ (but, for example, monotonic bounded sequences converge). Hence

$$f_1 = q_1 \cdots q_m,$$

with q_i irreducible. So we can write

$$f = p_1 p_2 \cdots p_n q_1 q_2 \cdots q_m,$$

a product of irreducibles where $p_i \in R \subseteq R[X]$ are constant polynomials and the q_i are non-constant, i.e. $\deg(q_i) > 0$.

Uniqueness: We will first deal with the p_is. Note that

$$c(f) = p_1 p_2 \cdots p_n$$

is a unique factorisation of the content, up to reordering and associates, as R is a UFD. So cancelling the content, we only have to show that primitives can be factored uniquely.

Suppose we have two factorisations

$$f_1 = q_1 q_2 \cdots q_m = r_1 r_2 \cdots r_\ell.$$

Note that each q_i and each r_i is a factor of the primitive polynomial f_1, so are also primitive. This follows from that fact that $c(q_1 \cdots q_m)$ is an associate of $c(q_1) \cdots c(q_m)$ by one of the Corollary's above, and similarly for the r_i.

Now let F be the field of fractions of R, and consider $q_i, r_i \in F[X]$. Since F is a field, $F[X]$ is a Euclidean domain, hence principal ideal domain, hence unique factorisation domain.

By Gauss' lemma, since the q_i and r_i are irreducible in $R[X]$, they are also irreducible in $F[X]$. As $F[X]$ is a UFD, we find that $\ell = m$, and after reordering, r_i and q_i are associates, say

$$r_i = u_i q_i,$$

with $u_i \in F[X]$ a unit. What we want to say is that r_i is a unit times q_i in $R[X]$. Firstly, note that $u_i \in F$ as it is a unit. Clearing denominators, we can write

$$a_i r_i = b_i q_i \in R[X].$$

Taking contents, since r_i, q_i are primitives, we know a_i and b_i are associates, say

$$b_i = v_i a_i,$$

with $v_i \in R$ a unit. Cancelling a_i on both sides, we know $r_i = v_i q_i$ as required. □

The key idea is to use Gauss' lemma to say the reducibility in $R[X]$ is the same as reducibility in $F[X]$, as long as we are primitive. The first part about contents is just to turn everything into primitives.

Example 3.9. We know $\mathbb{Z}[X]$ is a UFD, and if R is a UFD, then $R[X_1, \cdots, X_n]$ is also a UFD.

This gives us examples of UFDs that are not PIDs, thus completing our classification of properties laid out in the previous section.

In such rings, we would also like to have an easy to determine whether something is reducible. Fortunately, we have the following criterion:

Proposition 3.10 (Eisenstein's criterion)**.** Let R be a UFD, and let

$$f = a_0 + a_1 X + \cdots + a_n X^n \in R[X]$$

be primitive with $a_n \neq 0$. Let $p \in R$ be irreducible (hence prime) be such that

1. $p \nmid a_n$;
2. $p \mid a_i$ for all $0 \leq i < n$;
3. $p^2 \nmid a_0$.

Then f is irreducible in $R[X]$, and hence in $F[X]$ where $F = \mathrm{Frac}(R)$.

Proof. See Problem Sheet 2. □

Example 3.11. Consider the polynomial $X^n - p \in \mathbb{Z}[X]$ for p a prime. Apply Eisenstein's criterion with p, and observe all the conditions hold. This is certainly primitive, since this is monic. So $X^n - p$ is irreducible in $\mathbb{Z}[X]$, hence in $\mathbb{Q}[X]$. In particular, $X^n - p$ has no rational roots, i.e. $\sqrt[n]{p}$ is irrational (for $n > 1$).

Example 3.12. Consider a polynomial

$$f = X^{p-1} + X^{p-2} + \cdots + X^2 + X + 1 \in \mathbb{Z}[X],$$

where p is a prime number. If we look at this, we notice Eisenstein's criteria does not apply. What should we do? We observe that

$$f = \frac{X^p - 1}{X - 1}.$$

So it might be a good idea to let $Y = X - 1$. Then we get a new polynomial

$$\hat{f} = \hat{f}(Y) = \frac{(Y+1)^p - 1}{Y} = Y^{p-1} + \binom{p}{1} Y^{p-2} + \binom{p}{2} Y^{p-3} + \cdots + \binom{p}{p-1}.$$

When we look at it hard enough, we notice Eisenstein's criteria can be applied — we know $p \mid \binom{p}{i}$ for $1 \leq i \leq p-1$, but $p^2 \nmid \binom{p}{p-1} = p$. So $\hat{f}$ is irreducible in $\mathbb{Z}[Y]$.

Now if we had a factorisation

$$f(X) = g(X)h(X) \in \mathbb{Z}[X],$$

then we get

$$\hat{f}(Y) = g(Y+1)h(Y+1)$$

in $\mathbb{Z}[Y]$. So f is irreducible.

Note that this implies that none of the roots of f are rational (but we knew that already as the roots ζ_p^i are not even real).

3.2 Algebraic integers

Previously we defined rings like $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-5}]$ as the explicit subrings $\{a + bi : a \in \mathbb{Z}\}$ and $\{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$ of $\mathbb{C}$. How should we define $\mathbb{Z}[\zeta_n]$ for $\zeta_n = e^{2\pi i/n} \in \mathbb{C}$ a root of unity? How should we define $\mathbb{Z}[2i]$? The aim of this section will be to systematise this notion through the concept of algebraic integers.

Definition 3.13. An $\alpha \in \mathbb{C}$ is called an algebraic integer if it is a root of a monic polynomial in $\mathbb{Z}[X]$, i.e. there is a monic $f \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$.

We can immediately check that this is a sensible definition: not all complex numbers are algebraic integers, since there are only countably many polynomials with integer coefficients, hence only countably many algebraic integers, but there are uncountably many complex numbers.

Definition 3.14. For α an algebraic integer, we write $\mathbb{Z}[\alpha] \leq \mathbb{C}$ for the smallest subring containing α.

We can also construct $\mathbb{Z}[\alpha]$ by taking it as the image of the map $\phi : \mathbb{Z}[X] \to \mathbb{C}$ given by $g \mapsto g(\alpha)$. In particular, ϕ induces an isomorphism

$$\mathbb{Z}[X]/I \cong \mathbb{Z}[\alpha], \quad I = \ker \phi.$$

Note that I is non-empty since there exists a (monic) $f \in I$ since α is an algebraic integer.

Proposition 3.15. Let $\alpha \in \mathbb{C}$ be an algebraic integer and let $\phi : \mathbb{Z}[X] \to \mathbb{C}$ be the ring homomorphism given by $f \mapsto f(\alpha)$. Then the ideal

$$I = \ker(\phi)$$

is principal, and equal to (f_α) for some irreducible monic f_α.

Definition 3.16. Let $\alpha \in \mathbb{C}$ be an algebraic integer. Then the *minimal polynomial* is a polynomial f_α is the irreducible monic such that $I = \ker(\phi) = (f_\alpha)$.

Note that defining the minimal polynomial over $\mathbb{Q}[X]$ is straightforward since $\mathbb{Q}[X]$ is a principal ideal domain. Since $\mathbb{Z}[X]$ is not a principal ideal domain, there is no immediate guarantee that I is generated by one polynomial. What we are proving here is therefore much more powerful than the merely the existence of a minimal polynomial over $\mathbb{Q}[X]$.

Proof of Proposition 3.15. By definition, there is a monic $f \in \mathbb{Z}[X]$ such that $f(a) = 0$. So $f \in I$. So $I \neq 0$. Now let $f_\alpha \in I$ be such a polynomial of minimal degree. We may suppose that f_α is primitive. We want to show that $I = (f_\alpha)$, and that f_α is irreducible.

Let $h \in I$. We pretend we are living in $\mathbb{Q}[X]$. Then we have the Euclidean algorithm. So we can write

$$h = f_\alpha q + r,$$

with $r = 0$ or $\deg r < \deg f_\alpha$. This was done over $\mathbb{Q}[X]$, not $\mathbb{Z}[X]$. We now clear denominators. We multiply by some $a \in \mathbb{Z}$ to get

$$ah = f_\alpha(aq) + (ar),$$

where now $(aq), (ar) \in \mathbb{Z}[X]$. We now evaluate these polynomials at α. Then we have

$$ah(\alpha) = f_\alpha(\alpha)aq(\alpha) + ar(\alpha).$$

We know $f_\alpha(\alpha) = h(\alpha) = 0$, since f_α and h are both in I. So $ar(\alpha) = 0$. So $(ar) \in I$. As $f_\alpha \in I$ has minimal degree, we cannot have $\deg(r) = \deg(ar) < \deg(f_a)$. So we must have $r = 0$.

Hence we know

$$ah = f_\alpha \cdot (aq)$$

is a factorization in $\mathbb{Z}[X]$. This is almost right, but we want to factor h, not ah. Again, taking contents of everything, we get

$$ac(h) = c(ah) = c(f_\alpha(aq)) = c(aq),$$

as f_α is primitive. In particular, $a \mid c(aq)$. This, by definition of content, means (aq) can be written as $a\bar{q}$, where $\bar{q} \in \mathbb{Z}[X]$. Cancelling, we get $q = \bar{q} \in \mathbb{Z}[X]$. So we know

$$h = f_\alpha q \in (f_\alpha).$$

So we know $I = (f_\alpha)$.

To show f_α is irreducible, note that

$$\frac{\mathbb{Z}[X]}{(f_\alpha)} \cong \frac{\mathbb{Z}[X]}{\ker \phi} \cong \mathrm{im}(\phi) = \mathbb{Z}[\alpha] \leq \mathbb{C}.$$

Since $\mathbb{C}$ is an integral domain, so is $\mathrm{im}(\phi)$. So we know $\mathbb{Z}[X]/(f_\alpha)$ is an integral domain. So (f_α) is prime. So f_α is prime, hence irreducible.

If this final line looks magical, we can unravel this proof as follows: suppose $f_\alpha = pq$ for some non-units pq. Then since $f_\alpha(\alpha) = 0$, we know $p(\alpha)q(\alpha) = 0$. Since $p(\alpha), q(\alpha) \in \mathbb{C}$, which is an integral domain, we must have, say, $p(\alpha) = 0$. But then $\deg p < \deg f_\alpha$, so $p \notin I = (f_\alpha)$. Contradiction. □

Example 3.17.

1. We know $\alpha = i$ is an algebraic integer with $f_\alpha = X^2 + 1$.
2. Also, $\alpha = \sqrt{2}$ is an algebraic integer with $f_\alpha = X^2 - 2$.
3. More interestingly, $\alpha = \frac{1}{2}(1 + \sqrt{-3})$ is an algebraic integer with $f_\alpha = X^2 - X - 1$.
4. The polynomial $X^5 - X + d \in \mathbb{Z}[X]$ with $d \in \mathbb{Z}_{\geq 0}$ has precisely one real root α, which is an algebraic integer. It is a theorem from Galois Theory that this α cannot be constructed from integers via $+, -, \times, \div, \sqrt[n]{\cdot}$. It is also a theorem that degree 5 polynomials are the smallest degree for which this can happen (the prove involves writing down formulas analogous to the quadratic formula for degree 3 and 4 polynomials).

Lemma 3.18. Let $\alpha \in \mathbb{Q}$ be an algebraic integer. Then $\alpha \in \mathbb{Z}$.

Proof. Let $f_\alpha \in \mathbb{Z}[X]$ be the minimal polynomial, which is irreducible. In $\mathbb{Q}[X]$, the polynomial $X - \alpha$ must divide f_α. However, by Gauss' lemma, we know $f \in \mathbb{Q}[X]$ is irreducible. So we must have $f_\alpha = X - \alpha \in \mathbb{Z}[X]$. So α is an integer. □

Remark 3.19. It turns out the collection of all algebraic integers form a subring of $\mathbb{C}$. This is not at all obvious: given $f, g \in \mathbb{Z}[X]$ monic such that $f(\alpha) = g(\alpha) = 0$, there is no easy way to find a new monic h such that $h(\alpha + \beta) = 0$.

3.3 Noetherian rings and Hilbert's basis theorem

We now revisit the idea of Noetherian rings, something we have briefly mentioned when proving that PIDs are UFDs.

Definition 3.20. A commutative ring is *Noetherian* if for any chain of ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots,$$

there is some N such that $I_N = I_{N+1} = I_{N+2} = \cdots$.

This condition is known as the *ascending chain condition*.

Example 3.21. Every finite ring is Noetherian. This is since there are only finitely many possible ideals.

Example 3.22. Every field is Noetherian. This is since there are only two possible ideals.

Example 3.23. The ring $\mathbb{Z}[X_1, X_2, X_3, \cdots]$ is not Noetherian. This has the chain of strictly increasing ideals

$$(X_1) \subseteq (X_1, X_2) \subseteq (X_1, X_2, X_3) \subseteq \cdots.$$

We previously showed that principal ideal domains are Noetherian. This can be generalised as follows.

Definition 3.24. An ideal I is *finitely generated* if it can be written as $I = (r_1, \cdots, r_n)$ for some $r_1, \cdots, r_n \in R$.

Proposition 3.25. A commutative ring is Noetherian if and only if every ideal is finitely generated.

Every PID trivially satisfies this condition. So we know every PID is Noetherian.

Proof. See Problem Sheet 3. □

When we have developed some properties or notions, a natural thing to ask is whether it passes on to subrings and quotients.

If R is Noetherian, does every subring of R have to be Noetherian? The answer is no. For example, since $\mathbb{Z}[X_1, X_2, \cdots]$ is an integral domain, we can take its field of fractions, which is a field, hence Noetherian, but $\mathbb{Z}[X_1, X_2, \cdots]$ is a subring of its field of fractions. However, the property of being Noetherian is closed under quotients.

Proposition 3.26. Let R be a Noetherian ring and $I \subseteq R$ an ideal. Then R/I is Noetherian.

Proof. Consider the quotient map

$$\begin{aligned} \pi : R &\to R/I \\ x &\mapsto x + I. \end{aligned}$$

We can prove this result by finitely generated or ascending chain condition. We go for the former. Let $J \subseteq R/I$ be an ideal. We want to show that J is finitely generated. Consider the inverse image $\pi^{-1}(J)$. This is an ideal of R, and is hence finitely generated, since R is Noetherian. So $\pi^{-1}(J) = (r_1, \cdots, r_n)$ for some $r_1, \cdots, r_n \in R$. Then J is generated by $\pi(r_1), \cdots, \pi(r_n)$. □

This gives us many examples of Noetherian rings. But there is one important case we have not tackled yet: polynomial rings. We know $\mathbb{Z}[X]$ is not a PID, since $(2, X)$ is not principal. However, $(2, X)$ is at least finitely generated. We might try to construct some non-finitely generated ideal, but we are bound to fail since $\mathbb{Z}[X]$ is Noetherian by the following theorem.

Theorem 3.27 (Hilbert basis theorem)**.** Let R be a Noetherian ring. Then so is $R[X]$.

Proof. In the proof, we will use both the ascending chain condition and the fact that all ideals are finitely generated.

Let $I \subseteq R[X]$ be an ideal. We want to show it is finitely generated. Since we know R is Noetherian, we want to generate some ideals of R from I. How can we do this? One idea is to take all constants of I, i.e. $I \cap R$. But we can do better. We can consider all linear polynomials, and take their leading coefficients. Thinking for a while, this is indeed an ideal.

In general, for $n = 0, 1, 2, \cdots$, we let

$$I_n = \{r \in R : \text{there is some } f \in I \text{ such that } f = rX^n + \cdots\} \cup \{0\}.$$

Then it is easy to see, using the strong closure property, that each ideal I_n is an ideal of R. Moreover, they form a chain, since if $f \in I$, then $Xf \in I$, by strong closure. So $I_n \subseteq I_{n+1}$ for all n.

By the ascending chain condition of R, we know there is some N such that $I_N = I_{N+1} = \cdots$. Now for each $0 \leq n \leq N$, since R is Noetherian, we can write

$$I_n = (r_1^{(n)}, r_2^{(n)}, \cdots, r_{k(n)}^{(n)}).$$

Now for each $r_i^{(n)}$, we choose some $f_i^{(n)} \in I$ with $f_i^{(n)} = r_i^{(n)} X^n + \cdots$.

We now claim the polynomials $f_i^{(n)}$ for $0 \leq n \leq N$ and $1 \leq i \leq k(n)$ generate I.

Suppose not. We pick $g \in I$ of minimal degree not generated by the $f_i^{(n)}$.

There are two possible cases. If $\deg g = n \leq N$, suppose

$$g = rX^n + \cdots.$$

We know $r \in I_n$. So we can write

$$r = \sum_i \lambda_i r_i^{(n)}$$

for some $\lambda_i \in R$, since that's what generating an ideal means. Then we know

$$\sum_i \lambda_i f_i^{(n)} = rX^n + \cdots \in I.$$

But if g is not in the span of the $f_i^{(j)}$, then $g - \sum_i \lambda_i f_i^{(n)}$ isn't either. But this has a lower degree than g. This is a contradiction.

Now suppose $\deg g = n > N$. Since $I_n = I_N$, the same proof works. We write

$$g = rX^n + \cdots.$$

But we know $r \in I_n = I_N$. So we know

$$r = \sum_I \lambda_i r_i^{(N)}.$$

Then we know

$$X^{n-N} \sum_i \lambda_i f_i^{(N)} = rX^n + \cdots \in I.$$

Hence $g - X^{n-N} \sum \lambda_i f_i^{(N)}$ has smaller degree than g, but is not in the span of $f_i^{(j)}$. □

As an aside, let F be a field and let $\mathcal{E} \subseteq F[X_1, \cdots, X_n]$ be any set of polynomials. We view this as a set of equations $f = 0$ for each $f \in \mathcal{E}$. The claim is that to solve the potentially infinite set of equations $\mathcal{E}$, we actually only have to solve finitely many equations.

Consider the ideal $(\mathcal{E}) \subseteq F[X_1, \cdots, X_n]$. By the Hilbert basis theorem, there is a finite list $f_1, \cdots, f_k$ such that

$$(f_1, \cdots, f_k) = (\mathcal{E}).$$

We want to show that we only have to solve $f_i(x) = 0$ for these f_i. Given $(\alpha_1, \cdots, \alpha_n) \in F^n$, consider the homomorphism

$$\begin{aligned} \phi_\alpha : F[X_1, \cdots, X_n] &\to F \\ X_i &\mapsto \alpha_i. \end{aligned}$$

Then we know $(\alpha_1, \cdots, \alpha_n) \in F^n$ is a solution to the equations $\mathcal{E}$ if and only if $(\mathcal{E}) \subseteq \ker(\varphi_\alpha)$. By our choice of f_i, this is true if and only if $(f_1, \cdots, f_k) \subseteq \ker(\varphi_\alpha)$. By inspection, this is true if and only if $(\alpha_1, \cdots, \alpha_n)$ is a solution to all of $f_1, \cdots, f_k$. So solving $\mathcal{E}$ is the same as solving $f_1, \cdots, f_k$. This is useful in, say, algebraic geometry.

4 Modules

Finally, we are going to look at modules. Recall that to define a vector space, we first pick some base field F. We then defined a vector space to be an abelian group V with an action of F on V (i.e. scalar multiplication) that is compatible with the multiplicative and additive structure of F.

In the definition, we did not at all mention division in F. So in fact we can make the same definition, but allow F to be a ring instead of a field. We call these *modules*. Unfortunately, most results we prove about vector spaces do use the fact that F is a field. So many linear algebra results do not apply to modules, and modules have much richer structures.

4.1 Basic definitions and examples

For the purposes of this section, R will be an arbitrary (possibly non-commutative) ring.

Definition 4.1. Let R be a ring. An *R-module* is a set M together with function $+ : M \times M \to M$ and $\cdot : R \times M \to M$, and a given element $0_M \in M$, such that the following holds:

- $(M, +)$ is an abelian group with identity 0_M
- The operation $\cdot : R \times M \to M$ satisfies:
  1. $(r_1 + r_2) \cdot m = (r_1 \cdot m) + (r_2 \cdot m)$
  2. $r \cdot (m_1 + m_2) = (r \cdot m_1) + (r \cdot m_2)$
  3. $r_1 \cdot (r_2 \cdot m) = (r_1 \cdot r_2) \cdot m$, and
  4. $1_R \cdot m = m$.

Remark 4.2. Strictly, this is the definition of a *left* R-module but we will call them R-modules in this course. This is similar to how we took ideal to mean left ideals. We can define a right R-module to be an abelian group M with an operation $\cdot : M \times R \to R$ satisfying analogous conditions.

Remark 4.3. There are two different additions going on: addition in the ring and addition in the module, and similarly two notions of multiplication. However, it is easy to distinguish them since they operate on different things. If needed, we can make them explicit by writing, say, $+_R$ and $+_M$.

This is slightly cumbersome and so we will restate it in slightly more compact notation. For an abelian group A, recall that $\mathrm{End}(A) = \{f : A \to A \mid f \text{ is a group homomorphism}\}$ is the endomorphism ring of A with operations $(f + g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(g(x))$.

Definition 4.4. Let R be a ring. An *R-module* is an abelian group M equipped with a ring homomorphism $\varphi : R \to \mathrm{End}(M)$. Given $r \in R$ and $m \in M$, we write $r \cdot m$ to denote $\varphi(r)(m)$.

Proposition 4.5. Show that these two definitions are equivalent.

Proof. See Problem Sheet 3. □

The idea is that, if $\varphi : R \to \mathrm{End}(M)$ is a ring homomorphism, then this corresponds to an R-action via: $r \cdot m := \varphi(r)(m)$ for $r \in R$ and $m \in M$.

We can imagine modules as rings acting on abelian groups, just as groups can act on sets. Hence we might say "R acts on M" to mean M is an R-module.

Example 4.6. Let F be a field. An F-module is precisely the same as a vector space over F (the axioms are the same). In particular, F^n is an F-module for any $n \geq 1$. Here F^n denotes the abelian group $(F,+)\times\cdots\times(F,+)$ with the F-action defined by $r\cdot(r_1,\cdots,r_n) := (rr_1,\cdots,rr_n)$.

An example which better demonstrates what modules actually are, and which you should bear in mind when thinking of modules, is an ideal.

Example 4.7. Let $I \subseteq R$ be an ideal. Then I is an R-module (i.e. the underlying abelian group $(I,+)$ is an R-module) via

$$r \cdot a := r \cdot_R a.$$

Example 4.8. Let R be a ring. Then R is an R-module (i.e. the underlying abelian group $(R,+)$ is an R-module) with R-action given by multiplication in R. Explicitly, we mean the R-module (M,ψ) where $M := (R,+)$ denotes the abelian group underlying the ring R and $\psi : R \to \mathrm{End}(M)$, $r \mapsto (m \mapsto r \cdot_R m)$ for all $r \in R$ and $m \in M = R$.

More generally, R^n is an R-module via

$$r \cdot (r_1,\cdots,r_n) = (rr_1,\cdots,rr_n).$$

This works in the same way as the example F^n given above.

Example 4.9. If $I \subseteq R$ is a two-sided ideal, then R/I is an R-module via

$$r \cdot (a+I) := (r \cdot_R a) + I.$$

In the above examples, we had different objects each with their own algebraic structures: rings and ideals. When we take some like a ring or an ideal and say that "it" is an R-module then we are really referring to its underlying abelian group.

Example 4.10. A $\mathbb{Z}$-module is precisely the same as an abelian group. For A an abelian group, we have

$$\begin{aligned} \mathbb{Z}\times A &\to A \\ (n,a) &\mapsto \underbrace{a+\cdots+a}_{n \text{ times}}, \end{aligned}$$

where we adopt the notation

$$\underbrace{a+\cdots+a}_{-n \text{ times}} = \underbrace{(-a)+\cdots+(-a)}_{n \text{ times}},$$

and adding something to itself 0 times is just 0. This definition is essentially forced upon us since, by the axioms of a module, we must have $(1,a) \mapsto a$. Then we must send, say, $(2,a) = (1+1,a) \mapsto a+a$.

An alternate perspective, using the second definition, is to note that there is only one choice the action $\mathbb{Z} \to \mathrm{End}(M)$. This is because, as we saw earlier in the course, there is a unique ring homomorphism to any ring from $\mathbb{Z}$ (which determines the characteristic of the ring).

4.2 Constructions of modules

Definition 4.11. Let $M_1, M_2, \cdots, M_k$ be R-modules. The *direct sum* is the R-module

$$M_1 \oplus M_2 \oplus \cdots \oplus M_k,$$

which is the set $M_1 \times M_2 \times \cdots \times M_k$, with addition given by

$$(m_1, \cdots, m_k) + (m_1', \cdots, m_k') = (m_1 + m_1', \cdots, m_k + m_k'),$$

and the R-action given by

$$r \cdot (m_1, \cdots, m_k) = (rm_1, \cdots, rm_k).$$

We've been using one example of the direct sum already, namely

$$R^n = \underbrace{R \oplus R \oplus \cdots \oplus R}_{n \text{ times}}.$$

Definition 4.12. Let M be an R-module. A subset $N \subseteq M$ is an *R-submodule* if it is a subgroup of $(M, +, 0_M)$, and if $n \in N$ and $r \in R$, then $rn \in N$. We write $N \leq M$.

Example 4.13. We know R itself is an R-module. Then a subset of R is a submodule if and only if it is an ideal.

Example 4.14. A subset of an F-module V, where F is a field, is an F-submodule if and only if it is a vector subspace of V.

Definition 4.15. Let $N \leq M$ be an R-submodule. The *quotient module* M/N is the set of N-cosets in $(M, +, 0_M)$, with the R-action given by

$$r \cdot (m + N) = (r \cdot m) + N.$$

It is easy to check this is well-defined and is indeed a module.

Remark 4.16. Note that modules are different from rings and groups. In groups, we had subgroups, and we have some really nice ones called normal subgroups. We are only allowed to quotient by normal subgroups. In rings, we have subrings and ideals, which are unrelated objects, and we only quotient by ideals. In modules, we only have submodules, and we can quotient by arbitrary submodules.

Definition 4.17. A function $f : M \to N$ between R-modules is an *R-module homomorphism* if it is a homomorphism of abelian groups, and satisfies

$$f(r \cdot m) = r \cdot f(m)$$

for all $r \in R$ and $m \in M$.

An *isomorphism* is a bijective homomorphism, and two R-modules are *isomorphic* if there is an isomorphism between them.

Note that this gives a meaningful equivalence relation on the class of ideals beyond just two ideals being equal as subsets of the ring.

Example 4.18. If F is a field and V, W are F-modules (i.e. vector spaces over F), then an F-module homomorphism is precisely an F-linear map.

These operations all take R-modules to R-modules. However, various constructions of new rings from old also yield corresponding constructions on the modules over those rings.

Definition 4.19. If R_1 and R_2 are rings, M_1 is an R_1-module and M_2 is an R_2-module, then $M_1 \times M_2$ is an $(R_1 \times R_2)$-module with action $(r_1, r_2) \cdot (m_1, m_2) := (r_1 m_1, r_2 m_s)$. This will be verified on Problem Sheet 3.

Definition 4.20. Let R be a commutative ring, let $S \subseteq R$ be a multiplicative submonoid and let M be an R-module. The *localisation of M by S*, denoted $S^{-1}M$, is the set of equivalence classes of pairs (m, s) for $m \in M$ and $s \in S$ where $(m, s) \sim (m', s')$ if there exists $t \in S$ such that $r(ms' - m's) = 0$. This is an $S^{-1}R$-module with the natural structure of an abelian group and with $S^{-1}R$-action given by $(r, t) \cdot (m, s) := (rm, ts)$ for $(r, t) \in S^{-1}R$ and $(m, s) \in S^{-1}M$. As usual, we often write $\frac{m}{s}$ to denote (m, s).

It is straightforward to see that, given an ideal $I \subseteq R$, the localisation $S^{-1}I \subseteq S^{-1}R$ as an ideal is isomorphism as an $S^{-1}R$-module to the localisation of I as a module.

4.3 Basic theory of modules

As for groups and rings, we also have three isomorphism theorems. The proofs are similar to the cases of groups and rings and so will be omitted for brevity.

Theorem 4.21 (First isomorphism theorem)**.** Let $f : M \to N$ be an R-module homomorphism. Then

$$\ker(f) = \{m \in M : f(m) = 0\} \leq M$$

is an R-submodule of M. Similarly,

$$\operatorname{im}(f) = \{f(m) : m \in M\} \leq N$$

is an R-submodule of N. Then

$$\frac{M}{\ker(f)} \cong \operatorname{im}(f).$$

Note that, unlike the situation for rings, the fact that $\operatorname{im} f \leq N$ is a submodule means that one further module $N/\operatorname{im} f$ arises for an R-module homomorphism $f : M \to N$. This leads to a new concept (which also exists for abelian groups):

Definition 4.22. Let $f : M \to N$ be an R-module homomorphism. The *cokernel* of f, is

$$\operatorname{coker}(f) = N/\operatorname{im}(f).$$

Theorem 4.23 (Second isomorphism theorem)**.** Let $A, B \leq M$. Then

$$A + B = \{m \in M : m = a + b \text{ for some } a \in A, b \in B\} \leq M,$$

and

$$A \cap B \leq M.$$

We then have

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

Remark 4.24. More generally, for submodules $A_1, \cdots, A_n$,

$$A_1 + \cdots + A_n = \{a_1 + \cdots + a_n : a_i \in A_i\} \leq M$$

is an R-submodule.

Theorem 4.25 (Third isomorphism theorem)**.** Let $N \leq L \leq M$. Then $L/N \leq M/N$ and

$$\frac{M}{L} \cong \left(\frac{M}{N}\right) / \left(\frac{L}{N}\right).$$

Finally, we have a correspondence

$$\{\text{submodules of } M/N\} \longleftrightarrow \{\text{submodules of } M \text{ which contain } N\}$$

The proofs of each isomorphism theorem is similar to the case of rings. They will be omitted for brevity but are a useful exercise. It is also worth thinking about what these mean in the cases where R is a field, and modules are vector spaces.

We will now define what it means for an R-module to be finitely generated.

Definition 4.26. Let M be an R-module, and $m \in M$. The *submodule generated by* m is

$$Rm = \{r \cdot m \in M : r \in R\}.$$

In order to understand this, consider the R-module homomorphism

$$\begin{aligned} \varphi : R &\to M \\ r &\mapsto rm. \end{aligned}$$

This is clearly a homomorphism. Then we have

$$\begin{aligned} \operatorname{im}(\varphi) &= Rm, \\ \ker(\varphi) &= \{r \in R : r \cdot m = 0\}. \end{aligned}$$

We now have a new concept that was not present in rings and groups.

Definition 4.27. Let M be an R-module, and $m \in M$. The *annihilator* of m is

$$\operatorname{Ann}(m) = \{r \in R : r \cdot m = 0\}.$$

This is a two-sided ideal since $\operatorname{Ann}(m) = \ker(\varphi)$. Moreover, we have that:

$$Rm \cong R/\operatorname{Ann}(m).$$

As we mentioned, rings acting on modules is like groups acting on sets. We can think of this as the analogue of the orbit-stabilizer theorem.

Example 4.28. Suppose R is an integral domain and $I = (m) \subseteq R$ is a principal ideal with $m \neq 0$. Then I is an R-module and we have $\operatorname{Ann}(m) = \{0\}$. To see this, note that the R-action in I is just multiplication in R. So $r \in \operatorname{Ann}(m)$ if and only if $r \cdot m = 0 \in R$ if and only if $r = 0$, since $m \neq 0$ and R is an integral domain.

Hence, as an R-module, we have $I = Rm \cong R/\operatorname{Ann}(m) = R/\{0\} \cong R$. So all principal ideals in R are isomorphic as R-modules. This demonstrates that, for ideals in R, we have two very different notions of equivalence: being the same ideal, and being isomorphic as R-modules.

In general, we can generate a submodule with many elements.

Definition 4.29. An R-module M is *finitely generated* if there is a finite list of elements $m_1, \cdots, m_n$ such that

$$M = Rm_1 + Rm_2 + \cdots + Rm_n = \{r_1m_1 + r_2m_2 + \cdots + r_nm_n : r_i \in R\}.$$

This is in some sense analogous to the idea of a vector space being finite-dimensional. However, it behaves much more differently.

While this definition is rather concrete, it is often not the most helpful characterisation of finitely generated modules. Instead, we use the following lemma:

Lemma 4.30. An R-module M is finitely generated if and only if there is a surjective R-module homomorphism $f : R^n \twoheadrightarrow M$ for some finite n.

Proof. If

$$M = Rm_1 + Rm_2 + \cdots + Rm_n,$$

we define $f : R^n \to M$ by

$$(r_1, \cdots, r_k) \mapsto r_1 m_1 + \cdots + r_n m_n.$$

It is clear that this is an R-module homomorphism. This is by definition surjective. So done.

Conversely, given a surjection $f : R^n \twoheadrightarrow M$, we let

$$m_i = f(0, 0, \cdots, 0, 1, 0, \cdots, 0),$$

where the 1 appears in the ith position. We now claim that

$$M = Rm_1 + Rm_2 + \cdots + Rm_n.$$

So let $m \in M$. As f is surjective, we know

$$m = f(r_1, r_2, \cdots, r_n)$$

for some r_i. We then have

$$\begin{aligned} & f(r_1, r_2, \cdots, r_n) \\ &= f((r_1, 0, \cdots, 0) + (0, r_2, 0, \cdots, 0) + \cdots + (0, 0, \cdots, 0, r_n)) \\ &= f(r_1, 0, \cdots, 0) + f(0, r_2, 0, \cdots, 0) + \cdots + f(0, 0, \cdots, 0, r_n) \\ &= r_1 f(1, 0, \cdots, 0) + r_2 f(0, 1, 0, \cdots, 0) + \cdots + r_n f(0, 0, \cdots, 0, 1) \\ &= r_1 m_1 + r_2 m_2 + \cdots + r_n m_n. \end{aligned}$$

So the m_i generate M. □

This view is a convenient way of thinking about finitely generated modules. For example, we can immediately prove the following corollary:

Corollary 4.31. Let $N \leq M$ be R-modules. If M is finitely generated, then M/N is finitely generated.

Proof. Since m is finitely generated, we have some surjection $f : R^n \twoheadrightarrow M$. Moreover, we have the surjective quotient map $q : M \twoheadrightarrow M/N$. Then we get the following composition

$$R^n \xrightarrow{f} M \xrightarrow{q} M/N,$$

which is a surjection, since it is a composition of surjections. So M/N is finitely generated. □

It is very tempting to believe that if a module is finitely generated, then its submodules are also finitely generated. However, we have:

Example 4.32. A submodule of a finitely generated module need not be finitely generated.

We let $R = \mathbb{C}[X_1, X_2, \cdots]$. We consider the R-module $M = R$, which is finitely generated (by 1). A submodule of the ring is the same as an ideal. Moreover, an ideal is finitely generated as an ideal if and only if it is finitely generated as a module. We pick the submodule

$$I = (X_1, X_2, \cdots),$$

which we have already shown to be not finitely generated.

Example 4.33. For a complex number α, the ring $\mathbb{Z}[\alpha]$ (i.e. the smallest subring of $\mathbb{C}$ containing α) is a finitely generated as a $\mathbb{Z}$-module if and only if α is an algebraic integer.

The proof will not be given here. This gives an easier way to prove that algebraic integers are closed under addition and multiplication, since it is easier to argue about whether $\mathbb{Z}[\alpha]$ is finitely generated.

4.4 Free and projective modules

We will now consider the simplest classes of modules: free modules, stably free modules and projective modules.

Definition 4.34. Given a set S, define the *free module over* S to be the R-module

$$R^{(S)} = \bigoplus_{i \in S} R = \{(x_i)_{i \in S} \in \prod_{s \in S} R : x_i = 0 \text{ for all but finitely many } i\}$$

with coordinate wise addition and R-action. An R-module M is *free* if $M \cong R^{(S)}$ for some S.

Proposition 4.35. Let R be a non-trivial ring. Then the free module $R^{(S)}$ is finitely generated if and only if S is finite.

Proof. ($\Leftarrow$): If $|S| = n$, then $R^{(S)} \cong R^n = R \oplus \cdots \oplus R$. In particular, there is an isomorphism $f : R^n \to R^{(S)}$ (which is, in particular, surjective).

($\Rightarrow$): Suppose S is infinite and $R^{(S)} = Rm_1 + \cdots + Rm_n$. For each $1 \leq r \leq n$, write $m_r = (x_i^{(r)})_{i \in S} \in R^{(S)} \subseteq \prod_{s \in S} R$ and define $S_r = \{i \in S : x_i^{(r)} \neq 0\} \subseteq S$. By assumption, S_r is finite and so $S_1 \cup \cdots \cup S_n \subseteq S$ is finite. This implies that $S \setminus (S_1 \cup \cdots \cup S_n)$ is non-empty since S is infinite. Let $s \in S \setminus (S_1 \cup \cdots \cup S_n)$. Then $x_s^{(r)} = 0$ for all r and so the sth component of $r_1 m_1 + \cdots + r_n m_n$ is $r_1 0 + \cdots + r_n 0 = 0$. Since R is non-trivial, there exists $a \in R \setminus \{0\}$. Let $\alpha = (x_i) \in R^{(S)}$ where $x_s = a$ and $x_i = 0$ for $i \geq a$. Then $a \notin Rm_1 + \cdots + Rm_n$ and so $Rm_1 + \cdots + Rm_n \subsetneq R^{(S)}$. □

Given this, a finitely generated R-module is free if and only if $M \cong R^n$ for some n. We refer to R^n as the *free module of rank* n.

Proposition 4.36. Let F be a field. If M is an F-module, then M is a free F-module.

Proof. See Problem Sheet 4. □

We can think of free modules as natural generalisations of vector spaces of fields. We will now explore a more abstract definition of free modules using a universal property.

Definition 4.37. A subset $S \subseteq M$ generates M *freely* if

(i) S generates M as an R-module, i.e. $R \cdot S = M$

(ii) Any set function $\psi : S \to N$ to an R-module N extends to an R-module map $\theta : M \to N$.

Note that if θ_1, θ_2 are two such extensions, we can consider $\theta_1 - \theta_2 : M \to N$. Then $\theta_1 - \theta_2$ sends everything in S to 0. So $S \subseteq \ker(\theta_1 - \theta_2) \leq M$. So the submodule generated by S lies in $\ker(\theta_1 - \theta_2)$ too. But this is by definition M. So $M \leq \ker(\theta_1 - \theta_2) \leq M$, i.e. equality holds. So $\theta_1 - \theta_2 = 0$. So $\theta_1 = \theta_2$. So any such extension is unique.

Thus, what this definition tells us is that giving a map from M to N is exactly the same thing as giving a function from S to N.

Definition 4.38. An R-module M is *free* if it is freely generated by some subset $S \subseteq M$. A set S with this property is called a *basis* for M.

Proposition 4.39. The two definitions of free module are equivalent.

We will first show the following.

Lemma 4.40. Suppose M and N be R-modules such that M is freely generated by $S \subseteq M$ and N is freely generated by $T \subseteq N$. If there exists a bijection $S \cong T$, then $M \cong N$ as R-modules.

Proof. There are injective functions $i_M : S \to M$ (by inclusion) and $i_N : S \to N$ (composing $S \cong T$ with inclusion). Since S generates M freely, i_N extends to a map $\theta_N : M \to N$, i.e. $\theta_N \circ i_M = i_N$. Similarly, i_M extends to $\theta_M : N \to M$, i.e. $\theta_M \circ i_N = i_M$. In order to show these are R-module isomorphisms, it suffices to prove that $\theta_M \circ \theta_N = \mathrm{id}_M$ and $\theta_N \circ \theta_M = \mathrm{id}_N$.

Note that $\theta_M \circ \theta_N \circ i_M = i_M$ and $\mathrm{id}_M \circ i_M = i_M$. Hence $i_M : S \to M$ extends to both $\theta_M \circ \theta_N : M \to M$ and $\mathrm{id}_M : M \to M$. However, as shown above, this extension must be unique and so $\theta_M \circ \theta_N = \mathrm{id}_M$. Similarly $\theta_N \circ \theta_M = \mathrm{id}_N$ and so $M \cong N$. □

Remark 4.41. Given $\theta_M \circ \theta_N \circ i_M = i_M$, it does not follow immediately that $\theta_M \circ \theta_N = \mathrm{id}_M$ since i_M is not surjective. It only implies that they are equal when restricted to $\mathrm{im}(i_M)$.

Proof of Proposition 4.39. Let S be a set and consider the module $R^{(S)} = \bigoplus_{i \in S} R$. We can view $S \subseteq M$ by identifying each $s \in S$ with the element $(x_i)_{i \in S}$ which has $x_s = 1$ and $x_i = 0$ for $i \neq s$. It follows immediately from the definition that S generates M as an R-module. Now suppose $\psi : S \to N$ is a function. Then the map $\theta : R^{(S)} \to N$ sending $(x_i)_{i \in S} \mapsto \sum_{i \in S} x_i \cdot \psi(s)$ is an R-module homomorphism which extends ψ. Note that the sum is finite and hence well-defined.

Conversely, suppose M is an R-module and $S \subseteq M$ generates M freely. Since $R^{(S)}$ has this property, we have $M \cong R^{(S)}$ by the lemma above. □

We will now use this new definition to formulate free modules in a way more similar to what we do in linear algebra.

Definition 4.42. Let $m_1, \cdots, m_n \in M$. Then $\{m_1, \cdots, m_n\}$ is *linearly independent* if

$$\sum_{i=1}^{n} r_i m_i = 0$$

implies $r_1 = r_2 = \cdots = r_n = 0$.

Proposition 4.43. For a subset $S = \{m_1, \cdots, m_n\} \subseteq M$, the following are equivalent:

(i) S generates M freely.

(ii) S generates M and the set S is linearly independent.

(iii) Every element of M is *uniquely* expressible as

$$r_1 m_1 + r_2 m_2 + \cdots + r_n m_n$$

for some $r_i \in R$.

Proof. The fact that (ii) and (iii) are equivalent is something we would expect from what we know from linear algebra, and in fact the proof is the same. So we only show that (i) and (ii) are equivalent.

Let S generate M freely. If S is not independent, then we can write

$$r_1 m_1 + \cdots + r_n m_n = 0,$$

with $r_i \in R$ and, say, r_1 non-zero. We define the set function $\psi : S \to R$ by sending $m_1 \mapsto 1_R$ and $m_i \mapsto 0$ for all $i \neq 1$. As S generates M freely, this extends to an R-module homomorphism $\theta : M \to R$.

By definition of a homomorphism, we can compute

$$\begin{aligned} 0 &= \theta(0) \\ &= \theta(r_1 m_1 + r_2 m_2 + \cdots + r_n m_n) \\ &= r_1 \theta(m_1) + r_2 \theta(m_2) + \cdots + r_n \theta(m_n) \\ &= r_1. \end{aligned}$$

This is a contradiction. So S must be independent.

To prove the other direction, suppose every element can be uniquely written as $r_1 m_1 + \cdots + r_n m_n$. Given any set function $\psi : S \to N$, we define $\theta : M \to N$ by

$$\theta(r_1 m_1 + \cdots + r_n m_n) = r_1 \psi(m_1) + \cdots + r_n \psi(m_n).$$

This is well-defined by uniqueness, and is clearly a homomorphism. So it follows that S generates M freely. □

Remark 4.44. To prove (i) $\Leftrightarrow$ (ii), we could have also used the fact that (i) is equivalent to having $M \cong R^n$ for some n (since the two definitions of free module given above are equivalent).

Example 4.45. The $\mathbb{Z}$-module $\mathbb{Z}/2\mathbb{Z}$ is not free. Suppose $\mathbb{Z}/2\mathbb{Z}$ were generated by some $S \subseteq \mathbb{Z}/2\mathbb{Z}$. Then this can only possibly be $S = \{1\}$. Then this implies there is a homomorphism $\theta : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$ sending 1 to 1. But it does not send $0 = 1 + 1$ to $1 + 1$, since homomorphisms send 0 to 0. So $\mathbb{Z}/2\mathbb{Z}$ is not freely generated.

Example 4.46. The set $\{2, 3\} \in \mathbb{Z}$ generates $\mathbb{Z}$. However, they do not generate $\mathbb{Z}$ freely, since

$$3 \cdot 2 + (-2) \cdot 3 = 0.$$

Recall from linear algebra that if a set S spans a *vector space* V, and it is not independent, then we can just pick some useless vectors and throw them away in order to get a basis. However, this is no longer the case in modules. Neither 2 nor 3 generate $\mathbb{Z}$.

Definition 4.47. Let M be a finitely generated R-module. We have shown that there is a surjective R-module homomorphism $\varphi : R^n \to M$ for some n. We call the R-submodule $\ker(\varphi) \leq R^n$ the *relation module* for those generators.

Definition 4.48. A finitely generated R-module M is *finitely presented* if there exists a surjective homomorphism $f : R^n \to M$ such that $\ker(f)$ is a finitely generated R-module.

Note that, if $\ker(f)$ is finitely generated, then there exists a surjective R-module homomorphism $g : R^m \twoheadrightarrow \ker(f)$ for some m. Let $\varphi : R^m \to R^n$ be $i \circ g$ where $i : \ker(f) \to R^n$ is the inclusion map. Then we have:

$$\operatorname{coker}(\varphi) = R^n / \operatorname{im}(g) = R^n / \ker(f) \cong \operatorname{im}(f) = M.$$

Similarly to vector spaces, maps between finitely generated free R-modules correspond to matrices over R.

Proposition 4.49. Let $\varphi : R^m \to R^n$ be an R-module homomorphism. Let $e_1, \cdots, e_m \in R^m$ and $v_1, \cdots, v_n \in R^n$ be the standard basis elements. Let $\varphi(e_j) = \sum_{i=1}^{n} A_{ij} e_i$ for some $A_{ij} \in R$ and let $A = (A_{ij}) \in M_{m \times n}(R)$ be the corresponding $n \times m$ matrix. Then $\varphi(r) = r \cdot A$ where $\cdot$ represents right matrix multiplication of the row vector $r = r_1 e_1 + \cdots + r_n e_n$ by A.

Remark 4.50. If R is commutative, we can choose A such that $\varphi(r) = A \cdot r$ is left matrix multiplication.

In particular, every finitely presented R-module M is the cokernel of a map $\varphi : R^m \to R^n$ which can be expressed as an $m \times n$ matrix A. We will often write φ_A to denote the map corresponding to A.

A natural question we might ask is if $n \neq m$, then can we ever have $R^n \cong R^m$? In vector spaces, they obviously must be different, since basis and dimension are well-defined concepts.

Definition 4.51. We say that a ring R has the *invariant basis number property* (IBN) if $R^n \cong R^m$ are isomorphic as R-modules if and only if $n = m$.

Proposition 4.52. Non-trivial commutative rings have the invariant basis number property.

Proof. See Problem Sheet 4. □

Clearly the trivial ring $R = \{0\}$ does not have IBN. It is possible to construct examples of non-commutative rings which do not have IBN (see Problem Sheet 4). On the other hand, most reasonable classes of non-commutative rings are known to have IBN such as group rings $R[G]$ for R a commutative ring and G a group.

We will now consider classes of modules which are as close to being free modules as possible.

Definition 4.53. An R-module M is *stably free* if there exists n such that $M \oplus R^n$ is a free module. An R-module M is *projective* if there exists an R-module N such that $M \oplus N$ is a free R-module.

It follows from the definitions that free $\Rightarrow$ stably free $\Rightarrow$ projective. However, the converse need not hold:

Example 4.54. Let R_1 and R_2 be non-zero rings and let $R = R_1 \times R_2$. Then R_1 is an R-module with the natural R_1-action and the trivial R_2-action. Similarly R_2 is an R-module. As R-modules, we have that $R_1 \oplus R_2 \cong R$ is free. Hence R_1 and R_2 are projective R-modules.

However, R_1 and R_2 not free. To see this, note that the elements of $R_2 \subseteq R$ act trivially on R_1 but would act non-trivially on any free module.

A specific example is $\mathbb{Z}/6 \cong \mathbb{Z}/2 \times \mathbb{Z}/3$ which has projective modules $\mathbb{Z}/2$ and $\mathbb{Z}/3$ as described above. We can see they are not even stably free since they are both finitely generated but finitely generated stably free $\mathbb{Z}/6$-modules must have order 6^n for some n.

Example 4.55. Let $R = \mathbb{R}[X, Y, Z]/(X^2 + Y^2 + Z^2 - 1)$. Consider

$$M = \{(a, b, c) \in R^3 : aX + bY + cZ = 0 \in R\}$$

which is an R-submodule of R^3. Then $M \oplus R \cong R^3$ but $M \ncong R^2$. The proof will not be given here[1] and uses the hairy ball theorem from topology: there does not exists a non-vanishing continuous vector field on the sphere S^2.

On the other hand, free modules, stably free modules and projective modules often coincide for nice classes of rings.

Example 4.56. If R is a principal ideal domain, then projective modules over $R[X_1, \cdots, X_n]$ are free. Whilst this might sound like a simple statement to prove, it was an open problem for a long time and was resolved independently by Daniel Quillen and Andrei Suslin in 1976. Quillen was awarded a Fields Medal for his proof in 1978. The case where R is a field is particularly important since it has far reaching consequences in algebraic geometry.

4.5 Noetherian modules

Let R be a ring and let M be an R-module. We say M is *Noetherian* if every increasing infinite chain

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$$

of R-submodules N_i of M is eventually constant. (That is, for any such chain, we have $N_i = N_{i+1}$ for all sufficiently large i.) A ring R is Noetherian if R is Noetherian as an R-module.

Since the R-submodules of R are just the ideals of R, a ring R is Noetherian if every increasing infinite chain:

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

of ideals I_j of R is eventually constant.

The following result about Noetherian R-modules is fundamental:

Theorem 4.57. An R-module M is Noetherian if and only if every R-submodule of M is finitely generated.

Proof. Suppose first that M is Noetherian, and let N be an R-submodule of M. Choose an element n_0 of N, and let N_0 be the R-submodule of N generated by n_0. If N_0 is all of N, then N is finitely generated. Otherwise, choose n_1 in $N \setminus N_0$, and let N_1 be the R-submodule of N generated by n_0 and n_1. If N is not finitely generated, we may continue this process indefinitely, choosing for each i an n_i in $N \setminus N_{i-1}$ (which is nonempty since N is not finitely generated), and letting N_i be generated by $n_0, \dots, n_i$. In this way we obtain a strictly increasing infinite chain

$$N_0 \subsetneq N_1 \subsetneq N_2 \subsetneq \dots$$

of submodules of M, contradicting the fact that M is Noetherian.

Conversely, suppose that every R-submodule of M is finitely generated, and let

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq N_3 \subseteq \dots$$

be an increasing chain. We must show that this chain is eventually constant. Let N be the union of the submodules N_i; note that N is an R-submodule of M. Thus N is finitely generated, say by $n_0, \dots, n_r$. Since N is the union of the N_i, there exists $j_1, \dots, j_r$ such that n_i is in N_{j_i} for

[1] See `https://kconrad.math.uconn.edu/blurbs/linmultialg/stablyfree.pdf` for details.

all i. Let j be the largest of the j_i. Then N_j contains $n_0, \ldots, n_r$ so it contains N. In particular for any $i \geq j$ we have $N_j \subset N_i \subset N \subset N_j$, so $N = N_i = N_j$ for all such i and the chain is constant after N_j. □

Corollary 4.58. Let R be a Principal Ideal Domain. Then R is Noetherian.

Proof. Every ideal of R is principal, hence finitely generated. □

The goal of this section is to prove the following crucial theorem:

Theorem 4.59. Any finitely generated module over a Noetherian ring is Noetherian.

We proceed in several steps. First note:

Proposition 4.60. Let M be a Noetherian R-module. Then, for any submodule $N \leq M$, both N and M/N are Noetherian.

Proof. Since M is Noetherian, any submodule of M is finitely generated, and thus any submodule of N is finitely generated. Given a submodule J of M/N, let $\tilde{J}$ be its preimage in N under the natural map $M \to M/N$. Then $\tilde{J}$ is finitely generated, and the image of a generating set for $\tilde{J}$ in J is a generating set for J. □

Proposition 4.61. Let M be an R-module, let N be a Noetherian submodule of M, and suppose that M/N is Noetherian. Then M is Noetherian.

Proof. See Problem Sheet 4. □

Corollary 4.62. If M and N are Noetherian R-modules, then so is $M \oplus N$.

Proof. We have a surjection $M \oplus N \to N$ taking (m, n) to n. Its kernel K is the set of pairs of the form $(m, 0)$, which is isomorphic to M, and hence Noetherian. The surjection $M \oplus N \to N$ descends to an isomorphism $(M \oplus N)/K \cong N$, so that $(M \oplus N)/K$ is Noetherian. Thus $M \oplus N$ is Noetherian. □

Corollary 4.63. If R is Noetherian, then any free R-module of finite rank is Noetherian.

Proof. A free R-module of rank s is the direct sum of s copies of R, each of which is Noetherian as an R-module when R is Noetherian. □

Proof of Theorem 4.59. Let M be a finitely generated R-module, and let $m_1, \ldots, m_s$ be a set of generators for M. Then if F is a free R-module of rank s, with generators $e_1, \ldots, e_s$, we have a surjection of F onto M taking e_i to m_i for all i. Let K be the kernel. Then M is isomorphic to F/K, and F is a Noetherian R-module, so M is Noetherian as well. □

One nice consequence of this is as follows.

Corollary 4.64. Let R be a Noetherian ring. Then every finitely generated R-module is finitely presented.

Proof. Let M be a finitely generated R-module. Then there exists a map $f : R^n \twoheadrightarrow M$ and we want to show that $\ker(f) \leq R^n$ is finitely generated. Since R^n is a finitely generated R-module, Theorem 4.59 (or, more specifically, Corollary 4.63) implies that R^n is a Noetherian module. Theorem 4.57 then implies that $\ker(f)$ is finitely generated. □

Since principal ideal domains are Noetherian, this implies that finitely generated modules over principal ideal domains are finitely presented.

4.6 Modules over principal ideal domains

The aim of this section will be to prove the following theorem.

Theorem 4.65 (Classification of finitely generated modules over a PID)**.** Let R be a principal ideal domain. If M is a finitely generated R-module, then there exists $n, r \geq 0$ and elements $d_1, \cdots, d_r \in R$ such that

$$M \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r).$$

Furthermore, we can assume that $d_1 \mid d_2 \mid \cdots \mid d_r$.

In less compact notation, the last part says that $d_1 \mid d_2$, $d_2 \mid d_3$, $\cdots$, $d_{r-1} \mid d_r$. It can be shown that, if we choose the d_i to satisfy these conditions, then the n and the d_i are unique. We will not prove this in this course.

The outline for the proof is as follows.

(1) Let R be a PID and let M be a finitely generated R-module. Since R is Noetherian, the result in the previous section implies that M is finitely presented.

(2) Since M is finitely presented, we have that $M \cong \operatorname{coker}(\varphi_A)$ where $\varphi_A : R^m \to R^n$ is an R-module homomorphism represented by an $m \times n$ matrix A over R.

(3) If $B = PAQ$ for invertible matrices P and Q, then $\operatorname{coker}(\varphi_A) \cong \operatorname{coker}(\varphi_B)$ as R-modules. We call A and B *equivalent matrices*.

(4) Every matrix over a PID can be transformed into a rectangular diagonal matrix using a sequence of row and column operations (i.e. a square diagonal matrix with additional rows or columns which are zero). This is called *Smith normal form*.

(5) Let A be a rectangular diagonal matrix whose diagonal contains non-zero entries $d_1, \cdots, d_r$ and n copies of 0. Then $\operatorname{coker}(\varphi_A) \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$.

We have already completed parts (1) and (2) in the previous sections. We will now focus on parts (3), (4) and (5).

Definition 4.66. Two $m \times n$ matrices A and B over R are *equivalent* if there exists invertible matrices $P \in \mathrm{GL}_n(R)$ and $Q \in \mathrm{GL}_m(R)$ such that

$$B = PAQ.$$

For the corresponding maps $\varphi_A, \varphi_B : R^m \to R^n$, this is equivalent to the existence of R-module isomorphisms $f : R^n \to R^n$ and $g : R^m \to R^m$ such that $\varphi_B = f^{-1} \circ \varphi_A \circ g$. In particular, we can take $f = \varphi_{P^{-1}}$ and $g = \varphi_Q$.

We will now resolve part (3) of the proof.

Proposition 4.67. Let A and B be $m \times n$ matrices over a ring R. If A and B are equivalent, then $\operatorname{coker}(\varphi_A) \cong \operatorname{coker}(\varphi_B)$ are isomorphic as R-modules.

This is true since equivalent matrices correspond to the same map $\varphi : R^m \to R^n$ with different choices of bases for R^m and R^n. We will now check the details more carefully.

Proof. We have maps $\varphi_A, \varphi_B : R^m \to R^n$. Since A and B are equivalent, there exists R-module isomorphisms $f : R^n \to R^n$ and $g : R^m \to R^m$ such that $f \circ \varphi_B = \varphi_A \circ g$ and so $\operatorname{im}(\varphi_A) = \operatorname{im}(\varphi_A \circ g) = \operatorname{im}(f \circ \varphi_B) = f(\operatorname{im}(\varphi_B))$.

Consider the map $\bar{f} : R^n \to R^n/\operatorname{im}(\varphi_A)$ which is the composition of f with the quotient map $\bar{\cdot} : R^n \to R^n/\operatorname{im}(\varphi_A)$. We claim that $\ker(\bar{f}) = \operatorname{im}(\varphi_B)$. First note that $\bar{f}(\operatorname{im}(\varphi_B)) = \overline{f(\operatorname{im}(\varphi_B))} = \overline{\operatorname{im}(\varphi_A)} = \{0\}$ and so $\operatorname{im}(\varphi_B) \subseteq \ker(\bar{f})$. Conversely, if $\bar{f}(r) = 0$, then $f(r) \in \operatorname{im}(\varphi_A)$ implies $r \in \operatorname{im}(f^{-1} \circ \varphi_A) = \operatorname{im}(f^{-1} \circ \varphi_A \circ g) = \operatorname{im}(\varphi_B)$ and so $\ker(\bar{f}) \subseteq \operatorname{im}(\varphi_B)$. Since $\bar{f}$ is surjective, the first isomorphism theorem now implies that $R^n/\operatorname{im}(\varphi_B) \cong R^n/\operatorname{im}(\varphi_A)$, i.e. $\operatorname{coker}(\varphi_B) \cong \operatorname{coker}(\varphi_A)$. □

In practice, finding equivalent matrices typically amounts to performing row or column operations.

Definition 4.68. *Elementary row operations* on an $m \times n$ matrix A with entries in R are operations of the form

(i) Add $c \in R$ times the ith row to the jth row. This may be done by multiplying by the following matrix on the left:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & c & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix},$$

where c appears in the ith column of the jth row.

(ii) Swap the ith and jth rows. This can be done by left-multiplication of the matrix

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

Again, the rows and columns we have messed with are the ith and jth rows and columns.

(iii) We multiply the ith row by a *unit* $c \in R$. We do this via the following matrix:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

Notice that if R is a field, then we can multiply any row by any non-zero number, since they are all units.

We also have elementary column operations defined in a similar fashion, corresponding to right multiplication of the matrices. Notice all these matrices are invertible.

The following is immediate, though worth stating.

Proposition 4.69. Let A be an $m \times n$ matrix over R. If B is obtained from A by elementary row and column operations, then A and B are equivalent.

The aim is now to find, for each matrix, a matrix equivalent to it that is as simple as possible. Recall from linear algebra that if R is a field, then we can put any matrix into the form

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

via elementary row and column operations. This is no longer true when working with rings. For example, over $\mathbb{Z}$, we cannot put the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

into that form, since no operation can turn the 2 into a 1.

Definition 4.70. We say that an $m \times n$ matrix over R is in *Smith normal form* if it has the form

$$\begin{pmatrix} d_1 & & & & & & \\ & d_2 & & & & & \\ & & \ddots & & & & \\ & & & d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix},$$

with the d_i all non-zero and

$$d_1 \mid d_2 \mid d_3 \mid \cdots \mid d_r.$$

We would like to show that matrices over a principal ideal domains R can be put in Smith normal form.

Theorem 4.71 (Smith normal form)**.** Every $m \times n$ matrix over a principal ideal domain R is equivalent to a matrix in Smith normal form.

Note that, since R is a PID, R is a UFD and so greatest common divisors exist. Whilst gcd is only defined up to units, we will fix a choice each time we use it.

Proof. Let $A = (a_{ij})$ be an $m \times n$. When we say that we modify A, we mean that we replace A by an equivalent matrix. For ease of notation, we will still use A to denote this matrix.

If $A = 0$, then done. So suppose $A \neq 0$. We start by proving the following.

Claim 1. Given two entries a_{ij} and a_{kl} in the same row ($j = l$) or column ($i = k$), we can modify A so that $\gcd(a_{ij}, a_{kl})$ appears as an entry in the matrix.

Proof. Suppose the two entries are in the same column, i.e. we a_{ij} and a_{ik}. Since invertible 2×2 matrices can be extended to invertible $m \times m$ matrices, it suffices to show that any vector

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

can be be left-multiplied by an invertible 2×2 matrix P to obtain a vector with an entry $\gcd(a, b)$. Since R is a PID, we have that $(a, b) = (d)$ for some $d \in R$. It follows easily that $d = \gcd(a, b)$. Since $d \in (a, b)$, there exists $x, y \in R$ such that $xa+yb = d$. Note that multiplying by a matrix if the form

$$P = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

changes a to d. We have $\det(P) = xw - yz$. Since $xa + yb = \gcd(a, b)$, we must have that $\gcd(x, y) = 1$ and so there exists $u, v \in R$ such that $xu + yv = 1$ and so choosing $z = -v$ and $w = u$ gives a matrix with $\det(P) = 1$ and so which is invertible.

If the entries of A are in the same row, then we can do the same but with right-multiplication by an invertible matrix instead. Alternatively, apply the above the transpose A^T to get a matrix P such that $P \cdot A^T$ has the required entry. Then $AP^T = (PA^T)^T$ has the required entry. □

This will now be used to show the following, which will be the basis for an algorithm which we will describe to modify A to put it in Smith normal form.

Claim 2. We can modify A so that $a_{11} \mid a_{i1}$ and $a_{11} \mid a_{1j}$ for all i, j.

Proof. For $r \in R \setminus \{0\}$, let $\delta(r) \in \mathbb{Z}_{\geq 0}$ denote the number of factors which appear in the factorisation of r into irreducible elements, i.e. if $r = p_1^{n_1} \cdots p_t^{n_t}$ for p_i irreducible, then $\delta(r) = \sum_{i=1}^{t} n_i$. Since R is a PID, R is a UFD and so this is well-defined.

Suppose A is not of this form. Then there exists a non-zero entry a_{ij}. Using row and column operations, we can move this entry to the top left corner and so we can assume that $a_{11} \neq 0$. Let $\alpha_1 = a_{11}$. If A is still not of the required form, then there exists a_{ij} with $i = 1$ or $j = 1$ such that $a_{11} \nmid a_{ij}$. By Claim 1, we can modify A so that $\alpha_2 := \gcd(a_{11}, a_{ij})$ appears as an entry in the matrix. Using row and column operations, we can modify A to get that α_2 is in the top left corner. Note that $\alpha_2 \mid \alpha_1$ but α_2 and α_1 are not associates since $\alpha_2 \mid a_{ij}$ and $\alpha_1 \nmid a_{ij}$. This implies that $\delta(\alpha_2) < \delta(\alpha_1)$. By the well-ordering principle, this process terminates. That is, we

eventually obtain an entry α_t in the top left corner such that $\alpha_t \mid a_{ij}$ for all i, j such that $i = 1$ or $j = 1$. □

Suppose A is in the form given in Claim 2 and let $d_1 = a_{11}$. Since $d_1 \mid a_{1j}$ for all j, we can subtract appropriate multiples of the first column from others so that $a_{1j} = 0$ for $j \neq 1$. After these transformations, it is still true that $a_1 \mid a_{i1}$ for all i and so we can do the same thing with rows so that the first row is cleared. Then we have a matrix of the form:

$$A = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & C & \\ 0 & & & \end{pmatrix}.$$

We can now apply the same process to C. If $C = 0$, then we are done and have $r = 1$. If not, we can put C in the above form with d_2 in the top left corner. By induction on n, repeating this process puts the matrix in a diagonal rectangular form with diagonal entries $d_1, \cdots, d_r, 0 \cdots, 0$.

It remains to show that we can modify this matrix to get that $d_1 \mid d_2 \mid \cdots \mid d_r$. Pick $1 \leq i, j \leq r$ with $i \neq j$ and consider the following row and column operations on 2×2 matrices

$$\begin{pmatrix} d_i & 0 \\ 0 & d_j \end{pmatrix} \sim \begin{pmatrix} d_i & \gcd(d_i, d_j) \\ 0 & d_j \end{pmatrix} \sim \begin{pmatrix} d_i & \gcd(d_i, d_j) \\ -d \cdot d_i & 0 \end{pmatrix} \sim \begin{pmatrix} \gcd(d_i, d_j) & 0 \\ 0 & d \cdot d_i \end{pmatrix}$$

where d is such that $d_j = d \cdot \gcd(d_i, d_j)$. These operations can be extended to operations on $m \times n$ matrices and show that, given d_i, d_j on the diagonal, we can modify A to still be diagonal rectangular and have a diagonal entry $\gcd(d_i, d_j)$. We can now apply the same argument that we used in the proof of Claim 2.Let $\alpha_1 = d_1$. If $d_1 \nmid d_i$ for some $i \geq 2$, then use the procedure above to get $\alpha_2 = \gcd(d_1, d_i)$ and then use row and column operations to move it into the place of d_1. We have $\alpha_2 \mid \alpha_1$ but $\alpha_1 \nmid \alpha_2$ (since $d_1 \nmid d_i$) and so $\delta(\alpha_2) < \delta(\alpha_1)$. By the well ordering process, this process terminates and so we eventually obtain a sequence where $d_1 \mid d_i$ for all i. We can repeat the same process for $d_2 \nmid d_i$ for some $i \geq 3$. Note that $d_1 \mid d_i$ for all i still holds after using the operations described to modify $d_2, \cdots, d_r$. Hence this process eventually results in a sequence such that $d_1 \mid d_2 \mid \cdots \mid d_r$, as required. □

If R is a Euclidean domain, then this can be done is a reasonably straightforward manner using row and column operations.

Exercise 4.72. Show that, if R is a Euclidean domain, then every $m \times n$ matrix can be transformed using row and column operations into a matrix in Smith normal form. Deduce that, for all n, every matrix in $\mathrm{GL}_n(R)$ is the product of elementary matrices.

Does this hold for principal ideal domains? In order to find an example where this does not work, we need R to be a principal ideal domain but not a Euclidean domain. We gave exactly one such example in this course which was the ring $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$.

Exercise 4.73. Let $R = \mathbb{Z}[\alpha]$ where $\alpha = \frac{1+\sqrt{-19}}{2}$. Prove that the following matrix in $\mathrm{GL}_2(R)$ is not the product of elementary matrices:

$$\begin{pmatrix} 3 - \alpha & 2 + \alpha \\ -3 - 2\alpha & 5 - 2\alpha \end{pmatrix}$$

This can be viewed as a further extension of Q11 on Problem Sheet 1.

This shows that, in our proof that matrices over principal ideals domains are equivalent to matrices in Smith normal form, it is necessary to use something other than row and column operations somewhere in the proof. We used such matrices in the proof of Claim 1.

Remark 4.74. We say that a ring is *generalised Euclidean* (GE) if, for all $n \geq 1$, every matrix in $\mathrm{GL}_n(R)$ is the product of elementary matrices. The above shows that, if R is a Euclidean domain, then R is generalised Euclidean. Furthermore, $\mathbb{Z}[\frac{1+\sqrt{-19}}{2}]$ is an example of a ring which is a principal ideal domain but which is not generalised Euclidean.

This completes part (4) of the proof. We will now resolve part (5).

Proposition 4.75. Let A be an $m \times n$ matrix over a ring R which is in Smith normal form, i.e.

$$A = \begin{pmatrix} d_1 & & & & & & \\ & d_2 & & & & & \\ & & \ddots & & & & \\ & & & d_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

for some non-zero elements $d_1, \cdots, d_r \in R$. Then $\mathrm{coker}(\varphi_A) \cong R^{n-r} \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$.

Proof. Define a surjective R-module homomorphism $f : R^n \twoheadrightarrow R/(d_1) \oplus \cdots R/(d_r) \oplus R \oplus \cdots \oplus R$ to be the diagonal map which is the quotient map $R \to R/(d_i)$ on the first r components and the identify map $R \to R$ on the next $n - r$ components. Note that $\mathrm{im}(\varphi_A) = (d_1) \oplus (d_2) \oplus \cdots (d_r) \oplus 0 \oplus \cdots \oplus 0$, and it is straightforward to see that this coincides with $\ker(f)$. Hence, by the first isomorphism theorem, we have

$$\mathrm{coker}(\varphi_A) = R^n / \mathrm{im}(\varphi_A) = R^n / \ker(f) \cong \mathrm{im}(f) \cong R^{n-r} \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$$

as required. □

This completes the proof of our our main theorem. We will recap on the steps of the proof.

Theorem 4.76 (Classification of finitely generated modules over a PID)**.** Let R be a principal ideal domain. If M is a finitely generated R-module, then there exists $n, r \geq 0$ and elements $d_1, \cdots, d_r \in R$ such that

$$M \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r).$$

Furthermore, we can assume that $d_1 \mid d_2 \mid \cdots \mid d_r$.

Proof. (1) Let R be a PID and let M be a finitely generated R-module. Since R is Noetherian, the result in the previous section implies that M is finitely presented.

(2) Since M is finitely presented, we have that $M \cong \mathrm{coker}(\varphi_A)$ where $\varphi_A : R^m \to R^n$ is an R-module homomorphism represented by an $m \times n$ matrix A over R.

(3) If A and B are equivalent matrices, then $\mathrm{coker}(\varphi_A) \cong \mathrm{coker}(\varphi_B)$ as R-modules.

(4) Since R is a PID, every matrix over a R is equivalent to a matrix B in Smith normal form.

(5) Suppose B is a matrix whose diagonal contains non-zero entries $d_1, \cdots, d_r$ and n copies of 0. Then $M \cong \operatorname{coker}(\varphi_A) \cong \operatorname{coker}(\varphi_B) \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$. □

Remark 4.77. There are another type of decomposition for modules over principal ideal domains that we can derive from this one. For each, if $d \in R \setminus \{0\}$ has the form $d = p_1^{n_1} \cdots p_m^{n_m}$ for $p_i \in R$ prime, then we can show the following using the Chinese remainder theorem for rings:

$$R/(d) \cong R/(p_1^{n_1}) \oplus \cdots \oplus R/(p_m^{n_m}).$$

In particular, if M is a finitely generated R-module where R is a PID, then we have:

$$M \cong R^n \oplus R/(p_1^{n_1}) \oplus \cdots \oplus R/(p_t^{n_t})$$

for some $n, t, n_i \geq 0$ and some prime elements $p_i \in R$. This is known as *prime decomposition*.

In the remainder of this section, we will give a few consequences of this theorem and briefly discuss how to approach the classification of modules over rings which are not principal ideal domains.

Firstly note that the classification of finitely generated modules over principal ideal domains is even interesting in the case where $R = \mathbb{Z}$, i.e. where R-modules are just abelian groups.

Corollary 4.78 (Classification of finitely generated abelian groups)**.** If G is a finitely generated abelian group, then there exists $n, r \geq 0$ and positive integers $d_1, \cdots, d_r$ such that

$$G \cong \mathbb{Z}^n \oplus \mathbb{Z}/(d_1) \oplus \cdots \oplus \mathbb{Z}/(d_r).$$

Furthermore, we can assume that $d_1 \mid d_2 \mid \cdots \mid d_r$.

Proof. Put $R = \mathbb{Z}$ in the above theorem. If $d_i < 0$, then we can replace it with $-d_i$ since (d_i) and $(-d_i)$ are equal ideals in $\mathbb{Z}$. □

It is important to note that our proof was constructive. In particular, we described an explicit algorithm that can be used to obtain the integers $n, r \geq 0$ and positive integers $d_1, \cdots, d_r \in \mathbb{Z}$.

Example 4.79. Let A be the abelian group generated by a, b, c with relations

$$\begin{aligned} 2a + 3b + \ \ c &= 0, \\ a + 2b \ \ \ \ \ \ &= 0, \\ 5a + 6b + 7c &= 0. \end{aligned}$$

In other words, we have

$$A = \frac{\mathbb{Z}^3}{\langle (2,3,1), (1,2,0), (5,6,7) \rangle}.$$

We would like to get a better description of A. It is not even obvious if this module is the zero module or not. To work out a good description, We consider the matrix

$$X = \begin{pmatrix} 2 & 1 & 5 \\ 3 & 2 & 6 \\ 1 & 0 & 7 \end{pmatrix}.$$

Using row and column operations, we can put this in Smith normal form. We obtain:

$$X' = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So we know

$$A \cong \frac{\mathbb{Z}}{(1)} \oplus \frac{\mathbb{Z}}{(1)} \oplus \frac{\mathbb{Z}}{(3)} \cong \frac{\mathbb{Z}}{(3)} \cong C_3.$$

We can also apply the theorem in the case $R = F[X]$ where F is a field, since R is a Euclidean domain and so a principal ideal domain.

Let A be an $n \times n$ matrix over F with corresponding map $\varphi_A : F^n \to F^n$. Then there is a corresponding $F[X]$-module M_A with abelian group F^n and, for $f = a_m X^m + \cdots + a_1 X + a_0$ and $x \in F^n$, the $F[X]$-action is given by

$$f \cdot x := a_m \cdot \varphi_A^m(x) + \cdots + a_1 \cdot \varphi_A(x) + a_0.$$

Conversely, suppose M is a finitely generated $F[X]$-module whose underlying F-module is finitely generated and so of the form F^n for some n. The action by $X \in F[X]$ gives a map $X\cdot : F^n \to F^n$. If this map is denote by φ_A, then we have that $M \cong M_A$ as above.

Hence there is a one-to-one correspondence between $n \times n$ matrices A over F and finitely generated $F[X]$-modules M_A whose underlying F-module is F^n. This means that decompositions of $F[X]$-modules can be interpreted as giving decompositions for matrices over F.

Example 4.80. Let F be a field and let A be an $n \times n$ matrix over F. Since $F[X]$ is a PID, there exists $n, r \geq 0$ and $f_1, \cdots, f_r \in F[X] \setminus \{0\}$ such that

$$M_A \cong F[X]^n \oplus F[X]/(f_1) \oplus \cdots \oplus F[X]/(f_r).$$

Since the underlying F-module of M_A is finitely generated, we must have that $n = 0$. We can assume the f_i are monic since $(df_i) = (f_i)$ for all $d \in F^\times$.

If $f = X^m + a_{m-1}X^{m-1} + \cdots + a_1 X + A_0$, then it can be shown that $\mathbb{F}[X]/(f) \cong M_{c(f)}$ where $c(f)$ is the following $m \times m$ matrix over F:

$$c(f) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{m-1} \end{pmatrix}.$$

known as the *companion matrix* for f. Hence we have that

$$M_A \cong M_{c(f_1)} \oplus \cdots \oplus M_{c(f_r)} \cong M_{c(f_1) \oplus \cdots \oplus c(f_r)}$$

where $\oplus$ denotes block addition of matrices. That is, every $n \times n$ matrix A is equivalent to a

block diagonal matrix of the form:

$$c(f) = \begin{pmatrix} c(f_1) & 0 & \cdots & 0 \\ 0 & c(f_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c(f_r) \end{pmatrix}.$$

This is known as Rational canonical form.

Prime decomposition of modules over PIDs leads to further types of matrix decomposition. Over $\mathbb{C}[X]$, this decomposition takes a particularly simple form since the prime elements are just $X - \lambda$ for $\lambda \in \mathbb{C}$. In particular, we need only determine the matrix corresponding to $(X - \lambda)^n$ which is a Jordan block. The associated decomposition is Jordan normal form.

Given a general form for a module over a principal ideal domain R, we can now determine when R-modules have particular properties. One property which tends to be particularly elusive is the property of being a projective R-module. We have:

Corollary 4.81. Let R be a principal ideal domain. Then finitely generated projective R-modules are free.

Proof. Let M be a finitely generated projective R-module. Then, by the classification of finitely generated modules over a PID, we have that

$$M \cong R^n \oplus R/(d_1) \oplus \cdots \oplus R/(d_r)$$

for some $d_1, \cdots, d_r \in R \setminus \{0\}$. Since M is projective, there exists an R-module N such that $M \oplus N \cong R^{(S)}$ is a free module for some set S. If $d_1, \cdots, d_r \in R^\times$, then $M \cong R^n$ is free. If not, then there exists an injective R-module homomorphism $i : R/(d) \hookrightarrow R^{(S)}$. Let $i(1) = (x_i)_{i \in S} \in R^{(S)}$. Since i is injective, $i(1) \neq 0$ and so there exists $s \in S$ such that $x_s \neq 0$. Since $d \cdot i(1) = i(d) = 0$, we must have that $dx_s = 0 \in R$. Since R is an integral domain and $x_s \neq 0$, this implies that $d = 0$ which is a contradiction. □

Remark 4.82. In fact, for R a principal ideal domain, all projective R-modules are free. There is a theory of infinitely generated projective modules, which we will not give here, that implies that such modules are free for a much wider class of rings than principal ideal domains.

We conclude by asking what happens when R is not a principal ideal domain. Consider the case where $R = \mathbb{Z}[\alpha]$ where $\alpha \in \mathbb{C}$ is an algebraic integer (or, more generally, R can be any Dedekind domain). If $F = \mathrm{Frac}(R)$, then a *fractional ideal* is an R-submodule $I \subseteq F$ such that $r \cdot I \subseteq R$ for some $r \in R$. For rings of this form, we can consider the *ideal class group*

$$C(R) = \{I : I \text{ a fractional ideal in } R\}/\sim$$

where $I \sim J$ if there exists $u \in F^\times$ such that $I = u \cdot J \subseteq F$. This is an abelian group under the operation $(I, J) \mapsto I \cdot J$. It can be shown that $C(R)$ is trivial if and only if R is a principal ideal domain and so, in some sense, $C(R)$ measures the failure of R to be a principal ideal domain.

It turns out that the ideal class group is just a special case of a much more general construction which works for all rings. For an arbitrary ring R, define the *projective class group*

$$\widetilde{K}_0(R) = \{P : P \text{ a finitely generated projective } R\text{-module}\}/\cong_{\mathrm{st}}$$

where $P \cong_{\text{st}} Q$ if and only if there exists n, m such that $P \oplus R^n \cong Q \oplus R^m$. This is an abelian group under the operation $(P, Q) \mapsto P \oplus Q$ and is one of the basic objects of algebraic K-theory. Note that a projective R-module P is stably free if and only if $[P] = 0 \in \widetilde{K}_0(R)$ and so, in some sense, $\widetilde{K}_0(R)$ measures the failure of projective R-modules to be stably free. If R is a principal ideal domain, then the above implies that $\widetilde{K}_0(R)$ is trivial. More generally, if $R = \mathbb{Z}[\alpha]$ for $\alpha \in \mathbb{C}$ an algebraic integer (or, more generally, R a Dedekind domain), then it can be shown that $\widetilde{K}_0(R) \cong C(R)$ as abelian groups.

There are many deep questions about $\widetilde{K}_0(R)$ which have remained unanswered for decades, and which remain at the forefront of modern research. One such question is: does there exists a torsion-free group G (i.e. every non-zero element has infinite order) such that $\widetilde{K}_0(\mathbb{Z}[G])$ is non-trivial? That is, does there exists a non-stably free projective $\mathbb{Z}G$-module over a torsion-free group G? For example, if G is finitely generated abelian group, then $G \cong \mathbb{Z}^n$ (by Corollary 4.78) and so $\mathbb{Z}[G] \cong \mathbb{Z}[X_1, X_1^{-1}, \cdots, X_n, X_n^{-1}]$. Since $\mathbb{Z}$ is a PID, the Quillen-Suslin theorem (Example 4.56) implies that projective modules are free over $\mathbb{Z}[X_1, \cdots, X_n]$. This can be used to deduce that projective modules are free over $\mathbb{Z}[X_1, X_1^{-1}, \cdots, X_n, X_n^{-1}]$ and so $\widetilde{K}_0(\mathbb{Z}[G])$ is trivial in this case. More generally, the Farrell-Jones conjecture in algebraic K-theory predicts that such groups do not exist and this has been verified for many large classes of torsion free groups (not just finitely generated abelian groups). However, whether or not the prediction made by the Farrell-Jones conjecture actually holds remains a mystery.