# Rings and Modules

## 6CCM350A, Fall 2020[1]

**Salman Ahmad Faris**

*King's College London*

*E-mail:* salman.faris@kcl.ac.uk

[1] *based on lectures by Nicholas Shepherd-Barron. These notes, however, have been altered strongly after the lectures. In particular, some proofs were filled in by me; and some of the contents were adjusted so that organization is prioritized and things are coherent. All errors are surely mine, feel free to email me if you spot any.*

# Contents

**Epilogue.** Let's be honest. This course is actually a course on **commutative algebra**. What is this you ask? It is the study of applications of rings and its modules to number theory and geometry. This gives rise to two fundamentals branch of mathematics — algebraic number theory and algebraic geometry. Do not expect that this is a course on simply learning the basics of rings and modules. We will still give a fundamental treatment on them before moving on to the fun stuff. Oh and one more thing, linear algebra will be assumed throughout. You have been warned!

**Notation.** We will mention only a few conventions as specific ones will be mentioned throughout the notes. If $X, Y$ are sets, we prefer $X \backslash Y$ over $X - Y$ to mean set difference. If $f : X \to Y$ is a map of sets, the *direct image* of $f$ will be written $f^{\text{img}}(X)$ and the *preimage* of $f$ will be denoted $f^{\text{preimg}}(Y)$. When things are clear, we may resort to the standard convention of just writing $f(X)$ and $f^{-1}(Y)$, but we avoid doing so. Our writing philosophy is always to value clarity over simplicity. Finally, the symbol **何!** means that we have reached a contradiction, mainly in proofs.

**References.** Together with the lectures, the following books were useful references in the making of this notes:

(1). M. Artin, *Algebra*, Prentice-Hall.

(2). M. F. Atiyah, I. G. MacDonald, *Introduction to Commutative Algebra*, CRC Press.

(3). D. S. Dummit, R. M. Foote, *Abstract Algebra*, John Wiley and Sons.

(4). I. Kaplansky, *Commutative Rings*, The University of Chicago Press.
(mostly for G-domains)

(5). D. Chua, *Part IB — Groups, Rings and Modules.*

# 1 Rings

Rings are not that far from groups. In group theory, we study an object $G$ with an algebraic structure defined by a binary operation $\star$ and call the pair $(G, \star)$ a group . In ring theory, we will investigate an object $R$ with an algebraic structure defined by two binary operations, $+, \times$. We then call the triplet $(R, +, \times)$ a ring. We start with some definitions.

## 1.1 Basic definitions

Let us start with a formal definition of rings.

**Definition** (*Ring*)**.** A **ring** $R$ is a set with two binary operations $+$ and $\cdot$ which satisfies the following axioms:

1. $(R, +)$ is an abelian group with identity $0_R$.
2. $\cdot$ is associative and distributive over addition.
3. $\exists 1_R \in R$ such that $x \cdot 1_R = 1_R \cdot x = x$ for all $x \in R$.

We may sometimes write $(R, +, \cdot)$ for a ring. We will refer to $0_R$ as the additive identity and $1_R$ as the multiplicative identity of $R$.

these operations are called *addition* and *multiplication* respectively.

**Remark.** We insist that $1_R$ is in $R$. Some books don't necessarily agree with this convention. They define rings without $1_R$ and call rings with it a *unital ring*. Some notation remarks:

- We may write $rs$ instead of $r \cdot s$ for $r, s \in R$ just because it is easier to read (and less painful to write in LaTeX).
- We will write $-r$ for the additive inverse of the ring element $r$, and treat addition with additive inverses by simply writing $s - r$ instead of $s + (-r)$ where $s \in R$.

**Example.** The set of continuous real-valued functions of a real variable $x$ forms a ring, with addition and multiplication being:

$$(f+g)(x) = f(x) + g(x), \qquad (fg)(x) = f(x)g(x).$$

**Remark.** Let $R$ be a ring. Note that it is possible that $0_R = 1_R$. But if so, things don't get very interesting. Suppose $R$ has this property, and let $r \in R$ be arbitrary. Then we have

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

and so $R = \{0_R\}$. We say that $R$ is the **zero ring** and denote this by 0. A ring that is not the zero ring is called a **nonzero ring**.

We have effectively prove the following in the last remark.

**Proposition 1.1.** $R$ is a ring such that $1_R = 0_R$ *if and only if* $R$ is the zero ring.

We will be most interested in the case that elements of a ring commute with each other. This is because these type of rings behave like $\mathbb{Z}$. As we have established, we want to study number theory and geometry using rings. So we want to deduce general theorems about commutative rings and apply it to the case when the ring is $\mathbb{Z}$.

**Definition** (*Commutative ring*)**.** Let $R$ be a ring. We say $R$ is **commutative** if for all $r, s \in R$, we have $r \cdot s = s \cdot r$.

**Remark.** Very Important Remark: We will only consider commutative rings from here on out. Thus, by **ring**, we will mean a **commutative ring with multiplicative identity**.

This is the second[1] warning in this course. If you want to memorize things in here, read *rings* as *commutative rings*. There is a chance that the theorems we will prove holds for non-commutative rings as well. But probabilistically speaking, reduce your chances of being wrong by following our advice.

**Definition** (*Subring*)**.** Let $(R, +, \cdot)$ be a ring and $S \subseteq R$. We say $S$ is a **subring** of $R$ if $(S, +, \cdot)$ is a ring itself and write $S \leqslant R$.

with $1_S = 1_R$ and $0_S = 0_R$.

**Strong note.** Remember that we demand that there is a multiplicative identity in any ring, including subrings! Here is a handy criterion to check for subring.

**Theorem 1.1** (*Subring criterion*)**.** Let $(R, +, \cdot)$ be a ring and $S \subseteq R$ non-empty. Then $S$ is a subring of $R$ if

$$a, b \in S \implies a - b \in S \text{ and } a \cdot b \in S,$$

[1] *first being linear algebra is assumed throughout.*

and $1_R \in S$.

**Example.** $\mathbb{Z} \leqslant \mathbb{Q} \leqslant \mathbb{R} \leqslant \mathbb{C}$ are all (commutative) rings with the usual addition and multiplication, and usual 0 and 1.

**Definition** (*Unit*). A **unit** in a ring $R$ is an element $u$ such that there exists $v \in R$ such that $u \cdot v = 1_R$.

A unit is basically an element which has a multiplicative inverse in the ring.

**Example.** The only units in $\mathbb{Z}$ are $\pm 1$.

**Fact** (*Always units*). $1_R$ and $-1_R$ are always units in any ring.

**Example.** Any nonzero element $a/b \in \mathbb{Q}$ is a unit with inverse $b/a$.

Note that multiplicative inverses are not required to exist in a ring. If they do exist (for all nonzero elements), then every nonzero element is a unit and the ring is called a field.

**Definition** (*Field*). Let $K$ be a nonzero ring. We say $K$ is a **field** if every nonzero element of $K$ is a unit.

We could have an alternative definition when we define domains.

**Example.** $\mathbb{Q} \leqslant \mathbb{R} \leqslant \mathbb{C}$ are all fields. In the context of number theory, we will be most interested in the field $\mathbb{Q}$ and its "extensions". In the context of geometry, we are interested in "algebraically closed" fields.

One of the most important rings are polynomial rings. Before we can say what they are, we have to barrage you with some new definitions.

**Definition** (*Polynomial*). Let $R$ be any ring. A **polynomial** over $R$ (in the indeterminate $X$) is an expression of the form

$$f = a_0 + a_1 X + \ldots + a_n X^n = \sum_{i=0}^{n} a_i X^i$$

where $a_i \in R$.

over $R$ is also equivalent to saying having *coefficients in* $R$.

**Remark.** We will usually omit the "in the indeterminate" part. An indeterminate is equivalently called a *variable*.

**Definition** (*Degree of a polynomial*). Let $f$ be a polynomial over a ring $R$. We say that $f$ has **degree** $n$ if we can write

$$f = a_0 + a_1 X + \ldots + a_n X^n$$

with $a_n \neq 0_R$. In this case, we will write $\deg f = n$.

it should naturally be understood that the powers of $X$ are written in such a way that $n$ is the biggest integer.

A nice class of polynomials are the monic ones.

**Definition** (*Monic*). Let $f$ be a polynomial over a ring $R$. We say that $f$ is **monic** if $\deg f = n$ and $a_n = 1_R$.

**Definition** (*Polynomial ring*)**.** Let $R[X]$ be the set of all polynomials over a ring $R$ in an indeterminate $X$. It can be made into a ring by defining addition and multiplication in the following way: Let $f = \sum_i^n a_i X^i$ and $g = \sum_j^m b_i X^j$. Write $N = \max\{m, n\}$. Then

$$f + g = \sum_{i=0}^{N} (a_i + b_i) X^i, \quad \text{and} \quad f \cdot g = \sum_{i=0}^{m+n} \left( \sum_{j=0}^{i} a_j b_{i-j} \right) X^i.$$

Moreover, $0_{R[X]} = 0_R$ and $1_{R[X]} = 1_R$. In fact, $R$ can be viewed as constant polynomials in $R[X]$.

Of course, we expect some common sense here. In the multiplication, $a_i = 0$ whenever $i > \deg f = n$ and $b_j = 0$ whenever $j > \deg g = m$.

One useful fact about polynomial rings is that they have long division. That is, you can divide one polynomial into another and get a quotient-remainder form just like integers. There is one caveat: the dividing polynomial has to be monic!

**Theorem 1.2** (*Long division*)**.** Let $R$ be a ring. Suppose that $f, g \in R[X]$ such that $g$ is **monic**. Then there exist polynomials $q, r \in R[X]$ such that

$$f = g \cdot q + r,$$

with $\deg r < \deg g$ or $r = 0_R$.

**Proof.** Suppose that $\deg f = n$ and $\deg g = m$. Then we may write

$$f(X) = \sum_{i=0}^{n} a_i X^i, \qquad g(X) = \sum_{i=0}^{m} b_i X^i,$$

we view $g$ as fixed throughout this proof. In particular $\deg g = m$ is fixed.

where the leading coefficients $a_n$ and $b_m$ are nonzero.

If $f = 0_R$, then just take $q = r = 0_R$ and we are done. So assume that $f \neq 0_R$. If $n < m$, then take $q = 0$ and $r = f$ and we are done. So assume that $n \geqslant m$. We will now proceed by (strong) induction on $n$ given this assumption.

**Base case.** If $n = 0$. Then $f(X) = \alpha$ and $g(X) = \beta$ for some $\alpha, \beta \in R$. But we assumed that $g$ is monic, so $\beta = 1_R$. So taking $q = \alpha$ and $r = 0$, we are done.

**Inductive hypothesis.** Suppose the result is true for all degrees strictly less than $n$ i.e. it is true for all polynomials $h \in R[X]$ such that $\deg h < n$.

That is, such a $h$ can be "long divided".

**Inductive step.** Let $\lambda = a_n X^{m-n}$ and consider the polynomial

$$f_1 = f - \lambda g.$$

Since $g$ is monic, we are able to kill the $X^n$ coefficient; and so $\deg f_1 < n$. Since $\deg f_1 < n$, the inductive hypothesis tells us that there exist $q_1, r_1 \in R[X]$ such that

Convince yourself that this is true. Just write it on paper.

$$f_1 = q_1 g + r_1,$$

with $\deg r_1 < \deg g$ or $r_1 = 0$. But then

$$f = \lambda g + f_1 = \lambda g + (q_1 g + r_1) = (\lambda + q_1) g + r_1.$$

So taking $q = \lambda + q_1$ and $r = r_1$ we are done. ■

By relaxing the condition that $f, g \in K[X]$ where $K$ is a field, we can get the theorem without needing $g$ to be monic. If we assumed that $R$ is an integral domain[2], then the polynomials $q, r$ are unique.

## 1.2 Homomorphisms and ideals

Just like in groups, we have ring homomorphisms which we will similarly use to characterise when two rings are "equivalent". Moreover, a ring homomorphism reduces to a group homomorphism of abelian groups.

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

(i). $\phi(x+y) = \phi(x) + \phi(y)$,

(ii). $\phi(x \cdot y) = \phi(x) \cdot \phi(y)$,

(iii). $\phi(1_R) = 1_S$.

If $\phi$ is bijective, we call it a (ring) **isomorphism**.

that is, $\phi$ is a homomorphism of abelian groups.

**Remark.** Note that condition (i) implies that $\phi(0_R) = 0_S$.

**Remark.** If $S \leqslant R$, then the inclusion map $\iota : S \to R$ is a ring homomorphism.

**Definition** (*Ring algebra*)**.** If $\phi : R \to S$ is a homomorphism of rings, then we say that $S$ is an **$R$-algebra.**

We will meet ring algebras again properly after we study modules. Just defining it this way does not do it justice, there is a nice structure that we can put on $S$ using the homomorphism $\phi$.

**Definition** (*Kernel*)**.** The **kernel** of a homomorphism $\phi : R \to S$, denoted $\ker \phi$, is the set $\{r \in R \mid \phi(r) = 0\}$.

As expected from groups, the kernel tells us when a homomoprhism is injective.

**Proposition 1.2.** A homomorphism $\phi : R \to S$ is injective *if and only if* $\ker \phi = \{0_R\}$.

**Proof.** The forward direction is obvious so it remains to prove the converse. Assume $\ker \phi = \{0_R\}$. Now suppose $\phi(r) = \phi(s)$ for some $r, s \in R$. Then $\phi(r) - \phi(s) = 0_R$ and hence $\phi(r - s) = 0_R$. By hypothesis, $r - s = 0_R$ which implies $r = s$. ■

**Definition** (*Image*)**.** The **image** of a homomorphism $\phi : R \to S$, denoted im $\phi$, is the set $\{s \in S \mid s = \phi(r) \text{ for some } r \in R\}$.

We will see that kernels and images play an important role. But let us first talk about *ideals*, a special subset of a ring.

**Definition** (*Ideal*)**.** Let $(R, +, \cdot)$ be a ring. We say that $I \subseteq R$ is an **ideal** if

1. $(I, +)$ is a subgroup of $(R, +)$,

Apparently, there is a semantic conflict between "ideals of" and "ideals in" a ring. We're Asian and we're not native speakers of English. The point is, we don't care. So we will just write these interchangeably, favouring the first (since it makes more sense to us).

[2] *we will encounter what this is later on.*

2. For all $r \in R$ and all $x \in I$, $rx \in I$.

We then write $I \lhd R$. We say $I$ is a **proper ideal** if $I \neq R$.

Two immediate candidates of ideal in any ring should pass your mind now.

**Lemma 1.1.** Let $R$ be a ring. Then $R$ and $\{0\}$ are ideals of $R$.

**Proof.** This is too easy. ■

From now on, we will call $\{0\}$ the **zero ideal** and $R$ the **unit ideal** in any ring $R$. The reason $R$ is called the unit ideal may be unclear for now. But we will see that it is because of the fact that it is the ideal "generated" by the unit element $1_R$. We may also write 0 instead of $\{0\}$ for the zero ideal. Forgive us for this.

**Remark** (*Ideals vs subrings*)**.** An obvious question to ask is what are the differences between ideals and subrings? Well, ideals satisfy a stronger condition.

- An ideal is closed under multiplication of elements in the ideal with **any** elements in the ring.
- A subring is closed under multiplication of elements only in the subring.

It is tempting to say that ideals are subrings, but they are not; at least most of the time. This has to do with us insisting that $1_R$ is part of the ring axioms. If we remove this condition, then we get that every ideal are subrings for free.

**Slogan.** Ideals are almost never subrings (at least in our case).

It turns out that they are subrings whenever they are **not** proper.

**Lemma 1.2.** Every proper ideal $I \lhd R$ does not contain $1_R$.

**Proof.** Suppose $1_R \in I$. Then for any $r \in R$, we have $1_R \cdot r = r \in I$. 何! ■

Let us look at some examples.

**Example.** Fix $n \in \mathbb{Z}$. Then $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

We know that $n\mathbb{Z}$ is an additive subgroup of $\mathbb{Z}$. Now consider $r \in \mathbb{Z}$ and $x \in n\mathbb{Z}$.

$$r \cdot x = r \cdot (nz) = n(r \cdot z) \in n\mathbb{Z}$$

for some $z \in \mathbb{Z}$. So indeed $n\mathbb{Z} \lhd \mathbb{Z}$.

**Example.** $\mathbb{Z}$ is a subring of $\mathbb{R}$, but not an ideal. Why? Consider the unit element $1 \in \mathbb{Z}$ which is also the unit element in $\mathbb{R}$. Now is $1 \cdot \pi \in \mathbb{Z}$?

**Example.** Let $R$ be the ring of infinitely differentiable functions $\mathbb{R}^n \to \mathbb{R}$. The set of functions vanishing on some fixed subset of $\mathbb{R}^n$ is an ideal of $R$.

**Proposition 1.3.** Every proper ideal $I \lhd R$ does not contain units of $R$.

**Proof.** Suppose $I \lhd R$ is proper and let $u \in I$ be a unit. By definition of unit, there exists $v \in R$ such that $u \cdot v = 1_R$. By definition of an ideal, $u \cdot v \in I$ and so $1_R \in I$. 何! This contradicts Lemma 1.2. ■

The following slogan is thus worth remembering.

**Slogan.** $\exists \, \text{unit}_R \in I \iff I = R$.

Let us take a detour and consider the ideals of $\mathbb{Z}$.

**Proposition 1.4.** $I$ is an ideal of $\mathbb{Z}$ *if and only if* $I = n\mathbb{Z}$ for some $n \in \mathbb{Z}$.

The only ideals of $\mathbb{Z}$ are $n\mathbb{Z}$.

If we believe that every subgroup of $(\mathbb{Z}, +)$ is of the form $n\mathbb{Z}$, we are done. Let's take a more direct approach by considering a special arsenal in $\mathbb{Z}$, the Euclidean algorithm.

**Proof.** We have shown that $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$, so it remains to prove that they give all the ideals. We may assume that $I \neq \{0\}$ for otherwise, $I = 0\mathbb{Z}$ is of the form desired with $n = 0$, and we are done. Since $I \neq \{0\}$ , then $I$ must contain at least one nonzero element, say, $x$. Since $I$ is an additive subgroup, $-x \in I$. So, $I$ has at least one positive and one negative element. Let $n$ be the smallest[3] positive element in $I$. We now claim that $I = n\mathbb{Z}$.

Since $n \in I$, we get for free that $nx \in I$ for all $x \in \mathbb{Z}$ and so $n\mathbb{Z} \subseteq I$.

Now, let $y \in I$. We want to show that $y \in n\mathbb{Z}$. By the Euclidean algorithm, there exist $q, r \in \mathbb{Z}$ such that $y = n \cdot q + r$ where $r < n$ or $r = 0$. Since $n \in I$ and $q \in \mathbb{Z}$, $q \cdot n \in I$ and so $r = y - n \cdot q \in I$. But, we assumed that $n$ is the smallest positive element in $I$. So[4] $r = 0$, and we are done as $y = n \cdot q \in n\mathbb{Z}$. ■

It seems that we can generalize this idea of ideals being generated by one element.

**Definition** (*Generated ideal*)**.** Let $R$ be a ring and $a \in R$. The **(principal) ideal generated by** $a$, denoted $(a)$ is the set

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

If $a_1, \ldots, a_k \in R$, then the **ideal generated by** $a_1, \ldots, a_k$ is the set

$$(a_1, \ldots, a_k) = \left\{ \sum_{i=1}^{k} a_i \cdot r_i \mid r_i \in R \right\}.$$

Of course, we have to check that it is in fact, an ideal.

**Lemma 1.3.** Let $R$ be a ring and $a \in R$. Then $(a)$ is an ideal.

**Proof.** Let $(R, +, \cdot)$ be a ring and $a \in R$. We want to first show that $((a), +)$ is a subgroup of $(R, +)$. We will use the two-step subgroup criterion. Firstly, $(a)$ is non-empty as $0_R = 0_R \cdot a \in (a)$.

**Step 1.** Let $x, y \in (a)$. Then $x = a \cdot r$ and $y = a \cdot s$ for some $r, s \in R$. And so $x + y = a \cdot r + a \cdot s = a \cdot (r + s) \in (a)$.

[3] *smallest is guaranteed by the well-ordering principle.*

[4] *note that we cannot conclude this before showing that $r \in I$, as we use the fact that $n$ is smallest in $I$.*

**Step 2.** Now, suppose $x \in (a)$. Then $x = a \cdot r$ for some $r \in R$. By definition of a ring, $-r \in R$. So, by associativity of $\cdot$ we have $-x = -(a \cdot r) = a \cdot (-r) \in (a)$. By the two-step subgroup criterion, we are done.

Finally, we check the second ideal axiom. Let $r \in R$ and $x \in (a)$. Then $x = a \cdot s$ for some $s \in R$. Now, observe that $r \cdot x = r \cdot (a \cdot s) = (r \cdot s) \cdot a \in (a)$. ■

It is an easy exercise to verify that the ideal generated by finitely many elements of $R$ is an ideal. Just copy the arguments above.

**Definition** (*Principal ideals*)**.** Let $R$ be a ring. We say an ideal $I \lhd R$ is a **principal ideal** if $I = (a)$ for some $a \in R$.

As you can guess by now, principal ideals are quite analogous to cyclic subgroups. In some sense, they have the same feel.

**Remark.** Equivalently, $I$ is the set of all multiples of $a$ or further equivalently the set of all things in $R$ that can be divisible by $a$.

Principal is just a fancy way to really say:

**Slogan.** Principal $\implies$ generated by a single element.

Thinking of rings with ideals that are all principal is easy. We have shown that $\mathbb{Z}$ contains only principal ideals. It is not too hard see hat this is also true for $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$. Here is an example of a ring with an ideal that is not principal.

**Example.** The ideal $(2, X) \lhd \mathbb{Z}[X]$ is not principal.

*Proof.* Suppose that $I = (2, X)$ is principal. Then there exists $f \in \mathbb{Z}[X]$ such that $I = (f)$. Since $2 \in I$, then there exists $g \in \mathbb{Z}[X]$ such that $2 = fg$. Considering their degrees, we have that

$$0 = \deg 2 = \deg(fg) = \deg f + \deg g.$$

Since degree of a nonzero polynomial is positive or zero, then $\deg f = \deg g = 0$. This implies that $f, g$ are constant polynomials, say, $f = a$ and $g = b$ for some $a, b \in \mathbb{Z}$. Thus, we may now write $2 = ab$. The only solutions (in $\mathbb{Z}$) to this equation are the pairs $(a, b) = (1, 2), (-1, -2), (2, 1), (-2, -1)$.

In the case that $(a, b) = (1, 2), (-1, -2)$. We have $a = \pm 1$ and so $f = \pm 1$. Since we assumed $I$ is generated by $f$, we have $I = (1)$. This implies that there exists $r, s \in \mathbb{Z}[X]$ such that $2r(X) + Xs(X) = 1$. Looking at constants of $r$ and $s$ (by evaluation at $X = 0$), we have $2r(0) = 1$ and so $r(0) = 1/2$ contradicting the fact that $r(0) \in \mathbb{Z}$. 何!

On the other hand, if $(a, b) = (2, 1), (-2, -1)$, we have $I = (2)$. So there exist $r, s \in \mathbb{Z}[X]$ such that $2r(X) + Xs(X) = 2$; and so $Xs(X) = 2t(X)$ where $t(X) = 1 - r(X)$. This implies that $t(X) = Xs(X)/2$ contradicting the fact that $1, r(X) \in \mathbb{Z}[X]$. 何!

An important result is the following.

**Lemma 1.4.** Let $\phi : R \to S$ be a homomorphism of rings. Then $\ker \phi \lhd R$.

**Proof.** Since $\phi : R \to S$ is a ring homomorphism, it is a group homomorphism in addition; and we know that kernel of a group homomorphism is a subgroup. So, it remains to show the second ideal axiom.

Suppose $r \in R$ and $x \in \ker \phi$. We want to show that $r \cdot x \in \ker \phi$. This is easy:

$$\phi(r \cdot x) = \phi(r) \cdot \phi(x) = \phi(r) \cdot 0_S = 0_S.$$

We are done. Note that there are two different multiplications going on here. ■

The reason we define ideals the way they are defined is so that we can quotient out by them, just like how we can quotient out normal subgroups. If $R$ is a ring, and $I \lhd R$, the cosets of $(I, +)$ are subsets of the form

$$a + I = \{a + x \mid x \in I\}$$

where $a \in R$. From group theory we know that the set of cosets $R/I = \{a + I \mid a \in R\}$ forms a group under coset addition

when we write $R/I$ here, we view $R, I$ all as groups under addition.

$$(a + I) + (b + I) = (a + b) + I,$$

because $(R, +)$ is abelian and so $I$ is normal. It turns out, we can make it into a ring.

**Definition** (*Quotient ring*)**.** Let $R$ be a ring and $I \lhd R$. The **quotient ring** $R$ modulo $I$, denoted $R/I$, is the set of cosets $\{a + I \mid a \in R\}$ with coset addition (as defined above) and coset multiplication:

$$(a + I) \cdot (b + I) = a \cdot b + I,$$

and $0_{R/I} = 0_R + I$ and $1_{R/I} = 1_R + I$.

**Remark** (Some notation)**.** Let $R$ be a ring and $I \lhd R$. Let $a, b \in R$. We write $a \equiv b \pmod{I}$ if $a = b + i$ for some $i \in I$; or equivalently if $a - b \in I$. If moreover the ideal $I$ is principal, say $I = (r)$ for some $r \in R$, we can get even lazier by simply writing $a \equiv b \pmod{r}$ instead of $a \equiv b \pmod{(r)}$.

This is consistent with congruence relation in the integers: $a \equiv b \pmod{n}$ (should be mod $n\mathbb{Z}$ but this is principal generated by $n$) means $a = b + nz$ for some $z \in \mathbb{Z}$; and $nz \in n\mathbb{Z}$.

We'll have to show that this definition of a ring is well-defined. In particular, we want to verify this new of way of multiplication.

**Theorem 1.3.** Let $R$ be a ring and $I \lhd R$. Then $R/I$ is a ring.

**Proof.** As discussed earlier, the group $(R/I, +)$ where $+$ is coset addition is well-defined as $(R, +)$ is abelian and hence $(I, +)$ is normal. It remains to show that coset multiplication is well-defined. That is, if $a + I = a' + I$ and $b + I = b' + I$, then $ab + I = a'b' + I$.

Suppose $a + I = a' + I$ and $b + I = b' + I$. Then $a = a' + j$ and $b = b' + k$ for some $j, k \in I$. Now, observe that

It is understood that $a, a', b, b' \in R$ by definition of the set of cosets.

$$ab + I = (a' + j)(b' + k) + I = (a'b' + a'k + jb' + jk) + I.$$

But $a'k, jb', jk \in I$ by the second ideal axiom. So, in fact $ab + I = a'b' + I$ as desired.

It is an easy exercise to check that $0_{R/I} = 0_R + I$ and $1_{R/I} = 1_R + I$. ■

Observe that there is a natural map $R \to R/I$ which has the same feel of an inclusion map. This map is simply the map that take elements to its coset representative.

**Definition** (*Projection map*). Let $R$ be a ring and $I \lhd R$. Then the map $\pi : R \to R/I$ defined by $\pi(a) = a + I$ is called the **projection map**.

This map is sometimes known as the *canonical* or *natural* map.

One important fact about this map is the following.

**Proposition 1.5.** The projection map $\pi : R \to R/I$ is a surjective homomorphism with $\ker \pi = I$.

**Proof.** Checking that $\pi$ is a homomorphism is trivial. $\pi$ is surjective as any coset $a+I$ is the image of $a \in R$. Now, if $x \in \ker \pi$, then $\pi(x) = 0_R + I$. But $x + I = 0_R + I$ if and only if $x \equiv 0_R \pmod{I}$ i.e. $x = 0_R + j$ for some $j \in I$. That is, if and only if $x \in I$. ■

**Corollary 1.1.** Every ideal is the kernel of a ring homomorphism, and vice-versa.

Note that we have proved the "vice-versa" part earlier.

Let's try to "extrapolate" a useful result from group theory: the isomorphism theorems.

**Theorem 1.4** (*First isomorphism theorem*). Let $\phi : R \to S$ be a ring homomorphism. Then $R/\ker\phi \cong \operatorname{im}\phi$ by the map $a + \ker\phi \longmapsto \phi(a)$ induced by $\phi$.

The important conclusion is that, $R/\ker\phi \cong \operatorname{im}\phi$.

This theorem is possibly the most spammed theorem in this course.

**Proof.** Consider the map $\Phi : R/\ker\phi \to \operatorname{im}\phi$ defined by $\Phi(a + \ker\phi) = \phi(a)$. Let's first check that it is well-defined. If $a + \ker\phi = b + \ker\phi$, then $a \equiv b \pmod{\ker\phi}$ i.e. $a - b \in \ker\phi$ and so $\phi(a - b) = 0_R$. This implies that

$$\Phi(a + \ker\phi) = \phi(a) = \phi(b) = \Phi(b + \ker\phi),$$

proving that $\Phi$ is **well-defined**. It is easy to check that $\Phi$ is a **homomorphism**. To check that $\Phi$ is **injective**, consider $a + \ker\phi \in \ker\Phi$. Then $\phi(a) = \Phi(a + \ker\phi) = 0_S$. This implies $a = a - 0_R \in \ker\phi$, i.e. $a \equiv 0_R \pmod{\ker\phi}$. Thus, $\ker\Phi = \{0_R + \ker\phi\} = \{0_{R/I}\}$ and so $\Phi$ is injective. Finally, $\Phi$ is **surjective** since any $\phi(a) \in \operatorname{im}\phi$ is the image of some $a + \ker\phi \in R/\ker\phi$. ■

Alternative way to prove injectivity. Suppose $\Phi(a + \ker\phi) = \Phi(b + \ker\phi)$. Then $\phi(a) = \phi(b)$ i.e. $\phi(a - b) = 0_R$. So, $a - b \in \ker\phi$. This implies that $a \equiv b \pmod{\ker\phi}$ i.e. $a + \ker\phi = b + \ker\phi$.

**Theorem 1.5** (*Correspondence theorem*). Let $R$ be a ring and $I \lhd R$. Let $\pi : R \to R/I$ be the projection map. Then there is a bijective correspondence:

$$\{\text{ideals of } R \text{ that contain } I\} \longleftrightarrow \{\text{ideals of } R/I\},$$

given by $J \longmapsto \pi^{\text{img}}(J)$ and $\pi^{\text{preimg}}(\mathfrak{J}) \longleftarrow \mathfrak{J}$.

Note the arrows' direction carefully.

This theorem, which is also known as the *fourth isomorphism theorem*, gives a super nice correspondence between ideals of $R$ and $R/I$. Usually, it is harder to talk about ideals of $R/I$. But this theorem tells us that it suffices to talk about ideals of $R$ that contain $I$; and whatever result that holds for it, should hold for ideals of $R/I$. Mindblowing!

**Proof.** Let $\pi : R \to R/I$ be the projection map $a \mapsto a + I$. We first refresh our memory regarding set theory.

Let $J \lhd R$. The direct image of $J$ under $\pi$ is

$$\pi^{\text{img}}(J) = \{\pi(j) \mid j \in J\} = \{j + I \mid j \in J\} \subseteq R/I.$$

Let $\mathfrak{J} \lhd R/I$. The preimage of $\mathfrak{J}$ under $\pi$ is

$$\pi^{\text{preimg}}(\mathfrak{J}) = \{r \in R \mid \pi(r) \in \mathfrak{J}\} \subseteq R.$$

We will first prove two claims that basically say that the two maps are well-defined.

**Claim 1: If $J \lhd R$, then $\pi^{\text{img}}(J) \lhd R/I$.**

*Proof.* We will use the two-step subgroup criterion. Since $0_R \in J$, then $\pi^{\text{img}}(J)$ is non-empty as $0_R + I = \pi(0_R) \in \pi^{\text{img}}(J)$. Now suppose $j + I, k + I \in \pi^{\text{img}}(J)$ where $j, k \in J$. Then $j + k \in J$ and so

$$(j + I) + (k + I) = (j + k) + I \in \pi^{\text{img}}(J).$$

For the second step, take $j + I \in \pi^{\text{img}}(J)$. Since $j \in J$, then $-j \in J$ and so

$$-(j + I) = -j + I \in J.$$

Thus, $\pi^{\text{img}}(J)$ is an additive subgroup of $R/I$. We now have to check that it satisfies the second ideal axiom. Suppose $j + I \in \pi^{\text{img}}(J)$ and $a + I \in R/I$. Since $J \lhd R$ and $a \in R$, then $ja \in J$ and so

$$(j + I)(a + I) = ja + I \in \pi^{\text{img}}(J). \quad \blacksquare$$

**Claim 2: If $\mathfrak{J} \lhd R/I$, then $\pi^{\text{preimg}}(\mathfrak{J}) \lhd R$.**

*Proof.* Again, use the two-step subgroup criterion. Since $0_R + I \in \mathfrak{J}$, then $0_R = \pi^{-1}(0_R+I) \in \pi^{\text{preimg}}(J)$ so the preimage is non-empty. Now suppose $r, s \in \pi^{\text{preimg}}(\mathfrak{J})$. Then $\pi(r), \pi(s) \in \mathfrak{J}$ and so

$$\pi(r + s) = \pi(r) + \pi(s) \in \mathfrak{J}.$$

where we have used the fact that $\pi$ is a homomorphism. We need not to though as it was sufficient to see from definitions: $\pi(r + s) = (r + s) + I = (r + I) + (s + I) \in I$.

This implies that $r + s \in \pi^{\text{preimg}}(\mathfrak{J})$. For the second step, take $a \in \pi^{\text{preimg}}(\mathfrak{J})$. Then $-a \in \pi^{\text{preimg}}(\mathfrak{J})$ since

$$\pi(-a) = -a + I = -(a + I) \in \mathfrak{J}.$$

Thus, $\pi^{\text{preimg}}(\mathfrak{J})$ is an additive subgroup of $R$. It remains to prove that is satisfies the second ideal axiom. Suppose $j \in \pi^{\text{preimg}}(\mathfrak{J})$ and $a \in R$. Then $\pi(j) \in \mathfrak{J}$. So,

$$\pi(aj) = aj + I = (a + I)(j + I) \in \mathfrak{J},$$

since $a + I \in R/I$ and $j + I \in \mathfrak{J}$, then the product is in $\mathfrak{J}$ because it is an ideal of $R/I$.

and thus, $aj \in \pi^{\text{preimg}}(\mathfrak{J})$. $\blacksquare$

We finally show that the maps are in fact bijections.

**Claim 3: The correspondence maps $J \longmapsto \pi^{\text{img}}(J)$ and $\pi^{\text{preimg}}(\mathfrak{J}) \longleftarrow\!\shortmid \mathfrak{J}$ are inverse maps of each other, and hence are bijections.**

*Proof.* We want to show that $\pi^{\text{img}}(\pi^{\text{preimg}}(\mathfrak{J})) = \mathfrak{J}$ and $\pi^{\text{preimg}}(\pi^{\text{img}}(J)) = J$ for any $\mathfrak{J} \lhd R/I$ and any $J \lhd R$ which contains $I$.

By elementary set theory, we know that $J \subseteq \pi^{\text{preimg}}(\pi^{\text{img}}(J))$ and $\pi^{\text{img}}(\pi^{\text{preimg}}(\mathfrak{J})) \subseteq \mathfrak{J}$. Moreover, since $\pi$ is surjective, then $\mathfrak{J} = \pi^{\text{img}}(\pi^{\text{preimg}}(\mathfrak{J}))$. So it remains to show that $\pi^{\text{preimg}}(\pi^{\text{img}}(J)) \subseteq J$.

Let $J \lhd R$ which **contains** $I$ i.e. $I \subseteq J$. Consider $a \in \pi^{\text{preimg}}(\pi^{\text{img}}(J))$. Then $\pi(a) \in \pi^{\text{img}}(J)$. So by definition, there exists $j \in J$ such that $\pi(a) = \pi(j)$. Since $\pi$ is a homomorphism, $\pi(a-j) = 0_R$ and so $a - j \in \ker \pi = I \subseteq J$. Now, because $j \in J$ and we've just shown that $a - j \in J$, then $a = (a-j) + j \in J$, as desired. ■

**Strong note.** Remember that $\pi^{-1}$ isn't (usually) a map. It is just the preimage of a set.

We have shown when we define the projection map that $\ker \pi = I$.

As expected, next comes the second and third isomorphism theorem which is not very useful for us, at least in this notes. We will prove them anyways because they are quite easy.

**Theorem 1.6** (*Second isomorphism theorem*)**.** Let $R$ be a ring, $S \leqslant R$ and $J \lhd R$. Then we have that $J \lhd (S+J) \leqslant R$. Moreover, $S \cap J \lhd S$ and

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

**Proof.** We have to prove 4 things, the last of them depending on the preceding two.

**Claim 1:** $(S+J) \leqslant R$.

*Proof.* We will use the subring criterion. $S + J$ is non-empty as $0_{S+J} = 0_S + 0_J = 0_R + 0_R = 0_R \in S + J$. If $s + j, s' + j' \in S + J$, then

$$(s+j) - (s'+j') = (s-s') + (j-j') \in S+J, \text{ and}$$

$$(s+j)(s'+j') = ss' + (sj' + js' + jj') \in S + J$$

Since $S, J \subseteq R$, we have $sj', js', jj' \in J$ because $J \lhd R$.

Finally, $1_{S+J} = 1_S$, so $S + J$ is a subring of $R$. ■

**Claim 2:** $J \lhd (S+J)$.

*Proof.* $J$ is trivially an additive subgroup of $S + J$. To check the second ideal axiom, just take $j \in J$ and $s + j' \in S + J$. Then $j(s+j') = js + jj' \in J$ because $J \lhd R$ and so $js, jj' \in J$. ■

by identifying $J$ with $0 + J = \{0_S + j : j \in J\}$.

**Claim 3:** $S \cap J \lhd S$.

*Proof.* Since $0_R \in S, J$, then $0_R \in S \cap J$ so $S \cap J$ is non-empty. If $a, b \in S \cap J$, then $a \in S, J$ and $b \in S, J$. Since $S \leqslant R$ and $J \lhd R$, so $-b \in S, J$. It follows that $a - b \in S \cap J$. Additive subgroup check done. Let's check the second ideal axiom: suppose $a \in S \cap J$ and $s \in S$. Then $a \in S, J$. It follows that $as \in S$ since $S \leqslant R$ and $as \in J$ since $J \lhd R$. That is, $as \in S \cap J$. ■

**Claim 4: The isomorphism.**

Idea: Find a surjective homomorphism $S \to (S+J)/J$ whose kernel is $J$.

*Proof.* Consider the map $\Phi : S \to (S+J)/J$ defined by $\Phi(s) = s + J$. Checking that $\Phi$ is a **homomorphism** is easy.

Let us check that $\Phi$ is **surjective**. Let $(s+j) + J \in (S+J)/J$. Since $s \in S$, we have $\Phi(s) = s + J$. But $s \equiv s + j \pmod{J}$ (because $j \in J$) i.e. $s + J = (s+j) + J$. So, in fact $\Phi(s) = (s+j)+J$. So we have found our element in $S$ whose image is $(s+j)+J$, and this was arbitrary; thus, $\operatorname{im}\Phi = (S+J)/J$.

Finally, let us consider the **kernel**. We have

$$\begin{aligned}\ker\Phi &= \{s \in S \mid \Phi(s) = 0_S + J\}\\ &= \{s \in S \mid s + J = 0_S + J\}\\ &= \{s \in S \mid s \in J\}\\ &= S \cap J.\end{aligned}$$

By the first isomorphism theorem, we are done. ■

**Theorem 1.7** (*Third isomorphism theorem*)**.** Let $R$ be a ring and $I, J \lhd R$ such that $I \subseteq J$. Then $J/I \lhd R/I$ and

$$\frac{R/I}{J/I} \cong R/J.$$

**Proof.** It is easy to check that $J/I \lhd R/I$: just run the two-step subgroup criterion and since $J \lhd R$, we get the second ideal axiom almost for free.

Idea of the proof: use the first isomorphism theorem. That is, find a surjective homomorphism $R/I \to R/J$ whose kernel is $J/I$.

It is easy to verify that $I \lhd J$ so modding out $J$ by $I$ is fine.

Consider the map $\Phi : R/I \to R/J$ defined by $\Phi(a+I) = a+J$. Let us check that this map is indeed **well-defined**. Suppose $a+I = b+I$. Then $a-b \in I \subseteq J$. So, $\Phi(a+I) = a+J = b+J = \Phi(b+I)$, awesome.

This map is **surjective** since every $a+J \in R/J$ is the image of $a+I \in R/I$.

Finally, let us consider the **kernel**. Suppose $a+I \in \ker\Phi$. This is true iff $a+J = \Phi(a+I) = 0_R + J$ i.e. if and only if $a = a - 0_R \in J$. So, $\ker\Phi = J/I$. The first isomorphism theorem completes the proof. ■

## 1.3 Domains, maximal and prime ideals

We will start with two definitions.

**Definition** (*Zero divisor*)**.** Let $R$ be a ring and $z \in R$. We say $z$ is a **zero divisor** if $z \neq 0_R$ and there exists $r \in R$ with $r \neq 0_R$ such that $z \cdot r = 0_R$.

**Definition** (*Integral domain*)**.** Let $R$ be a nonzero ring. We say $R$ is an **integral domain** if $r \cdot s = 0_R$ implies either $r = 0_R$ or $s = 0_R$ for any $r, s \in R$.

Equivalently, $R$ is an integral domain iff every element of $R$ is not a zero divisor. One nice property of integral domains is the following.

**Lemma 1.5** (*Cancellation law*)**.** Let $R$ be an integral domain and $r, s, t \in R$. If $r \cdot s = r \cdot t$, and that $r \neq 0_R$, then $s = t$.

**Proof.** If $r \cdot s = r \cdot t$, then $r \cdot (s-t) = 0_R$. Since $R$ is an integral domain and $r \neq 0_R$, then $s - t = 0_R$. The result follows. ■

So in integral domains, we can do our favorite "dividing on both sides" that we can do in $\mathbb{R}$ and $\mathbb{C}$. But what are these? These are fields.

**Example.** Any field $K$ is an integral domain.

*Proof.* Let $r, s \in K$ such that $r \cdot s = 0_K$. WLOG, assume that $r \neq 0_K$. Then there is an inverse $r^{-1}$ so that $r^{-1} \cdot r = 1_K$. Multiplying this by $s$, we have $r^{-1} \cdot (r \cdot s) = s$. But $r \cdot s = 0_R$, so $s = 0_R$.

**Example.** Any subring of a field is an integral domain.

So the tower of number system subrings $\mathbb{C} \leqslant \mathbb{R} \leqslant \mathbb{Q} \leqslant \mathbb{Z}$ are all integral domains.

This definition of integral domain is good, at least, from a number theory point of view. As we will see later, it allows us to talk about factorization more generally: think fundamental theorem of arithmetic but now in a general setting. This is nice as the FTA is a big building block of number theory.

**Proposition 1.6.** Let $R$ be an integral domain. Then $R[X]$ is an integral domain.

**Proposition 1.7.** Any finite integral domain is a field.

By finite integral domain, we mean it is an integral domain with only finitely many elements.

A nice characterization of fields is given by their ideals.

**Lemma 1.6** (*Ideals of a field*)**.** Let $K$ be nonzero ring.

(i). If $K$ is a field, then its only ideals are $K$ and $\{0\}$.

(ii). If $K$ has exactly two ideals, then $K$ is a field.

**Proof.** (i). Let $K$ be a field. Suppose $I \lhd K$ such that $I \neq \{0\}$. Then there exists $x \in I$ such that $x \neq 0_K$. Since $K$ is a field, $x$ is a unit. But every proper ideal of $K$ does not contain units of $K$, so $I$ must be proper i.e. $I = K$.

(ii). Suppose that $K$ has exactly two ideals. Then these ideals must be the unit ideal $K$ and $\{0\}$. Now let $x \in K$ such that $x \neq 0_K$. We want to show that it is a unit. Consider the ideal $I = (x)$. It cannot be $\{0\}$ because $x \in I$. So it must be that $I = K$. This implies that there exist $k \in K$ such that $k \cdot x = 1_K$. By definition $x$ is thus a unit. Since $x$ was arbitary, we are done. ■

We thus have the following handy slogan.

**Slogan.** $K$ is a field $\iff$ its (only) ideals are $K$ and $\{0\}$.

Remember the Correspondence Theorem? It says that there is a relation between ideals of $R$ containing $I \lhd R$ and the ideals of $R/I$. We proved that if $K$ is a field, then it has exactly two ideals. So if $K = R/I$ is a field, then it has exactly two ideals $R/I$ and $\{0_R\}$. The Correspondence Theorem then tells us that there are exactly two ideals of $R$ containing $I$. These ideals are precisely $I$ and $R$. Such ideals are special, so we give it a name.

**Definition** (*Maximal ideal*)**.** An ideal $\mathfrak{m} \lhd R$ is **maximal** if $\mathfrak{m} \neq R$ but $\mathfrak{m}$ is not contained in any other ideals other than $\mathfrak{m}$ and $R$.

In other words, if $J \lhd R$ is an ideal such that $\mathfrak{m} \subseteq J$, then either $J = \mathfrak{m}$ or $J = R$.

The name *maximal* should not be odd. The definition really tells us that an ideal is *maximal* if it is maximal w.r.t set inclusion within a set containing all/some proper ideals. The motivation of our definition for maximal ideals immediately gives the following result.

**Theorem 1.8.** Let $\mathfrak{m} \lhd R$ be an ideal. Then

$$\mathfrak{m} \text{ is maximal} \iff R/\mathfrak{m} \text{ is a field.}$$

**Proof.** Suppose $R/\mathfrak{m}$ is a field. Then it has exactly two ideals $R/\mathfrak{m}$ and $\{0\}$. By the correspondence theorem, $\pi^{-1}(R/\mathfrak{m}) = R$ and $\pi^{-1}(\{0_{R/\mathfrak{m}}\}) = \mathfrak{m}$ are the only ideals of $R$ containing $\mathfrak{m}$. This is exactly the definition of $\mathfrak{m}$ being maximal. By noting that the argument works backwards as well, we are done. ■

Recall that $0_{R/\mathfrak{m}} = 0_R + \mathfrak{m}$. The preimage of the set containing $0_R/\mathfrak{m}$ is thus everything in $\mathfrak{m}$ i.e. $\mathfrak{m}$.

**Proposition 1.8.** The zero ideal of $R$ is maximal $\iff$ $R$ is a field.

**Proof.** If $R$ is a field, then its only ideals are $\{0\}$ and $R$. So necessarily $\{0\}$ is not contained in any other ideals and hence is maximal.

Conversely, if $\{0\}$ is maximal. Then it is only contained in the ideals $\{0\}$ and $R$ by definition. But since $\{0\}$ is a subset of all ideals, this implies that all the ideals of $R$ are in fact only $\{0\}$ and $R$. So $R$ must be a field. ■

Remember that an ideal is an additive subgroup of the ring so it necessarily contains 0.

Let us give one more definition which will be useful.

**Definition** (*Prime ideal*)**.** An ideal $\mathfrak{p} \lhd R$ is **prime** if $\mathfrak{p} \neq R$ and if $r, s \in R$ is such that $r \cdot s \in \mathfrak{p}$, then either $r \in \mathfrak{p}$ or $s \in \mathfrak{p}$.

**Notation.** Note that we have used $\mathfrak{m}$ for maximal ideals and $\mathfrak{p}$ for prime ideals. These are the fraktur $m$ and $p$ respectively; and it is a convention that maximal and prime ideals are written using fraktur symbols. We will follow this convention.

Note that prime ideals is not a generalization of the prime numbers. However, the two idea are connected. One way to see this connection is the following.

**Proposition 1.9.** A nonzero ideal $n\mathbb{Z} \lhd \mathbb{Z}$ is prime if and only if $n$ is prime.

We have a similar (sounding) result for maximal ideals.

**Theorem 1.9.** Let $\mathfrak{p} \lhd R$ be an ideal. Then

$$\mathfrak{p} \text{ is prime} \iff R/\mathfrak{p} \text{ is an integral domain.}$$

**Proof.** ($\implies$). Let $\mathfrak{p} \lhd R$ be prime. Now suppose $a + \mathfrak{p}, b + \mathfrak{p} \in R/\mathfrak{p}$ is such that

$$(a + \mathfrak{p}) \cdot (b + \mathfrak{p}) = 0_{R/\mathfrak{p}} = 0_R + \mathfrak{p}.$$

Let us evaluate the lhs. The product gives $a \cdot b + \mathfrak{p}$, and so $ab \equiv 0 \pmod{\mathfrak{p}}$ i.e. $ab \in \mathfrak{p}$. But $\mathfrak{p}$ is prime, so we have that either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. This implies that either $a = 0_R + \mathfrak{p}$ or $b = 0_R + \mathfrak{p}$, as desired.

($\impliedby$). Suppose $R/\mathfrak{p}$ is an integral domain. This just means that if $a \cdot b \equiv 0 \pmod{\mathfrak{p}}$, then either $a = 0_R + \mathfrak{p}$ or $b = 0_R + \mathfrak{p}$. But unpacking this definition, this is precsiely saying that if $ab \in \mathfrak{p}$, then either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. So $\mathfrak{p}$ is prime. ■

The standard ID definition gives $(a + \mathfrak{p}) \cdot (b + \mathfrak{p}) = 0_R + \mathfrak{p}$ implies that $a + \mathfrak{p} = 0_R + \mathfrak{p}$ or $b + \mathfrak{p} = 0_R + \mathfrak{p}$. But $(a + \mathfrak{p}) \cdot (b + \mathfrak{p}) = a \cdot b + \mathfrak{p}$. And remember our notation that $x + \mathfrak{p} = y + \mathfrak{p}$ iff $x \equiv y \pmod{\mathfrak{p}}$ iff $x - y \in \mathfrak{p}$.

This is a really nice theorem as we will talk about prime ideals a lot later on. We will mostly use the forward direction of the theorem; that is an ideal is prime that quotienting out by it gives a domain.

**Corollary 1.2.** Every maximal ideal is prime.

**Proof.** Let $\mathfrak{m} \lhd R$. If $\mathfrak{m}$ is maximal then $R/\mathfrak{m}$ is a field, which is an integral domain. So $\mathfrak{m}$ is prime. ■

The converse is not true in general. But they are true when every ideal is principal in the ring. More generally, they are true when there is "unique factorization". We will see these results later on.

**Proposition 1.10.** The zero ideal of $R$ is prime $\iff$ $R$ is an integral domain.

**Proof.** The map $\phi : R \mapsto R/0$ defined by $r \mapsto r + 0_R$ is an isomorphism. Now apply Theorem 1.9 with $\mathfrak{p} = 0$. ■

### 1.3.1 Euclidean and principal ideal domains

We've seen the division algorithm (also known as long division) in at least two places now — our familiar $\mathbb{Z}$ and $R[X]$. It is then tempting to ask if there is a natural generalization of this. With integral domains, there is one.

**Definition** (*Euclidean domain*)**.** An integral domain $R$ is **Euclidean** if there is a map $\phi : R\backslash\{0\} \to \mathbb{N}$ such that for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $\phi(r) < \phi(b)$.

**Remark.** Some remarks on Euclidean domains.

- The definition says, we can divide $b$ into $a$ to get a quotient $q$ and a remainder $r$ and the remainder should be smaller than the divisor.
- We think of the map $\phi$ as a **size function**. That is, it measures the "size" of the element of the ring.

As we shall soon see, this definition is very important as it has very powerful consequences. The following example should not surprise you as this definition was really motivated by the division algorithm on the integers.

**Example.** $\mathbb{Z}$ is Euclidean.

The first step in any Euclidean proof is to always define a size function.

**Size function:** $\phi(n) = |n|$.

*Proof.* Idea: Draw $\mathbb{Z}$ as a subset of the real line, $\mathbb{R}$. This can be done by drawing a line and marking the integers. You will see that $\mathbb{Z}$ tesselates $\mathbb{R}$.

meaning $\mathbb{Z}$ divides $\mathbb{R}$ into non-overlapping unit intervals.

Suppose $a, b \in \mathbb{Z}$ with $b \neq 0$. Consider $x = a/b \in \mathbb{R}$. Then (looking at our drawing), there is a unit interval $[n, n+1]$ such that $x \in [n, n+1]$. So there is an integer $q = n$ or $q = n + 1$ such that

$$|x - q| \leqslant \frac{1}{2} \implies \left|\frac{a}{b} - q\right| \leqslant \frac{1}{2}.$$

Multiplying by $|b|$, we get $|a - bq| \leqslant |b|/2 < |b|$. Taking $r = a - bq$, we have $a = bq + r$ with $\phi(r) < \phi(b)$, as required. ■

Let us consider a special subgroup of the complex numbers $\mathbb{C}$.

**Example.** The ring $\mathbb{Z}[i] = \{x + iy \mid x, y, \in \mathbb{Z}\}$ is Euclidean.

As usual, the first step is to provide a size function.

**Size function:** $\phi(z) = |z|^2$.

*Proof.* Idea: As in the case of $\mathbb{Z}$, draw a good large diagram. You will see that $\mathbb{Z}[i]$ tesselates $\mathbb{C}$ into unit squares. Now in any unit square, if we take a random point $w$, then there is a vertex of the square whose distance from $w$ is at most $1/\sqrt{2}$ .

We'll have to show that $\phi$ works as a size function. Suppose $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. We want to divide $b$ into $a$ to get a quotient $q$ and a remainder $r$ such that $\phi(r) < \phi(b)$.

Consider the complex number $w = a/b$. Then $w$ is inside one of the unit squares (in the diagram). So there is a vertex $v$ of the unit square such that $\operatorname{dist}(v, w) \leqslant 1/\sqrt{2}$. Since $v$ is a vertex, $v \in \mathbb{Z}[i]$. Now take $q = v$. Then

$$\left|\frac{a}{b} - q\right| \leqslant \frac{1}{\sqrt{2}}.$$

If we square this and multiply by $|b|^2$, then we have

$$|a - bq|^2 \leqslant \frac{|b|^2}{2} < |b|^2.$$

That is, $\phi(a - bq) < \phi(b)$. So take $r = a - bq$, then $a = bq + r$ and $\phi(r) < \phi(b)$. ∎

This special ring will hold some interest for us later on. One important observation is that the polynomial rings over a field is always Euclidean.

**Theorem 1.10.** If $K$ is a field, then $K[X]$ is Euclidean.

**Proof.** We will prove that $K[X]$ is Euclidean with the following size function.

**Size function:** $\phi(f) = \deg f$.

We want to show that given $f, g \in K[X]$ with $f \neq 0$, we can write $g = qf + r$, where $q, r \in K[X]$; and either $r = 0$ **or** $\deg r < \deg f$.

Let $f, g \in K[X]$. Then we can write

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

Notice that since $a_n \neq 0$, we have $\deg f = n$.

with $a_i \in K$ and $a_n \neq 0$. Since $K$ **is a field** (and $a_n \neq 0$), we can further write

$$f = a_n \left(a_n^{-1} a_0 + a_n^{-1} a_1 X + \cdots + X^n\right) = a_n f_1.$$

with $f_1 = a_n^{-1}a_0 + a_n^{-1}a_1 X + \cdots + X^n$.

Observe that $f_1$ monic. So long division implies that there exist $q_1, r_1 \in K[X]$ such that $g = q_1 f_1 + r_1$ where $r_1 = 0$ or $\deg r_1 < \deg f_1 = \deg f = n$ . But $f_1 = a_n^{-1} f$ and so

$$g = q_1(a_n^{-1} f) + r_1.$$

So take $q = q_1 a_n^{-1}$ and $r = r_1$, and we are done. ∎

Remember that we are trying really hard to copy properties of $\mathbb{Z}$: we want to do number theory. Since $\mathbb{Z}$ is an integral domain and we have shown that all its ideals are principal, let us talk about rings that has this property.

**Definition** (*PID*)**.** A **principal ideal domain (PID)** is an integral domain in which every ideal is principal.

**Remark.** Here is a recall: *Every ideal is principal* means that for every ideal $I \subseteq R$, there exists $x \in I$ such that $I = (x)$.

**Dangerous bend.** Note that we never claimed that $x$ is unique! We only say that it exists.

**Example.** Any field $K$ is a PID.

*Proof.* If $K$ is a field, then it has exactly two ideals. The zero ideal is generated by 0 and the unit ideal (itself) is generated by 1. ■

It turns out that PIDs are the "**simplest**" rings after fields. We will see why.

**Theorem 1.11.** Every Euclidean domain is a PID.

**Proof.** Suppose $R$ is a Euclidean domain. In particular, there is a size function $\phi : R\backslash\{0\} \to \mathbb{N}$. Further suppose that $I$ is an ideal in $R$. We then have 2 cases.

**Boring case.** Suppose $I = \{0\}$ . Then $I = (0)$, so it is principal.

**Interesting case.** Suppose $I \neq \{0\}$. Now choose $x \in I - \{0\}$ such that $\phi(x)$ is minimal[5]. We now claim that $I = (x)$.

*Proof.* Since $x \in I$, we get $(x) \subseteq I$ for free. Conversely, suppose that $y \in I$. Since $R$ is Euclidean, there exist $q, r \in R$ such that $y = qx + r$ and either $r = 0$ or $\phi(r) < \phi(x)$. Writing in terms of $r$, we have[6] $r = y - qx \in I$. But $\phi(x)$ is minimal, so $\phi(r) < \phi(x)$ is impossible. This forces $r = 0$ and thus $y = qx \in (x)$. We have thus proved $I \subseteq (x)$. ■

**Example.** So $\mathbb{Z}$, $\mathbb{Z}[i]$ are PIDs. Also $K[X]$ whenever $K$ is a field.

When we talk about the many type of rings, it is good to remind ourselves the order of these rings. Right now, we have the following:

$$\text{ring} \supseteq \text{integral domain} \supseteq \text{PID} \supseteq \text{Euclidean domain} \supseteq \text{field}.$$

There's a much more general construction than the PID which is a domain whose ideals are principal. Instead of domain we just talk about general rings. Instead of ideals being generated by a single element, we talk about ideals being generated by finitely many elements. This is the notion of Noetherian rings.

## 1.4 Noetherian rings

Let us first formalize the generalization of principal ideals, called finitely generated ideals.

**Definition** (*f.g. ideals*)**.** Let $R$ be a ring. We say that $I \lhd R$ is **finitely generated** if $I = (a_1, \ldots, a_n)$ for some $a_i \in R$, $1 \leqslant i \leqslant n$.

We have seen what $(a_1, \ldots, a_n)$ looks like when we define generated ideals. Its elements are simply $R$-linear combinations of the $a_i$. Using the notion of finitely generated ideals, we can define Noetherian rings.

[5] *remember that $\phi(x) \in \mathbb{N}$ and any non-empty subset of $\mathbb{N}$ always has a minimal element.*

[6] *since $y \in I$ and $-qx \in I$.*

**Definition** (*Noetherian rings*)**.** A ring $R$ is **Noetherian** if every ideal in $R$ is finitely generated.

**N.B.** Noetherian rings are not integral domains in general! But those that we care about in this course are usually domains.

Noetherian rings are very important as they are more natural (in some sense) than PID. We shall see why when we encounter the Hilbert's basis theorem.

**Example.** Every PID is Noetherian.

*Proof.* Take $n = 1$ in the definition of f.g. ideals. ■

Consequently, every Euclidean domain and fields are Noetherian.

**Example.** $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{Z}[i]$ are Noetherian rings. If $K$ is a field, then it is Noetherian and so is $K[X]$.

There's another approach to the notion of Noetherian rings which depends on a more mysterious formulation.

**Definition** (*Ascending chain*)**.** An **ascending chain** of ideals in a ring $R$ is a sequence $(I_n)$ of ideals in $R$ which satisfies:

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq I_{n+1} \subseteq \cdots$$

such a sequence is viewed as an infinite sequence.

When the chain terminates or is stationary at some ideal, we say that it is stationary.

**Definition** (*ACC*)**.** We say that a ring $R$ has the **ascending chain condition (ACC)** if every ascending chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ in $R$ is **stationary**[7]. That is, there exists $N \geqslant 1$ such that $I_N = I_{N+1} = I_{N+2} = \cdots$. In this case, we say that the chain is *stationary at* $I_N$.

Note that *ascending chain* is a property of ideals whereas *ACC* is a property of rings.

**Lemma 1.7.** If $(I_n)$ is an ascending chains of ideals in $R$, then $\bigcup_{n\geqslant 1} I_n$ is also an ideal.

**Proof.** We have to show that $J = \bigcup_{n\geqslant 1} I_n$ satisfies the axioms of an ideal.

**1. Show that $J \leqslant R$.** Suppose $x, y \in J$. It suffices to show that $x - y \in J$. By definition, there exist $m, n \geqslant 1$ (integers) such that $x \in I_m$ and $y \in I_n$. Consider $p = \max\{m, n\}$. By hypothesis, we must have $I_m \subseteq I_p$ and $I_n \subseteq I_p$. This implies that $x, y \in I_p$ and hence their difference $x - y \in I_p \subseteq J$.

**2. Second ideal axiom.** Suppose $x \in J$ and $r \in R$. We need to show that $rx \in J$. By definiton, there exists $m \geqslant 1$ (integer) such that $x \in I_m$. Since $I_m$ is an ideal in $R$, $rx \in I_m \subseteq J$, as desired. ■

Here is where the ACC is important: it characterizes Noetherian ring, and vice-versa.

**Theorem 1.12.** $R$ is a Noetherian ring if and only if it satisfies the ACC.

**Proof.** **($\Rightarrow$).** Assume that $R$ is Noetherian. Now, suppose that

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

[7] *Other words for stationary are terminates, stabilizes and eventually constant*

is an ascending chains of ideals in $R$. By the preceding lemma, $J = \bigcup_{n\geqslant 1} I_n$ is an ideal and so it is finitely generated[8] i.e. there exist $x_1, x_2, \ldots, x_r \in J$ such that $J = (x_1, \ldots, x_r)$.

By definition of the union, for every $x_i \in J$, there exists $n_i \geqslant 1$ (integer) such that $x_i \in I_{n_i}$. Take $m = \max\{n_1, \ldots, n_r\}$. Then each $I_{n_i} \subseteq I_m$ and so $x_i \in I_m$ for all $i$ i.e. we have shown that $J \subseteq I_m$. But by definition $I_m \subseteq J$ and so we conclude that $J = I_m$. So, the chain $I_1 \subseteq I_2 \subseteq \cdots$ is stationary at $I_m$.

stationary because: for every $r \geqslant m$, $I_r \subseteq J \subseteq I_m$ . But by the ascending chains condition, $I_m \subseteq I_r$ since $m \leqslant r$.

$(\Leftarrow)$. Assume that $R$ satisfies the ACC but is not Noetherian. Now suppose $I \lhd R$ is an ideal such that it is not finitely generated. Then for any set of finite elements $x_1, \ldots, x_n \in R$, we have $I \neq (x_1, \ldots, x_n)$.

Now choose any $x_1 \in I$. Since $I$ is not finitely generated, $I \neq (x_1)$. So there exists $x_2 \in I \setminus (x_1)$. $I$ is not finitely generated so $I \neq (x_1, x_2)$. So there exists $x_3 \in I \setminus (x_1, x_2)$. Again, $I$ is not finitely generated so $I \neq (x_1, x_2, x_3)$; so on and so forth. Continuing in this way indefinitely, we have an ascending chain of ideals

$$(x_1) \subseteq (x_1, x_2) \subseteq (x_1, x_2, x_3) \subseteq \cdots,$$

which does not terminate. This contradicts the ACC on $R$. 何! ■

**Proposition 1.11.** Let $R$ be a ring and $I \lhd R$. If $R$ is Noetherian, then $R/I$ is Noetherian.

A handy fact for the proof of this proposition is that ideals of $R/I$ has the form $J/I$ for some $J \lhd R$. This was given by the correspondence theorem.

**Proof.** Suppose $\mathfrak{J}_1 \subseteq \mathfrak{J}_2 \subseteq \mathfrak{J}_3 \subseteq \cdots$ is an ascending chain of ideals in $R/I$. The correspondence theorem tells us that there there exist ideals $J_1, J_2, J_3, \ldots$ in $R$ such that $I \lhd J_i$ and $J_i/I = \mathfrak{J}_i$. Moreover, $J_i \subseteq J_{i+1}$ for all $i$ since $\mathfrak{J}_i \subseteq \mathfrak{J}_{i+1}$ for all $i$ and so

$$J_1 \subseteq J_2 \subseteq J_3 \subseteq \cdots,$$

is an ascending chain of ideals in $R$. But we assumed $R$ is Noetherian, so this chain is stationary, say, at $J_N$. Consequently[9], the earlier ascending chain in $R/I$ is stationary at $\mathfrak{J}_N = J_N/I$. Hence, $R/I$ is Noetherian. ■

It is trivial to prove that if $A/I \subseteq B/I$ then $A \subseteq B$. This is immediate from definition. So if $J_i/I \subseteq J_{i+1}/I$, then $J_i \subseteq J_{i+1}$.

Here's another characterization of Noetherian rings.

**Theorem 1.13.** $R$ is a Noetherian ring if and only if every non-empty set of ideals in $R$ has a maximal element.

That is, the non-empty set of ideals does not contain a maximal element w.r.t set inclusion.

**Proof.** $(\Leftarrow)$. Suppose for contradiction that $S$ is a non-empty set of ideals in $R$ which has no maximal element. Now choose any ideal $I_1 \in S$. Since $I_1$ is not maximal, there exist $I_2 \in S$ such that $I_2 \supseteq I_1$ and $I_1 \neq I_2$. Again, $I_2$ cannot be maximal so there exists $I_3 \in S$ such that $I_3 \supseteq I_2$ and $I_2 \neq I_3$; so on and so forth. Continuing in this way indefinitely, we have an ascending chain

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

This proof is similar to when we prove that any ring satisfying the ACC is Noetherian. This is because we now know that Noetherian rings has the ACC property.

[8] *we assumed $R$ is Noetherian!*

[9] *Because the chain is stationary at $J_N$, for all $k \geqslant N$, we have $J_k = J_N$. This implies that $\mathfrak{J}_k = J_k/I = J_N/I$ for all such $k$ as well i.e. stationary at $\mathfrak{J}_N$.*

which does not terminate. This contradicts the ACC on $R$. 何!

**( $\Rightarrow$).** Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ be an ascending chain of ideals in $R$ and consider $S = \{I_1, I_2, I_3, \ldots\}$. By hypothesis, $S$ has a maximal element, $I_N$ because every non-empty set of ideals contains a maximal element. By definition of an ascending chain, for any $k \geqslant N$, we have $I_N \subseteq I_k$. But $I_N$ is maximal (in $S$) so this implies that $I_k \subseteq I_N$ since $I_k \in S$. So this ascending chain terminates and so $R$ is Noetherian. ■

By restricting such non-empty set of ideals to proper ideals, we have the following result.

**Lemma 1.8.** Any nonzero Noetherian ring $R$ has a maximal ideal.

**Proof.** Let $S$ be the set of all proper ideals $I$ in $R$. Since $\{0\} \in S$, this set is non-empty. Since $R$ is Noetherian, $S$ thus has a maximal element w.r.t set inclusion. By definition[10] such an element is the maximal ideal. ■

This lemma is very important, and we will use it a lot of times later on. In fact, we can prove that every nonzero ring, not necessarily Noetherian, has a maximal ideal. However, this uses Zorn's lemma which we will not get into.

**Lemma 1.9.** Let $R$ be a ring and $J \lhd R[X]$. For any $n \in \mathbb{N}_0$, define the set

$$I_n = \left\{ \begin{array}{c} \text{leading coefficients of elements} \\ f \in J \text{ with } \deg f \leqslant n \end{array} \right\} \cup \{0\}.$$

Then $I_n \lhd R$; and $I_n \subseteq I_{n+1}$ for all $n$.

$I_n$ can also be written more explicitly as the set $\{r \in R \mid \exists f \in J \text{ s.t.} f = rX^n + \ldots\} \cup \{0\}$.

**Proof.** We have to prove two things.

**Claim 1:** $I_n \lhd R$. By definition $0 \in I_n$ so it is non-empty. Now suppose $f, g \in J$ such that $\deg f = d$ and $\deg g = e$, both of which are less than or equal to $n$, with leading coefficients $a$ and $b$ respectively. Now for any $r \in R$ consider the difference $ra - b$. If it is 0, we are done; so assume not. WLOG, we may assume that $d \leqslant e$. Then $ra - b$ is the leading coefficient of the polynomial $r \cdot X^{e-d} \cdot f - g$ which is in[11] $J$ . Thus, $ra - b \in I_n$. This simultaneously proves that $I_n \leqslant R$; and that it satisfies the second ideal axiom.

**Claim 2: $I_n \subseteq I_{n+1}$ for all $n$.** Let $r \in I_n$. Then there exists $f \in J$ such that $f = rX^n + \cdots$. Since the polynomial $X \in R[X]$, $Xf \in J$ because $J \lhd R[X]$. But $Xf = rX^{n+1} + \cdots$ and so $r$ is the leading coefficient of $Xf$. This implies that $r \in I_{n+1}$. Since $I_n$ was arbitrary, we are done. ■

We want to show that the $I_n$ gives an ascending chain of ideals. This should hint to where we are going to use $I_n$ later in the proof.

The second claim basically says that $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$ is an ascending chain of ideals. This bizarre-looking lemma will be used to prove the following punchline theorem of this chapter.

**Theorem 1.14** (*Hilbert's Basis Theorem*)**.** If $R$ is a Noetherian ring, then so is $R[X]$.

**Proof.** Suppose $J \lhd R[X]$. We want to show that it is finitely generated. For any $n \in \mathbb{N}_0$, consider the set

[10] *of $S$ and a maximal ideal.*

[11] *since $J$ is an ideal of $R[X]$.*

$$I_n = \left\{ \begin{array}{c} \text{leading coefficients of elements} \\ f \in J \text{ with } \deg f \leqslant n. \end{array} \right\} \cup \{0\}.$$

Then $I_n \lhd R$ and it gives an ascending chain of ideals by the preceding lemma. Since $R$ is Noetherian, it has the ACC so there exists $N \in \mathbb{N}_0$ such that the ascending chain terminates i.e. $I_N = I_{N+1} = \cdots$. Also, since $R$ is Noetherian, all its ideals are finitely generated. So for all $n \leqslant N$, we can write

$$I_n = (a_{n,1}, a_{n,2}, \dots, a_{n,M_n}),$$

for some $a_{n,i} \in R$ where $1 \leqslant i \leqslant M_n$ and $M_n \in \mathbb{N}$. By definition of $I_n$, there exists $f_{n,i} \in J$ such that $f_{n,i} = a_{n,i}X^e + \cdots$, where $\deg f = e \leqslant n$ for each $i$. We now claim that these $f_{n,i}$ forms a generating set for $J$.

**Claim: We can write** $J = (f_{n,1}, f_{n,2}, \dots, f_{n,M_n})$.

*Proof of claim.* Let $g \in J$ such that its leading coefficient is $\lambda \in R$. That is,

$$g = \lambda X^d + \cdots,$$

where $d = \deg g$. We will prove by strong induction on $d$, that $g$ is an $R[X]$-linear combination[12] of $f_{d,1}, f_{d,2}, \dots, f_{d,M_d}$.We have two cases to take care of, so we are doing an induction proof twice.

**Case 1:** $d \leqslant N$. The lowest possible value $d$ (and $N$) can go is 0.

**Base case:** $d = 0$. In this case $g = \lambda$. Since $\lambda \in I_0$, we can write

$$\lambda = \sum_{i=1}^{M_0} s_i a_{0,i} = \sum_{i=1}^{M_0} s_i f_{0,i},$$

The fact $a_{0,i} = f_{0,i}$ is straight from definition of $I_0$ being finitely generated by the $a_{0,i}$.

where $s_i \in R$. Since $g = \lambda$, we are done as we have shown that $\lambda$ is an $R[X]$-linear combination of $f_{0,i}$.

**Inductive hypothesis.** Assume that any polynomial $p \in J$ with $\deg p < d$ is an $R[X]$-linear combination of $f_{\deg p,j}$.

**Inductive step.** We will show that the result is true with degree $d$. Since $\lambda \in I_d$, we can write

$$\lambda = \sum_{i=1}^{M_d} s_i a_{d,i},$$

$I_d$ was finitely generated by these $a_{d,1}, \dots, a_{d,M_d}$ remember?

for some $s_i \in R$. Consequently, we have that

$$h = \sum_{i=1}^{M_d} s_i f_{d,i} = \lambda X^d + \cdots.$$

This is because $f_{d,i} = a_{d,i}X^d + \cdots$. Using our very recent fact about $\lambda$. we get this line of claim.

Since $J$ is an ideal, $h \in J$. And so the difference $g_1 = g - h \in J$. But since $g, h$ have the same leading coefficient $\lambda$, the $X^d$ terms cancel out and thus $\deg g_1 < d = \deg g$. By the inductive hypothesis applied on $g_1$, we are done.

**Case 2:** $d > N$. The proof is basically identical. The base case now[13] is $d = 1$, but this is easy. Assume that the result is true for all polynomials in $J$ with degree $< d$. Let's show that the result is true with degree $d$.

[12] *This will prove that $J$ is finitely generated by these $f_{d,j}$.*

[13] *since $N = 0$ is the lowest possible.*

In this case $I_d = I_N$. We have $\lambda \in I_d = I_N$. So we can write

$$\lambda = \sum_{i=1}^{M_N} s_i a_{N,i},$$

for some $s_i \in R$. Consequently, we have that

$$h = X^{d-N} \sum_{i=1}^{M_N} s_i f_{N,i} = \lambda X^d + \cdots \in J.$$

So the difference $g_1 = g - h \in J$ with $\deg g_1 < d$. Applying the inductive hypothesis, we are done. ■

We now know by the Hilbert's basis theorem that if $R$ is Noetherian, then so is $R[X]$. But if $R[X]$ is Noetherian, then so is $R[X][Y] = R[X, Y]$ again by the Hilbert's basis theorem. By induction, we thus have the following corollary.

**Corollary 1.3.** If $R$ is Noetherian, then so is $R[X_1, \ldots, X_n]$.

In particular, when $K$ is a field, then $K[X_1, \ldots, X_n]$ is Noetherian. This is an immensely important fact in things that we want to study later on.

# 2 Modules

Modules are inherently a generalization of vector spaces. Instead of giving some hand-wavery introduction, we will make it precise immediately and deduce how it is in fact a generalization. One more thing, remember that we said we assume linear algebra throughout? This is where it begins. If you forget your linear algebra, the path ahead may be quite painful.

## 2.1 What is a module?

As usual, let us begin by laying down some definitions and looking at some examples.

**Definition** (*Module*)**.** Let $R$ be a ring. An **$R$-module** or a **module over $R$** is a set $M$ together with

(1). a group law $+$ on $M$, for which $(M,+)$ is an abelian group; and

(2). a scalar multiplication $R \times M \to M$, defined by $(r,m) \mapsto r \cdot m$, for all $r \in R, m \in M$ which satisfies these axioms:

(i) $(r+s)\cdot m = r \cdot m + s \cdot m$,

(ii) $(rs)\cdot m = r \cdot (s \cdot m)$,

(iii) $r \cdot (m+n) = r \cdot m + r \cdot n$,

(iv) $1_R \cdot m = m$

(v) $0_R \cdot m = 0_M$

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

In one sentence, the module axioms tells us that we can form linear combinations of the form $r \cdot m + s \cdot n$ in $M$ for any $r, s \in R$ and any $m, n \in M$. That is, we can write elements of a module as an $R$-linear combination.

**Remark.** From now on, the $\cdot$ symbol is used to denote the scalar multiplication and we will omit the ring multiplication symbol by writing $rs$ for a product of $r, s \in R$.

**Remark.** Some readers would have notice that the (scalar) multiplication above is really just an action of $R$ on $M$. So (2) is really saying that there is an action on $M$ satisfying some properties.

Here is the relation between modules and vector spaces. Stare at the axioms and observe that these are exactly the vector space axioms. If $K$ is a field, then a $K$-module is precisely a $K$-vector space. What makes the study of modules more complicated than vector spaces is that some elements of the ring need not be units. Furthermore, we will see that by "doing linear algebra over rings" instead of over fields, there are very weird and unexpected consequences.

**Example** (*Zero module*)**.** Let $R$ be a ring. Then $M = (\{0_R\},+)$ is an $R$-module. It is called the **zero module** or *trivial module* and is denoted 0 when there is no confusion.

ironically, there are more than 4 things that we can write simply as 0 now: $0_R$, $0_M$, the zero ideal and now the zero module.

Everytime we talk about a module, we will define explicitly what is its scalar multiplication. In the above example, the scalar multiplication is the multiplication in $R$. Likewise for the example below.

**Example** (*A ring is a module over itself*)**.** Let $R$ be a ring. Then $M = (R, +)$ is an $R$-module with scalar multiplication

$$r \cdot m := rm,$$

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

Here are more examples.

**Example** (*Quotient ring is a module*)**.** Let $R$ be a ring and $I \lhd R$. Then $M = (R/I, +)$ is an $R$-module with scalar multiplication

$$r \cdot (s + I) := rs + I,$$

for all $r \in R$ and $s + I \in M$.

**Example** (*Polynomial ring is a module*)**.** Let $R$ be a ring. Then $M = (R[X], +)$ is an $R$-module with scalar multiplication

$$r \cdot f := rf,$$

for all $r \in R$ and $f \in R[X]$.

**Example** (*Abelian group $\Leftrightarrow$ $\mathbb{Z}$-module*)**.** Let $(G, +)$ be an abelian group. Then $M = (G, +)$ is a $\mathbb{Z}$-module with scalar multiplication $\mathbb{Z} \times G \to G$ defined by

$$n \cdot g = \underbrace{g + \cdots + g}_{n \text{ times}}$$

Note that the + operation here is the group law on $G$.

The following is an extremely important example.

**Example** (*Homomorphism image module*)**.** Let $\phi : R \to S$ be a ring homomorphism. Then $M = (S, +)$ is an $R$-module with scalar multiplication

$$r \cdot s := \phi(r)s,$$

for all $r \in r, s \in S$.

If you recall that we mentioned that $R$-algebra can be given a nice structure, this is that structure — a module structure. We will not discuss more about algebras right now as we are more concern about modules. But we promise you that we will take a look at it carefully later on. An equally important module is the free module of rank $n$.

**Example** (*Free module of rank $n$*)**.** Let $R$ be any ring and let $n \in \mathbb{Z}^+$. Define

$$R^n = \{(r_1, r_2, \ldots, r_n) \mid r_i \in R\}.$$

Then $M = (R^n, +)$ where + is componentwise addition is an $R$-module with scalar multiplication

$$\lambda \cdot \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} \lambda r_1 \\ \vdots \\ \lambda r_n \end{pmatrix},$$

for all $\lambda \in R$ and $(r_1, \ldots, r_n) \in R^n$.

We will define free modules properly later on when we talk about finitely generated modules. For now, let's look at submodules.

**Definition** (*Submodule*)**.** Let $R$ be a ring and $M$ an $R$-module. An **$R$-submodule** of $M$ is a subgroup $N \subseteq M$ under $+$ such that $r \cdot n \in N$ for all $r \in R$ and $n \in N$.

Like subrings, we shall write $N \leqslant M$ whenever $N$ is an $R$-submodule of $M$.

**Remark.** Some remarks on submodules.

- Submodules of $M$ are just subsets of $M$ which are themselves modules under the restricted operations.
- If $R = K$ is a field, submodules are the same as subspaces.
- Every $R$-module $M$ has the submodules $M$ and $0$.

0 is called the **trivial submodule**.

Here are three important examples.

Note that every time we utter the word *important* for something, it means that that thing will be used extensively later on.

**Example** (*Ideals are ring submodules*)**.** Let $R$ be a ring. Then the $R$-submodules of $R$ are the ideals of $R$.

**Example** (*Subspaces are field submodules*)**.** Let $K$ be a field. Then the $K$-submodules of a $K$-module (which is a $K$-vector space) are its $K$-vector subspaces.

**Example** (*Subgroups are $\mathbb{Z}$-submodules*)**.** Let $G$ be an abelian group. Then it is a $\mathbb{Z}$-module. Its $\mathbb{Z}$-submodules are its subgroups.

An analogous criterion of the subgroup criterion for submodules is the following.

**Proposition 2.1** (*Submodule criterion*)**.** Let $R$ be a ring and let $M$ be an $R$-module. Then $N \subseteq M$ is a submodule of $M$ if and only if

(i). $N \neq \varnothing$,

(ii). $m + r \cdot n \in N$ for all $r \in R$ and all $m, n \in N$

Like rings, we have module homomorphisms.

**Definition** (*Module homomorphism*)**.** Let $R$ be a ring and let $M, N$ be $R$-modules. A map $\phi : M \to N$ is an **$R$-module homomorphism** if

or just $R$-homomorphisms.

$$\phi(r \cdot m + s \cdot n) = r \cdot \phi(m) + s \cdot \phi(n),$$

for all $r, s \in R$ and $m, n \in M$. If such a map is bijective, then we call it an **isomorphism** (of $R$-modules). The set of all $R$-module homomorphisms from $M$ into $N$ is denoted $\mathrm{Hom}_R(M, N)$, or just $\mathrm{Hom}(M, N)$ when $R$ is clear.

$R$-module homomorphisms are also known as $R$-linear maps or linear transformations. This agrees with linear algebra. It is an easy exercise to check that $\mathrm{Hom}_R(M, N)$ is itself an $R$-module.

**Lemma 2.1.** $\mathrm{Hom}_R(M, N)$ is an $R$-module with addition $+$ defined by $(\phi + \psi)(x) = \phi(x) + \psi(x)$ and scalar multiplication $R \times \mathrm{Hom}_R(M, N) \to \mathrm{Hom}_R(M, N)$ defined by

$$r \cdot \phi(m) = (r\phi)(m),$$

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

**Proof.** Everything else is mundane and obvious. What is worth proving is showing that $r\phi \in \mathrm{Hom}_R(M,N)$ so that the scalar multiplication that we defined is well-defined. Now, we have $(r\phi)(m) = r \cdot \phi(m)$. But $\phi \in \mathrm{Hom}_R(M,N)$, so in fact $r \cdot \phi(m) = \phi(r \cdot m)$. So for any $m, n \in M$ and $x \in R$, we have

$$(r\phi)(x \cdot m + n) = \phi(r \cdot (x \cdot m + n)) = \phi(r \cdot (x \cdot m) + r \cdot n) = (r\phi)(x \cdot m) + (r\phi)(n),$$

which implies that $r\phi \in \mathrm{Hom}_R(M,N)$ as desired. ■

**Example.** Let $K$ be a field. Then $K$-homomorphisms are linear maps (of vector spaces).

**Example.** $\mathbb{Z}$-homomorphisms are homomorphisms of abelian groups.

Linear maps $M \to M$ of modules will hold some interest for us later on, so we will define it here.

**Definition** (*Endomorphism*)**.** Let $R$ be a ring and let $M$ be an $R$-module. The $R$-homomorphism $\phi : M \to M$ is called an **$R$-endomorphism** of $M$. The set of all $R$-endomorphisms of $M$ is denoted $\mathrm{End}_R(M)$ or just $\mathrm{End}(M)$ when $R$ is clear.

Of course, the set $\mathrm{End}_R(M)$ is also an $R$-module as it is just $\mathrm{Hom}_R(M,M)$. As expected, we have kernel and images for module homomorphisms.

**Definition.** Let $\phi : M \to N$ be an $R$-homomorphism. Then the **kernel** of $\phi$ is the set $\ker\phi = \{m \in M \mid \phi(m) = 0_N\}$; and the **image** of $\phi$ is the set $\operatorname{im}\phi = \phi^{\mathrm{img}}(M)$.

We have an analogous result for $R$-homomorphisms.

**Lemma 2.2.** Let $\phi : M \to N$ be a homomorphism of $R$-modules. Then $\ker\phi \subseteq M$ is an $R$-submodule of $M$ and $\operatorname{im}\phi \subseteq N$ is an $R$-submodule of $N$.

**Definition** (*Quotient modules*)**.** Let $N \leqslant (M,+)$ be an $R$-submodule. The **quotient module** $M$ modulo $N$, denoted $M/N$, is the set $\{m + N \mid m \in M\}$ with scalar multiplication $R \times M/N \to M/N$ defined by

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

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

It is an exercise to check that: (i) the scalar multiplication is well-defined; and (ii) it actually gives a module. We now define a new concept that was not found in groups and rings.

**Definition** (*Cokernel of module homomorphisms*)**.** Let $\phi : M \to N$ be an $R$-module homomorphism. The **cokernel of $\phi$** is the set

$$\operatorname{coker}\phi = N/\operatorname{im}\phi = \{n + \operatorname{im}\phi \mid n \in N\},$$

which is a quotient module.

As in rings (and groups), we have analogous results about the projection map and the isomorphism theorems.

**Definition** (*Projection map*). Let $R$ be a ring and $M$ an $R$-module; let $N \leqslant M$ be a submodule. The map $\pi : M \to M/N$ defined by $\pi(m) = m + N$ is called the **$R$-projection map**, or simply the projection map.

It is obvious why we are giving it the same name as nothing differs aside from the actors in play.

**Proposition 2.2.** Let $R$ be a ring. The projection map $\pi : M \to M/N$ is a surjective $R$-homomorphism with $\ker \pi = N$.

**Theorem 2.1** (*First isomorphism theorem*). Let $\phi : M \to N$ be an $R$-homomorphism. Then $M/\ker\phi \cong \operatorname{im}\phi$ by the map $m + \ker\phi \longmapsto \phi(m)$ induced by $\phi$.

**Theorem 2.2** (*Correspondence theorem*). Let $R$ be a ring and $M$ be an $R$-module. Let $\pi : M \to M/N$ be the projection map. Then there is a bijective correspondence:

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

given by $S \longmapsto \pi^{\text{img}}(S)$ and $\pi^{\text{preimg}}(\mathfrak{S}) \longleftarrow \mathfrak{S}$.

Note the arrows' direction carefully.

We will not prove these theorems as even the proofs are completely analogous to those of ring's. Proving it yourself is a good exercise to verify your understanding.

## 2.2 Finitely generated modules

As we already had something named *finitely generated* (cf. f.g. ideals), the definition for finitely generated modules should come as natural. In fact, the way we define it is identical: using a set of finite elements. But in modules, these set of finite generators tells us more if they have an extra property called *linearly independent* (in the usual sense of linear algebra). Let us start by defining these objects properly.

**Definition** (*Generating set, linear independence and basis*). Let $R$ be a ring and let $M$ be an $R$-module. Let $\mathcal{E} = (e_i)_{i\in I} \subseteq M$. Then

Note here that $I$ is any index set, possibly unordered and infinite.

(1). We say $\mathcal{E}$ is a **generating set** for $M$ if every $m \in M$ can be written as a sum

$$m = \sum_{i\in I} r_i \cdot e_i,$$

for some scalars $r_i \subseteq R$. The elements of $\mathcal{E}$ are then called **generators**.

In other words, we can write

$$M = \left\{ \sum_{i=1}^{k} r_i \cdot e_i \;\middle|\; r_i \in R \right\}.$$

(2). We say $\mathcal{E}$ is **linearly independent** if for any subset $\{e_1, \ldots, e_k\} \subseteq \mathcal{E}$ of distinct elements, we have

$$\sum_{i=1}^{k} r_i \cdot e_i = 0_M \implies r_1 = r_2 = \cdots = r_k = 0_R.$$

If $\mathcal{E}$ is both a generating set for $M$ and linearly independent, then we say that it is a **basis** for $M$.

Note that like in the case of vector spaces, bases need not be unique. However, there is a subtle difference. Moreover, the existence of bases for vector spaces is guaranteed[14], whereas modules need not have a basis.

**Definition** (*Finitely generated modules*)**.** Let $R$ be a ring and $M$ be an $R$-module. We say $M$ is **finitely generated** if there exists a finite generating set $\mathcal{E}$ for $M$. If $\#\mathcal{E} = 1$, say, $\mathcal{E} = \{m\}$, then $M$ is said to be **cyclic** and we write $Rm$.

The idea of finitely generated modules is analogous to the notion of finite dimensional vector spaces. However, as mentioned before, there is a subtle difference.

**Example.** Let $R$ be a ring. Then $R$ is a finitely generated $R$-module as $1_R$ is a generator. Consequently, the free module $R^n$ of rank $n$ is a finitely generated $R$-module.

**Example.** $\mathbb{Z}[i]$, the Gaussian integers, is a finitely generated $\mathbb{Z}$-module (abelian groups) as

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\},$$

with generators $1, i \in \mathbb{Z}[i]$. We can

**Example.** $\mathbb{R}^2$, the plane, is a finitely-generated $\mathbb{R}$-module (vector space) as

$$\mathbb{R}^2 = \left\{ x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;\middle|\; x, y \in \mathbb{R} \right\},$$

with generators $(1,0), (0,1) \in \mathbb{R}^2$.

Observe that there is a natural surjective homomorphism from $R$ onto the cyclic $R$-module $Rm$ given by

$$\phi : r \longmapsto r \cdot m.$$

By the first isomorphism theorem on modules, we thus have that $Rm \cong R/\ker\phi$. This kernel is special that we give it a name.

**Definition** (*Annihilator*)**.** Let $M$ be an $R$-module; and let $m \in M$. The **annihilator of $m$** is the set

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

where again we remind you that $\cdot$ means the scalar multiplication.

If $S \subseteq M$, then we define the **annihilator of $S$** to be

$$\operatorname{Ann}(S) = \{r \in R \mid r \cdot s = 0 \text{ for all } s \in S\} = \bigcap_{s \in S} \operatorname{Ann}(s).$$

So given a cyclic $R$-module $Rm$, we thus have

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

Conversely, if we have an $R$-module isomorphism $R/\operatorname{Ann}(m) \xrightarrow{\sim} Rm$, then the image[15] of $1_{R/\operatorname{Ann}(m)}$ generates $Rm$. If we further have that $R$ is Euclidean, then it is a PID and so in fact, $\operatorname{Ann}(m)$ (as an ideal) is principal. Thus, we have that $\operatorname{Ann}(m) = (r)$ for some $r \in R$.

**Definition** (*Free modules*)**.** An $R$-module $M$ is said to be **free** if it has a basis $\mathcal{E}$. The cardinality of the basis $\#\mathcal{E}$ is called the **rank** of $M$. If $\#\mathcal{E} < \infty$, we say $M$ is **free of finite rank**.

[14] *we accept the axiom of choice*

[15] *the image is the* $1$ *in* $Rm$*, standard homomorphism stuff.*

**Example.** Let $R$ be a ring. Then $R$ is a free module of rank one as $1_R$ is a basis. In fact, any unit in $R$ is a basis.

**Example.** Let $R$ be a ring. Then $R^n$, where $n \in \mathbb{N}$ is a free $R$-module of rank $n$ with basis

$$\begin{pmatrix} 1_R \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1_R \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1_R \end{pmatrix} \in R^n.$$

This is the free module of rank $n$ that we have said many times already. But the formal definition of a "free module of rank $n$" that is going on in our head needs the notion of direct sum modules.

**Definition** (*Direct sum module*). Let $M, N$ be $R$-modules. The **direct sum** of $M$ and $N$ is the set

$$M \oplus N = \{(m, n) \mid m \in M, n \in N\},$$

which is itself an $R$-module.

Note that if × is the Cartesian product of sets (no module structure), then in fact

$$M \oplus N = M \times N.$$

As we will see, however, they are not necessarily equal in the generalization that we will make. The direct sum of $R$-modules $M, N$ is itself a module with componentwise addition and scalar multiplication $R \times (M \oplus N) \to (M \oplus N)$ defined by

$$r \cdot (x, y) = (r \cdot x, r \cdot y).$$

Furthermore, by transitivity, we can easily extend the above definition to a finite direct sum of $R$-modules $M_1, M_2, \dots, M_n$ which will denote as $\bigoplus_{i=1}^{m} M_i$. In fact, we can make a definition so that an arbitrary family of modules makes sense and is still a module. If $(M_i)_{i \in I}$ is an arbitrary family of $R$-modules, where $I$ is some index[16] set, then we can define their direct sum by

$$\bigoplus_{i \in I} M_i = \{(x_i) \mid x_i \in M_i \text{ and only finitely many } x_i \text{ are nonzero}\}.$$

If we remove the condition that only finitely many $x_i$ are nonzero, we get the **direct product** of modules $\prod_{i \in I} M_i$ which is basically the Cartesian product of arbitrary modules, which can also be given a module structure. Consequently, if $I \subseteq \mathbb{N}$ is a finite set, then these two ideas coincide:

$$\bigoplus_{i=1}^{n} M_i = \prod_{i=1}^{n} M_i,$$

as mentioned earlier.

**Remark.** Here is a layman version. To say $P = M \oplus N$, means that $P$ can be entirely broken up into $M$ and $N$, which we can regard as separate pieces.

**Theorem 2.3** (*Free equivalence*). Let $M$ be an $R$-module. $M \cong \bigoplus_{i \in I} N_i$ for some $R$-modules $N_i$ such that $N_i \cong R$ for each $i$ *if and only if* $M$ is free.

Finally, we are in a position to formally define the free module of rank $n$.

[16] *Note that the index set $I$ need not be ordered or finite.*

**Definition** (*Free module of rank $n$*)**.** An $R$-module $M$ is called a **free module of rank $n$** if $M \cong \bigoplus_{i=1}^{n} R_i$. We will denote this module by $R^n$.

Note that $R^m \oplus R^n \cong R^{m+n}$.

This definition is consistent with our definition of rank, and our definition of a free module of rank $n$ earlier. So why do we go so far to formally give this definition? Notice that when say a "free module of rank $n$", we mean **any** module that is isomorphic to $R^n$, not necessarily **just** $R^n$ that we have defined earlier.

It turns out that weird things happen when we do linear algebra with rings instead of fields. If $V$ is a finite-dimensional vector space over a field $K$, then all its subspaces are also finite dimensional. But this is not true for modules. A finitely generated module $M$ over some ring $R$ does not necessarily have finitely generated submodules. Instead, we need to assume a bit more.

**Theorem 2.4.** Let $R$ be a Noetherian ring. If $M$ is a finitely generated $R$-module, then so is every of its submodules.

**Proof.** Just prove by induction on the generators of $M$ and use the correspondence theorem for modules. ■

This is consistent with what we know about vector spaces since fields are Noetherian.

**Slogan.** Noetherian f.g. modules have f.g. submodules

## 2.3 From modules to matrices

Since we are essentially doing linear algebra on rings, it is fair to guess that matrices is coming next. But before we discuss this transition, we will need some homological algebra.

To be fair, we don't actually necessarily need it now, but it makes our proof cleaner and we would need it later anyways.

**Definition** (*Exact sequences*)**.** Let $R$ be a ring. A sequence of $R$-modules and $R$-homomorphisms

$$\cdots \longrightarrow M_{i-1} \xrightarrow{f_i} M_i \xrightarrow{f_{i+1}} M_{i+1} \longrightarrow \cdots$$

is said to be **exact at $M_i$** if $\operatorname{im} f_i = \ker f_{i+1}$. The sequence is **exact** if it is exact at $M_i$ for all $i$.

**Remark.** It is possible that this sequence is infinite in one direction or even both.

Let $R$ be a ring. Recall that we write $0$ to mean the zero $R$-module $\{0_R\}$. Now note that for any $R$-module $M$, there are two unique $R$-module homomorphisms: $M \to 0$ defined by $m \mapsto 0$ for all $m \in M$; and $0 \to M$ defined by $0 \mapsto 0$. These maps are called the **zero $R$-homomorphisms**. With this out of the way, let's look at special cases of exact sequences.

**Proposition 2.3** (*Special case 1*)**.** Let $R$ be a ring. The sequence of $R$-modules

$$0 \longrightarrow A \xrightarrow{f} B$$

is exact if and only if $f$ is injective.

**Proof.** By definition of an exact sequence, we have $\{0\} = \operatorname{im} 0 = \ker f$ where 0 here is the zero $R$-homomorphism. So $f$ is injective. The converse is similar. ■

**Proposition 2.4** (*Special case 2*). Let $R$ be a ring. The sequence of $R$-modules

$$B \xrightarrow{g} C \longrightarrow 0$$

is exact if and only if $g$ is surjective.

**Proof.** By definition of an exact sequence, we have $\operatorname{im} g = \ker 0$ where 0 here is the zero $R$-homomorphism. But $\ker 0 = C$ by definition of the zero homomorphism. So $g$ is surjective. The converse is similar. ■

**Proposition 2.5** (*Special case 3*). Let $R$ be a ring. The sequence of $R$-modules

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is exact if and only if $f$ is injective, $g$ is surjective and $\operatorname{coker} f \cong C$.

**Proof.** If the sequence is exact, we get injectivity of $f$ and surjectivity of $g$ for free from the first two special cases by looking at parts of the current sequence. We also have that $\operatorname{im} f = \ker g$ which implies

$$\operatorname{coker} f = B/\operatorname{im} f = B/\ker g.$$

But by the first isomorphism theorem, $B/\ker g \cong \operatorname{im} g = C$. We are done. The converse is similar. ■

The equality $\operatorname{im} g = C$ comes from the fact that $g$ is surjective.

The third special case is so useful that we give it a name.

**Definition** (*Short exact sequences*). Let $R$ be a ring. A **short exact sequence** of $R$-modules is given by two $R$-homomorphisms $f : A \to B$ and $g : B \to C$ with

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

such that $\operatorname{im} f = \ker g$.

The last special case 3 gives us a novel way of thinking about short exact sequences. We can think of the module $B$ as an extension of $A$ by $C$. Let's come back to what we have earlier.

**Lemma 2.3.** Let $R$ be a ring and let $M$ be a finitely generated $R$-module. Then there is a surjective $R$-homomorphism $\phi : R^n \to M$ for some $n \in \mathbb{N}$.

**Proof.** Since $M$ is a finitely generated $R$-module, there exists a finite set

$m_1, \ldots, m_k \in M$ which generates $M$. Then define $\phi : R^n \to M$ to be the map

$$\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \longmapsto \sum_{i=1}^{k} r_i \cdot m_i.$$

This map is clearly an $R$-module homomorphism. By construction, it is surjective, so we are done. ∎

We can say more when $R$ is a Noetherian ring.

**Lemma 2.4** (*Describing modules via matrices*)**.** Let $R$ be a Noetherian ring and let $M$ be a finitely-generated $R$-module. Then there is an $R$-homomorphism $\pi : R^m \to R^n$ for some $m, n \in \mathbb{N}$; such that

$$\operatorname{coker} \pi \cong M.$$

**Proof.** Since $M$ is a finitely generated $R$-module, the preceding Lemma 2.3 implies that there exists a surjective $R$-homomorphism $\phi : R^n \to M$ for some $n \in \mathbb{N}$.

Now, since $R$ is Noetherian and $M$ is a finitely generated $R$-module, then $\ker \phi \subseteq R^n$ is also a finitely generated $R$-module. By the preceding Lemma 2.3 again, there is a surjective $R$-homomorphism $\psi : R^m \to \ker \phi$ for some $m \in \mathbb{N}$. What we have now is the map

$$R^m \to \ker \phi \subseteq R^n \to M.$$

Now consider $\pi = \iota \circ \psi$ where $\iota$ is the inclusion map $\ker \phi \hookrightarrow R^n$. Then we can *concatenate* these homomorphisms to get a sequence of homomorphisms

$$R^m \xrightarrow{\pi} R^n \xrightarrow{\phi} M \longrightarrow 0,$$

which is exact. In particular, we have $M \cong \operatorname{coker} \pi = R^n / \operatorname{im} \pi$. ∎

by special case 3 on exact sequences.

It is important to note that the homomorphism $\pi$ in the lemma above is not unique. Consequently, for a fix finitely-generated module $M$, there are many different descriptions[17] of $M$. Observe that having the homomorphism $R^n \xrightarrow{\pi} R^m$ is the same as having an $m \times n$ matrix, say, $\Pi$ with entries in $R$. This can be seen as follows. Since $R^m$ and $R^n$ are finitely generated ($R$-modules), they have a basis $\{f_1, \ldots, f_m\}$ and $\{e_1, \ldots, e_n\}$ respectively. So using our $R$-homomorphism, we have

$$\pi(e_j) = \sum_{i=1}^{m} a_{ij} f_i,$$

for some $a_{ij} \in R$ and this gives the matrix $\Pi = (a_{ij})$ explicitly written as

$$\Pi = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Recall that $M \cong \operatorname{coker} \pi$. So $M$ will be unchanged if we replace $\Pi$ by an *equivalent* matrix, say, $\Pi'$, where **equivalent** means we can move from $\Pi$ to $\Pi'$ by a sequence of elementary row and column operations. These elementary operations are:

[17] *as the cokernel of $\pi$ for the various possible $\pi$.*

1. Exchange 2 rows or 2 columns
2. Divide a row (or a column) by a unit in $R$.
3. Replace row (or column) $\mathcal{R}_i$ by $\mathcal{R}_i + a\mathcal{R}_j$ for any element $a \in R$ (not necessarily a unit) provided $i \neq j$.

We can also see $\Pi$ and $\Pi'$ being equivalent as the following.

**Definition** (*Equivalent matrix*)**.** Two matrices $A$ and $B$ with entries in a ring $R$ are **equivalent** if there exist invertible matrices $P^{-1}, Q$ (also with entries in $R$) such that

$$B = QAP^{-1}$$

**Remark.** Referring to the previous lemma, let $\pi : R^m \to R^n$ be an $R$-homomorphism of free modules where $R$ is a Noetherian. Then $R^m$ and $R^n$ has a basis, say, $\mathcal{M}$ and $\mathcal{N}$ which gives a matrix. Now observe that doing elementary operations is just changing these bases. In particular, column operations changes the basis $\mathcal{M}$ of $R^m$ and row operations changes the basis $\mathcal{N}$ of $R^n$.

From now on, we will denote the $i$-th row of a matrix by $\mathcal{R}_i$, the $j$-th column of a matrix by $\mathcal{C}_j$ and the $i, j$-th entry of a matrix by $(i, j)$; with clear adjustments to the notations if needed. Let's recall a definiton.

**Definition** (*Diagonal matrix*)**.** Let $\Pi = (a_{ij})$ be an $m \times n$ matrix with entries in a ring $R$. We say $\Pi$ is **diagonal** if $i \neq j$ implies $a_{ij} = 0$.

**Example.** Define the matrices,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 6 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} x^2 & 1 \\ 0 & x^3+1 \end{pmatrix}, \quad C = \begin{pmatrix} x & 0 & 0 \\ 0 & x^2 & 0 \\ 0 & 0 & x^3 \\ 0 & 0 & 0 \end{pmatrix}.$$

The matrix $A$ and $C$ are diagonal, whereas $B$ is not.

**Theorem 2.5** (*Smith normal form*)**.** Let $R$ be a Euclidean domain. Then any matrix with entries in $R$ is equivalent to a diagonal matrix $D = (a_{ij})$ such that with $d_i = a_{ii}$, then $d_i \neq 0$ for all $i$ and $d_1 \mid d_2 \mid \cdots \mid d_n$.

As per the name of the theorem, such a diagonal matrix will be called the **Smith normal form** of the matrix.

**Proof.** Let $R$ be a Euclidean domain with size function $\phi$. Let $\Pi$ be an $m \times n$ matrix with entries in $R$. If $\Pi$ is the zero matrix, then we are done as it is already in Smith normal form, so suppose not.

Otherwise consider the collection of all matrices equivalent to $\Pi$, and choose one that has a least[18] nonzero entry. If necessary, exchange two rows and exchange two columns so that this least entry is the $(1, 1)$ entry. We now claim the following.

**Lemma:** There exist elementary operations so that every other entry in $\mathcal{R}_1$ (and $\mathcal{C}_1$) is zero. In other words, we have $(1, j) = (i, 1) = 0$ for all $i, j \neq 1$.

*Proof of claim.* Look at the $(1, 2)$ entry. Suppose the $(1, 1)$ entry is $a$ and $(1, 2)$ entry is $b$. Since $R$ is Euclidean, we can divide $a$ into $b$ to have $b = aq + r$ with either $r = 0$

[18] *as measured by the size function on $R$.*

or $\phi(r) < \phi(a)$. But $a$ was minimal, so $r = 0$. Replacing $\mathcal{C}_2 \to \mathcal{C}_2 - q\mathcal{C}_1$, we have $r = 0$ at the $(1, 2)$ entry. We can do this for all other columns in $\mathcal{R}_1$ to have $(1, j) = 0$ except when $j = 1$ where we have $(1, 1) = a$. By the same logic, we can also do this for other rows at column $\mathcal{C}_1$ so that $(i, 1) = 0$ except at $i = 1$. ■

After these elementary operations, the original matrix $\Pi$ is now equivalent to some other matrix

$$\left[\begin{array}{c|ccc} a & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \Pi' & \\ 0 & & & \end{array}\right].$$

By induction, $\Pi'$ is equivalent to a matrix with diagonal form, one of:

$$\left[\begin{array}{ccc|c} a_2 & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & a_m & \end{array}\right] \quad \text{or} \quad \left[\begin{array}{ccc} a_2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & a_m \\ \hline & \mathbf{0} & \end{array}\right],$$

We implicitly proceeded by strong induction on the size of the matrix, viewing it as a $k \times k$ block matrix. The base case is obvious. We can either consider $k = 0$ or $k = 1$, both of which are already in Smith normal form for free.

with $a_2 \mid a_3 \mid \cdots \mid a_m$. It remains to show that $a \mid a_2$. Suppose $\Pi'$ is equivalent to the diagonal matrix on the left, that is assume $m \leqslant n$. Then, we can do the row operation (on $\Pi$) $\mathcal{R}_1 \to \mathcal{R}_1 + \mathcal{R}_2$ to have

$$\left[\begin{array}{c|cccc} a & a_2 & \cdots & \cdots & 0 \\ \hline 0 & \ddots & & & \\ \vdots & & \ddots & & \\ 0 & & & a_m & 0 \end{array}\right].$$

Since $R$ is Euclidean, we can write $a_2 = qa + r$ where either $r = 0$ or $\phi(r) < \phi(a)$. Then we can do the column operation $\mathcal{C}_2 \to \mathcal{C}_2 - q\mathcal{C}_1$ to get

$$\left[\begin{array}{c|cccc} a & r & \cdots & \cdots & 0 \\ \hline 0 & \ddots & & & \\ \vdots & & \ddots & & \\ 0 & & & a_m & 0 \end{array}\right].$$

But $a$ was minimal, so $r = 0$. Thus, we have $a \mid a_2$, and the matrix is still equivalent to a diagonal matrix; we are done. ■

## 2.4 Structure theorems

Using the Smith normal form theorem, we can prove the structure theorem for finitely generated modules over Euclidean domains. This gives a classification for such modules. But first we need two lemmas.

**Lemma 2.5.** Let $R$ be a ring and let $M_1, \ldots, M_n$ be $R$-modules with submodules $N_i \leqslant M_i$. Then

$$\frac{\bigoplus_i M_i}{\bigoplus_i N_i} \cong \bigoplus_i \frac{M_i}{N_i},$$

where $0 \leqslant i \leqslant n$.

**Proof.** Consider the $R$-homomorphism $\phi : \bigoplus M_i \to \bigoplus M_i/N_i$ defined by

$$(m_1, \dots, m_n) \longmapsto (m_1 + N_1, \dots, m_n + N_n).$$

This map is clearly surjective by construction. Moreover, $\ker \phi = \bigoplus N_i$. By the First Isomorphism Theorem, we get the claim. ■

Recall that we write $M = Rm$ for the cyclic $R$-module generated by the single element $m \in M$. Here's where it will hurt if you have not pay careful attention. Let $R$ be a ring, let $M$ be an $R$-module and let $d \in R$. If $m \in M$, then $d \cdot m \in M$ simply by definition. So it makes sense to talk about a module that is finitely generated by $d \cdot m$, which we will simply write $Rdm$ (instead of $Rd \cdot m$ which is thousand times more confusing.)

**Lemma 2.6.** Let $R$ be a ring with $d \in R$ nonzero and consider the cyclic $R$-modules $Rm, Rdm$. Then there is an isomorphism

$$R/(d) \cong Rm/Rdm.$$

**Proof.** Consider the $R$-homomorphism $\phi : R \to Rm/Rdm$ defined by

$$r \longmapsto rm + Rdm.$$

This map is clearly surjective by construction. Now if $x \in \ker \phi$, then $\phi(x) = xm + Rdm = 0_{Rm} + Rdm$. In other words, $xm \in Rdm$ and so $xm = rdm$ for some $r \in R$. But this just means $x = rd$ for some $r \in R$ or $x \in (d)$. And since all the implications are in fact iff statements, then $\ker \phi = (d)$. The First Isomorphism Theorem concludes the proof. ■

Using these two lemmas, we can now prove perhaps the top 3 most important theorem in this course.

**Theorem 2.6** (*Structure theorem for finitely generated modules over Euclidean domains*). Let $M$ be a finitely generated module over a Euclidean domain $R$. Then there exist $d_1, \dots, d_n \in R$ and $m \in \mathbb{N}$ such that

$$M \cong \bigoplus_{i=1}^{n} R/(d_i) \oplus R^m,$$

where $d_i \mid d_{i+1}$ for all $1 \leqslant i \leqslant n-1$.

The proof of this theorem is not too difficult. But it will definitely test your understanding so far.

**Proof.** Let $M$ be a finitely generated module over a Euclidean domain $R$. Since $R$ is Euclidean, it is a PID and hence, Noetherian. So $M \cong \operatorname{coker} \pi$ for some homomorphism $\pi : R^m \to R^n$ of free modules of finite rank. WLOG, we can assume that $n \geqslant m$. These free modules each has a basis which gives rise to a matrix $\Pi$ with entries in $R$. By Smith normal form, $\Pi$ can be put into an $n \times m$ matrix

$$\Pi' = \left[\begin{array}{cccc} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \\ \hline 0 & \cdots & \cdots & 0 \end{array}\right],$$

where $d_1 \mid d_2 \mid \cdots \mid d_m$, with new bases $\mathcal{E} = \{e_1, \dots, e_m\}$ for $R^m$ and $\mathcal{F} = \{f_1, \dots, f_n\}$ for $R^n$. That is, under $\pi$ we now have $\pi(e_i) = d_i f_i$ for $i = 1, \dots, m$. So we see that

$$R^n \cong \bigoplus_{i=1}^{n} Rf_i.$$

where $Rf_i$ is the cyclic $R$-module generated by $f_i$.

From this new description, it follows that $M$ is generated by $n$ elements $f_1, \dots, f_n$ subject to relations

$$d_1 f_1 = d_2 f_2 = \cdots = d_m f_m = 0_{R^n}.$$

These relations, in some sense, comes from the fact that we are talking about the cokernel of $\pi$ i.e. the kernel of $\pi^t$ where $t$ here means transpose.

This implies that $\pi(\mathcal{E}) = \{d_1 f_1, d_2 f_2, \cdots, d_m f_m\}$ forms a basis for the submodule $\operatorname{im} \pi \subseteq R^n$, and so we have an isomorphism

$$\operatorname{im} \pi \cong \bigoplus_{i=1}^{m} Rd_i f_i \oplus R0f_{m+1} \oplus \cdots \oplus R0f_n.$$

where $0f_{m+1} = \cdots = 0f_n = 0_{R^n}$, so all these last terms are in fact the zero ideal.

Since we had $M \cong \operatorname{coker} \pi = R^n / \operatorname{im} \pi$, the preceding Lemma 2.5 implies that

$$M \cong \bigoplus_{i=1}^{k} \frac{Rf_i}{Rd_i f_i} \oplus \underbrace{R \oplus \cdots \oplus R}_{n-m \text{ times}}.$$

where we got just $R$ since $R/(0) \cong R$.

The preceding Lemma 2.6 also gives $Rf_i / Rd_i f_i \cong R/(d_i)$ since $d_i \neq 0$, so we are done. ■

**Slogan.** F.g. modules over Euclidean domains is a direct sum of cyclic modules and a free module.

Let's look at a concrete example.

**Example.** Consider the ring $\mathbb{Z}$ and suppose $M$ is a $\mathbb{Z}$-module finitely generated by $m_1, m_2, m_3$ subject to 2 relations

$$\begin{aligned} 1 \cdot m_1 + 2 \cdot m_2 + 3 \cdot m_3 &= 0, \\ 4 \cdot m_1 + 5 \cdot m_2 + 6 \cdot m_3 &= 0. \end{aligned}$$

Then the matrix that describes $M$ is

$$\Pi = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Since $\mathbb{Z}$ is Euclidean, we can put this into Smith normal form. To do this, just scan through this matrix for the smallest possible nonzero element (here w.r.t the size

In the proof of the Smith normal form, what we did was scan the infinite set of all possible matrices equivalent to $\Pi$. This is not very efficient to do by hand. We did it in the proof because it was the easiest to describe abstractly. If we have a concrete matrix like $\Pi$ here, we can approach as in this example.

function $\phi(n) = |n|$). The smallest element here is thus 1. So we begin diagonalizing the matrix with the column operation $\mathcal{C}_2 \to \mathcal{C}_2 - 4\mathcal{C}_1$ to have

$$\begin{pmatrix} 1 & 0 \\ 2 & -3 \\ 3 & -6 \end{pmatrix}.$$

Then we do the row operation $\mathcal{R}_2 \to \mathcal{R}_2 - 2\mathcal{R}_1$ to have

$$\begin{pmatrix} 1 & 0 \\ 0 & -3 \\ 3 & -6 \end{pmatrix}.$$

And then we do the row operation $\mathcal{R}_3 \to \mathcal{R} - 3\mathcal{R}_1$ to have

$$\begin{pmatrix} 1 & 0 \\ 0 & -3 \\ 0 & -6 \end{pmatrix}.$$

Now we look at the submatrix $\begin{pmatrix} -3 \\ 6 \end{pmatrix}$ and play the same game again. We can do this by the row operation $\mathcal{R}_3 \to \mathcal{R}_3 + 2\mathcal{R}_2$ to have

$$\begin{pmatrix} 1 & 0 \\ 0 & -3 \\ 0 & 0 \end{pmatrix}.$$

We are finally in Smith normal form. So $M$ is now generated by $n_1, n_2, n_3$ subject to relations

$$\begin{aligned} 1 \cdot n_1 &= 0, \\ -3 \cdot n_2 &= 0, \end{aligned}$$

with $n_3$ free. These relations implies that $n_1 = -3 \cdot n_2$ so we can eliminate $n_1$ as a generator. Consequently, $M$ is generated by two generators $n_2, n_3$ with one relation $3 \cdot n_2 = 0$. Thus,

$$M \cong \mathbb{Z}/(3) \oplus \mathbb{Z},$$

where $\mathbb{Z}/(3)$ is generated by $n_2$ and $\mathbb{Z}$ is generated by $n_3$.

This example may look small, but if you understand every step then you understand everything about structure theorems.

**Remark.** One more thing, it is easy to observe that in general, an $m \times n$ matrix describes a module via $m$ generators and $n$ relations.

An immediate corollary is the Structure Theorem for finitely generated abelian groups. This is done by simply taking $R = \mathbb{Z}$ as abelian groups are $\mathbb{Z}$-modules. The Structure Theorem effectively tells us that all of finitely generated module over a Euclidean domain (in fact, PID) are of the direct sum above. So if we want to investigate more about it, we can just assume that it has the above form.

**Definition** (*Torsion submodule, element, torsion-free*)**.** Let $R$ be an **integral domain**

and $M$ be an $R$-module. Then the **torsion submodule of $M$** is the set

$$\operatorname{Tors}(M) = \{m \in M \mid \exists\, r \in R \text{ nonzero such that } r \cdot m = 0_M\}.$$

We call an element of $\operatorname{Tors}(M)$ a **torsion element of $M$**. We say $M$ is **torsion-free** if $\operatorname{Tors}(M) = \{0\} = 0$. We say $M$ is **torsion** if $M = \operatorname{Tors}(M)$.

**N.B.** We need $R$ to be an integral domain. Otherwise, the torsion submodule is... not a submodule.

That is, $M$ is torsion if for any $m \in M$, there exists a nonzero element $r \in R$ such that $r \cdot m = 0$.

**Exercise.** Prove that $\operatorname{Tors}(M)$ is a submodule of $M$ when $R$ is a domain.

**Example.** $\operatorname{Tors}(M)$ is not necessarily a submodule of $M$ if $R$ is not a domain. Consider the ring $R = \mathbb{Z}/6\mathbb{Z}$ which is not a domain and view it as a module, $M$ over itself. Then $2, 3 \in M$ are torsion elements annihilated by each other (viewed as $2, 3 \in R$). Now, $5 = 2 + 3 \in M$. However, 5 is not a torsion element of $M$ and this can be checked by inspection.

**Example.** $R$ as an $R$-submodule is torsion-free.

*Proof.* This is easy. Suppose $m \in R$ such that $rm = 0$ for some nonzero $r \in R$. But $R$ is an integral domain, and $r \neq 0$ and so this forces $m = 0$. In other words $\operatorname{Tors}(R)$ is trivial. ■

Note importantly that $R$ is an integral domain here! Otherwise, the notion of torsion submodule does not even makes sense in the first place.

**Example.** $R/(a)$ is torsion (as an $R$-module) for all nonzero $a \in R$.

*Proof.* Obvious. Consider $r + (a) \in R/(a)$. Then we have

$$a \cdot (r + (a)) = ra + (a) = 0_R + (a),$$

where the last equality is true since $ra \in (a)$. ■

**Lemma 2.7.** If $M$ is a module over an integral domain, then $M/\operatorname{Tors}(M)$ is torsion-free.

**Proof.** Let $M$ be a module over an integral domain $R$ and suppose that $M/\operatorname{Tors}(M)$ is not torsion-free. Then there exist a nonzero element $m + \operatorname{Tors}(M) \in \operatorname{Tors}(M/\operatorname{Tors}(M))$ i.e. an element $m + \operatorname{Tors}(M)$ such that

$$r \cdot (m + \operatorname{Tors}(M)) = (r \cdot m) + \operatorname{Tors}(M) = 0 + \operatorname{Tors}(M) = 0_{M/\operatorname{Tors}(M)},$$

for some nonzero $r \in R$. This implies that $r \cdot m \in \operatorname{Tors}(M)$. But by definition[19], this implies that there exists a nonzero $s \in R$ such that $s \cdot (r \cdot m) = (sr) \cdot m = 0_M$ . But $R$ is an integral domain and $r, s$ are both nonzero so this implies that $sr \neq 0$. Consequently, the product gives $m = 0_M$. 何! This contradicts the fact that $m + \operatorname{Tors}(M)$ was a nonzero torsion element. ■

[19] *of the torsion submodule of M*

**Lemma 2.8.** Let $M$ be a module over an integral domain $R$ and let $A, B$ be submodules of $M$ such that $M = A \oplus B$. Then

$$\operatorname{Tors}(M) = \operatorname{Tors}(A) \oplus \operatorname{Tors}(B).$$

The torsion submodule of a direct sum is the direct sum of torsion submodules. This is true even for an arbitrary index set.

**Proof.** Suppose $(a, b) \in M$. Then there is a nonzero $r \in R$ such that $r \cdot (a, b) = 0_M$. This implies that $r \cdot a = 0_A$ and $r \cdot b = 0_B$; and so $a \in \operatorname{Tors}(A)$ and $b \in \operatorname{Tors}(B)$. Accordingly, $(a, b) \in \operatorname{Tors}(A) \oplus \operatorname{Tors}(B)$.

Conversely, suppose $(a,b) \in \operatorname{Tors}(A) \oplus \operatorname{Tors}(B)$. Then there are nonzero $r, s \in R$ such that $r \cdot a = 0_A$ and $s \cdot b = 0_B$. Now consider

$$rs \cdot (a,b) = (s \cdot (r \cdot a), r \cdot (s \cdot b)) = (s \cdot 0_A, r \cdot 0_B) = (0_A, 0_B) = 0_M.$$

But $rs \neq 0$ as $R$ is an integral domain, and so $(a,b) \in \operatorname{Tors}(M)$. ■

A direct corollary to this lemma is that if $A$ and $B$ are torsion-free, then so is $M$.

**Proposition 2.6** (*Free ⇒ torsion-free*)**.** Let $M$ be a module over an integral domain $R$. If $M$ is free, then it is torsion-free.

**Proof.** Since $M$ is free, then $M \cong \bigoplus_i R$. Since $R$ is torsion-free (as $R$ is a domain) and the direct sum of torsion-free modules is torsion-free (by the preceding lemma), $M$ is torsion-free. ■

We will see that the converse is also true when $M$ is finitely generated. This is a direct corollary to the Structure Theorem. Before that, it is worth mentioning that an $R$-homomorphism sends torsion elements to torsion elements.

**Lemma 2.9** (*Torsion is sent to torsion*)**.** Let $\phi : M \to N$ be an $R$-homomorphism of an integral domain $R$. If $m$ is a torsion element of $M$, then $\phi(m)$ is a torsion element of $N$.

**Proof.** Suppose $\phi : M \to N$ is an $R$-homomorphism of modules. If $m \in M$ is a torsion element, then there is a nonzero $r \in R$ such that $r \cdot m = 0_M$. Consequently, $\phi(r \cdot m) = 0_N$ as it is a homomorphism. But $\phi(r \cdot m) = r \cdot \phi(m)$ and so the existence of this $r$ gives $\phi(m)$ as a torsion element of $N$. ■

This gives the following result. If $\phi : M \to N$ is an $R$-homomorphism, then it restricts to an $R$-homomorphism $\operatorname{Tors}(M) \to \operatorname{Tors}(N)$. So if $\phi$ is in fact an isomorphism, then $\operatorname{Tors}(M) \cong \operatorname{Tors}(N)$.

**Lemma 2.10.** Let $M$ be a module over an integral domain $R$. If $M \cong T \oplus F$ where $T$ is a torsion $R$-module and $F$ is a free $R$-module, then $T \cong \operatorname{Tors}(M)$.

**Proof.** Since $F$ is free over an integral domain, then it is torsion-free i.e. $\operatorname{Tors}(F) = 0$. Since $M \cong T \oplus F$, then Lemma 2.8 gives

$$\operatorname{Tors}(M) \cong \operatorname{Tors}(T) \oplus \operatorname{Tors}(F) \cong \operatorname{Tors}(T).$$

where we have identified $\operatorname{Tors}(T) \oplus 0 \cong \operatorname{Tors}(T)$.

But $T$ is torsion and so $T = \operatorname{Tors}(T)$. This gives the claim. ■

Two important corollaries to the Structure Theorem are the following.

**Corollary 2.1.** Let $M$ be a finitely generated module over a Euclidean domain $R$. Then $M$ has a direct sum decomposition $M \cong \operatorname{Tors}(M) \oplus F$ where $F$ is a free $R$-module.

In fact, in all its obviousness, $F \cong M/\operatorname{Tors}(M)$.

**Proof.** Recall that $R/(d)$ is torsion for $d \neq 0$. So the Structure Theorem says that $M$ is a decomposition of direct sum of torsion modules and a free module. But the preceding Lemma 2.10 implies that these torsion modules are in fact $\mathrm{Tors}(M)$. ■

**Corollary 2.2** (*Torsion-free ⇒ free*)**.** Let $M$ a finitely generated module over a Euclidean domain $R$. If $M$ is torsion-free, then $M$ is free.

**Proof.** By the preceding Corollary 2.1, $M \cong \mathrm{Tors}(M) \oplus F$ where $F$ is a free $R$-module. If $M$ is torsion-free, then $\mathrm{Tors}(M) = 0$ and $M$ is left with a free part. ■

It is important that $M$ is finitely generated here. A counter example is the following.

**Example.** $\mathbb{Q}$ is a torsion-free $\mathbb{Z}$-module that is not free.

Another corollary is the following.

**Corollary 2.3.** Let $M$ be a finitely generated **free** module over a Euclidean domain $R$. Then every $R$-submodule $N \leqslant M$ is free.

**Proof.** Since $M$ is finitely generated, then any submodule $N \subseteq M$ is also finitely-generated. Since $M$ is free over a domain, then it is torsion-free i.e. $\mathrm{Tors}(M) = 0$. But since $\mathrm{Tors}(N) \subseteq \mathrm{Tors}(M)$, $N$ is also torsion-free and so is free by Corollary 2.2. ■

**Slogan.** Submodules of a free f.g. module over Euclidean are free.

This last corollary leads us to the submodule structure theorem.

**Theorem 2.7** (*Submodule structure theorem*)**.** Let $M$ be a finitely generated **free** module over a Euclidean domain $R$; and let $N \leqslant M$ be a submodule of $M$. Then there exists a basis $\{e_1, \dots, e_m\}$ for $M$ and $d_1, \dots, d_m \in R$ such that $\{d_1 e_1, \dots, d_m e_m\}$ is a basis for $N$ (after deletion of terms that are zero).

The proof for this theorem should feel like a deja vu, we are really proving it in the same fashion as proving the structure theorem for finitely generated modules.

**Proof.** Since $M$ is a finitely generated free module over a Euclidean domain, the preceding Corollary 2.3 implies that $N$ is free. So $M \cong R^m$ and $N \cong R^n$ for some $m, n \in \mathbb{N}$. WLOG[20], we can assume that $m \geqslant n$. Consider the inclusion $\iota : N \hookrightarrow M$ which is an $R$-homomorphism. Since $M$ and $N$ are free, they both have a basis which under $\iota$ gives rise to a matrix $\Pi$. Smith normal form implies that we can put $\Pi$ into a diagonal matrix

$$\Pi' = \left[\begin{array}{cccc} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \hline 0 & \cdots & \cdots & 0 \end{array}\right],$$

with new basis $\{f_1, \dots, f_n\}$ for $N$ and $\{e_1, \dots, e_m\}$ for $M$. That is, under $\iota$ we now have $\iota(f_i) = d_i e_i$ for $i = 1, \dots, n$. This gives $\{d_1 e_1, \dots, d_m e_m\}$ as basis for $N$ (after deleting terms that are zero). ■

The $f_i$ already gives a basis for $N$ and $\iota$ is just the inclusion map. Since $\iota(f_i) = d_i e_i$, we have in fact $f_i = d_i e_i$, and so the $d_i e_i$ gives a basis for $N$.

[20] *in fact, this is a truth but it requires a proof.*

**Remark.** Note that all we really need was the Smith normal form here.

### 2.4.1 Structure theorem for modules over $k[X]$

This section would feel more of a digression rather than a whole section. Let's talk about the structure theorem for modules when the Euclidean domain is $k[X]$ where $k$ is a field. Suppose $V$ is a finite-dimensional vector space over $k$ and $\phi$ is an endomorphism of $V$. Observe that $V$ is a $k[X]$-module if we define scalar multiplication in the following way:

$$\left(\sum_{i=1}^{n} a_i X^i\right) \cdot v = \sum_{i=1}^{n} a_i \phi^i(v).$$

That is, we can think of the scalar multiplication really as $X \cdot v = \phi(v)$.

**Proposition 2.7.** Let $V$ be a finite-dimensional vector space over a field $k$ and let $\phi$ be an endomorphism of $V$. Then $V$ is a $k[X]$-module with scalar multiplication defined above.

If we write $A = k[X]$, by the structure theorem, we have an isomorphism

$$V \cong A/(a_1) \oplus \cdots \oplus A/(a_r) \oplus A^n,$$

where $a_1 \mid a_2 \mid \cdots \mid a_r$, all elements of $A$, and $n \geqslant 0$. Since $V$ is finite-dimensional as a $k$-vector space, then we must have $n = 0$, for otherwise $V$ has to be infinite dimensional. So we in fact have

$$V \cong A/(a_1) \oplus \cdots \oplus A/(a_r).$$

The question now is, what does these quotient modules look like? That is, if $a \in A$ is nonzero, then what does $A/(a)$ look like?

Firstly, observe that $A/(a)$ is a $k$-vector space with an action of $X$. That is, $X : A/(a) \to A/(a)$ defines a $k$-linear map. Suppose that

$$a = r_0 + r_1 X + \cdots + r_{m-1} X^{m-1} + X^m,$$

We know from our introduction to modules that if $R$ is a ring and $I \lhd R$, then $R/I$ has a $R$-module structure.

where $\deg a = m$. Then $\left\{1, X, X^2, \ldots, X^{m-1}\right\}$ is a $k$-basis of $A/(a)$. The matrix of $X$ with respect to this basis is

$$M = \begin{pmatrix} 0 & 0 & \cdots & 0 & -r_0 \\ 1 & 0 & \cdots & 0 & -r_1 \\ 0 & 1 & \cdots & 0 & -r_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -r_{m-1} \end{pmatrix}.$$

Now go back to $V \cong A/(a_1) \oplus \cdots \oplus A/(a_r)$, then the matrix of $X = \phi$ is in block form: there are blocks $M_1, \ldots, M_r$ such that each block looks like $M$. This is called the **rational canonical form** of $X$. And over any field, a linear map $\phi$ can be put into such a form.

## 2.5 Noetherian modules

Before we talk about more interesting things, let's make a definition of Noetherian modules.

**Definition** (*Noetherian modules*)**.** An $R$-module of a ring $R$ is said to be **Noetherian** if every of its submodule is finitely generated.

Just like Noetherian rings, the Noetherian modules satisfies the ACC conditions on its submodules. We are in fact interested in a few basic results.

**Proposition 2.8** (*Submodule of noetherian is noetherian*)**.** Let $M$ be a Noetherian $R$-module. Then every submodule of $M$ is Noetherian.

**Proof.** If $M$ is Noetherian, then it itself viewed as a submodule is finitely generated. Consequently, any of its submodule (i.e. viewed as submodule of a submodule) is finitely generated so they are Noetherian themselves. ■

**Proposition 2.9.** $R$ is a Noetherian ring *if and only if* it is a Noetherian $R$-module.

**Proof.** If $R$ is a Noetherian ring, then all its ideals are finitely generated. If $R$ is viewed as an $R$-module, its submodules are its ideals. Moreover, the idea of being finitely generated coincides[21]. So all its submodules are finitely generated i.e. $R$ is a Noetherian $R$-module. Now, note that the implications also work backwards, we are done. ■

We will state without proof the following two propositions.

**Proposition 2.10.** Let

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

be a short exact sequence of $R$-modules. Then $B$ is Noetherian *if and only if* both $A$ and $C$ are Noetherian.

**Proposition 2.11.** Let $R$ be a Noetherian ring and let $M$ be an $R$-module. Then $M$ is a Noetherian module *if and only if* it is finitely generated (as an $R$-module).

[21] *to see this, just rewrite what it means to be f.g. as an ideal and a submodule.*

# 3 Rings finite over subrings

So we are done with establishing the foundation that we need to discuss more interesting things. From here on out, we will be mostly applying what we have, taking a detour in the next chapter to discuss unique factorization, and then get back on track to prove very useful facts that gives rise to algebraic number theory and algebraic geometry.

## 3.1 A few definitions

Here, we will start to use modules as a tool.

**Definition** (*Finite over subring*)**.** Let $R$ be a ring and $S \leqslant R$ a subring. Then $R$ is **finite over $S$** if $R$ is a finitely generated $S$-module.

Note that this definition is much stronger than saying that $R$ is finitely generated as just a ring over $S$ (or equivalently, saying that $R$ is an $S$-algebra). To see this, consider the following example.

**Example.** The polynomial ring $S[X]$ is finitely generated (by a single generator element $X$) as a ring over $S \leqslant S[X]$ but **not** as an $S$-module.

*Proof.* Suppose for contradiction that $S[X]$ is finitely generated as an $S$-module by $f_1, \dots, f_r$. Now choose $N > \deg f_i$ for all $i$ and consider $X^N$. Since everything in $S[X]$ is an $S$-linear combination of $f_1, \dots, f_r$, we can write

Note that $f_i \in S[X]$ and so are polynomials.

$$X^N = \sum_{i=1}^{r} a_i f_i,$$

for some $a_i \in S$. But then we can rewrite this as

$$X^N - \sum_{i=1}^{r} a_i f_i = 0.$$

By hypothesis, the lhs is a polynomial in $X$ of degree $N$ since there are no terms in $\sum a_i f_i$ that can cancel $X^N$. But the rhs is zero, so the above implies that polynomials of degree $N$ are identically 0 in $S[X]$. 何!

The converse, however, is surely true. If $R$ is finitely generated as a module over $S$, it is finitely generated as a ring over $S$.

**Lemma 3.1** (*Tower Law*)**.** Let $A, B, C$ be rings such that $A \leqslant B \leqslant C$. If $C$ is finite over $B$ and $B$ is finite over $A$, then $C$ is finite over $A$.

**Proof.** Suppose $x_1, \dots, x_m \in B$ generates $B$ as an $A$ module; and suppose $y_1, \dots, y_n \in C$ generates $C$ as a $B$-module. Then all possible products $\{x_i y_j\}$ generates $C$ as an $A$-module. ■

**Definition** (*Integral over subring*)**.** Let $R$ be a ring and $S \leqslant R$. An element $r \in R$ is **integral over $S$** if there exists a monic polynomial $f \in S[X]$ such that $f(r) = 0$. If every element of $R$ is integral over $S$, we say that **$R$ is integral over $S$**.

where $0 = 0_S$ as we are saying that $f(r)$ is the zero polynomial with coefficients in $S$.

In number theory, we would be most interested in the case $S = \mathbb{Z}$ and $R = \mathbb{C}$ i.e. complex numbers integral over the integers. Such an integral element has a special name and is called an **algebraic integer**. This is a special case of the following.

**Definition** (*Algebraic over subfields*)**.** Let $L$ be a field and $K \leqslant L$ be a subfield. An element $\alpha \in L$ is said to be **algebraic over $K$** if it is integral over $K$.

So an algebraic element is an integral element, we just need fields. The following lemma characterizes the interplay between finite and integral over subrings.

**Lemma 3.2.** Let $R$ be a ring and $S \leqslant R$. Then $\lambda \in R$ is integral over $S$ *if and only if* $S[\lambda]$ is finite over $S$.

**Proof.** $(\Rightarrow)$. Suppose $\lambda \in R$ is integral over $S$. Then there exists a monic polynomial $f(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_0$ with coefficients in $S$ such that $f(\lambda) = 0$. Accordingly, $\lambda^n$ can[22] be written as an $S$-linear combination of lower powers of $\lambda$. This implies that $\left\{1, \lambda, \ldots, \lambda^{n-1}\right\}$ generates $S[\lambda]$ over $S$ or in other words, $S[\lambda]$ is finite over $S$, as desired.

$(\Leftarrow)$. Suppose $S[\lambda]$ is finite over $S$. That is, $S[\lambda]$ is finitely generated as an $S$-module, say, by $s_1, \ldots, s_k$. Each $s_i$ is a polynomial in the variable $\lambda$ with coefficients in $S$ and with degree $n_i$. Now consider $N \in \mathbb{N}$ such that $N > \max_{1 \leqslant i \leqslant k} n_i$. So $\lambda^N$ is an $R$-linear combination of $s_1, \ldots, s_k$. This gives a monic polynomial of degree $N$ whose root is $\lambda$. ■

**Remark.** Remembering this would make your life easier. As per the proof above, always remember that $S[\lambda]$ over $S$ is finitely generated by $\left\{1, \lambda, \ldots, \lambda^{n-1}\right\}$ for some $n \in \mathbb{N}$. We are not saying it's unique; but we are saying that it's the most obvious generating set.

**Theorem 3.1.** Let $R$ be a ring and $S \leqslant R$. If $\lambda_1, \ldots, \lambda_n \in R$ are integral over $S$, then $S[\lambda_1, \ldots, \lambda_n]$ is finite over $S$.

**Proof.** We prove by induction on $n$. The base case $n = 1$ is the preceding lemma.

Inductive step. Define $S_k := S[\lambda_1, \ldots, \lambda_k]$ for any $k \in \mathbb{N}$. Then we can write $S_n = S_{n-1}[\lambda_n]$. Clearly, $S_n$ is finite over $S_{n-1}$. But the induction hypothesis gives that $S_{n-1}$ is finite over $S$. By the tower law, we thus have $S_n = S[\lambda_1, \ldots, \lambda_n]$ is finite over $S$, as desired. ■

**Corollary 3.1.** Let $S$ be a Noetherian ring where $S \leqslant R$, and let $x, y \in R$ such that they are integral over $S$. Then $x + y$, $x - y$ and $x \cdot y$ are integral over $S$.

**Proof.** By the preceding theorem, since $x, y$ are integral over $S$, then $S[x, y]$ is finite over $S$. Denote $\star$ to be one of addition, subtraction or multiplication. Then observe that $S[x \star y] \subseteq S[x, y]$. Since $S$ is a Noetherian ring, then Hilbert's basis theorem implies that $S[x \star y]$ is Noetherian. By Proposition 2.9 and Proposition 2.11 combined, $S[x \star y]$ is a finitely generated $S$-module; and so $x \star y$ is integral over $S$. ■

[22] *what this means is the following: we can write $f(X) = X^n + g(X)$ for some degree $n-1$ polynomial $g \in S[X]$. So $0 = f(\lambda) = \lambda^n + g(\lambda)$ and thus $\lambda^n = -g(\lambda)$.*

**Corollary 3.2.** Let $S$ be a Noetherian ring where $S \leqslant R$. Then the set

$$D = \{x \in R \mid x \text{ is integral over } S\}$$

is a subring of $R$.

**Proof.** We only need to show that if $x, y \in D$ then $x+y$, $x-y$ and $xy$ are all elements of $D$. But this is the previous corollary. ■

The ring $D$ holds a lot of interest for us so we give it a name.

**Definition** (*Integral closure*)**.** Let $R$ be a ring and $S \leqslant R$ such that $S$ is Noetherian. The ring

$$D = \{x \in R \mid x \text{ is integral over } S\},$$

is called the **integral closure of $S$ in $R$**. If $D = S$, we say that $S$ is **integrally closed** in $R$. If $D = R$, we recover the definition of $R$ being integral over $S$.

Recall that fields are Noetherian, so it should be well expected that there is an analogous notion of the above for fields. Basically, we just replace every instance of "integral" with "algebraic".

**Definition** (*Algebraic closure*)**.** Let $L$ be a field and $K \leqslant L$ is a subfield. Then

$$D = \{x \in L \mid x \text{ is algebraic over } K\},$$

is called the **algebraic closure of $K$ in $L$**. If $D = K$, we say that $K$ is **algebraically closed** in $L$. If $D = L$, then we say that **$L$ is algebraic over $K$**.

**Example.** The complex numbers $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$ (in $\mathbb{C}$). Moreover, $\mathbb{C}$ is algebraically closed (in $\mathbb{C}$). This is the so-called Fundamental Theorem of Algebra.

An equivalent definition for algebraically closed is the following. The proof is straight from definitions.

**Proposition 3.1.** Let $K$ be a field. Then $K$ is said to be algebraically closed if every non-constant polynomial in $K[X]$ has a root in $K$.

An important theorem regarding algebraically closed fields is the following.

**Theorem 3.2.** If $K$ is an algebraically closed field, then $K$ is an infinite field.

**Proof.** Suppose $K$ is an algebraically closed field that is finite. Now consider the polynomial $f \in K[X]$ defined by

$$f(X) = 1 + \prod_{a \in K} (X - a).$$

Surely, $f$ is non-constant and is well-defined. But now observe that for all $a \in K$, we have $f(a) = 1$. So there does not exist an element in $K$ such that $f$ vanishes. In other words, $f$ has no root in $K$. 何! ■

## 3.2 Towards algebraic number theory

Fun time has arrived. Our goal is now to prove that in certain circumstances, this integral closure $D$ of $S$ in $R$ is in fact finite over $S$. For us, we only care when $S = \mathbb{Z}$ and $R$ is a field $K$ such that $\mathbb{Q} \subseteq K$ (i.e. it is a field extension of $\mathbb{Q}$) and $K$ is finite over $\mathbb{Q}$. That is, $R$ is a **finite field extension** of $\mathbb{Q}$. For example, let $\alpha \in \mathbb{C}$ which is integral over $\mathbb{Q}$ and consider $K = \mathbb{Q}[\alpha]$.

**Definition** (*Number field*)**.** Let $K \supseteq \mathbb{Q}$ be a field. We say $K$ is an (algebraic) **number field** if $K$ is finite over $\mathbb{Q}$.

That is, $K$ is finitely generated as a $\mathbb{Q}$-module i.e. we can write informally $K = \sum_i \mathbb{Q}k_i$ for finite generators $k_i$ of $K$. Since $\mathbb{Q}$ is a field, $K$ is essentially a finite dimensional $\mathbb{Q}$-vector space and its finite generators gives a finite basis of it.

**Definition** (*Degree of number field*)**.** Let $K$ be a number field. The **degree** of $K$ denoted $\dim_{\mathbb{Q}} K$ or $[K : \mathbb{Q}]$ is the dimension of $K$ as a $\mathbb{Q}$-vector space. By definition of a number field $[K : \mathbb{Q}] < \infty$ is always true.

i.e. it is the cardinality of its basis.

**Example** (*Most obvious example*)**.** $\mathbb{Q}$ itself is a number field with degree 1.

**Example.** The Gaussian field $\mathbb{Q}[i] = \{a + bi \mid a, b \in \mathbb{Q}\}$ is a number field.

**Example** (*Shocking non-example*)**.** $\mathbb{R}$ and $\mathbb{C}$ are **not** number fields as they are not finitely generated over $\mathbb{Q}$.

*Slogan*: How on earth are you going to write transcendetal numbers as a $\mathbb{Q}$-linear combination?

*Proof.* Suppose for contradiction that $R$ is finitely generated by $r_1, \dots, r_n$ over $\mathbb{Q}$. So any element $r \in R$ can be written as a $\mathbb{Q}$-linear combination

$$r = \sum_{i=1}^{n} q_i r_i,$$

for some $q_i \in \mathbb{Q}$. Since $R$ is a field, this linear combination can be written uniquely. That is, there is a bijection $\mathbb{Q}^n \to \mathbb{R}$ given by $(q_1, \dots, q_n) \longmapsto \sum q_i r_i$. 何! This is a contradiction as $\mathbb{Q}^n$ is countable but $\mathbb{R}$ is uncountable. Since $\mathbb{C} \supseteq \mathbb{R}$, the idea holds for $\mathbb{C}$. ■

More interesting number fields arise as some complex number $\alpha \in \mathbb{C}$ adjoint to $\mathbb{Q}$.

**Definition** (*Ring of integers*)**.** Let $K$ be a number field. The **ring of integers** of $K$, denoted $\mathcal{O}_K$, is the integral closure of $\mathbb{Z}$ in $K$.

So $\mathcal{O}_K$ is basically the set $\{x \in K \mid x \text{ is integral over } \mathbb{Z}\}$.

**Example.** The simplest ring of integers is $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$, the integers itself.

**Example.** The second simplest ring of integers is $\mathcal{O}_{\mathbb{Q}[i]} = \mathbb{Z}[i]$, the Gaussian integers.

If $K = \mathbb{C}$, we call the elements of $\mathcal{O}_{\mathbb{C}}$ *algebraic integers* as we've mentioned previously.

**Definition** (*Algebraic integers*)**.** An element of $\mathcal{O}_{\mathbb{C}}$ is called an **algebraic integer**. They are $\alpha \in \mathbb{C}$ such that there is a monic polynomial $f \in \mathbb{Z}[X]$ with $f(\alpha) = 0$.

Before we move to more theory, let's lay down a few more definitions.

**Definition** (*Irreducible*)**.** Let $R$ be an integral domain. An element $r \in R$ is said to be **irreducible** in $R$ if it satisfies

(1). $r$ is nonzero.

(2). $r$ is not a unit.

(3). If $r = st$ for some $s, t \in R$, then either $s$ or $t$ is a unit.

We say $r$ is **reducible** if it is not irreducible.

Actually, the condition that $r$ is nonzero is superfluous. The condition (2) and (3) implies that an irreducible cannot be 0 as it's not a unit and if $0 = st$, then $s = t = 0$ both of which are not units.

Note that we define irreducibility of an element on a general integral domain and not on number fields. We can think of irreducible elements as generalization of the prime numbers in $\mathbb{Z}$. Accordingly, they are the first natural example.

**Example.** The prime numbers $p$ and its additive inverse $-p$ are the irreducible elements of $\mathbb{Z}$.

**Example.** Let $K$ be a field and let $f \in K[X]$ be a nonzero polynomial. Here's a handy cheat sheet of when $f$ is irreducible in $K[X]$:

| $\deg f$ | $f$ irreducible in $K[X]$ |
|---|---|
| 0 | Never |
| 1 | Always |
| 2 or 3 | Sometimes[23] |
| $\geqslant 4$ | Hard |

To see why this is the case for degree 0 and 1 polynomials, let's look a bit deeper.

**Proposition 3.2** (*Units in polynomial rings over field*)**.** Let $K$ be a field. Then a nonzero polynomial $f \in K[X]$ is a unit *if and only if* $\deg f = 0$.

That is, units in $K[X]$ are nonzero constant polynomials. In other words, all nonzero elements of $K$.

**Proof.** ($\Leftarrow$). Let $f \in K[X]$ be a unit in $K[X]$. Then there exists $g \in K[X]$ such that $fg = 1$. This implies that

$$\deg f + \deg g = \deg(fg) = \deg 1 = 0,$$

so $\deg f = \deg g = 0$.

($\Rightarrow$). Suppose $f \in K[X]$ is nonzero and such that $\deg f = 0$. Then it is a nonzero constant polynomial i.e. $f \in K \setminus \{0\}$. But $K$ is a field and so $f$ is a unit in $K$. Consequently, there is an element $g \in K \subseteq K[X]$ such that $fg = 1$. So $f$ is a unit in $K[X]$. ■

**Slogan.** Units of field$[X]$ are precisely the units of field; and vice-versa.

Consequently, we now know that degree 0 polynomials are reducible in $K[X]$ as they are units. How about degree 1 polynomials?

**Proposition 3.3.** Let $K$ be a field and let $f \in K[X]$ such that $\deg f = 1$. Then $f$ is irreducible.

[23] *If there does not exists $k \in K$ such that $f(k) = 0$.*

**Proof.** Suppose $f = gh$ for some $g, h \in K[X]$. Then

$$1 = \deg f = \deg(gh) = \deg(g) + \deg(h),$$

where the last equality is due to the fact that $K$ is a field. So $\deg g$ and $\deg h$ can be a combination of either 0 and 1 or 1 and 0, whichever with degree 0 being a unit. ■

**Proposition 3.4.** Let $K$ be a field and let $f \in K[X]$ such that $\deg f$ is 2 or 3. Then $f$ is irreducible *if and only if* for any $\lambda \in K$, $f(\lambda) \neq 0$.

Note that the notion of irreducibility depends on the domain $R$.

**Example.** $f(X) = X^2 + 1$ is irreducible in $\mathbb{R}[X]$ but is reducible in $\mathbb{C}[X]$ as $f(i) = 0$.

Next, we talk about minimal polynomials.

**Definition** (*Minimal polynomial*)**.** Let $K \supseteq \mathbb{Q}$ be a number field and let $\alpha \in K$ be integral over $\mathbb{Q}$. The **minimal polynomial** of $\alpha$, denoted $m_\alpha$, is the (unique) monic polynomial with coefficients in $\mathbb{Q}$ of least degree among all other polynomials in $\mathbb{Q}[X]$ that are satisfied by $\alpha$.

The definition of minimal polynomial is well-defined as $\alpha$ being integral over $\mathbb{Q}$ implies the existence of at least one such monic polynomial.

**Lemma 3.3** (*Min. polynomial is irreducible*)**.** Let $K$ be a number field and let $\alpha \in K$ be integral over $\mathbb{Q}$. Then $m_\alpha$ is irreducible.

**Proof.** Suppose not. Then there exists nonzero $g, h \in \mathbb{Q}[X]$ such that $m_\alpha = gh$ where both $g, h$ are not units. Clearly, the degrees of $g$ and $h$ are lower than $m_\alpha$ for otherwise one of them is a unit. Now $m_\alpha(\alpha) = g(\alpha)h(\alpha) = 0$. Since $\mathbb{Q}[X]$ is an integral domain, this implies that $g(\alpha) = 0$ or $h(\alpha) = 0$. 何! This contradicts the minimality of $m_\alpha$. ■

**Lemma 3.4** (*Min. polynomial is unique*)**.** Let $K$ be a number field and let $\alpha \in K$ be integral over $\mathbb{Q}$. Then $m_\alpha$ is unique.

**Proof.** Suppose $g \in K[X]$ is a monic polynomial of minimal degree different than $m_\alpha$ such that $g(\alpha) = 0$. By the division algorithm, there exist polynomials $q, r \in K[X]$ such that $g = qm_\alpha + r$ with $\deg r < \deg m_\alpha$ or $r = 0$. If $r \neq 0$, then evaluating at $\alpha$ gives $r(\alpha) = 0$. 何! This contradicts minimality of $m_\alpha$, so $r = 0$. Consequently, $g = qm_\alpha$. But $g, m_\alpha$ are both monic of the same degree, so $q = 1$. ■

We would be most interested in the case $K = \mathbb{Q}[\alpha]$ for some $\alpha \in \mathbb{C}$. We will blindly assume that there always exist such an $\alpha$ for now, but the existence of such $\alpha$ is guaranteed by the so-called *Primitive Element Theorem*. Let us recall that the obvious generating set for $S[\alpha] \leqslant S$ when it is finite over $S$ is given by $\{1, \alpha, \dots, \alpha^{n-1}\}$ for some $n \in \mathbb{N}$. If we put $S = \mathbb{Q}$, then we are talking about vector spaces so this generating set in fact gives a basis.

**Proposition 3.5.** Let $K = \mathbb{Q}[\alpha]$ be a number field where $\alpha \in \mathbb{C}$. Then there is an isomorphism

$$K \cong \mathbb{Q}[X]/(m_\alpha).$$

If moreover that $[K:\mathbb{Q}] = n$, then $\deg m_\alpha = n$ and $m_\alpha$ has $n$ distinct roots in $\mathbb{C}$.

**Proof.** Consider the evaluation homomorphism $\phi : \mathbb{Q}[X] \to K$ given by $f(x) \mapsto f(\alpha)$. This map is surjective by construction. Now observe that

$$f \in \ker \phi \iff f(\alpha) = 0 \iff m_\alpha \mid f(x).$$

That is, $\ker \phi = (m_\alpha)$ the principal ideal generated by $m_\alpha$. By the First Isomorphism Theorem, we have $K \cong \mathbb{Q}[X]/(m_\alpha)$, as desired. We leave the second part to the reader. ■

**Example.** Consider $K = \mathbb{Q}[\sqrt{-5}]$. Then $[K:\mathbb{Q}] = 2$ as $K$ is finitely generated by 1 and $\sqrt{-5}$. The minimal polynomial of $\sqrt{-5}$ is $m_{\sqrt{-5}} = X^2 + 5$ whose roots are $\sqrt{-5}$ and its conjugate.

**Definition** (*Embeddings*)**.** Let $F$ be a field. An **embedding** of a field $K$ in $F$ is a ring homomorphism $\sigma : K \to F$.

Note that we have defined embeddings on fields, not necessarily on number fields.

An embedding is necessarily injective.

**Lemma 3.5.** An embedding $\sigma : K \hookrightarrow F$ is injective.

Note that we have used the $\hookrightarrow$ notation instead of $\to$. This is a nice piece of notation to mean that the map is injective.

**Proof.** Since $K$ is a field, its only ideals are $\{0\}$ and itself. Consequently[24], $\ker \sigma$ is one of these ideals. Since $\sigma$ is a homomorphism, $\sigma(1_K) = 1_F$ and not $0_F$. This implies that $\ker \sigma \neq K$ as $1_K \notin \ker \sigma$. So $\ker \sigma$ is trivial and thus $\sigma$ is injective. ■

**Proposition 3.6.** Let $K = \mathbb{Q}[\alpha]$ be a number field where $\alpha \in \mathbb{C}$ with $[K : \mathbb{Q}] = n$. Then there are exactly $n$ distinct embeddings $\sigma_1, \ldots, \sigma_n : K \hookrightarrow \mathbb{C}$.

**Proof.** Let $\zeta_1, \ldots, \zeta_n$ be the $n$ (distinct) roots of $m_\alpha$ in $\mathbb{C}$. Then we have $n$ embeddings $\sigma_i : K \hookrightarrow \mathbb{C}$ which (necessarily) restricts to the identity map on $\mathbb{Q}$ and maps $\alpha \mapsto \zeta_i$. To see this, let

$$m_\alpha = \sum_{j=0} a_j X^j,$$

with $a_j \in \mathbb{Q}$. By definition

$$m_\alpha(\alpha) = \sum_{j=0} a_j \alpha^j = 0$$

Applying $\sigma_i$ to both sides of the second equality, we have

$$\sigma_i\left(\sum_{j=0}^{n} a_j \alpha^j\right) = \sum_{j=0}^{n} \sigma_i(a_j)\sigma_i(\alpha)^j = \sum_{j=0}^{n} a_j \sigma_i(\alpha)^j = 0,$$

[24] *as we know that the kernel of a ring homomorphism is an ideal.*

where we have used the fact that $\sigma_i$ restricts to the identity map on $\mathbb{Q}$, so $\sigma_i(a_j) = a_j$ as $a_j \in \mathbb{Q}$. This implies that $\sigma_i(\alpha)$ is a root of $m_\alpha$. So $\sigma_i(\alpha)$ is one of $\zeta_i$. If $\tau$ is any other embedding $K \hookrightarrow \mathbb{C}$, it is easy to see by the above argument that it must coincide with $\sigma_i$ for some $i$. So these $\sigma_i$ gives all the embeddings. ■

A slightly different approach is the following. We know that if $K = \mathbb{Q}[\alpha]$ has degree $[K : \mathbb{Q}] = n$, then $K$ has a basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$. So if we know what happens to $\alpha$, we know what happens to $K$. To see this, any element of $K$ can be written uniquely as

$$\sum_{j=0}^{n-1} q_j \alpha^j,$$

for some $q_j \in \mathbb{Q}$. If $\zeta_1, \ldots, \zeta_n$ are all the roots of $m_\alpha$ in $\mathbb{C}$, then we obtain $n$ embeddings $\sigma_1, \ldots, \sigma_n$ simply by

$$\sum_{j=0}^{n-1} q_j \alpha^j \xrightarrow{\sigma_i} \sum_{j=0}^{n-1} q_j \zeta^j,$$

where $\zeta = \zeta_i$ for some $1 \leqslant i \leqslant n$. If $\tau$ is any other embedding, the same argument as our first proof gives that $\tau(\alpha)$ is a root of $m_\alpha$ and thus, $\tau$ must coincide with one of $\sigma_i$.

**Example.** Consider again $K = \mathbb{Q}[\sqrt{-5}]$. The two embeddings $K \hookrightarrow \mathbb{C}$ are

$$a + b\sqrt{-5} \longmapsto \begin{cases} a + \sqrt{-5}\, b, \\ a - \sqrt{-5}\, b. \end{cases}$$

By using the previous notation, here $\zeta_1 = \sqrt{-5}$ and $\zeta_2 = -\sqrt{-5}$, both of which are the two roots of $m_{\sqrt{-5}} = X^2 + 5$.

The notion of embeddings is important in field theory, in particular, in the development towards Galois theory. Let us now recall the notion of a characteristic polynomial from linear algebra.

**Definition** (*Char. poly. of matrix*)**.** Let $A$ be an $n \times n$ matrix with coefficients in some ring $R$. The **characteristic polynomial** of $A$ is the polynomial

$$\chi_A(t) = \det(tI - A).$$

Recall that any linear map of vector spaces (even finitely generated modules) gives rise to a matrix just by looking at coefficients of the basis's image. We will be particularly interested in one specific map.

**Definition.** Let $V$ be a vector space over some field $K$, and let $\alpha \in V$. Consider the linear map

$$\begin{aligned} \mu_{V,\alpha} : V &\to V \\ v &\mapsto \alpha \cdot v \end{aligned}$$

of $K$-vector spaces, and let $M_\alpha$ be the matrix of $\mu_{V,\alpha}$ in some basis. We will usually just write $\mu_\alpha$ when $V$ is clear. The **characteristic polynomial of $\alpha$**, denoted $\chi_{V,\alpha}$ or just $\chi_\alpha$, is defined to be the characteristic polynomial of $M_\alpha$. The **norm** of $\alpha$ is defined to be

$$N_{V/K}(\alpha) = \det M_\alpha \in K,$$

and the **trace** of $\alpha$ is defined to be

$$\operatorname{Tr}_{V/K}(\alpha) = \operatorname{Tr}(M_\alpha) \in K.$$

Again, we drop the subscript $V/K$ when $K$ is clear.

The notation $V/K$ actually has a very precise meaning. For now, just treat this as just a notation that has nothing to do with quotienting out stuff.

Taking $V = \mathbb{Q}[\alpha]$ for some $\alpha \in \mathbb{C}$ which gives a number field and putting $K = \mathbb{Q}$, we can now talk about characteristic polynomials, norms and traces of elements of number fields.

Let $K = \mathbb{Q}[\alpha]$ be a number field where $\alpha \in \mathbb{C}$. Recall that $K$ has the canonical basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Further suppose we have the minimal polynomial of $\alpha$

$$m_\alpha = X^n + a_{n-1}X^{n-1} + a_1 X + a_0, \tag{3.1}$$

where $a_i \in \mathbb{Q}$. Then you can convince yourself that the matrix of $\mu_{K,\alpha}$ (as in the last definition above) with respect to this canonical basis is

$$M_\alpha = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}. \tag{3.2}$$

Observe that the diagonal terms are all zero except the $(n,n)$ entry which contains $-a_{n-1}$. This matrix should be familiar (if not, cf. rational canonical form).

We now claim that the characteristic polynomial of $\alpha$ is exactly the minimal polynomial $m_\alpha$. This can be easily done using the explicit matrix we have now.

**Proposition 3.7.** Let $K = \mathbb{Q}[\alpha]$ be a number field where $\alpha \in \mathbb{C}$. Then $\chi_{K,\alpha} = m_{\mathbb{Q},\alpha}$.

**Proof.** Just proof by induction on the size of the matrix. Compute the characteristic polynomial of (3.2) and expand along the top row. We then have a linear combination of a similar-looking matrix and an upper-triangular matrix whose determinant is easy. By induction, the determinant of the similar-looking matrix is now in a nice polynomial form. Expanding things, we see that it really coincides with $m_\alpha$. ■

**Definition** (*Conjugate*)**.** Let $K = \mathbb{Q}[\alpha]$ where $\alpha \in \mathbb{C}$. The **conjugate elements** of $\alpha$ are the roots of $m_\alpha$ in $\mathbb{C}$. We may also simply call this the conjugates.

Recall that there are $n$ different embeddings $\sigma_1, \ldots, \sigma_n : K \hookrightarrow \mathbb{C}$ where $\sigma_i(\alpha)$ is a root of $m_\alpha$ in no particular order. So the conjugate elements of $\alpha$ are $\sigma_1(\alpha), \ldots, \sigma_n(\alpha)$.

**Example.** Let $K = \mathbb{Q}[\sqrt{-5}]$. There are two embeddings $\sigma_1, \sigma_2$ on $K$ as $[K : \mathbb{Q}] = 2$ and the conjugates of $\sqrt{-5}$ are $\sigma_1(\sqrt{-5}) = \sqrt{-5}$ and $\sigma_2(\sqrt{-5}) = -\sqrt{-5}$.

Since we now know that $m_\alpha = \chi_\alpha$, we see that the conjugate elements of $\alpha$ are really the eigenvalues of the $\mathbb{Q}$-linear map $\mu_\alpha : K \to K$. In particular, we have that

$$\operatorname{Tr}(\alpha) = \sum_{j=1}^{n} \sigma_j(\alpha).$$

But if $m_\alpha$ is given explicitly by (3.1), then we also have

$$\mathrm{Tr}(\alpha) = -a_{n-1}.$$

**Proposition 3.8.** Let $K = \mathbb{Q}[\alpha]$ be a number field where $\alpha \in \mathbb{C}$. Then if

$$\chi_\alpha = X^n + a_{n-1}X^{n-1} + a_1 X + a_0,$$

we have $\mathrm{Tr}(\alpha) = -a_{n-1}$ and $N(\alpha) = (-1)^n a_0$.

Let $K = \mathbb{Q}[\alpha]$ where $\alpha \in \mathbb{C}$ and now suppose $\beta \in K$. Put $L = \mathbb{Q}[\beta]$. Clearly, we have $\mathbb{Q} \subseteq L \subseteq K$ and $L$ is a number field; so $L \cong \mathbb{Q}[X]/(m_{\mathbb{Q},\beta})$. Now stare at the minimal polynomial $m_{\mathbb{Q},\beta}$. By our previous result, we have $\chi_{L,\beta} = m_{\mathbb{Q},\beta}$, but what about $\chi_{K,\beta}$?

**Lemma 3.6.** Let $L \subseteq K$ be number fields where $L = \mathbb{Q}[\beta]$ for some $\beta \in K$. Then

$$\chi_{K,\beta}(X) = m_{\mathbb{Q},\beta}(X)^{[K:L]}.$$

Consequently, we have that $\mathrm{Tr}(\beta) = \sum_{j=1}^n \sigma_j(\beta)$ where $\sigma_i(\beta)$ are the roots of $m_\beta$ in $\mathbb{C}$. Note closely here that the power is $[K:L] = \dim_L K$ meaning the dimension of $K$ as a vector space over $L$.

**Proof.** Let $\{e_1, \dots, e_n\}$ be a $\mathbb{Q}$-basis for $L$ and let $M_\beta$ be the matrix for $\mu_{L,\beta} : L \to L$ with respect to this basis. Let $\{f_1, \dots, f_m\}$ be an $L$-basis for $K$. Using the tower law, all possible products $\{e_i f_j\}$ for $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$ gives a $\mathbb{Q}$-basis for $K$. The matrix for $\mu_{K,\beta} : K \to K$ with respect to this new basis is the $m \times m$ block matrix

$$\begin{pmatrix} M_\beta & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & M_\beta & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & M_\beta \end{pmatrix}.$$

It is then easy to conclude that

$$\chi_{K,\beta}(X) = \det(XI - M_\beta)^m = \chi_{L,\beta}(X)^m = \chi_{L,\beta}(X)^{[K:L]}.$$

By using the result that $\chi_{L,\beta} = m_{\mathbb{Q},\beta}$, we are done. ■

Recall that the ring of integers $\mathcal{O}_K \leqslant K$ is the integral closure of $\mathbb{Z}$ in the number field $K$. We now turn our attention to show that $\mathcal{O}_K$ is a finitely generated $\mathbb{Z}$-module.

### 3.2.1 Is $\mathcal{O}_K$ a finitely generated $\mathbb{Z}$-module?

Suppose that $B \leqslant \mathcal{O}_K$ and that[25] $B$ is a finitely generated free $\mathbb{Z}$-module of rank $n$. We can now define something similar to what we did for vector spaces. Since $B$ is a ring, every element $\beta \in B$ defines a $\mathbb{Z}$-linear map $\mu_\beta : B \to B$ defined by $x \mapsto \beta x$. Then $\mathrm{Tr}(\beta) \in B$ and $\mathrm{Tr}(\beta) = \sum_{j=1}^n \sigma_j(\beta)$, where $\sigma_i$ are the embeddings $K \hookrightarrow \mathbb{C}$, since $\sigma_i(\beta)$ are the eigenvalues of $\mu_\beta$. We are now in a position to discuss a fundamental definition which characterizes subrings of $\mathcal{O}_K$.

[25] *the advanced reader should easily spot that we are assuming* $K = \mathrm{Frac}(B)$ *here. This notion will be discussed closely later.*

**Definition** (*Discriminants*)**.** Let $B \leqslant \mathcal{O}_K$ be a finitely generated free $\mathbb{Z}$-module of rank $n$. If $(x_1, \ldots, x_n) \in B^n$, then define its **discriminant** by

$$D(x_1, \ldots, x_n) = \det \operatorname{Tr}(x_i x_j),$$

which is a determinant of an $n \times n$ matrix.

on the rhs, the matrix we are taking determinant of is

$$A = (a_{ij}) = (\operatorname{Tr}(x_i x_j))$$

Our first result about discriminants is the following.

**Lemma 3.7.** Let $B \leqslant \mathcal{O}_K$ be a finitely generated free $\mathbb{Z}$-module of rank $n$; and let $(x_1, \ldots, x_n) \in B^n$. Suppose $(y_1, \ldots, y_n) \in B^n$ with $y_i = \sum_j a_{ij} x_j$ where $a_{ij} \in \mathbb{Z}$. Then

$$D(y_1, \ldots, y_n) = (\det(a_{ij}))^2 D(x_1, \ldots, x_n).$$

**Proof.** The main ingredient is the linearity of the Tr, anything else is just unpacking definitions. We have

$$\operatorname{Tr}(y_p y_q) = \operatorname{Tr}\left(\sum_{i,j} a_{pi} a_{qj} x_i x_j\right) = \sum_{i,j} a_{pi} a_{qj} \operatorname{Tr}(x_i x_j).$$

Viewed as a matrix, we have

$$(\operatorname{Tr}(y_p y_q)) = (a_{pi})(\operatorname{Tr}(x_i x_j))(a_{jq})^\top,$$

where $^\top$ denotes the matrix transpose. Now take determinants and we are done. ■

An immediate result is the fact that it makes sense to talk about **the** discriminant of the subring $B$ instead of **a** discriminant.

**Corollary 3.3.** Let $B \leqslant \mathcal{O}_K$ be a finitely generated free $\mathbb{Z}$-module of rank $n$. If $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$ are both $\mathbb{Z}$-basis of $B$, then

$$D(x_1, \ldots, x_n) = D(y_1, \ldots, y_n).$$

So the discriminant is independent of the choice of basis, and so this defines the **discriminant of $B$**, denoted $D_B$.

**Proof.** Using the preceding Lemma 3.7, we can write

$$D(y_1, \ldots, y_n) = (\det A)^2 D(x_1, \ldots, x_n),$$

where $A = (a_{ij})$ is an $n \times n$ matrix over $\mathbb{Z}$. But here, $A$ is nothing else than the change of basis matrix[26], so it is invertible. Since it is invertible over $\mathbb{Z}$, then $\det A = \pm 1$. We're taking the square of $\det A$, so we get the claim. ■

**Proposition 3.9.** Let $B \leqslant \mathcal{O}_K$ be a finitely generated free $\mathbb{Z}$-module of rank $n$; and let $\{x_1, \ldots, x_n\}$ be a $\mathbb{Z}$-basis of $B$. Then

$$D_B = (\det \Sigma)^2,$$

[26] *recall how $y_i$ were dependent on $x_i$ in its definition in the previous lemma. And both gives a basis, so...*

where $\Sigma = (\sigma_i(x_j))$ is an $n \times n$ matrix.

**Proof.** Unpacking definitions. we have $D_B = D(x_1, \dots, x_n) = \det \operatorname{Tr}(x_i x_j)$. Now $\operatorname{Tr}(x_i x_j) = \sum_{k=1}^n \sigma_k(x_i x_j)$, so we have

$$D_B = \det\left(\sum_{k=1}^n \sigma_k(x_i x_j)\right) = \det\left(\sum_{k=1}^n \sigma_k(x_i)\sigma_k(x_j)\right),$$

where we have used that $\sigma_k$ is a homomorphism. But by definition $\sigma_k(x_i)$ is really just $\Sigma_{ki}$, the $ki$-th entry of $\Sigma$. So we have

$$D_B = \det\left(\sum_{k=1}^n \Sigma_{ki}\Sigma_{kj}\right) = \det\left(\sum_{k=1}^n \Sigma_{ik}^\top \Sigma_{kj}\right).$$

But now notice that everything in the parentheses on the rhs is really just the $ij$-th entry of $\Sigma^\top\Sigma$. So we have

$$D_B = \det\left(\Sigma^\top \Sigma\right) = (\det \Sigma)^2,$$

as desired. ■

Recall that the **index** of a subgroup $H$ in $G$ (its supergroup) is the number of left cosets of $H$ in $G$. Also, recall that $\mathbb{Z}$-modules are really just abelian groups.

**Proposition 3.10.** Let $B \leqslant C$ be subrings of $\mathcal{O}_K$, both of which are finitely generated free $\mathbb{Z}$-modules of rank $n$. Further suppose that the index of $B$ in $C$ is $r$. Then $D_B = r^2 D_C$.

**Slogan.** The smaller the subring, the bigger the discriminant. It is bigger by a factor of the index.

Let's prove this theorem using our good ol' friend the submodule structure theorem.

**Proof.** By the submodule structure theorem, there exists a $\mathbb{Z}$-basis $\{y_1, \dots, y_n\}$ of $C$ and integers $r_1, \dots, r_n \in \mathbb{Z}$ such that $\{r_1 y_1, \dots, r_n y_n\}$ is a $\mathbb{Z}$-basis of $B$. So there is an isomorphism

$$C/B \cong \mathbb{Z}/(r_1) \oplus \cdots \oplus \mathbb{Z}/(r_n).$$

We emphasize that $C/B$ has only the additive (quotient) group structure here. Not a ring!

But the index of $B$ in $C$ is $r$ and so $\#C/B = r$. The isomorphism above gives us $r = \prod_i r_i$. Now by the preceding Proposition 3.9,

$$D_B = \det(\sigma_i(r_j y_j))^2 = \det(r_j\, \sigma_i(y_j))^2,$$

where we have used the fact that $\sigma_i$ is a homomorphism. Using multilinearity of the determinant, we can pull each $r_i$ out. We thus have

$$D_B = \left(\prod_{i=1}^n r_i\right)^2 \det(\sigma_i(y_j))^2 = r^2 D_C,$$

where we have again used Proposition 3.9, but now for $D_C$, to get the final equality. ■

For the discriminants to be useful to us, we would need $D_B \neq 0$. We will see why very soon. Using Dedekind's lemma, we can prove that this is in fact true.

**Lemma 3.8** (*Dedekind*)**.** The embeddings $\sigma_1, \ldots, \sigma_n : K \hookrightarrow \mathbb{C}$ are linearly independent over $K$. That is, if

$$\sum_{i=1}^{n} a_i \sigma_i(z) = 0,$$

for some $a_i \in K$, for all $z \in K$, then $a_i = 0$ for all $i$.

**Proof.** Suppose not. Then pick the shortest non-trivial linear relation

$$\sum_{i=1}^{q} a_i \sigma_i(z) = 0,$$

for all $z \in K$ where $q \leqslant n$, $a_i \neq 0$ for all $i$ and $q$ is minimal. Certainly, $q \geqslant 2$. Now for all $y, z \in K$, we have

$$0 = \sum_{i=1}^{q} a_i \sigma_i(yz) = \sum_{i=1}^{q} (a_i \sigma_i(y))\sigma_i(z).$$

But we also have $\sum_{i=1}^{q} a_i \sigma_i(z) = 0$. So we have

$$\sum_{i=1}^{q} (a_i \sigma_1(y))\sigma_i(z) = 0.$$

If we subtract these two sums we have, we get

$$\sum_{i=2}^{q} a_i(\sigma_i(y) - \sigma_1(y))\sigma_i(z) = 0.$$

This is a shorter[27] non-trivial linear relation since $\sigma_1(y) \neq \sigma_i(y)$ for all $y$ and all $i \geqslant 2$. 何! ■

**Proposition 3.11.** Let $B \leqslant \mathcal{O}_K$ be a finitely generated free $\mathbb{Z}$-module of rank $n$. Then we have $D_B \neq 0$.

**Proof.** Suppose that $\mathcal{B} = \{x_1, \ldots, x_n\}$ is a $\mathbb{Z}$-basis of $B$. For contradiction, suppose $D_B = 0$. Then the rows of the matrix $(\sigma_i(x_j))$ are linearly dependent over $K$. That is, there exist $a_1, \ldots, a_n \in K$ not all 0, such that

$$\sum_{i=1}^{n} a_i \sigma_i(x_j) = 0,$$

for all $1 \leqslant j \leqslant n$. But $\mathcal{B}$ is also a $\mathbb{Q}$-basis of $K$. 何! This implies that

$$\sum_{i=1}^{n} a_i \sigma_i(z) = 0,$$

for all $z \in K$, and this contradicts Dedekind's lemma. ■

[27] *we are summing from $i = 2$ to $q$. The relation we started with summed from $i = 1$ to $q$.*

Finally, we are now in a position to prove the thing we care about.

**Theorem 3.3.** $\mathcal{O}_K$ is a finitely generated $\mathbb{Z}$-module.

**Proof.** Suppose not. Then there is an infinite ascending chain

$$B_1 \subsetneq B_2 \subsetneq B_3 \subsetneq \cdots$$

of subrings of $\mathcal{O}_K$ such that each $B_i$ is a finitely generated free $\mathbb{Z}$-module of rank $n$. But

$$D_{B_i} = \#(B_{i+1}/B_i)^2 D_{B_{i+1}}$$

where $\#(B_{i+1}/B_i)$ is the index of $B_{i+1}$ as a subgroup in $B_i$. So the sequence of discriminants

$$|D_{B_1}|, |D_{B_2}|, |D_{B_3}|, \ldots$$

is a strictly decreasing sequence of nonzero positive integers. 何! This contradicts the preceding Proposition 3.11. ■

Here $B_{i+1}/B_i$ is a quotient group. Recall the proposition smaller subring, bigger discriminant.

**Corollary 3.4.** $\mathcal{O}_K$ is a Noetherian ring.

**Proof.** Since $\mathcal{O}_K$ is a finitely generated $\mathbb{Z}$-module, it is finitely generated as a ring over $\mathbb{Z}$. By the Hilbert's basis theorem, we are done. ■

### 3.2.2 Integers in quadratic fields

Let's calculate a special case of $\mathcal{O}_K$ when $K = \mathbb{Q}[\sqrt{d}]$ where $d \in \mathbb{Z}$ such that $d \neq 0, 1$ and is squarefree. That is, when $K$ is a **quadratic field**. Take $B = \mathbb{Z}[\sqrt{d}]$. Since $\sqrt{d} \in \mathcal{O}_K$, then $B \leqslant \mathcal{O}_K$. We have two embeddings $\sigma_1, \sigma_2 : K \hookrightarrow \mathbb{C}$ which maps

$$\sigma_1(\sqrt{d}) = \sqrt{d}, \quad \sigma_2(\sqrt{d}) = -\sqrt{d}.$$

Moreover, $\left\{1, \sqrt{d}\right\}$ is a $\mathbb{Z}$-basis of $B$. And so

$$D_B = \det \begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{pmatrix}^2 = (-2\sqrt{d})^2 = 4d.$$

Now suppose that the index of $B$ in $\mathcal{O}_K$ is $r$. Then $4d = D_B = r^2 D_{\mathcal{O}_K}$ and so $r^2 \mid 4d$. Since $d$ is squarefree, $r \mid 2$, so $r$ is either 1 or 2. This implies that $2\mathcal{O}_K \subseteq B$. Therefore, if $x \in \mathcal{O}_K$, then

$$x = \frac{a + b\sqrt{d}}{2},$$

$2\mathcal{O}_K \subseteq B$, and so $2x = a + b\sqrt{d}$ for some $a, b \in \mathbb{Z}$.

where $a, b \in \mathbb{Z}$. Now $B$ and $\mathcal{O}_K$ are rings. So we can add or subtract an element of $B$ to $x$ without changing the question of whether $x \in \mathcal{O}_K$. In particular, we can shift[28] $a, b$ by multiples of 2 (because 2 is the denominator). So we may assume that $a, b$ are either 0 or 1. Now, $x \in \mathcal{O}_K$ is true if and only if its minimal polynomial $m_x$ has $\mathbb{Z}$-coefficients. So, we have $m_x(T) = (T - \sigma_1(x))(T - \sigma_2(x))$ where

$$\sigma_1(x) = \frac{a + b\sqrt{d}}{2}, \quad \sigma_2(x) = \frac{a - b\sqrt{d}}{2},$$

[28] *if we had $k \in \mathbb{N}$ in the denominator instead of 2, we could shift $a, b$ by multiples of $k$ instead. In this case, we can assume that $a, b$ are either one of $0, 1, \ldots, k-1$.*

since the zeros of $m_x$ are the conjugate of $x$. So it follows that

$$m_x(T) = T^2 - aT + \frac{a^2 - b^2 d}{4}.$$

we got these coefficients simply by considering sum and product of roots. Remember Vieta's formula from high school algebra?

Thus, we have that

$$x \in \mathcal{O}_K \iff a \in \mathbb{Z} \text{ and } a^2 - b^2 d \equiv 0 \pmod 4.$$

The former condition is something we already knew. So now we focus on the latter condition $a^2 - b^2 d \equiv 0 \pmod 4$. First, we consider the two cases of $a$.

(i). If $a = 0$, then $b^2 d \equiv 0 \pmod 4$. But $d \not\equiv 0 \pmod 4$ since $d$ is squarefree, so $4 \mid b^2$ and hence, $2 \mid b$. Since $b$ is either 0 or 1, we have $b = 0$.

(ii). If $a = 1$, then $b^2 d \equiv 1 \pmod 4$. Since squares are 0 or 1 modulo 4, it follows that $b = 1$ and $d \equiv 1 \pmod 4$.

Now consider the two cases of either $d \equiv 1 \pmod 4$ and $d \not\equiv 1 \pmod 4$.

(1). If $d \equiv 1 \pmod 4$ then either $a = b = 0$ or $a = b = 1$. Now remember that $a$ and $b$ were shifted by multiples of 2. So we conclude that

$$x = \frac{a + b\sqrt{d}}{2} \in \mathcal{O}_K \iff a \equiv b \pmod 2.$$

Since $a = b = 0$ or $a = b = 1$, and they can be shifted by multiples of 2, it follows that if $a$ is even iff $b$ must be even; likewise if one of them is odd.

So in this case

$$\mathcal{O}_K = \left\{ \frac{a + b\sqrt{d}}{2} \;\middle|\; a, b \in \mathbb{Z} \text{ and } a \equiv b \bmod 2 \right\}.$$

So, $\mathcal{O}_K$ has a $\mathbb{Z}$-basis $\left\{1, \frac{1+\sqrt{d}}{2}\right\}$. Thus, $D_{\mathcal{O}_K} = d$ (by computation), so the index $r$ before is 1. Thus, viewed as a ring, we have $\mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{d}}{2}]$.

(2). In the other case that $d \not\equiv 1 \pmod 4$, then $a = b = 0$. Recall that $a, b$ were shifted by multiples of 2. So

$$x = \frac{a + b\sqrt{d}}{2} \in \mathcal{O}_K \iff a, b \equiv 0 \pmod 2.$$

Then as a ring, we have $\mathcal{O}_K = Z[\sqrt{d}] = B$ with $D_{\mathcal{O}_K} = 4d$.

In summary, we have the following.

**Theorem 3.4.** Let $K = \mathbb{Q}[\sqrt{d}]$, where $d \in \mathbb{Z}$ such that $d \neq 0, 1$ and is squarefree. Then

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right], & \text{if } d \equiv 1 \pmod 4, \\ \mathbb{Z}\left[\sqrt{d}\right], & \text{if } d \not\equiv 1 \pmod 4. \end{cases}$$

In general, it is often the case that $\mathcal{O}_K$ is bigger than expected.

**Example.** If we consider $K = \mathbb{Q}[\sqrt{-5}]$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ since $-5 \not\equiv 1 \pmod 4$.

## 3.3 More on finite ring extensions

Let us now talk about more consequences regarding rings that are finite over subrings. In particular, we will discuss the Cayley-Hamilton theorem (known from linear algebra) in a more general setting. We will also talk about Nakayama's lemma and its consequences which will be very useful later on. Recall that an endomorphism of an $R$-module $M$ is just an $R$-linear map $M \to M$, and the (non-commutative) ring of $R$-endomorphisms of $M$ is denoted $\mathrm{End}_R(M)$.

**Definition** (*Polynomial ring in $\phi$ with coefficients in $R$*)**.** Let $R$ be a ring, let $M$ be a finitely generated $R$-module and suppose $\phi \in \mathrm{End}_R(M)$. Let $f \in R[X]$ be the polynomial

$$f(X) = \sum_{i=0}^{d} a_i\, X^i.$$

Then to $f$, we associate an $R$-linear map $f_\phi : M \to M$ given by

$$f_\phi(m) = \sum_{i=0}^{d} a_i\, \phi^i(m).$$

where $\phi^i$ just means $\phi$ composed with itself $i$ times (and $\phi^0$ is just the identity map).

The set of all $R$-linear maps $f_\phi$ arising in this way is denoted as $R[\phi]$.

In other words, if $\phi \in \mathrm{End}_R(M)$, then

$$R[\phi] = \left\{ \sum a_i\, \phi^i : a_i \in R \right\} \subseteq \mathrm{End}_R(M).$$

**Proposition 3.12.** $R[\phi]$ is a commutative subring of $\mathrm{End}_R(M)$.

Why is this proposition important? As we will see, we will consider square matrices over $R[\phi]$ and we would want to compute their determinants — which is only well-defined if the square matrix is defined over a commutative ring.

**Proposition 3.13.** If $P$ is a square matrix with entries in a commutative ring, then $P^{\mathrm{adj}}P = (\det P) \cdot \mathbb{I}$ where $P^{\mathrm{adj}}$ is the adjugate of $P$ and $\mathbb{I}$ is the identity matrix.

We need one more definition before talking about the Cayley-Hamilton theorem. If $R$ is a ring, $I \lhd R$ and $M$ is an $R$-module, then the set

$$IM = \left\{ \sum_{i=1}^{n} a_i m_i \;\middle|\; n \in \mathbb{N}, a_i \in I, m_i \in M \right\},$$

is a well-defined submodule of $M$. It is an easy exercise to check that $IM$ is indeed closed under addition and under multiplication by scalars (in $R$).

if $I$ is not an ideal, we may not get closure under scalar multiplication.

**Theorem 3.5** (*Cayley-Hamilton*)**.** Let $R$ be a ring and $I \lhd R$. Let $M$ be a finitely generated $R$-module, and let $\phi$ be an endomorphism of $M$ such that $\phi^{\mathrm{img}}(M) \subseteq IM$. Then $\phi$ satisfies an equation

$$\phi^n + a_{n-1}\phi^{n-1} + \cdots + a_1\phi + a_0 = 0,$$

where $a_i \in I$.

This theorem has been asked to be proved in the exam quite a few times.

**Proof.** Since $M$ is finitely generated, there is a generating set $\{m_1, \dots, m_n\}$ of $M$. So there exists $a_{ij} \in I$ such that

$$\phi(m_i) = \sum_{j=1}^{n} a_{ij} m_j, \qquad (*)$$

for every $1 \leqslant i \leqslant n$. The idea now is to *write this equation as a matrix equation.* Consider the following:

(i). Put the generators into a vector by putting $\boldsymbol{m} = (m_1, \dots, m_n)^\top$.

for emphasis, $\boldsymbol{m}$ here is a column vector.

(ii). Put the coefficients $a_{ij}$ into a matrix by putting $Q = (a_{ij})$ which is an $n \times n$ matrix;

(iii). Then consider $P = \phi\,\mathbb{I} - Q$ where $\mathbb{I}$ is the $n \times n$ identity matrix.

Then we can rewrite $(*)$ as a matrix equation

$$P\boldsymbol{m} = \boldsymbol{0}.$$

Observe that $P$ is an $n \times n$ matrix with entries in the commutative ring $R[\phi]$. So by Proposition 3.13, it has an adjugate $P^{\text{adj}}$ such that $P^{\text{adj}}P = (\det P) \cdot \mathbb{I}$. If we left multiply our matrix equation above by $P^{\text{adj}}$, we then get

$$(\det P) \cdot \mathbb{I}\boldsymbol{m} = P^{\text{adj}}P\boldsymbol{m} = \boldsymbol{0}.$$

So we have that $(\det P)m_i = 0$ for all $i$. But $\det P \in R[\phi] \subseteq \operatorname{End}_R(M)$. So these two facts imply that $\det P$ is a linear map $M \to M$ such that it kills every generator $m_i$ of $M$. In other words, $\det P = 0_{\operatorname{End}_R(M)}$, the zero endomorphism. But what is $\det P$? Well, we first write $P$ explicitly:

$$P = \phi\,\mathbb{I} - Q = \begin{pmatrix} \phi - a_{11} & -a_{12} & \cdots & -a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \phi - a_{nn} \end{pmatrix}.$$

So computing $\det P$ via expanding the cofactors, we have that

$$\det P = \phi^n + a_{n-1}\phi^{n-1} + \cdots + a_1\phi + a_0,$$

where each $a_i$ is a $\mathbb{Z}$-polynomial in the $a_{ij}$ i.e. $a_i \in \mathbb{Z}[a_{ij}]$. But $a_{ij} \in I$ by hypothesis, so $a_i \in I$, as desired. ■

A corollary of Cayley-Hamilton is that if $M$ is finitely generated over $R$, and there is an equality $IM = M$, then there exists an element of $R$ such that it annihilates the whole of $M$.

**Corollary 3.5** (*Nakayama's lemma*)**.** Suppose $R$ is a ring and $I \lhd R$. Let $M$ be a finitely generated $R$-module such that $IM = M$. Then there exists $x \in R$ such that $x \equiv 1_R \pmod{I}$ and $xM = 0$.

With the Cayley-Hamilton's theorem, this is immediate.

**Proof.** The identity map $\iota : M \to M$ is an endomorphism of $M$ such that $\iota^{\text{img}}(M) \subseteq IM$. By the Cayley-Hamilton theorem, there exist $a_i \in I$ such that

$$\iota^n + a_{n-1}\iota^{n-1} + \cdots + a_0 = 0.$$

So taking $x = 1_R + a_{n-1} + \cdots + a_0$ which is an element of $R$, we have that $xm = 0$ for all $m \in M$. Moreover, since $a_i \in I$, $x \equiv 1_R \pmod{I}$, as desired. ■

Let's move into a slightly different direction.

**Lemma 3.9.** Suppose that $S \leqslant R$ rings such that $R$ is finite over $S$. If $R$ and $S$ are integral domains, then $R$ is a field *if and only if* $S$ is a field.

**Proof.** Suppose $S$ is a field. Let $x \in R\backslash\{0\}$. Then $x$ defines an $S$-linear map

$$\phi_x : R \to R, \quad r \mapsto xr.$$

Since $R$ is a domain, $\phi_x$ is injective. By hypothesis, $R$ is a finite-dimensional vector space over $S$, so every injective $S$-linear map $R \to R$ is surjective. So there exists $y \in R$ such that $\phi_x(y) = xy = 1$. This implies that $y = x^{-1}$ in $R$, so $R$ is a field.

Conversely, suppose $R$ is a field. Let $x \in S\backslash\{0\}$. Then $y = x^{-1}$ exists in $R$ by assumption. Moreover, since $R$ is finite over $S$, then[29] $S[y]$ is finite over $S$ and so $y$ is integral over $S$. Thus, there exists a polynomial satisfying

$$y^n + a_{n-1}y^{n-1} + \cdots + a_0 = 0,$$

where $a_i \in S$. If we multiply the above equation by $x^{n-1}$ on both sides, we get

$$y + \underbrace{a_{n-1} + \cdots + a_0 x^{n-1}}_{\text{element of } S} = 0.$$

By closure of $S$, we must have $y \in S$; so $S$ is a field. ■

Recall that an ideal $\mathfrak{m} \lhd R$ is maximal if and only if $R/\mathfrak{m}$ is a field. This result together with Nakayama's lemma gives a shortcut to prove the following two results.

**Lemma 3.10.** Suppose that $S \leqslant R$ rings such that $R$ is finite over $S$. Suppose that $\mathfrak{p} \lhd R$ is a prime ideal. Then

$$\mathfrak{p} \text{ is maximal in } R \iff \mathfrak{p} \cap S \text{ is maximal in } S.$$

**Proof.** Now $S/(\mathfrak{p} \cap S) \subseteq R/\mathfrak{p}$ where both are integral domains. Moreover, $R/\mathfrak{p}$ is finite over $S/(\mathfrak{p} \cap S)$. So recalling a result of ideals and fields, we have

$$\begin{aligned} \mathfrak{p} \text{ is maximal in } R &\iff R/\mathfrak{p} \text{ is a field}, \\ &\iff S/(\mathfrak{p} \cap S) \text{ is a field}, \\ &\iff \mathfrak{p} \cap S \text{ is maximal in } S, \end{aligned}$$

where we have used the previous Lemma 3.9 in the second iff statement. ■

[29] *think of extending $S$ more into $R$ here. But the main point is, it still sits in $R$.*

**Theorem 3.6.** Suppose that $S \leqslant R$ are integral domains such that $R$ is finite over $S$. If $S$ is Noetherian, then for every maximal ideal $\mathfrak{m} \lhd S$, there is a maximal ideal $\mathfrak{n} \lhd R$ such that $\mathfrak{n} \cap S = \mathfrak{m}$.

Note that since $S$ is Noetherian, then $R$ is also Noetherian here.

**Proof.** Suppose that $\mathfrak{m} \lhd S$ maximal such that $\mathfrak{m}R = R$. Then by Nakayama's lemma[30], there exists $x \in S$ such that $x \equiv 1_S \pmod{\mathfrak{m}}$ and $xR = 0$. Since $R$ is a domain, $x = 0_S$. 何! This contradicts the fact that $x \equiv 1_S \pmod{\mathfrak{m}}$. So $\mathfrak{m}R \neq R$ i.e. a proper ideal of $R$. Then, by choosing any $\mathfrak{n} \lhd R$ maximal such that $\mathfrak{n} \supseteq \mathfrak{m}R$, we are done. ■

## 3.4 Rings and modules of fractions

We will now discuss the generalization of the field of fractions to rings. This discussion together with consequences of Nakayama's lemma will help us to prove some interesting things.

**Definition** (*Multiplicative subsets*)**.** Suppose that $A$ is a ring and $S \subseteq A$ is a **subset**. Then $S$ is **multiplicative** if $1_A \in S$ and $S$ is closed under multiplication (of $A$).

$S$ is a subset! Not a subgroup. Not a subring. It has no additional structure.

**Remark.** In this section (and in oncoming sections which utilizes this section's results), we will use the convention of using $A, B$ as rings instead of $R, S$ as we keep $S$ to denote a (special) subset of the ring.

Let's look at examples to get a better picture.

**Example.** If $A$ is an integral domain, then $S = A \setminus \{0\}$ is multiplicative.

**Example.** Let $A$ be a ring and $\mathfrak{p} \lhd A$ is prime. Then $S = A \setminus \mathfrak{p}$ is multiplicative.

In fact, $S = A \setminus \mathfrak{p}$ is multiplicative if and only if $\mathfrak{p}$ is prime.

**Example.** Let $A$ be a ring and $a \in A$. Then the subset $S = \left\{1, a, a^2, a^3, \ldots\right\}$ of powers of $a$ is multiplicative.

### 3.4.1 Rings of fractions

Now let $A$ be a ring such that $S \subseteq A$ is multiplicative. Then define a relation $\sim$ on $A \times S$ by $(a, s) \sim (b, t)$ if there exists $x \in S$ such that $x(at - bs) = 0$.

**Lemma 3.11.** The relation $\sim$ defined above is an equivalence relation.

**Proof.** Let $A$ be the ring and $S \subseteq A$ be multiplicative. We must check three things: reflexive, symmetric and transitive.

**Reflexive.** $1(as - as) = 0$, so $(a, s) \sim (a, s)$ with $x = 1_A$.

**Symmetric.** Suppose $(a, s) \sim (b, t)$ . So there exists $x \in S$ such that $x(at - bs) = 0$. Then $x(bs - at) = 0$ and so $(b, t) \sim (a, s)$.

**Transitive.** Suppose $(a, s) \sim (b, t)$ and $(b, t) \sim (c, u)$. Then there exist $x, y \in S$ such that $x(at - bs) = 0$ and $y(bu - ct) = 0$. Now, if we multiply the first by $yu$ and the second by $xs$, we get

$$xyu(at - bs) = 0, \quad xys(bu - ct) = 0.$$

[30] *here, $R$ is finite over $S$, so is a finitely generated $S$-module. The setting in the Nakayama's lemma statement here is that $I = \mathfrak{m}$, $R = S$ and $M = R$.*

So we have an equality

$$xyuat - xyubs = xysbu - xysct \implies xytau - xytcs = 0,$$

where the implication is due to $A$ being commutative. Thus, $xyt(au - cs) = 0$. But $z = xyt \in S$, so $z$ gives $(a, s) \sim (c, u)$. ■

So we can define the set of equivalence classes $S^{-1}A$ under $\sim$. That is $S^{-1}A = (A \times S)/\sim$. And write $a/s$ (or $s^{-1}a$) for the equivalence class (which is an element of $S^{-1}A$) that contains $(a, s)$.

**Proposition 3.14.** There exists a ring[31] structure on $S^{-1}A$ by defining addition and multiplication in the following way

$$\frac{a}{s} + \frac{b}{t} = \frac{at + bs}{s + t}, \quad \frac{a}{s}\frac{b}{t} = \frac{ab}{st}.$$

Notice that these operations are really how you've been dealing with usual fractions in $\mathbb{R}$ in your entire life.

**Proof.** We first need to prove that these operations are well-defined. That is, we need to show that (1) if $(c, u) \sim (b, t)$, then $a/s + c/u = a/s + b/t$ and (2) $a/s \cdot c/u = a/s \cdot b/t$.

**Addition.** Showing (1) is equivalent to proving $(au + cs)/su = (at + bs)/st$. That is, $(au + cs, su) \sim (at + bs, st)$, given that $(b, t) \sim (c, u)$. So it is enough to verify that

$$z(au + cs)st = z(at + bs)su,$$

given that $z(ct - bu) = 0$ for some $z \in S$. But now notice that we have

$$\text{lhs} = zs(aut + cst) = zs(aut + bus) = zsu(at + bs) = \text{rhs},$$

here "lhs" and "rhs" are left hand side and right hand side respectively. I know it's a bit confusing given that there are so many alphabets here (we even spelled bus!).

so we are done.

**Multiplication.** We want $a/s \cdot c/u = a/s \cdot bt$. That is, we need $(ac, su) \sim (ab, st)$, given that $(c, u) \sim (b, t)$. So it is enough to verify that

$$z(ac \cdot st) = z(ab \cdot su),$$

given that $zct = zbu$. But now observe that

$$\text{lhs} = zsact = zsabu = \text{rhs},$$

and we are done.

**Ring axioms.** Now we need to check the ring axioms, but this is immediate. We have $(a/s) \cdot (1/1) = a/s$. We also have $a/s + 0/s = a/s$. We have $-(a/s) = (-a)/s$. Finally, we have distributivity of multplication over addition,

$$\frac{a}{s}\left(\frac{b}{t} + \frac{c}{u}\right) = \frac{ab}{st} + \frac{ac}{su},$$

where $a, b, c \in A$ and $s, t, u \in S$. ■

Let's give the ring a name.

[31] *as usual, by ring we mean unital (has a unit) and commutative.*

**Definition** (*Ring of fractions*)**.** Let $A$ be a ring and $S \subseteq R$ is multiplicative. The ring $S^{-1}A$ is called the **ring of fractions of $A$ with respect to $S$**, with $0_{S^{-1}A} = 0/s$ and $1_{S^{-1}A} = s/s$.

Observe that there is a ring homomorphism $\phi : A \to S^{-1}A$ defined by $\phi(x) = x/1$. While it is tempting to say that it is injective, it is not true in general. The ring of fractions then has a **universal property** in the following sense.

**Proposition 3.15** (*Universal property*)**.** Let $\psi : A \to B$ be a ring homomorphism such that $\psi(s)$ is a unit in $B$ for all $s \in S$. Then there exists a unique ring homomorphism $\Psi : S^{-1}A \to B$ such that $\psi = \Psi \circ \phi$.

where $\phi$ is the ring homomorphism defined above.

The universal property says that we can fill in "?" in the diagram below

$$\begin{tikzcd} A \arrow[r, "\psi"] \arrow[dr, "\phi"'] & B \\ & S^{-1}A \arrow[u, "?"'] \end{tikzcd}$$

with a map $\Psi$, such that the diagram commutes. A proof of the universal property can be found in Atiyah-MacDonald [pg. 37]. Here is the field of fractions for those new to it.

**Definition** (*Field of fractions*)**.** Let $A$ be an **integral domain** and let $S = A\backslash\{0\}$ which is multiplicative. Then $S^{-1}A$ is the called the **field of fractions** of $A$ and is denoted $\mathrm{Frac}(A)$.

As its name goes, it is a field. In fact, it is the smallest field that contains $A$ and this is verified by the universal property of $\mathrm{Frac}(A)$.

some people just take this "smallest field" description as its definition.

**Example.** Let $A$ be an integral domain. Then for every multiplicative $T \subseteq A$, we have $A \subseteq T^{-1}A \subseteq \mathrm{Frac}(A)$.

Here is a concrete example of the above example.

**Example.** Consider the ring $\mathbb{Z}$ and let $p \in \mathbb{Z}$ prime. Then $T = \mathbb{Z}\backslash(p)$ is multiplicative. And we have

$$\mathbb{Z} \subseteq T^{-1}\mathbb{Z} \subseteq \mathrm{Frac}(\mathbb{Z}) = \mathbb{Q}.$$

### 3.4.2 Modules of fractions

If $M$ is an $A$-module, then we can define a similar object like $S^{-1}A$. The way we do it is identical. Define an equivalence relation $\sim$ on $M \times S$ by $(m, s) \sim (n, t)$ if there exists $u \in S$ such that $u(tm - sn) = 0$. The proof is exactly like before. Then define $S^{-1}M = (M \times S)/\sim$. The equivalence class of $(m, s)$ is written $m/s$ or $s^{-1}m$. Now here's the mind-bending part.

**Proposition 3.16.** Let $A$ be a ring, let $S \subseteq A$ be multiplicative and let $M$ be an $A$-module. Then $S^{-1}M$ is an $S^{-1}A$-module with addition and scalar multiplication.

$$\frac{m}{s} + \frac{n}{t} = \frac{tm + sn}{st}, \qquad \frac{a}{s} \cdot \frac{m}{t} = \frac{am}{st}.$$

**Proof.** Prove it yourself. ■

Here are some basic facts.

**Proposition 3.17.** Let $A$ be a ring, let $S \subseteq A$ be multiplicative and let $M$ be an $A$-module. If $N \subseteq M$ is an $A$-submodule of $M$, then

(1). $S^{-1}N$ is a $(S^{-1}A)$-submodule of $S^{-1}M$,

(2). $S^{-1}(M/N) \cong (S^{-1}M)/(S^{-1}N)$,

**Proposition 3.18.** Let $A$ be a ring, let $S \subseteq A$ be multiplicative and let $M, N$ be $A$-modules. $S^{-1}(M \oplus N) \cong (S^{-1}M) \oplus (S^{-1}N)$.

**Example.** Let $A$ be an integral domain and consider its field of fractions $K = \mathrm{Frac}(A)$. If $M$ is an $A$-module, then $S^{-1}M$ is a $K$-vector space. Let's look at specific examples.

In hand-wavy words, "the formation of $S^{-1}M$ converts the $A$-module $M$ into a $K$-vector space."

- $\mathrm{Tors}(M)$ is an $A$-module and we have $S^{-1}(\mathrm{Tors}(M)) = 0$. So while the fact $S^{-1}M$ is a $K$-vector space sounds like a good thing, we can actually lose too much data.
- If $M = A^n$ is a direct sum, then $S^{-1}M = K^n$. So, if $M \cong A^n$ and $M \cong A^m$, then $K^n \cong S^{-1}M \cong K^m$, so $m = n$, which is okay (and good even). So, when $A$ is a domain, the rank of a free $A$-module makes sense.

cf. Invariant Basis Theorem.

Finally, we arrive at the fun part. Let's talk about the relation between prime ideals and rings of fractions. Recall that if $I \lhd A$, then there is a correspondence between ideals in $A/I$ and ideals in $A$ that contain $I$. Moreover, this correspondence extends naturally to give a correspondence between prime ideals in $A/I$ and prime ideals in $A$ that contain $I$.

Here is where our discussion on rings and modules of fractions comes in. Consider the ring $S^{-1}A$. We now know (from Proposition 3.17) that every ideal $I \lhd A$ gives an ideal in $S^{-1}A$ since ideals are really submodules of $A$, where $A$ is viewed as a module over itself. Now there is an interesting observation for when the ideals are prime.

**Proposition 3.19** (*First correspondence theorem for prime ideals*)**.** Let $A$ be a ring and let $S \subseteq A$ be multiplicative. Then there is a bijective correspondence:

$$\{\mathfrak{p} \lhd A \mid \mathfrak{p} \text{ is prime, } \mathfrak{p} \cap S = \varnothing\} \longleftrightarrow \{\mathfrak{q} \lhd S^{-1}A \mid \mathfrak{q} \text{ is prime}\}.$$

That is, a correspondence between the prime ideals in $A$ that are disjoint from $S$, and the prime ideals of $S^{-1}A$.

**Proof.** Omitted. It's "easy". ■

We have seen that if $\mathfrak{p} \lhd A$ is prime, then $S = A \backslash \mathfrak{p}$ is multiplicative. For this particular (very interesting) setting, denote $A_{\mathfrak{p}} = S^{-1}A$ for rings, and $M_{\mathfrak{p}} = S^{-1}M$ for modules.

$S = A \backslash \mathfrak{p}$ is $S$ minus $\mathfrak{p}$ and **not** modulo $\mathfrak{p}$!

**Corollary 3.6** (*Second correspondence theorem for prime ideals*)**.** Let $A$ be a ring, let $\mathfrak{p} \lhd A$ be prime and let $S = A \backslash \mathfrak{p}$. Then there is a bijective correspondence:

$$\{\mathfrak{q} \lhd A \mid \mathfrak{q} \text{ is prime and } \mathfrak{q} \subseteq \mathfrak{p}\} \longleftrightarrow \{\mathfrak{q}_{\mathfrak{p}} \lhd A_{\mathfrak{p}} \mid \mathfrak{q}_{\mathfrak{p}} \text{ is prime}\}.$$

That is, a correspondence between the prime ideals in $A$ that are contained in $\mathfrak{p}$ and the prime ideals in $A_{\mathfrak{p}}$.

This is quite analogous to the case of the quotient ring $A/I$ where we had a correspondence

$$\{\text{ideals of } R \text{ that contain } I\} \longleftrightarrow \{\text{ideals of } R/I\}.$$

cf. Correspondence Theorem (for ideals).

From the point of view of the "combinatorics of the lattice of prime ideals", this is a very useful result. It tells us that if we are interested in prime ideals in $A$ that are contained in $\mathfrak{p}$, then we can form this quotient ring $A_{\mathfrak{p}} = S^{-1}A$ and look at the primes ideals in here. In particular, $A_{\mathfrak{p}}$ has a unique maximal ideal $\mathfrak{m}_{\mathfrak{p}}$ and so a prime ideal $\mathfrak{q}$ with $\mathfrak{q} \subseteq \mathfrak{p}$ then corresponds to $\mathfrak{m}_{\mathfrak{p}} \subseteq \mathfrak{p}_{\mathfrak{p}}$.

the notation of $\text{ideal}_{\text{ideal}}$ (e.g. $\mathfrak{m}_{\mathfrak{p}}$) corresponds to $\text{module}_{\text{prime ideal}}$ here. We have simply used our notation $M_{\mathfrak{p}} = S^{-1}M$ since $\mathfrak{p}$ is prime and ideals are modules. The most confusing of all is $\mathfrak{p}_{\mathfrak{p}}$. This is just $S^{-1}\mathfrak{p}$ with $\mathfrak{p} \lhd A$ prime, viewed as an $A$-module and $S = A \backslash \mathfrak{p}$.

### 3.4.3 Some applications of rings of fractions

So how do rings and modules of fractions comes to the big picture? Here is how they are applied

**Lemma 3.12.** Let $A$ be a Noetherian ring and let $S \subseteq A$ be multiplicative. Then $S^{-1}A$ is Noetherian.

**Proof.** Suppose $J \lhd S^{-1}A$. Then consider

$$I = \{x \in A \mid \exists\, s \in S \text{ with } x/s \in J\} \subseteq A.$$

Then it is easy to check that $I$ is an ideal in $A$, and $J = IS^{-1}A = S^{-1}I$. Since $A$ is Noetherian, then $I$ is finitely generated, say, by $\{x_1, \ldots, x_n\}$. But then $J$ is finitely generated by $\{x_1/1, \ldots, x_n/1\}$. ■

**Proposition 3.20.** Let $A \leqslant B$ be rings such that $A$ is Noetherian and $B$ finite over $A$. Then for every prime ideal $\mathfrak{p} \lhd A$, there exists a prime ideal $\mathfrak{q} \lhd B$ such that $\mathfrak{q} \cap A = \mathfrak{p}$.

We have already proved this for maximal ideals in Theorem 3.6. With the power of rings of fractions, we can prove that the same thing holds if we replace "maximal" with "prime".

**Proof.** Suppose $\mathfrak{p} \lhd A$ is prime and set $S = A \backslash \mathfrak{p}$. This is multiplicative in $A$ **and** in $B$. So consider $A_{\mathfrak{p}} = S^{-1}A$ and $B_{\mathfrak{p}} = S^{-1}B$. Clearly, we have $A_{\mathfrak{p}} \subseteq B_{\mathfrak{p}}$. Moreover, it is easy to verify that

(i). $A_{\mathfrak{p}}$ is Noetherian (previous lemma),

(ii). $B_{\mathfrak{p}}$ is finite over $A$,

(iii). $A_{\mathfrak{p}}$ has a unique maximal ideal $\mathfrak{m} = S^{-1}\mathfrak{p}$ (by (i), we can apply Lemma 1.8).

It is an exercise to check that (ii) indeed holds.

Then by Theorem 3.6, there is a maximal ideal $\mathfrak{n} \lhd B_{\mathfrak{p}}$ with $\mathfrak{n} \cap A_{\mathfrak{p}} = \mathfrak{m}$. But every maximal ideal is prime, so $\mathfrak{n}$ is prime. So by the first correspondence theorem, there exists a unique prime ideal $\mathfrak{q} \lhd B$ such that $S^{-1}\mathfrak{q} = \mathfrak{n}$. We then have an equality $(S^{-1}\mathfrak{q}) \cap (S^{-1}A) = S^{-1}\mathfrak{p}$. With some work, it is easy to see that $\mathfrak{q} \cap A = \mathfrak{p}$, as desired. ■

We had $\mathfrak{n} \cap A_{\mathfrak{p}} = \mathfrak{m}$. This equality is just unpacking definitions

## 3.5 Towards algebraic geometry

Algebraic geometry is the study of geometry using the methods of commutative algebra. The object that bridges these two seemingly separated fields is the celebrated Hilbert's Nullstellensatz. Remember that we have been talking about this fancy object called ring algebras and we said that we will give a proper treatment of it some time later? This is that time.

**Definition** (*Comm. ring algebra I*)**.** Let $\phi : R \to A$ be a ring homomorphism. The ring $A$, is said to be an **$R$-algebra** if $A$ is given the scalar multiplication

$$r \cdot a = \phi(r)a,$$

for all $r \in R$ and $a \in A$, so that $A$ is an $R$-module. We denote this as $(A, \phi)$, as the scalar multiplication is really defined by $\phi$.

Recall that our definition of a ring homomorphism must send $1_R$ to $1_A$.

It suffices to say that $A$ is an $R$-algebra if $A$ is a ring "equipped" with a ring homomorphism $\phi : R \to A$. This is how we defined it earlier. Now if $R$ is a field $k$, then[32] $\phi$ is injective. So a $k$-algebra is basically a ring such that $k$ is a subring.

**Remark.** On notation. We will usually use the letters $A, B$ instead of $R, S$ for ring algebras as it simply makes sense: what is the first letter in the word algebra?

**Example.** For any ring $R$, the polynomial ring $R[X_1, \dots, X_n]$ is an $R$-algebra. This is immediate by giving it the inclusion map $\iota : R \hookrightarrow R[X_1, \dots, X_n]$, mapping $r \in R$ to the constant polynomials $r \in R[X_1, \dots, X_n]$.

**Lemma 3.13.** Every (nonzero commutative) ring is a $\mathbb{Z}$-algebra.

This is not very surprising.

**Proof.** If $R$ is any ring, then there exists a unique homomorphism $\mathbb{Z} \to R$; this is the map $n \mapsto n \cdot 1_R$. ■

There is also an equivalent definition of algebras, which is in some sense more natural to think about. It is an exercise to prove their equivalence.

**Definition** (*Comm. ring algebra II*)**.** Let $R, A$ be rings. Then $A$ is an $R$-algebra if it is an $R$-module equipped with an associative $R$-bilinear map $\langle \cdot, \cdot \rangle : A \times A \to A$ such that $\langle \cdot, \cdot \rangle$ is unital: there is $a \in A$ such that $\langle a, x \rangle = \langle x, a \rangle = x$ for all $x \in A$.

To get from this definition to the first definition, we get our ring homomorphism $\phi : R \to A$ by the natural map $\phi(r) = r \cdot 1_A$ where $\cdot$ here is the $R$-module action on $A$.

**Definition** (*Subalgebra*)**.** Let $(A, \phi)$ be an $R$-algebra. Then $B \subseteq A$ is called an **$R$-subalgebra** if for every $x, y \in B$, then $x + y$, $x \cdot y \in B$ and $rx \in B$ for all $r \in R$. That is, it is an $R$-submodule of $A$, and it is closed under multiplication in $B$.

where we have identified "multiplication" in $B$ with the scalar multiplication inherited from $A$, defined by $\phi$. In other words, this multiplication is the bilinear map in definition II of algebras.

$B$ being a submodule implies that it is closed under addition and scalar multiplication. For it to be additionally a subalgebra, it has to be closed under multiplication as well. Here is a concrete example.

[32] *recall that when we talk about rings, we always mean that it is nonzero. In particular, $A$ is not the zero ring. Now if $R$ is a field, what do its ideals tell us?*

**Example.** The complex numbers $\mathbb{C} = \mathbb{R}[i]$ equipped with the inclusion map $\iota : \mathbb{R} \to \mathbb{C}$ is an $\mathbb{R}$-algebra. The line $\mathbb{R} \leqslant \mathbb{C}$ which is a subring is an $\mathbb{R}$-subalgebra.

Just by thinking about $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{Q}$, you can construct a lot of algebras.

**Example.** The ring of square matrices over any field $k$ equipped with the bilinear map of matrix multiplication is a $k$-algebra.

**Definition** (*Finitely generated algebras*)**.** Let $A$ be an $R$-algebra, equipped with the ring homomorphism $\phi : R \to A$. Then $A$ is said to be **finitely generated** if there exists a finite generating set $\{x_1, \ldots, x_n\} \subseteq A$ such that every $a \in A$ can be written as

$$a = \sum_{i=1}^{n} f_i x_i,$$

for some $f_i \in \phi(R)$.

In other words, every element of $A$ can be written as a polynomial in $x_1, \ldots, x_n$ with coefficients in $\phi(R)$. Actually, we can say with coefficients in $R$, as long as we remember how we defined scalar multiplication earlier: $r \cdot a = f(r)a$.

Finitely generated $R$-algebras are also called finitely generated as rings over $R$. Sounds familiar?

If $A$ is an $R$-algebra, then $A$ being finitely generated as an $R$-algebra is not the same as $A$ being finitely generated as an $R$-module! In literature, the latter is what we called a **finite $R$-algebra** which is (confusing and) **not** the same as a finitely generated $R$-algebra.

Here is a nice way to see this:

1. $A$ is a **finite** $R$-algebra if there is a generating set $\{x_1, \ldots, x_n\}$ such that

$$A = Rx_1 + \ldots + Rx_n = \left\{ \sum_{i=1}^{n} r_i x_i \;\middle|\; r_i \in R \right\}.$$

This notationally agrees with how we wrote finitely generated modules before. In particular, when we wrote cyclic modules.

2. $A$ is a **finitely generated** $R$-algebra if there exists a generating set $\{x_1, \ldots, x_n\}$ such that $A = R[x_1, \ldots, x_n]$.

However, observe that finite implies finitely generated but the converse is false.

**Theorem 3.7** (*Noether's normalization lemma*)**.** Let $k$ be a field, and let $R \supseteq k$ be a finitely generated $k$-algebra. Then there is a $k$-subalgebra $S \leqslant R$ such that $S$ is a polynomial ring over $k$ and $R$ is finite over $S$.

It makes sense to write $S \leqslant R$ as subalgebras are manifestly subrings.

If you are asked to reproduce the proof of this theorem in the exam, and you did not revise for it, then skip the question. The proof is too clever.

**Proof.** Let $\{x_1, \ldots, x_r\}$ be a generating set for $R$ over $k$. Then $R = k[x_1, \ldots, x_r]$. We will prove by induction on $r$ that there is a subring $S$ satisfying our conclusion above.

**Base case.** If $r = 1$, then $R = k[x]$. If $x$ satisfies no relation, then $R$ is a polynomial ring and we can take $S = R$. If $x$ satisfies a relation, then we can assume WLOG that it is a monic relation[33], so $x$ is integral over $k$. Consequently, $R = k[x]$ is finite over $k$ and so take $S = k$.

**Inductive hypothesis.** Suppose that the result holds for $r \leqslant n - 1$. That is, for $R = k[x_1, \ldots, x_{r-1}]$, there is such a subring $S$.

**Inductive step.** Suppose that $R$ is generated by $\{x_1, \ldots, x_r\}$ i.e. $R = k[x_1, \ldots, x_r]$

[33] *that is, $x$ satisfies a monic polynomial with coefficients in $k$.*

where $x_i \in k$. WLOG[34], we can assume that there exists a nonzero polynomial

$$f(T_1, \dots, T_n) = \sum_{\boldsymbol{n}} \left( a_{\boldsymbol{n}} \prod_{i=1}^{r} T_i^{n(i)} \right) \in k[T_1, \dots, T_n],$$

such that $f(x_1, \dots, x_r) = 0$; where $\boldsymbol{n} = (n(1), \dots, n(r))$ such that the $n(i)$ are not all zero, and $a_{\boldsymbol{n}} \in k$ is nonzero. That is, we have a non-trivial relation

$$\sum_{\boldsymbol{n}} \left( a_{\boldsymbol{n}} \prod_{i=1}^{r} x_i^{n(i)} \right) = 0. \qquad (\dagger)$$

Now fix an integer $g \in \mathbb{Z}$ such that $g > n(i)$ for all $1 \leqslant i \leqslant r$. For all $2 \leqslant i \leqslant r$, define

$$z_i = x_i - x_1^{g^{i-1}}. \qquad (\dagger\dagger)$$

Then we have an equality

$$R = k[x_1, \dots, x_r] = k[x_1, z_2, \dots, z_r],$$

since each $x_i$ is a polynomial in $x_1, z_2, \dots, z_r$, and each $z_i$ is a polynomial in $x_1, x_2, \dots, x_r$. From $(\dagger\dagger)$, we have $x_i = z_i + x_1^{g^{i-1}}$. Substituting this into $(\dagger)$ we have

$$\sum_{\boldsymbol{n}} \left( a_{\boldsymbol{n}} x_1^{n(1)} \prod_{i=2}^{r} \left( z_i + x_1^{g^{i-1}} \right)^{n(i)} \right) = 0. \qquad (*)$$

If we expand the product above, we can see that the highest term in $x_1$ is

$$a_{\boldsymbol{m}} x_1^p,$$

for some nonzero $a_{\boldsymbol{m}} \in k$ (which is free of $z$ terms), with $\boldsymbol{m} = (m(1), \dots, m(r))$ and where $p = m(1) + m(2)g + \dots + m(r)g^{r-1}$. Since $g > m(i)$ for all $i$, then each $p$ gives[35] a unique base $g$ number. So $x_1$ satisfies a monic polynomial with coefficients in $S = k[z_2, \dots, z_r]$. That is, $x_1$ is integral over $S$. But then by $(\dagger\dagger)$, $x_2, \dots, x_r$ are all integral over $S$. Since $x_1, \dots, x_r$ generates $R$, we can conclude that $R$ is integral over $S$. This implies that $R$ is finite over $S$. But $S$ is generated by $r-1$ generators, so the inductive hypothesis completes the inductive step. ■

This proof is so mind-bending that it really demands more explanation. From the eyes of an amateur mathematician, it definitely feels like magic. From the eyes of the layman, it is definitely WTF. Here is my take on it based on Nick's explanation in lectures. It is purposely a bit verbose so bear with me.

**Remark.** As in all proofs, when you lose focus of what you want to prove, you are lost in the sea of arguments. Let's talk about the inductive step. We originally had $R = k[x_1, \dots, x_r]$ and we want to show that it contains a (polynomial) subring for which $R$ is finite over it. As usual in an induction proof, we want to put $R$ into a form where we can use the induction hypothesis.

**Overall picture.** The clever part in the proof is $(\dagger\dagger)$, which is really just a non-linear change of variables. By doing such a simple thing, we are able to force an equality $R = k[x_1, z_2, \dots, z_r]$. This equality tells us that it is enough to show that $x_1$ satisfies a monic polynomial with coefficients in $S = k[z_2, \dots, z_n]$. Why? Because if that

[34] *For if there is no such relation, then $R$ is a polynomial ring and we are done*

[35] *Observe that for different vectors $\boldsymbol{n}$, $p$ is different (because it is a sum of $m(i)$ terms). This is because they have different expansions in base $g$. But if $g > m(i)$ for all $i$, which we assumed, then $p$ is necessarily unique.*

happens, then $x_1$ is integral over $S$. So what? Well, let's rewrite (††), we have

$$x_i = z_i + x_1^{g^{i-1}},$$

for $2 \leqslant i \leqslant r$. The $z_i$ is integral over $S$ for free: it satisfies the polynomial $T - z_i \in S[T]$. Since we now know that $x_1$ is integral over $S$, then $x_i$ is integral over $S$ for all $2 \leqslant i \leqslant r$ as well! Since these $x_i$ generates $R$, then **everything** in $R$ is integral over $S$. That's the idea of the proof.

**The protagonist.** Now let's talk about the monic polynomial that $x_1$ satisfies. It comes from the fact that we look at the *highest term in* $x_1$ which is $a_{\boldsymbol{m}} x_1^p$. What's so special about it? Since we want $x_1$ to satisfy a monic polynomial and we have $(*)$, it is natural to try manipulate $(*)$ in some way so that it gives the polynomial that we want. This is a standard trick in mathematics. But the key point is, we want $x_1$ to satisfy a monic polynomial **with coefficients in** $S = k[z_2, \ldots, z_n]$. So we can't immediately conclude that $f$ (after change of variables) gives such a polynomial. But if we play with $(*)$ long enough, we realize that the highest term in $x_1$ has some good properties. The exponent of $x_1$ is dependent on the index $\boldsymbol{n}$, so let's say that maximality of $x_1$ occurs at $\boldsymbol{n} = \boldsymbol{m} = (m(1), \ldots, m(r))$ as in the proof with $p = m(1) + \cdots + m(r)g^{r-1}$. Then the good properties of the $x_1$ term at $\boldsymbol{m}$, which is $a_{\boldsymbol{m}} x_1^p$ are:

f as in the one that we have in the proof above.

- $a_{\boldsymbol{m}}$ is nonzero[36].
- $a_{\boldsymbol{m}}$ is independent of $z$ terms!
- $p$ is *truly* maximal (since $g > m(i)$ for all $i$).

Since $g > m(i)$ for all $i$, this highest term $a_{\boldsymbol{m}} x_1^p$ in $x_1$ is **really** the highest term (that is, $p$ is maximal). If we did not assume this, cancellation might happen. When this happen, it may not be the highest term anymore. But by assuming this, this number $p$ is **unique** in base $g$, so we don't need to worry about cancellation.

But then $a_{\boldsymbol{m}}$ is a unit in $S$, since it is essentially a constant with respect to $S$. So after dividing by $a_{\boldsymbol{m}}$, we get a monic polynomial in $x_1$ of degree $p$ with coefficients in $S$. We don't care that the degree of this monic polynomial is as big as Sasuke's Susanoo (see Naruto ep. 358), we just want some monic polynomial that $x_1$ satisfies, and we have found one; we are done.

**Remark.** Observe that the key step in proving the theorem (or rather, the inductive step) is making the (very) non-linear change of variables. The problem is when $k$ is finite. If $k$ is infinite, it is sufficient to prove the theorem by a suitable *linear* change of variables. Finite fields are trickier, so we need to be more clever.

Recall that the most basic example of $R$-algebras, where $R$ is a ring is the polynomial ring $A = R[X_1, \ldots, X_n]$ over $R$. Clearly it is finitely generated by $\{X_1, \ldots, X_n\}$ where we view $X_i \in A$ for all $A$. Consequently, we can apply the Noether's normalization lemma to $A$. In fact, it is the one that we are most interested in.

**Corollary 3.7** (*Weak Nullstellensatz*). Let $k$ be a field.

(1). If $A$ is a finitely generated $k$-algebra and $\mathfrak{m} \lhd A$ is maximal, then $A/\mathfrak{m}$ is finite over $k$.

(2). If $k$ is algebraically closed, and $A = k[X_1, \ldots, X_n]$, then every maximal ideal $\mathfrak{m} \lhd A$ is of the form $\mathfrak{m} = (X_1 - \ell_1, \ldots, X_n - \ell_n)$ for some $\ell_i \in k$.

[36] *We got this for free from* $f$.

**Proof.** The strategy is to apply the Noether's normalization lemma to $A/\mathfrak{m}$ to prove part (1). Then use part (1) to prove part (2).

**(1).** By the Noether's normalization lemma, there is a subring $R = k[X_1, \dots, X_n]$ of $A/\mathfrak{m}$, which is a polynomial ring over $k$, such that $A/\mathfrak{m}$ is finite over $R$. But since $A/\mathfrak{m}$ is a field, by a preceding lemma, $R$ is also a field. But when is $k[X_1, \dots, X_n]$ is a field? When $n = 0$, so $R = k$; and $A/\mathfrak{m}$ is finite over $R = k$.

**(2).** Since $A = k[X_1, \dots, X_n]$ is a finitely generated $k$-algebra, part (1) implies that any element $x \in A/\mathfrak{m}$, where $\mathfrak{m} \lhd A$ is maximal, is algebraic over $k$. Since $k$ is algebraically closed, then $x \in k$. Consequently, $k \to A/\mathfrak{m}$ is an isomorphism (by composition). Now let $\pi : A \to A/\mathfrak{m} \cong k$ be the canonical projection $a \to a + \mathfrak{m}$. This map is surjective, so there is $\ell_i \in k$ such that $\pi(X_i) = \ell_i$. So $X_i - \ell_i \in \ker \pi = \mathfrak{m}$ and thus, the ideal $\mathfrak{j} = (X_1 - \ell_1, \dots, X_n - \ell_n)$ generated by $X_i - \ell_i$ is contained in $\mathfrak{m}$. But if we evaluate every polynomial in $A$ at $X_i = a_i$, we see that $A/\mathfrak{j} \cong k$. This implies that $\mathfrak{j}$ is maximal i.e. $\mathfrak{j} \supseteq \mathfrak{m}$; and so $\mathfrak{j} = \mathfrak{m}$. ■

It is intuitively clear that if a polynomial $f \in \mathbb{C}[X]$ vanishes at every point in $\mathbb{C}$, then $f$ vanishes identically. That is, it is the zero polynomial over $\mathbb{C}$. In general this is true provided that we are talking about **infinite** fields like $\mathbb{C}$. Despite it's obviousness, it is not true for finite fields (think zero divisors).

**Lemma 3.14.** Let $k$ be an infinite field and let $c \in k[X_1, \dots, X_n]$ such that $c$ vanishes at every point of $k^n$. Then $c$ is identically zero; it is the zero polynomial over $k$.

**Proof.** We prove by induction on the number of variables $n$.

**Base case.** If $n = 1$, then $c \in k[X]$ and has infinitely many zeros (as it vanishes at every point of $k$). But a polynomial of degree $d$ has at most $d$ zeros (since $k$ is a field.) So $c$ is identically zero.

**Inductive hypothesis.** Suppose that the result holds for all polynomials in $n - 1$ variables $X_1, \dots, X_{n-1}$.

**Inductive step.** Let $c \in k[X_1, \dots, X_n]$ such that it vanishes at every point of $k^n$, and write

$$c = \sum_{i=1}^{d} c_i X_n^i,$$

where $d \in \mathbb{N}$ and $c_i \in k[X_1, \dots, X_{n-1}]$. Now fix $\boldsymbol{a} = (a_1, \dots, a_{n-1}) \in k^{n-1}$ and put $\gamma_i = c_i(\boldsymbol{a}) \in k$, the evaluation of $c_i$ at $\boldsymbol{a}$. Then further put

$$\gamma = \sum_{i=1}^{d} \gamma_i X_n^i,$$

and notice that $\gamma \in k[X_n]$. But then $\gamma(b) = 0$ for all $b \in K$ (see margin note for explanation) and so by the base case, we know $\gamma = 0$ identically. So $\gamma_i = c_i(\boldsymbol{a}) = 0$ for all $\boldsymbol{a} \in k^{n-1}$. Therefore, $c_i$ is identically zero by the inductive hypothesis and thus $c$ is identically zero; we are done. ■

observe that $\gamma$ is really just $c$ with the first $n-1$ variables fixed. This justifies why $\gamma$ vanishes at all point of $k$, because we have assumed $c$ to vanish at every point of $k^n$.

This proof looks longer than it should be. If one is skilled enough to see the reduction from $c$ to $\gamma$ just by fixing points of $k^{n-1}$, then it is almost immediate. Now recall that an algebraically closed field $k$ is an infinite field, for otherwise we can always add $1_k$ to any polynomial that (we think) is satisfied by all elements of $k$. Consequently, we can apply the previous lemma to $k$.

**Theorem 3.8** (*Strong Nullstellensatz*)**.** Let $k$ be an algebraically closed field; and let $\mathfrak{p}$ be a maximal ideal in $R = k[X_1, \ldots, X_n]$. Then

$$\mathfrak{p} = \bigcap_{\mathfrak{m} \supseteq \mathfrak{p}} \mathfrak{m},$$

where $\mathfrak{m} \lhd R$ is maximal.

**Proof.** Let $B = R/\mathfrak{p}$. By the Noether's normalization lemma, there is a subring $A = k[X_1, \ldots, X_n]$ of $B$ such that $B$ is finite over $A$. Now suppose $f \in \bigcap_{\mathfrak{n} \lhd B} \mathfrak{n}$, where $\mathfrak{n}$ is maximal. We want to show that $f = 0$, the zero polynomial. Suppose not. Then $f$ is integral over $A$ and so satisfies a monic polynomial of minimal degree with coefficients in $A$:

$$f^p + c_{p-1} f^{p-1} + \cdots + c_1 f + c_0 = 0,$$

where $c_i \in A$. By minimality of the degree and the fact that $f \neq 0$, we have $c_0 \neq 0$. Now suppose $\mathfrak{m} \lhd A$ is maximal. Since $A$ is Noetherian[37], Theorem 3.6 tells us that there exists a maximal ideal $\mathfrak{n} \lhd B$ such that $\mathfrak{n} \cap A = \mathfrak{m}$. Since we assumed $f$ is in the intersection of maximal ideals in $B$, we have $f \in \mathfrak{n}$. This implies that

$$c_0 = f(-f^{p-1} - \cdots - c_1) \in \mathfrak{n}.$$

But $c_0 \in A$, and so $c_0 \in \mathfrak{n} \cap A = \mathfrak{m}$. Consequently, $c_0 \in \bigcap_{\mathfrak{m} \lhd A} \mathfrak{m}$ where $\mathfrak{m}$ is maximal. That is, $c_0$ is a polynomial in $n$ variables that vanishes at every point of $k^n$. The previous lemma implies that $c_0 = 0$ identically. 何! ■

**Definition** (*Nilpotent*)**.** Let $R$ be a ring. An element $r \in R$ is said to be **nilpotent** if there exists $n \in \mathbb{N}$ such that $r^n = 0_R$. The element $n$ is then called the **index** of $r$.

**Definition** (*Reduced ring*)**.** Let $R$ be a ring. We say that $R$ is **reduced** if $0_R$ is the only nilpotent element in $R$.

**Definition** (*Radical*)**.** Let $R$ be a ring and $I \lhd R$. The **radical** of $I$ is the set

$$\sqrt{I} = \{r \in R \mid r^n \in I \text{ for some } n \in \mathbb{N}\},$$

which is also an ideal in $R$. We say that $I$ is **radical** if $I = \sqrt{I}$.

Note that $I \subseteq \sqrt{I}$, so $I$ is radical iff $I \supseteq \sqrt{I}$.

We have defined quite a bunch of new things. Here is where they are related.

**Lemma 3.15.** Let $R$ be a ring and let $I \lhd R$. Then $I$ is radical *if and only if* $R/I$ is reduced.

**Proof.** ($\Rightarrow$). Let $I$ be radical and let $x + I \in R/I$ be nilpotent. Then there exists $n \in \mathbb{N}$ such that $(x+I)^n = 0_{R/I}$. But this implies that $x^n + I = 0_{R/I}$ or equivalently $x^n \in I$. Since $I$ is radical, this implies that $x \in I$ and so $x + I = 0_{R/I}$.

[37] *since $k$ is a field and hence, Noetherian*

($\Leftarrow$). This is just reversing our argument above. Suppose that $R/I$ is reduced, and let $x \in R$ be such that $x^n \in I$ for some $n \in \mathbb{N}$. We want to show that $x \in I$. By definition

$$0_{R/I} = 0_R + I = (0_R + x^n) + I = x^n + I = (x+I)^n.$$

But since $R/I$ is reduced, then we are forced to have $x + I = 0_{R/I}$ or equivalently $x \equiv 0 \pmod{I}$ or further equivalently $x \in I$. ■

**Lemma 3.16.** Let $k$ be an algebraically closed field and let $R = k[X_1, \dots, X_n]$. Then for every radical ideal $I \lhd R$, we have

$$I = \bigcap_{\mathfrak{m} \supseteq I} \mathfrak{m},$$

where $\mathfrak{m} \lhd R$ is maximal.

To prove this, it suffices to prove that $I$ is the intersection of prime ideals $\mathfrak{p}$ that contain it. Why? Because if $\mathfrak{p} = \bigcap_{\mathfrak{m} \supseteq \mathfrak{p}} \mathfrak{m}$, then by elementary set theory we have

$$I = \bigcap_{\mathfrak{p} \supseteq I} \mathfrak{p} = \bigcap_{\mathfrak{p} \supseteq I} \bigcap_{\mathfrak{m} \supseteq \mathfrak{p}} \mathfrak{m} = \bigcap_{\mathfrak{m} \supseteq I} \mathfrak{m}.$$

**Proof.** Suppose $I$ is radical. Then $R/I$ is reduced. We want to show that if $x \in R/I$ is nonzero, then there exists a prime ideal $\mathfrak{p} \lhd R/I$ such that $x \notin \mathfrak{p}$. It suffices to talk about $R$ in place of $R/I$ because of the correspondence theorem of quotient rings extended to the prime ideals case. So suppose that $x \in R$ is nonzero. Now consider the set

$$S = \{J \lhd R \mid x^n \notin J \text{ for all } n \in \mathbb{N}\}.$$

Since $x \neq 0$, then[38] $x$ is not nilpotent and so $x^k \notin (0)$ for all $k \in \mathbb{N}$. This implies that $S$ is non-empty as $(0) \in S$. Now since $R$ is Noetherian, Lemma 1.8 implies that $S$ has a maximal element, say, $\mathfrak{m} \lhd R$. So $x^k \notin \mathfrak{m}$ for all $k \in \mathbb{N}$ by definition. But note that for any $J \supsetneq \mathfrak{m}$, then $x^j \in J$ for some $j \in \mathbb{N}$ because $\mathfrak{m}$ is maximal. We now claim that $\mathfrak{m}$ is prime.

Suppose not. Then there exists $r, s \in R$ such that $r, s \notin \mathfrak{m}$ but $rs \in \mathfrak{m}$. But $r, s \notin \mathfrak{m}$ means that $(\mathfrak{m} + (r)) \supsetneq \mathfrak{m}$ and $(\mathfrak{m} + (s)) \supsetneq \mathfrak{m}$. Since $\mathfrak{m}$ is maximal in $S$, there exists $a, b \in \mathbb{N}$ such that $x^a \in \mathfrak{m} + (r)$ and $x^b \in \mathfrak{m} + (s)$. Then we can write $x^a = m + rp$ and $x^b = n + sq$ where $m, n \in \mathfrak{m}$ and $p, q \in R$. Then

$$x^{a+b} = \underbrace{(mn + msq + nrp)}_{\in \mathfrak{m}} + \underbrace{rs}_{\in \mathfrak{m}} pq,$$

so $x^{a+b} \in \mathfrak{m}$. 何! ■

[38] *we assumed $I$ is radical, so $R/I$ is reduced.*

# 4 Unique factorization

In abstract algebra, unique factorization is usually mentioned in the foundation stage when discussing rings. We do not follow this convention to emphasize that the mathematics we developed so far is really independent of unique factorization. We wanted to highlight the real star: Noetherian rings. Of course, some of the rings that we have discussed do have the "unique factorization" property, whatever that means for now. For example, fields and PIDs have this property.

We will start with some basic definitions that are crucial in the discussion of factorization. We have defined units and irreducible before, but for the sake of organization, we will define these again.

**Definition** (*Unit*)**.** Let $R$ be a ring. We say $u \in R$ is a **unit** if there exists $v \in R$ such that $u \cdot v = 1_R$.

All the definitions below require integral domains.

**Definition** (*Irreducible*)**.** A nonzero element $r$ of an integral domain $R$ is an **irreducible** if $r$ is not a unit, and if $r = st$ implies one of $s, t$ is a unit.

**Definition** (*Divide*)**.** Let $R$ be a integral domain, and let $r, s \in R$. We say $r$ **divides** $s$, denoted $r \mid s$, if there exists $t \in R$ such that $s = rt$.

Equivalently, we have $r \mid s \iff s \in (r) \iff (s) \subseteq (r)$.

**Definition** (*Prime*)**.** Let $R$ be an integral domain. An element $r \in R$ is **prime** if $r \neq 0_R$, is not a unit and if $r \mid st$ implies $r \mid s$ or $r \mid t$.

Note that the notion of prime really depends on the ring.

**Example.** All nonzero elements in $\mathbb{R}$ is a unit, so $5 \in \mathbb{R}$ is not prime. But $5 \in \mathbb{Z}$ is prime.

This is also a good place to remind you that in general, "prime numbers" is not equal to "prime elements" or even irreducibles. In $\mathbb{Z}$ this is true but not in, say, $\mathbb{R}$ as demonstrated in the example above.

## 4.1 Factorization in integral domains

Now prime implies irreducible, but the reverse implication is false in general.

**Lemma 4.1.** Let $R$ be an integral domain. If $r \in R$ is prime, then it is irreducible.

**Proof.** Let $r \in R$ be prime and assume that $r = st$ for some $s, t \in R$. Because $r \mid r$, then $r \mid st$. Since $r$ is prime, $r \mid s$ or $r \mid t$. Suppose $r \mid s$. Then by definition $s = ru$ for some $u \in R$. So we have $r = rut$. By the cancellation law ($R$ is a domain), we have $ut = 1$. This implies that $t$ is a unit. The same holds similar if we assumed $r \mid t$. ■

As you would have guess[39], prime ideals have a relation with prime elements of a domain.

[39] *or tempted to guess, given that we have been talking about prime ideals for a whole chapter now.*

**Lemma 4.2.** Let $R$ be an integral domain and let $r \in R$. Then $r$ is prime if and only if $(r)$ is a nonzero prime ideal.

**Proof.** $(\Rightarrow)$. Suppose $r$ is prime and consider the principal ideal $(r)$. Suppose $ab \in (r)$. Then $r \mid ab$. Since $r$ is prime, $r \mid a$ or $r \mid b$. Consequently, $a \in (r)$ or $b \in (r)$. So $(r)$ is a prime ideal and is nonzero since $r \neq 0_R$ by definition.

$(\Leftarrow)$. Suppose $(r)$ is a nonzero prime ideal. Now suppose $r \mid ab$ for some $a, b \in R$. Then $ab \in (r)$. But $(r)$ is prime so by definition $a \in (r)$ or $b \in (r)$. Equivalently, $r \mid a$ or $r \mid b$. Moreover, by definition of a prime ideal, $(r)$ is proper and so $r$ necessarily is not a unit. So $r$ is prime. ■

We've been throwing the term unique factorization a lot. Here is what it means.

**Definition** (*UFD*)**.** An integral domain $R$ is a **unique factorization domain** (or simply UFD) if

(1). Every nonzero $r \in R$ can be written as $r = x_1 \cdots x_n$ where each $x_i$ is irreducible;

(2). If moreover we also have $r = y_1 \cdots y_m$ where each $y_j$ is irreducible, then $m = n$ and (after reordering if necessary), then each $y_i = u_i x_i$ with $u_i \in R$ a unit.

$r$ can be essentialy uniquely factorized into a product of irreducibles.

**Example.** Any field $k$ is a UFD. It has no irreducible elements because any nonzero element is a unit. So axiom (1) vacuously holds.

**Example.** $\mathbb{Z}$ is a UFD since it has the Fundamental Theorem of Arithmetic.

In Noetherian domains, we get (1) for free but not (2). That is, we don't necessarily have the uniqueness part. But whenever we have uniqueness, we get the nice fact that being irreducible is equivalent to being prime.

**Proposition 4.1.** Let $R$ be a Noetherian integral domain. Then

(1). Every nonzero $r \in R$ is a product of irreducibles.

(2). $R$ is a UFD if and only if every irreducible element is prime.

**Proof.** See lecture notes. ■

**Remark.** Noetherian is actually too strong. We need ACC on principal ideals (ACCP) only not on all ideals. Also, statement (2) is true for general UFDs, not necessarily Noetherian UFDs (which is stronger).

**Slogan.** UFD *if and only if* irreducible $\iff$ prime.

The following proposition tells us that *highest common factors* (otherwise known as greatest common divisors) and lowest common multiples exists in a unique factorization domain. We could have defined it for general domains, but their existence are not always guaranteed.

**Proposition 4.2** (*hcf-lcm*)**.** Let $R$ be a UFD, and let $r, s \in R$. Suppose $r = \prod p_i^{e_i}$ and

in lectures, Nick for some reason assumed that $R$ is a Noetherian UFD.

$s = \prod p_i^{f_i}$ where $p_i$ is irreducible and $e_i, f_i \geqslant 0$. Then the following elements

$$\operatorname{hcf}(r,s) = \prod_i p_i^{\min(e_i, f_i)},$$

$$\operatorname{lcm}(r,s) = \prod_i p_i^{\max(e_i, f_i)},$$

of $R$ exist and are called the **highest common factor** and **lowest common multiple** of $r$ and $s$ respectively. This notion extends naturally for any finite set of elements.

Note that the highest common factor is only defined up to a unit. That is, if $g, h$ are both the highest common factor $\operatorname{hcf}(a,b)$ of $a, b \in R$, then $g = uh$ for some unit $u \in R$. This is not surprising considering that unique factorization is also *unique* up to units. So as far as factorization is concern, units don't matter.

think of the units in $\mathbb{Z}$. Now allow the hcf to be negative, what do you get?

**Definition** (*Coprime elements*)**.** Let $R$ be a unique factorization domain. Then $r, s \in R$ are said to be **coprime** if $\operatorname{hcf}(r,s) = 1_R$. We say that $r_1, \ldots, r_n \in R$ are **pairwise coprime** if any two of them are coprime.

Recall that every maximal ideal is prime. It turns out that in principal ideal domains, the converse is true.

**Proposition 4.3.** If $R$ is a principal ideal domain, then every nonzero prime ideal $\mathfrak{p} \lhd R$ is maximal.

**Proof.** Suppose $\mathfrak{p} \neq 0$ is prime but not maximal. Let $J \lhd R$ be any ideal such that

$$\mathfrak{p} \subsetneq J \subsetneq R, \qquad (*)$$

By hypothesis, every ideal in $R$ is principal, so write $\mathfrak{p} = (x)$ and $J = (y)$ for some $x, y \in R$. Now $\mathfrak{p} \subsetneq J$ implies that $x \in (y)$, and so $x = yz$ for some $z \in R$. Since $\mathfrak{p} = (x)$ is a prime ideal, $x$ is prime and so is irreducible. Therefore one of $y$ or $z$ is a unit in $R$. If $y$ is a unit, then $J = (y) = R$. 何! If $z$ is a unit, then $(x) = (y)$ and so $\mathfrak{p} = J$. 何! Both of which contradicts $(*)$. ■

recall that if $\mathfrak{m} \lhd R$ is maximal, then if $\mathfrak{m} \subseteq J \lhd R$, then $J = \mathfrak{m}$ or $J = \mathfrak{m}$.

**Slogan.** $R$ is PID *implies* (prime $\iff$ maximal ideals).

**Proposition 4.4** (PID $\Rightarrow$ UFD)**.** Every principal ideal domain is a unique factorization domain.

The converse is definitely **not true**. $\mathbb{Z}[X]$ is a unique factorization domain, but we know that it is not a principal ideal domain (cf. $(2, X) \lhd \mathbb{Z}[X]$ example). The idea of the proof is similar to the preceding proposition. We use the fact that every Noetherian domain has unique factorization if and only if every irreducible element of it is prime (see Proposition 4.1).

and PIDs are Noetherian rings which are also integral domains — Noetherian domains!

**Proof.** Suppose that $R$ is a principal ideal domain. Let $x \in R$ be irreducible. We need to show that it is prime. It suffices to show that $(x)$ is maximal[40]. Suppose not. Then let $J \lhd R$ such that $(x) \subsetneq J \subsetneq R$. Since every ideal in $R$ is principal, we can

[40] *as maximal implies prime, and prime principal ideals implies prime generators.*

write $J = (y)$ for some $y \in R$. Since $(x) \subsetneq J = (y)$, this means that $x = yz$ for some $z \in R$. But $x$ is irreducible so either $y$ or $z$ is a unit. If $y$ is a unit, $J = (y) = R$. 何! If $z$ is a unit, then $(x) = (y)$. 何! ■

**Proposition 4.5.** Let $R$ be a PID and let $r \in R$. Then the following are equivalent:

(i). $r \in R$ is irreducible.

(ii). $(r)$ is a maximal ideal in $R$.

(iii). $R/(r)$ is a field.

**Proof.** (ii) $\Leftrightarrow$ (iii) is immediate by the correspondence theorem on ideals. So it suffices to prove (i) $\Leftrightarrow$ (ii). But this is easy with the results we have so far:

$$r \text{ is irreducible} \overset{\text{UFD}}{\iff} r \text{ is prime} \overset{\text{ID}}{\Leftrightarrow} (r) \text{ is a prime ideal} \Leftrightarrow (\mathrm{r}) \text{ is a maximal ideal},$$

where the last forward implication is due to Proposition 4.3. The last reverse implication is due to Corollary 1.2. ■

**Proposition 4.6.** Let $K$ be a field and let $f \in K[X]$. Then

$$(f) \lhd K[X] \text{ is maximal} \Leftrightarrow f \text{ is irreducible in } K[X].$$

**Proof.** Since $K$ is a field, $K[X]$ is Euclidean and hence a PID. So we can apply the preceding Proposition 4.5. ■

**Proposition 4.7.** Let $R$ be a principal ideal domain and let $a_1, \ldots, a_n \in R$. Let $h = \operatorname{hcf}(a_1, \ldots, a_n)$. Then

$$(h) = (a_1, \ldots, a_n).$$

That is, the ideal generated by $a_1, \ldots, a_n$ is equal to the principal ideal generated by their hcf.

The notion of hcf is well-defined here as we now know PIDs are UFDs.

**Proof.** Suppose $(c) = (a_1, \ldots, a_n)$ for some $c \in R$. We want to show that $(c) = (h)$.

$(\Leftarrow)$. Now $h \mid a_i$ for all $i$, so there exists $d_i \in R$ such that $a_i = hd_i$. By definition, we can write $c = \sum e_i a_i$ for some $e_i \in R$. Therefore, $c = \sum e_i d_i h$. This implies that $c \in (h)$ or equivalently, $(c) \subseteq (h)$.

$(\Rightarrow)$. By definition, $a_i \in (c)$ and so $c \mid a_i$ for all $i$. But $h$ is the **highest** common factor of $a_i$ and so $c \mid h$. That is, $h \in (c)$ or equivalently, $(h) \subseteq (c)$. ■

Recall that if we have two ideals $I, J$ of a ring $R$, then their sum

$$I + J = \{i + j \mid i \in I, j \in J\},$$

is also an ideal in $R$. Moreover if $I = (a_1, \ldots, a_n)$ and $J = (b_1, \ldots, b_m)$, then $I + J = (a_1, \ldots, a_n, b_1, \ldots, b_m)$.

**Definition** (*Coprime ideals*)**.** Let $R$ be a ring and let $I, J \lhd R$. Then $I$ and $J$ are **coprime** if $I + J = R$. A collection of ideals $I_1, \dots, I_n \lhd R$ are **pairwise coprime** if any two are coprime.

Coprime ideals are also known as comaximal ideals. There are two things name coprime so they better be related.

**Proposition 4.8.** Let $R$ be a principal ideal domain and let $x, y \in R$. Then

$$(x) \text{ and } (y) \text{ are coprime} \iff x \text{ and } y \text{ are coprime.}$$

**Proof.** We have proved that in a PID, $(\operatorname{hcf}(x, y)) = (x, y)$. Now $(x, y) = (x) + (y)$ simply by definition of the sum ideal. So

$$(x), (y) \text{ are coprime ideals} \iff (x, y) = R \iff (\operatorname{hcf}(x, y)) = R \iff \operatorname{hcf}(x, y) = 1_R.$$

But the last statement just means $x, y$ are coprime, so we are done. ■

**Lemma 4.3.** Let $R$ be a ring such that $I_1, \dots, I_n \lhd R$ are pairwise coprime. Then $I_1$ and $\bigcap_{k \geqslant 2} I_k$ are coprime.

**Proof.** We can write

$$1_R = i_{k1} + j_k,$$

for some $i_{k1} \in I_1$ and some $j_k \in I_k$ (for $k \geqslant 2$) by the pairwise coprime assumption. We can then rewrite $1_R - i_{k1} = j_k$. So we have an equality of products

$$\prod_{k \geqslant 2} (1_R - i_{k1}) = \prod_{k \geqslant 2} j_k.$$

After expanding the products, the lhs looks like $1_R - r$ for some $r \in I_1$. The rhs evidently is an element of $\bigcap_{k \geqslant 2} I_k$. Consequently,

$$1_R = \underbrace{r}_{\in I_1} + \underbrace{\prod_{k \geqslant 2} j_k}_{\in \bigcap_{k \geqslant 2} I_k}.$$

The elements were arbitrary so the ideals $I_1$ and $\bigcap_{k \geqslant 2} I_k$ are coprime, as desired. ■

Next, we prove the Chinese Remainder Theorem.

**Theorem 4.1** (Chinese Remainder Theorem)**.** Let $R$ be a ring and let $I_1, \dots, I_n \lhd R$ be pairwise coprime. Then there is an isomorphism

$$R \Big/ \left( \bigcap_{k=1}^{n} I_k \right) \xrightarrow{\sim} \bigoplus_{k=1}^{n} R/I_k,$$

of rings and $R$-modules.

In fact, it is also an isomorphism of $R$-algebras.

Note that this is more general than the standard Chinese Remainder Theorem (CRT). Usually one impose the condition that $R$ is a principal ideal domain and talk about pairwise coprime elements of $R$ (instead of pairwise coprime ideals).

**Proof.** We will prove by induction on $n$.

**Base case.** If $n = 2$. Let $\phi : R \to R/I_1 \oplus R/I_2$ be the homomorphism defined by $x \mapsto (x \bmod I_1, x \bmod I_2) = (x + I_1, x + I_2)$. We now claim[41] that $\phi$ is surjective. By assumption, $I_1 + I_2 = R$. So there exists $i_1 \in I_1$ and $i_2 \in I_2$ such that $i_1 + i_2 = 1$. Now suppose we are given $(a + I_1, b + I_2) \in R/I_1 \oplus R/I_2$. Then we can write

$$\begin{aligned} a &= a \cdot 1 = ai_1 + ai_2, \\ b &= b \cdot 1 = bi_1 + bi_2. \end{aligned}$$

Consider $x = bi_1 + ai_2$. Then we have

$$\begin{aligned} x + I_1 &= ai_2 + bi_1 + I_1 = ai_2 + I_1 = ai_2 + ai_1 + I_1 = a + I_1 \\ x + I_2 &= bi_1 + ai_2 + I_2 = bi_1 + I_2 = bi_1 + bi_2 + I_2 = b + I_2. \end{aligned}$$

Therefore, $\phi(x) = (a + I_1, b + I_2)$, so $\phi$ is indeed surjective. It is also easy to see that $\ker \phi = I_1 \cap I_2$. By the first isomorphism theorem, we then have $R/\ker\phi \cong \operatorname{im}\phi$. That is $R/(I_1 \cap I_2) \cong R/I_1 \oplus R/I_2$.

**Inductive hypothesis.** Suppose the theorem holds for $n-1$ pairwise coprime ideals.

**Inductive step.** By the preceding Lemma 4.3, $I_1$ and $\bigcap_{k=2}^n I_k$ are coprime. So by the base case, there is an isomorphism

$$R\Big/\left(\bigcap_{k=1}^n I_k\right) = R\Big/\left(I_1 \cap \bigcap_{k=2}^n I_k\right) \xrightarrow{\sim} R/I_1 \oplus R\Big/\bigcap_{k=2}^n I_k.$$

But now, stare at the rhs; the inductive hypothesis implies that

$$R\Big/\bigcap_{k=2}^n I_k \cong \bigoplus_{k=2}^n R/I_k,$$

so the inductive step is complete. ■

**Proposition 4.9.** Let $R$ be a principal ideal domain and let $z_1, \ldots, z_n \in R$. Let $h = \operatorname{hcf}(z_1, \ldots, z_n)$ and $\ell = \operatorname{lcm}(z_1, \ldots, z_n)$. Then

(1). $(\ell) = \bigcap_{i=1}^n (z_i)$.

(2). $(h) = (z_1) + \cdots + (z_n)$.

**Proof.** For (1). Suppose $y \in R$. Then

$$y \in \bigcap_{i=1}^n (z_i) \iff z_i \mid y \text{ for all } i \iff \ell \mid y \iff y \in (\ell).$$

For (2). We have proved that $(h) = (z_1, \ldots, z_n)$; this was Proposition 4.7. But simply by definition $(z_1, \ldots, z_n) = (z_1) + \cdots + (z_n)$. ■

As mentioned previously, what we have proved is a much more general construction. The standard CRT is the following corollary.

[41] *here, we will use the coprime condition*

**Corollary 4.1** (*CRT in PID*)**.** Let $R$ be a principal ideal domain and let $x_1, \ldots, x_n \in R$ be nonzero and pairwise coprime. Then

$$R/(x_1, \ldots, x_n) \cong \bigoplus_{i=1}^{n} R/(x_i),$$

as rings and $R$-modules.

**Proof.** Apply Chinese Remainder Theorem. ■

The Chinese Remainder Theorem gives us a passage between our original structure theorem to an alternative version.

**Corollary 4.2** (Structure Theorem, alternative version)**.** Let $R$ be Euclidean and let $M$ be a finitely generated $R$-module. Then

$$M \cong \bigoplus_{i=1}^{n} R/(d_i) \oplus R^m,$$

where each $d_i$ is a prime power.

That is, $d_i = p_i^k$ for some prime $p_i$ and $k \in \mathbb{N}$.

**Proof.** We know from the usual structure theorem that

$$M \cong \bigoplus_{i=1}^{n} R/(x_i) \oplus R^m,$$

where $x_1 \mid x_2 \mid \cdots \mid x_n$. Now choose any of the $x_i$ and let $x = x_i$. Then we can write

$$x = \prod_{i=1}^{n} p_i^{e_i},$$

where the $p_i$ are distinct primes and and $e_i \geqslant 0$. By definition of the primes, they are pairwise coprime. So by the Chinese Remainder Theorem, we have

$$R/(x) \cong \bigoplus_{i=1}^{n} R/(p_i^{e_i}).$$

■

Note that this version of the structure theorem does not lead us to the submodule structure theorem.

**Example.** Consider the ring of integers $\mathbb{Z}$. Suppose $M = \mathbb{Z}/(2) \oplus \mathbb{Z}/(6)$. Now since $6 = 2 \cdot 3$, we have $M \cong \mathbb{Z}/(2) \oplus \mathbb{Z}/(2) \oplus \mathbb{Z}/(3)$.

## 4.2 Factorization in polynomial rings

The goal now is to prove that if $R$ is a unique factorization domain, then so is $R[X]$.

**Definition** (*Content*)**.** Let $R$ be a unique factorization domain and let $f \in R[X]$. Then the **content** of $f$, denoted, $c(f)$, is the highest common factor of its coefficients. If $c(f) = 1_R$, we say that $f$ is **primitive**.

Since the highest common factor is defined uniquely up to units, so is the content.

**Example.** Consider $f(X) = 2 + 6X^2$ and $g(X) = 2 + 3X^3$ defined over $\mathbb{Z}$. Then $c(f) = \operatorname{hcf}(2, 6) = 2$ and $c(g) = 1$.

**Lemma 4.4.** Let $R$ be a UFD and let $f \in R[X]$. Then $f$ can be written uniquely as $f = c(f)f_0$ where $f_0 \in R[X]$ is primitive.

**Proof.** Just pull out the primitive factor. ■

**Lemma 4.5.** Let $R$ be a UFD. If $f, g \in R[X]$ are primitive, then so is $fg$.

Recall that we have the fact that irreducible $\iff$ prime in general UFDs (not necessarily Noetherian UFDs). We will use that fact here.

**Proof.** Suppose not. Then there exists $a \in R$ irreducible such that $a \mid c(fg)$. Since $R$ is a UFD, $a$ is prime. Now consider the obvious[42] homomorphism $R \to R/(a)$. The ideal $(a)$ is prime by Lemma 4.2 since $a$ is prime. So $R/(a)$ is a domain by Theorem 1.9. Our homomorphism extends naturally to a homomorphism[43]

$$R[X] \xrightarrow{\pi} (R/(a))[X].$$

Since $R/(a)$ is a domain, Proposition 1.6 implies that $(R/(a))[X]$ is also a domain. Therefore[44], $\ker \pi$ is a prime ideal in $R[X]$. Since $a \mid c(fg)$, we have by definition that $\pi(fg) = 0$. And since $\pi$ is a homomorphism, so $\pi(f)\pi(g) = 0$. We are in a domain, so suppose $\pi(f) = 0$. This means that the coefficients of $f$ are 0 modulo $(a)$. In other words, $a \mid c(f)$. 何! This contradicts the assumption that $f$ is primitive. ■

**Lemma 4.6.** Let $R$ be a UFD and let $f, g \in R[X]$. Then $c(fg) = c(f)c(g)$.

**Proof.** We can write $f = c(f)f_0$ and $g = c(g)g_0$ where $f_0, g_0 \in R[X]$ are primitive. So $fg = c(f)c(g)f_0g_0$. By the preceding lemma, $f_0g_0$ is primitive. On the other hand, we can write uniquely $fg = c(fg)h_0$ where $h_0 \in R[X]$ is primitive. So $h_0$ and $f_0g_0$ coincides and $c(f)c(g) = c(fg)$ up to units. ■

**Proposition 4.10** (*Gauss's lemma*)**.** Let $R$ be a UFD and let $K = \operatorname{Frac}(R)$. If $f \in R[X]$ is irreducible and $\deg f \geqslant 1$, then $f$ is irreducible in $K[X]$.

The converse is true as well if $f$ is primitive.

**Proof.** Suppose $f = gh$ where $g, h \in K[X]$ such that $\deg g, \deg h \geqslant 1$ (so $g, h$ are not units). We can then write $g = G/a$ and $h = H/b$ where $G, H \in R[X]$ and $a, b \in R$ such that $\operatorname{hcf}(a, c(G)) = \operatorname{hcf}(b, c(H)) = 1$. So $f = (GH)/(ab)$. Since $f$ is irreducible in $R[X]$, then $c(f) = 1$. So $GH = abf$ and so $c(G)c(H) = ab$. But $a$ and $c(G)$ are coprime, likewise for $b$ and $c(H)$. So $a \mid c(H)$ and $b \mid c(G)$. So

$$f = \frac{GH}{ab} = \frac{G}{b} \cdot \frac{H}{a},$$

where $G/b, H/a \in R[X]$. 何! This contradicts the assumption that $f$ is irreducible in $R[X]$. ■

[42] *simply by taking $r \mapsto r + (a)$ i.e. reducing modulo $(a)$.*

[43] *this homomorphism takes a polynomial in $R[X]$ and reduces its coefficients modulo $(a)$.*

[44] $\operatorname{im} \pi = (R/(a))[X] \cong R[X]/\ker \pi$ *is a domain. So $R[X]/\ker \pi$ better be an integral domain and now apply Theorem 1.9.*

**Example.** Recall that $\mathbb{Q} = \mathrm{Frac}(\mathbb{Z})$. Any $f \in \mathbb{Z}[X]$ with $\deg f \geqslant 1$ irreducible is also irreducible in $\mathbb{Q}[X]$.

**Lemma 4.7.** Let $R$ be a unique factorization domain and let $K = \mathrm{Frac}(R)$ be its field of fractions. If $f \in R[X]$ is irreducible and $\deg f \geqslant 1$, then

$$R[X] \cap (f \cdot K[X]) = f \cdot R[X].$$

where $fK[X]$ is an ideal in $K[X]$ and $fR[X]$ is an ideal in $R[X]$.

**Proof.** We will first show that rhs $\subseteq$ lhs. Suppose $h \in f \cdot R[X]$. Then $h \in R[X]$ and $h \in fK[X]$. So done.

Now suppose $p \in R[X] \cap (f \cdot K[X])$. $f$ is irreducible and $\deg f \geqslant 1$, so $c(f) = 1$. So we can write $p = fg$ for some $g \in K[X]$. We can write $g = G/a$ where $G \in R[X]$, $a \in R$ and $a$ is coprime to $c(G)$. So $ap = fG$. This gives $a \cdot c(p) = c(p) \cdot c(G) = c(G)$. This implies[45] that $a$ is a unit. Then $g \in R[X]$, and we are done. ■

**Theorem 4.2.** If $R$ is a Noetherian UFD, then so is $R[X]$.

Actually, this is true for UFDs in general. But the proof for Noetherian UFD is easier.

**Proof.** Denote $R_X = R[X]$. By the Hilbert's Basis Theorem, $R_X$ is Noetherian. So it is enough to show that every irreducible $f \in R[X]$ is prime.

Case 1. Suppose $f \in R_X$ such that $\deg f = 0$ (i.e. constants). Then $f$ is prime in $R$. So $R/f$ is a domain. Then $R_X/(f \cdot R_X) \cong (R/fR)[X]$ which is a domain. So $fR_X$ is a prime ideal, and thus $f$ is prime, as desired.

Case 2. Suppose $f \in R_X$ such that $\deg f \geqslant 1$. Then by the preceding lemma, $R_X \cap (f \cdot K_X) = f \cdot R_X$, where $K_X = K[X]$ and $K = \mathrm{Frac}(R)$. So,

$$R_X/(f \cdot R_X) = R_X/(R_X \cap f \cdot K_X) \subseteq K_X/(f \cdot K_X).$$

which is a subring of $K_X/(f \cdot K_X)$. By Gauss's lemma, $f$ is irreducible in $K_X$ and $K_X$ is a unique factorization domain (because it is Euclidean). So $K_X/(f \cdot K_X)$ is a domain and so $R_X/(f \cdot R_X)$ is a domain. This implies that $f$ is prime, as desired. ■

We have seen so many types of domains, let us see a classification of what we have so far.

ring $\subsetneq$ domains $\subsetneq$ UFD $\subsetneq$ Noetherian UFD $\subsetneq$ PID $\subsetneq$ Euclidean domains $\subsetneq$ fields

Observe that we did not say UFDs are Noetherian domains. This is because there are UFDs that are not Noetherian. For example, consider the polynomial ring $R = k[X_1, \ldots]$ over some field $k$ in infinitely many indeterminates (not formal power series). Any $f \in R$ is a polynomial in **finite** indeterminates. So since $k$ is a UFD, $k[X]$ is a UFD. But the ideal $(X_1, \ldots)$ is not finitely generated.

## 4.3 Factorization in $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$

Let us focus our attention to polynomial rings $R[X]$ when $R$ is $\mathbb{Z}$ or $\mathbb{Q}$.

[45] $\mathrm{hcf}(a, c(G)) = 1$ *and* $a \mid c(G)$. *So* $a$ *must be a unit in* $R$.

**Theorem 4.3** (*Eisenstein's criterion*)**.** Let $p \in \mathbb{Z}$ be a prime number and let $f \in \mathbb{Z}[X]$ be the polynomial

$$f = \sum_{i=0}^{n} a_i X^i,$$

such that $p \nmid a_n$, $p \mid a_i$ for all $0 \leqslant i < n$, and $p^2 \nmid a_0$. Then $f$ is irreducible in $\mathbb{Z}[X]$; and hence in $\mathbb{Q}[X]$ by Gauss's lemma.

**Proof.** WLOG[46], we can assume that $f$ is primitive. It is enough to show that $f$ is irreducible in $\mathbb{Z}[X]$ as then Gauss's lemma completes the claim. Suppose $f = gh$ where $g, h \in \mathbb{Z}[X]$. Now write $\bar{f} = f \pmod p$. Then $\bar{f} = \overline{a_n} X^n$ since $p \mid a_i$ for all $0 \leqslant i < n$. By unique factorization in $(\mathbb{Z}/p\mathbb{Z})[X]$, we can write $\bar{g} = \beta X^r$ and $\bar{h} = \gamma X^{n-r}$ for some nonzero $\beta, \gamma \in \mathbb{Z}/p\mathbb{Z}$. In particular, the constant terms of $\bar{g}_0$ of $\bar{g}$ and $\bar{h}_0$ of $\bar{h}$ are 0 modulo $p$. But then $\bar{a}_0$ is divisible by $p^2$. 何! So either $r = 0$ or $n = r$ and thus by primitivity of $g$ and $h$ (since $f$ is primitive), either $g$ or $h$ is a unit in $\mathbb{Z}[X]$, so $f$ is irreducible. ■

Eisenstein's criterion is really an important tool, let us look at some examples.

**Example.** Claim: The polynomial $f = X^5 + 4X^2 + 8X + 2$ is irreducible in $\mathbb{Q}[X]$.

*Proof.* By Eisenstein's criterion with $p = 2$, we see that the claim is valid.

Sometimes, we cannot always invoke Eisenstein's criterion to prove irreducibility in $\mathbb{Q}[X]$, especially when one of the non-leading coefficients is 1. Fortunately, we can still invoke Gauss's lemma to turn it into a problem of irreducibility in $\mathbb{Z}[X]$ and work from there (which is usually easier).

**Example.** Claim: The polynomial $f = X^3 + X + 1$ is irreducible in $\mathbb{Q}[X]$.

*Proof.* By Gauss's lemma, it suffices to prove that $f$ is irreducible in $\mathbb{Z}[X]$. Suppose that $f$ is reducible in $\mathbb{Z}[X]$. Then since $\deg f = 3$, we can write $f$ as a product with at least one linear factor i.e. it is divisible by a linear factor. By inspection of coefficients of $f$, we see that such a linear factor must be of the form $X \pm 1$. But $f(1) = 3$ and $f(-1) = -1$, so $f$ does not vanish at $\pm 1$ which implies that $f$ is not divisible by $X \pm 1$. 何!

**Corollary 4.3.** Let $p \in \mathbb{Z}$ be a prime number and let $f \in \mathbb{Z}[X]$ be the polynomial

$$f = X^{p-1} + X^{p-2} + \cdots + X + 1.$$

Then $f$ is irreducible in $\mathbb{Q}[X]$.

**Proof.** Write $U = X - 1$ and let $g(U) = f(X)$. Now $(X-1)f(X) = X^p - 1$. So

$$Ug(U) = (U+1)^p - 1 = U^p + \sum_{i=0}^{p-1} \binom{p}{i} U^i = U^p + pU^{p-1} + \cdots + \binom{p}{2} U + p.$$

If we stare at it long enough, we have that $\binom{p}{i} \equiv 0 \pmod p$. Therefore, by Eisenstein's criterion, $g(U)$ is irreducible in $\mathbb{Q}[U]$. This implies that $f(X)$ is irreducible in $\mathbb{Q}[X]$. ■

Recall that $\binom{p}{i} = p!/(r!(p-i)!)$. The numerator is divisible by $p$ whereas the denominator is coprime to $p$.

We will now prove a basic result about polynomials over a field, and a result about

[46] *For otherwise, we can write $f = c(f) f_0$ for some $f_0 \in \mathbb{Z}[X]$ primitive. Since $c(f) \mid a_n$, then $p \nmid c(f)$ and so $f_0$ satisfies the theorem's hypothesis.*

groups, to deduce a number theoretic result.

**Lemma 4.8.** Let $K$ be a field and let $f \in K[X]$. Then there exist at most $\deg f$ roots of $f$.

**Proof.** Suppose $a_1, \dots, a_r \in K$ are distinct roots of $f$ and write $n = \deg f$. We want to show that $r \leqslant n$.

Fields are Euclidean domains, so for any $b \in K$, we can write $f(X) = (X-b)q + r$ for some $q, r \in K[X]$ where $\deg r < \deg(X-b)$ or $r = 0$. In both cases, we actually have $r \in K$. If we evaluate at $X = b$, then we see that $r = f(b)$. So we can rewrite $f(X) = (X-b)q + f(b)$. If evaluate at $b = a_i$, we have $f(X) = (X-a_i)q$ since $f(a_i) = 0$ by hypothesis. So $(X-a_i) \mid f$. Now the $a_i$ are distinct, so the $X - a_i$ are pairwise coprime. Since $K[X]$ is a UFD, then $f$ is divisible by $\prod_{i=1}^{r}(X-a_i)$. Therefore

$$r = \deg\left(\prod_{i=1}^{r}(X-a_i)\right) \leqslant \deg f = n,$$

as desired. ■

recall that if a polynomial $g$ divides another polynomial $f$, then necessarily $\deg g \leqslant \deg f$. This is where we got our inequality here.

**Lemma 4.9.** Let $K$ be a field and let $G$ be a **finite** subgroup of the multiplicative group $K^\times$. Then $G$ is cyclic.

The idea is to apply the structure theorem for modules over Euclidean domains with the domain being $\mathbb{Z}$.

**Proof.** $G$ is an abelian group so is a $\mathbb{Z}$-module. By the structure theorem,

$$G \cong \bigoplus_{i=1}^{r} \mathbb{Z}/(n_i),$$

where $n_1 \mid n_2 \mid \cdots \mid n_r$. Then $g^{n_r} = 1$ for all $g \in G$. In other words, every $g \in G$ is a zero of the polynomial $X^{n_r} - 1 \in K[X]$. By the preceding Lemma 4.8,

$$\#G \leqslant \deg(X^{n_r} - 1) = n_r.$$

But $\#G = \prod_i n_i$, so we conclude that $G \cong \mathbb{Z}/(n_r)$. ■

**Notation.** Write $\mathbb{F}_p = \mathbb{Z}/(p) = \mathbb{Z}/p\mathbb{Z}$ and denote $\mathbb{F}_p^\times = \mathbb{F}_p \backslash \{0\}$.

**Proposition 4.11.** Let $p \in \mathbb{Z}$ be an odd prime number. Then there exists $\sqrt{-1} \in \mathbb{F}_p$ if and only if $p \equiv 1 \pmod 4$.

**Proof.** The only elements $y \in \mathbb{F}_p$ with $y^2 = 1$ are $y = \pm 1$ since every such $y$ is a zero of $X^2 - 1$. So there exists $\sqrt{-1} \in \mathbb{F}_p$ if and only there exists an element of order 4 in $\mathbb{F}_p^\times$. But $\mathbb{F}_p^\times$ has order $p - 1$ and is cyclic (by the preceding Lemma 4.9), so there exists $\sqrt{-1} \in \mathbb{F}_p$ if and only if $4 \mid (p-1)$, which is true if and only if $p \equiv 1 \pmod 4$. ■

**Lemma 4.10.** The only units in $\mathbb{Z}[i]$ are $\pm 1$ and $\pm i$.

It is clear that these are units, so we just need to prove that these are the only units.

**Proof.** Suppose $z = a + ib \in \mathbb{Z}[i]$ with multiplicative inverse $z^{-1} = c + id \in \mathbb{Z}[i]$. By definition $zz^{-1} = 1$, and so applying absolute value and squaring we have

$$(a^2 + b^2)(c^2 + d^2) = 1.$$

But $a, b, c, d \in \mathbb{Z}$ and so their squares are positive integers. Therefore $a^2 + b^2 = 1$ and the only solutions to this equation are $(0, \pm 1)$ or $(\pm 1, 0)$. ■

**Theorem 4.4** (*Lagrange's theorem for two squares*)**.** Let $p \in \mathbb{Z}$ be an odd prime.

(1). If $p \equiv 3 \pmod 4$, then $p$ is prime in $\mathbb{Z}[i]$ and $p$ cannot be written as a sum of two squares.

(2). If $p \equiv 1 \pmod 4$, then $p = z\bar{z}$ in $\mathbb{Z}[i]$ where $z$ is prime.

(3). If $p \equiv 1 \pmod 4$, then $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.

**Proof.** We have a surjective homomorphism $\pi : \mathbb{Z}[X] \to \mathbb{Z}[i]$ given by the evaluation $f(X) \mapsto f(i)$. Now what is $\ker \pi$? We claim that $\ker \pi = (X^2 + 1)$. Let $f = X^2 + 1$. Then $f(i) = 0$ and so $f \in \ker \pi$. That is, $(f) = (X^2 + 1) \subseteq \ker \pi$. On the other hand, suppose $g \in \ker \pi$. Then by long division, there exists $q, r \in \mathbb{Z}[X]$ such that $g = fq + r$ and $\deg r < \deg f = 2$. So $r = aX + b$ for some $a, b \in \mathbb{Z}$. Since $g(i) = 0$, we have $r(i) = 0$. That is, $ai + b = 0$. 何! This is a contradiction as $i \notin \mathbb{Z}$.

By the first isomorphism theorem, we have $\mathbb{Z}[i] \cong \mathbb{Z}[X]/(X^2 + 1)$. And so

$$\mathbb{Z}[i]/(p) \cong \mathbb{Z}[X]/(X^2 + 1, p) \cong (\mathbb{Z}[X]/(p))/(X^2 + 1) \cong \mathbb{F}_p/(X^2 + 1).$$

We have proved that for odd primes $p \in \mathbb{Z}$, there exists $\sqrt{-1} \in \mathbb{F}_p$ if and only if $p \equiv 1 \pmod 4$. So $X^2 + 1$ factors in $\mathbb{F}_p[X]$ if and only if $p \equiv 1 \pmod 4$. So $p$ is prime in $\mathbb{Z}[i]$ if and only if $X^2 + 1$ is irreducible in $\mathbb{F}_p[X]$ if and only if $p \equiv 3 \pmod 4$. The proof of the theorem is now immediate.

(1). If $p \equiv 3 \pmod 4$. Then $p$ is prime in $\mathbb{Z}[i]$. Suppose $p = a^2 + b^2$. Then $p$ factorizes as $(a + ib)(a - ib)$. 何!

(2). If $p \equiv 1 \pmod 4$. Then $p$ is not prime in $\mathbb{Z}[i]$. Write $p = (a + ib)(c + id)$, which is a product of non-units. If we square $p$, we have

$$p^2 = |a + ib|^2 |c + id|^2 = (a^2 + b^2)(c^2 + d^2).$$

Now stare at the factorization. Since $a + ib$, $c + id$ are non-units, then $a^2 + b^2, c^2 + d^2 > 1$ by our preceding lemmas about units in $\mathbb{Z}[i]$. We have no choice but to take $p = a^2 + b^2 = c^2 + d^2$. ■

# 5 G-domain and Nullstellensatz (again)

We start with a lemma that sparks the idea of a G-domain.

**Lemma 5.1.** Let $R$ be an integral domain with field of fractions $K$. Then the following are equivalent

(1). $K$ is a finitely generated $R$-algebra.

(2). There exists $x \in R$ such that $K = R[x^{-1}]$.

**Proof.** Clearly, if $K = R[x^{-1}]$, then $K$ is a finitely generated $R$-algebra. Now suppose $K = R[a_1/b_1, \ldots, a_n/b_n]$. Then $K = R[1/x]$ with $x = b_1 \cdots b_n$. ■

## 5.1 One definition and many theorems

**Definition** (*G-domain*)**.** Let $R$ be an integral domain. We say that $R$ is a **G-domain**, if its field of fractions $\mathrm{Frac}(R)$ is a finitely generated $R$-algebra.

The G in G-domain is due to *Goldman.*

So our lemma above gives a first characterization of G-domains. It is a powerful lemma and we shall spam it until the end of this chapter.

**Example.** Every field is a G-domain.

It turns out, for our purpose, G-domains are something that we want to avoid. We want interesting rings (and fields are boring rings).

**Proposition 5.1.** Let $R$ be an integral domain with field of fractions $K$ and let $x \in R$ be nonzero. Then the following are equivalent:

(1). $K = R[x^{-1}]$.

(2). $x \in \mathfrak{p}$ for any nonzero prime ideal $\mathfrak{p} \lhd R$.

In fact, this says that every nonzero radical ideal in a G-domain $R$ contains $x$.

**Proof.** Suppose $\mathfrak{p} \lhd R$ is a nonzero prime ideal and let $b \in \mathfrak{p}$. Then $b^{-1} \in K = R[x^{-1}]$. So $b^{-1} = a_n x^{-n} + \cdots + a_0 = C/x^n$ for some $C \in R$. So $x^n = bc \in \mathfrak{p}$. But since $\mathfrak{p}$ is prime, $x \in \mathfrak{p}$. ■

**Lemma 5.2.** Let $R$ be a G-domain with field of fractions $K$ and let $R \subseteq T \subseteq K$ where $T$ is a ring. Then $T$ is also a G-domain.

**Proof.** This is obvious since $\mathrm{Frac}(T) = K$. So if $K = R[x^{-1}]$, then $K = T[x^{-1}]$. ■

**Proposition 5.2.** Suppose $R$ is a PID with infinitely many distinct primes elements. Then $R$ is **not** a G-domain.

Here, we will use the fact that PIDs are UFDs.

**Proof.** Suppose not. Then we can write $K = R[x^{-1}]$. By unique factorization, we can write $x = \prod_i p_i^{e_i}$ for some primes $p_i$ and $e_i \geqslant 0$ where $1 \leqslant i \leqslant n$. Then $(p_1), \ldots, (p_n)$ is a complete list of prime ideals in $R$ that contains $x$. But by Proposition 5.1, every nonzero prime ideal of $R$ contains $x$. 何! ■

This proposition tells us that G-domains are very rare. For example, $\mathbb{Z}$ is a PID with infinitely many primes. So $\mathbb{Z}$ is not a G-domain. On the contrary, it tells us that if $R$ is a G-domain, it is a PID with finitely many primes.

**Theorem 5.1.** Let $R$ be a integral domain and let $x \in R$. Then $R[x]$ is not a G-domain.

**Proof.** Suppose $R[x]$ is a G-domain. We have

$$R[x] \subseteq K[x] \subseteq K(X) = \mathrm{Frac}(K[x]) = \mathrm{Frac}(R[x]),$$

so by Lemma 5.2, $K[x]$ is a G-domain. But copying Euclid's argument[47], $K[x]$ has infinitely many primes. If $p_1, \ldots, p_n$ is a complete list, then consider the prime factorization of the product $p_1 \cdots p_n + 1$. ■

**Theorem 5.2.** Suppose $R \leqslant T$ are integral domains. Suppose $T = R[\alpha_1, \ldots, \alpha_k]$ where each $\alpha_i$ is algebraic over $R$. Then $R$ is a G-domain if and only if $T$ is a G-domain.

**Proof.** Let $K = \mathrm{Frac}(R)$ and let $L = \mathrm{Frac}(T)$. So $K \subseteq L$ and $L$ is finite over $K$.

$(\Rightarrow)$. Assume $R$ is a G-domain so we can write $K = R[x^{-1}]$ for some $x \in R$. Then since $T[x^{-1}] \subseteq L$, it is an integral domain and $T[x^{-1}] \supseteq K$. So $T[x^{-1}]$ is finite over $K$, so is a field. Thus, $T$ is a G-domain.

$(\Leftarrow)$. Assume $T$ is a G-domain so we can write $L = T[v^{-1}]$ for some $v \in T$. Then $v, \alpha_1, \ldots, \alpha_k$ are algebraic over $R$. So there exist equations

$$av^{-m} + \cdots = 0, \qquad b_i\alpha_i^{n_i} + \cdots = 0,$$

with coefficients in $R$. Now take $R_1 = R[a^{-1}, b_1^{-1}, \ldots, b_k^{-1}]$. So $R \subseteq R_1 \subseteq K$. Then $L = R[\alpha_1, \ldots, \alpha_k, v^{-1}]$. Therefore, $L = R_1[\alpha_1, \ldots, \alpha_k, v^{-1}]$. Now $\alpha_1, \ldots, \alpha_k, v^{-1}$ are integral over $R_1$. So $R_1$ is a field. But since we know $R \subseteq R_1 \subseteq K$, then in fact $R_1 = K$. So $K$ is finitely generated as an $R$-algebra. So $R$ is a G-domain. ■

$K$ is the smallest field that contains $R$.

**Corollary 5.1.** Let $R, T$ be an integral domain such that $T \subseteq R$. Let $x_1, \ldots, x_r \in T$ such that that $R[x_1, \ldots, x_r]$ is a G-domain. Then

(1). $x_i$ is algebraic over $R$ for all $1 \leqslant i \leqslant r$.

(2). $R$ is a G-domain.

Remark: The fact that $T$ is an integral domain containing $R$ implies that $R$ is an integral domain, but we emphasize that $R$ is an integral domain here to accomodate for all readers.

**Proof.** This is an immediate corollary from the preceding two theorems. ■

[47] *in proving infinitude of primes in* $\mathbb{Z}$

**Theorem 5.3.** Let $R$ be an integral domain. Then $R$ is a G-domain if and only if there exists $\mathfrak{m} \lhd R[X]$ maximal such that $\mathfrak{m} \cap R = 0$.

**Proof.** $(\Rightarrow)$. Since $R$ is a G-domain, we can write $K = R[x^{-1}]$. So then $K \cong R[x]/I$ for some ideal $I \lhd R[X]$. But $K$ is a field, so $I$ is maximal. Now consider the natural homomorphisms

$$R \to R[X] \to R[X]/I \cong K.$$

Then the composition homomorphism $\pi : R \to K$ is certainly injective. So its kernel is trivial. But what is its kernel? It is $R \cap I$.

$(\Leftarrow)$. Take $L = R[X]/\mathfrak{m}$. Then $L$ is a field and $R \cong R/(R \cap \mathfrak{m}) \subseteq L$. So $L$ is a finitely generated $R$-algebra. But $L$ is a field, so $L$ is a G-domain. By the preceding theorem, $R$ is a G-domain. ■

In particular, $L$ is generated by $X$ modulo $\mathfrak{m}$ as an $R$-algebra.

## 5.2 Application to number theory

The point of all this G-domain stuff is the following. The Nullstellensatz only tells us about maximal ideals in $K[X_1, \dots, X_n]$ where $K$ is a field. We can now say something about maximal ideals in $\mathbb{Z}[X_1, \dots, X_n]$.

**Theorem 5.4.** If $\mathfrak{m} \lhd \mathbb{Z}[X_1, \dots, X_n]$ is maximal, then $\mathbb{Z}[X_1, \dots, X_n]/\mathfrak{m}$ is a finite field.

This theorem tells us that every field that is finitely generated as a $\mathbb{Z}$-algebra is finite. Now since $\mathfrak{m}$ is maximal, it is clear that $\mathbb{Z}[X_1, \dots, X_n]/\mathfrak{m}$ is a field due to Theorem 1.8. We just need to prove that it is finite.

**Proof.** Let $\mathfrak{m} \lhd \mathbb{Z}[X_1, \dots, X_n]$ be maximal. Then there is an injective homomorphism

$$\mathbb{Z}/(\mathfrak{m} \cap \mathbb{Z}) \hookrightarrow \mathbb{Z}[X_1, \dots, X_n]/\mathfrak{m} \cong \mathbb{Z}/(\mathfrak{m} \cap \mathbb{Z})[x_1, \dots, x_n],$$

where $x_i \equiv X_i \bmod \mathfrak{m}$. Now $\mathbb{Z}[X_1, \dots, X_n]/\mathfrak{m}$ is a field, so $\mathbb{Z}/(\mathfrak{m} \cap \mathbb{Z})$ is a G-domain by a preceding corollary. Now $\mathbb{Z}$ is a PID with infinitely many primes, so is not a G-domain. Therefore $\mathfrak{m} \cap \mathbb{Z}$ is a nonzero prime ideal. So $\mathfrak{m} \cap \mathbb{Z} = (p)$ for some prime $p$. So in fact $\mathbb{Z}/(\mathfrak{m} \cap \mathbb{Z}) = \mathbb{Z}/(p) = \mathbb{F}_p$. Also, by the preceding corollary, each $x_i$ is algebraic over $\mathbb{F}_p$. Therefore, $\mathbb{Z}[X_1, \dots, X_n]/\mathfrak{m}$ is algebraic over $\mathbb{F}_p$ and so is finite. That is, it is a finite dimensional $\mathbb{F}_p$-vector space. ■

We can use the same technique to prove another version of the Nullstellensatz.

**Theorem 5.5.** Two statements.

(1). If $\mathfrak{m}$ is maximal in $\mathbb{Z}[X_1, \dots, X_n]$, then $\mathfrak{m} \cap \mathbb{Z}[X_1, \dots, X_i]$ is maximal in $\mathbb{Z}[X_1, \dots, X_i]$ for all $i \leqslant n$.

(2). There exists $p \in \mathbb{Z}$ prime and $f_i \in \mathbb{Z}[X_1, \dots, X_i]$ such that $\mathfrak{m} = (p, f_1, \dots, f_n)$.

**Proof.** Exercise. Non-trivial and examinable! ■

# 6 Symmetric functions

The symmetric group $S_n$ acts on the polynomial ring $A = \mathbb{Z}[X_1, \ldots, X_n]$ naturally by permuting the variables $X_1, \ldots, X_n$. If $\pi \in S_n$, define $\pi(X_i) = X_{\pi(i)}$ and $\pi(m) = m$ for all $m \in \mathbb{Z}$. In this sense, if $f \in A$, then we have $\pi(f) = f(X_{\pi(1)}, \ldots, X_{\pi(n)})$. The verification that this indeed defines an action is left as an exercise. For this chapter, we will always write $A = \mathbb{Z}[X_1, \ldots, X_n]$.

**Definition** (*Invariant*)**.** Let $f \in A$. We say $f$ is **invariant** or **$S_n$-invariant** if $\pi(f) = f$ for any $\pi \in S_n$. The set of invariants in $A$ is denoted $A^{S_n}$.

It is then easy to see that $A^{S_n} \subseteq A$ is in fact a ring. So we call this the **ring of invariants**.

**Example.** Sum $\sum_{i=1}^n X_i$ and product $\prod_{i=1}^n X_i$ of variables $X_i$ are invariants.

Motivated by the example above, we have an even more general construct.

**Definition** (*Elementary symmetric functions*)**.** The $k$-th **elementary symmetric function** $\sigma_k \in A$ is defined to be

$$\sigma_k = \sum_{i_1 < \cdots < i_k} X_{i_1} \cdots X_{i_k},$$

where $1 \leqslant k \leqslant n$.

Note that $n$ here comes from $A = \mathbb{Z}[X_1, \ldots, X_n]$. It then really makes sense that $1 \leqslant k \leqslant n$.

So our preceding example is really a special case of this as $\sigma_1 = \sum X_i$ and $\sigma_n = \prod X_i$.

**Lemma 6.1.** $\sigma_k \in A$ is invariant for all $k$.

**Proof.** Introduce a new variable $T$ and consider $p(X) = \prod_{i=1}^n (T - X_i) \in A[T]$ which has degree $n$ and its zeros are $X_1, \ldots, X_n$. Now $S_n$ also acts on $A[T]$ by

$$\pi\left(\sum_{i=1}^n a_i T^i\right) = \sum_{i=1}^n \pi(a_i) T^i.$$

That is, $\pi(T) = T$ for all $\pi \in S_n$. So $S_n$ permutes the terms $T - X_i$. Therefore $p$ is invariant and so the coefficients of $p$ are $S_n$-invariant. But if we expand $p$, we have

$$p = \sum_{i=1}^n (-1)^i \sigma_i T^{n-i} = T^n - \sigma_1 T^{n-1} + \cdots + (-1)^n \sigma_n.$$

So the coefficients of $p$ are, up to sign, the $\sigma_i$. So the $\sigma_i$ are $S_n$-invariant. ■

What this lemma really tells us is that $\mathbb{Z}[\sigma_1, \ldots, \sigma_n] \subseteq A^{S_n}$. Here's our first main theorem.

**Theorem 6.1** (*Newton*)**.** Let $A = \mathbb{Z}[X_1, \ldots, X_n]$. Then

(1). $A^{S_n} = \mathbb{Z}[\sigma_1, \ldots, \sigma_n]$.

(2). $\mathbb{Z}[\sigma_1, \ldots, \sigma_n]$ is a polynomial ring.

That is, there is no relation between $\sigma_1, \ldots, \sigma_n$.

To prove this theorem, we need to define the so-called **lexicographic order** $<_{\text{lex}}$ on the set of monomials $X_1^{e_1}\cdots X_n^{e_n}$ . We define this as follows:

$$\prod_{i=1}^{n} X_i^{p_i} <_{\text{lex}} \prod_{i=1}^{n} X_i^{q_i},$$

iff there exists $r \leqslant n$ such that $p_i = q_i$ for all $1 \leqslant i < r$ and $p_r < q_r$. Moreover, we write $M >_{\text{lex}} N$ if $N <_{\text{lex}} M$. Note that $<_{\text{lex}}$ is a total order: for any two monomials $M, N$ in $X_1, \dots, X_n$, then either $M <_{\text{lex}} N$ or $M = N$ or $M >_{\text{lex}} N$, and exactly only one is true. If the first and second case can occur simultaneously, we will write $M \leqslant_{\text{lex}} N$ (not very suprising huh). We are now ready to prove the theorem.

Lexicon is just a fancy name for a dictionary. The reason for the name is because it mimics the rule that governs the order of words in a dictionary of languages that are alphabetically ordered e.g. English, Malay, Arabic and not Japanese. In Japanese, words are ordered by how they sound (cf. Gojūon ordering).

**Proof.** **Proof of (1).** The preceding lemma already gives us one inclusion. We are left to show that $\mathbb{Z}[\sigma_1, \dots, \sigma_n] \supseteq A^{S_n}$.

Suppose $F \in A^{S_n}$ is a nonzero polynomial. We want to prove that $F$ is a polynomial in $\sigma_1, \dots, \sigma_n$. WLOG[48], we can assume that $F$ is homogeneous of degree $d$. Now $F$ contains only finitely many monomials with nonzero coefficients. Let $M = \min F$ be the smallest monomial that appears in $F$ with respect to the lexicographic order. Now observe that $\min(FG) = \min F \min G$. Suppose that $M$ has coefficient $\lambda \in \mathbb{Z}$ in $F$. Then for any $\pi \in S_n$, $\pi(M)$ has coefficient $\lambda$ in $F$. Therefore, $M \leqslant_{\text{lex}} \pi(M)$ for any $\pi \in S_n$. Now suppose $M = \prod_{i=1}^{n} X_i^{a_i}$ so that $d = \sum_i a_i$. Now consider the transposition $s = \begin{pmatrix} i & i+1 \end{pmatrix} \in S_n$. This acts on $M$ by exchanging the $i$-th monomial with the $(i+1)$-th monomial. That is,

$$s(M) = X_1^{a_1} \cdots X_{i-1}^{a_{i-1}} X_{i+1}^{a_i} X_i^{a_{i+1}} \cdots X_n^{a_n}.$$

But $M <_{\text{lex}} s(M)$. This implies that $a_i \geqslant a_{i+1}$, and therefore $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$. Put $a_{n+1} = 0$. Now consider

$$H = \prod_{i=1}^{n} \sigma_i^{a_i - a_{i+1}}.$$

Clearly, $H \in \mathbb{Z}[\sigma_1, \dots, \sigma_n]$. Also, $H$ is homogeneous since $\sigma_i$ is homogeneous of degree $i$. So

$$\deg H = \sum_{i=1}^{n} i(a_i - a_{i+1}) = \sum_{i=1}^{n} a_i i - \sum_{i=1}^{n} a_i(i-1) = \sum_{i=1}^{n} a_i = d.$$

Also, $\min H = \prod_{i=1}^{n} \min(\sigma_i)^{a_i - a_{i+1}}$. But $\min(\sigma_i) = X_1 \cdots X_i$ and so

$$\min H = \prod_{j=1}^{n} \left( \prod_{i=1}^{j} X_i \right)^{a_j - a_{j+1}} = \prod_{i=1}^{n} X_i^{a_i} = M.$$

So the coefficient of $M$ in $F - \lambda H$ is zero. Now put $F_1 = F - \lambda H$. Then $F_1$ is $S_n$-invariant, homogeneous of degree $d$, and $\min(F_1) >_{\text{lex}} \min F = M$. We can repeat this process to get other polynomials $F_i$. But since there are only finitely many monomials of degree $d$ in $X_1, \dots, X_n$, this process terminates.

**Proof of (2).** We want to prove that there is no relation between $\sigma_1, \dots, \sigma_n$. We prove by induction on $n$. The **base case** $n = 1$ is trivial. If $n = 1$, then $\sigma_1 = X_1$ so we are done. So assume that the result is true for $n - 1$ variables. We begin our **inductive step**. For $i = 1, \dots, n-1$, let $\tau_i = \sigma_i(X_1, \dots, X_{n-1})$. Really, we have

[48] *For otherwise, we can write*

$$F = \sum_{j=1}^{n} F_j,$$

*where $F_j$ is homogeneous of degree $j$. Then $F_j \in A^{S_n}$ for all $j$ because $S_n$ preserves degrees.*

$\tau_i = \sigma_i(X_1, \dots, X_{n-1}, 0)$. Then we have an equality on the ring of invariant

$$A[X_1, \dots, X_{n-1}]^{S_{n-1}} = \mathbb{Z}[\tau_1, \dots, \tau_{n-1}].$$

So suppose we have a relation $h(\sigma_1, \dots, \sigma_n) = 0$. Then set $X_n = 0$. We get

$$h(\tau_1, \dots, \tau_{n-1}, 0) = 0.$$

So $h(\sigma_1, \dots, \sigma_{n-1}, \sigma_n) = 0$ after setting $\sigma_n = 0$. Therefore $h(\sigma_1, \dots, \sigma_{n-1}, \sigma_n)$ is divisible by $\sigma_n$. Then divide by $\sigma_n$ to get a relation of lower degree. Continuing in this way, we have an infinite descent. 何! ■

The key point is that each $\tau_i$ is the $i$-th elementary symmetric function in the first $n-1$ variables; and there is **no relation** between the $\tau_i$.

# 7 Resultants and discriminants (again)

## 7.1 Two definitions and basic facts

The resultant is essentially the determinant of a very huge matrix called the Sylvester matrix. This is a matrix which simply contains the coefficients of two polynomials.

**Definition** (*Sylvester matrix*). Let $R$ be a ring; let $f = \sum_{i=0}^{n} a_i X^i$ and $g = \sum_{i=0}^{m} b_i X^i$ be nonzero polynomials over $R$. Then the **Sylvester matrix associated to $f$ and $g$** is the matrix

$$\mathfrak{S}(f,g) = \begin{bmatrix} a_0 & 0 & \cdots & 0 & b_0 & 0 & \cdots & 0 \\ a_1 & a_0 & \cdots & 0 & b_1 & b_0 & \cdots & 0 \\ a_2 & a_1 & \ddots & 0 & b_2 & b_1 & \ddots & 0 \\ \vdots & \vdots & \ddots & a_0 & \vdots & \vdots & \ddots & b_0 \\ a_n & a_{n-1} & \cdots & \vdots & b_m & b_{m-1} & \cdots & \vdots \\ 0 & a_n & \ddots & \vdots & 0 & b_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_{n-1} & \vdots & \vdots & \ddots & b_{m-1} \\ 0 & 0 & \cdots & a_n & 0 & 0 & \cdots & b_m \end{bmatrix}$$

which is an $(m+n) \times (m+n)$ matrix with coefficients in $R$.

The occurence of this matrix may be mysterious at first. The reason for taking its determinant is probably even more cryptic. But don't worry, we will see why it makes sense after two lemmas.

**Definition** (*Resultant*). Let $R$ be a ring; let $f = \sum_{i=0}^{n} a_i X^i$ and $g = \sum_{i=0}^{m} b_i X^i$ be nonzero polynomials over $R$. Then the **resultant** of $f$ and $g$ is defined to be

$$\operatorname{Res}(f,g) = \det \mathfrak{S}(f,g).$$

We will now prove the two lemmas that we mentioned.

**Lemma 7.1.** Let $R$ be a Noetherian UFD; and let $f, g \in R[X]$ be nonzero polynomials. Then

$$\deg \operatorname{hcf}(f,g) \geqslant 1 \iff \phi f - \gamma g = 0,$$

for some $\phi, \gamma \in R[X]$ such that $\deg \phi < \deg g$ and $\deg \gamma < \deg f$.

**Proof.** ($\Rightarrow$). Suppose $h = \operatorname{hcf}(f,g)$. Then we can write $f = f_1 h$ and $g = g_1 h$ for some polynomials $f_1, g_1 \in R[X]$. Since $\deg h \geqslant 1$ by hypothesis, then $\deg f_1 < \deg f$ and $\deg g_1 < \deg g$. Moreover, $g_1 f = g_1 f_1 h = f_1 g_1 h = f_1 g$. So take $\phi = g_1$ and $\gamma = f_1$.

($\Leftarrow$). For contradiction, assume that $f, g$ are coprime. Suppose we have a linear relation $\phi f - \gamma g = 0$ such that $\deg \phi < \deg g$ and $\deg \gamma < \deg f$. Then $f \mid \gamma g$. But $f, g$ are coprime and $R[X]$ is a UFD, so $f \mid \gamma$. 何! This contradicts the fact that $\deg \gamma < \deg f$. ■

**Lemma 7.2.** Let $R$ be a Noetherian UFD; and let $f, g \in R[X]$ be nonzero polynomials. Then

$$\deg \operatorname{hcf}(f,g) \geqslant 1 \iff \operatorname{Res}(f,g) = 0.$$

**Proof.** Suppose $\phi, \gamma \in R[X]$ defined by

$$\phi = c_0 X^{m-1} + \cdots + c_{m-1}, \qquad \gamma = d_0 X^{n-1} + \cdots + d_{n-1}.$$

Then $\phi f - \gamma g = 0$, or equivalently, $\phi g = \gamma g$ if and only if

$$\begin{gathered} c_0 a_0 = d_0 b_0 \\ c_0 a_1 + c_1 a_0 = d_0 b_1 + d_1 b_0 \\ \vdots \\ c_{m-1} a_n = d_{n-1} b_m, \end{gathered}$$

where we get these equality simply by comparing coefficients in $X^i$ for each $i$. We can rewrite this system of equations in matrix form

$$\mathfrak{S}(f,g) \cdot \begin{pmatrix} c_0 \\ \vdots \\ c_{m-1} \\ -d_0 \\ \vdots \\ -d_{n-1} \end{pmatrix} = \mathbf{0}.$$

So our system has a nonzero solution if and only if $\det \mathfrak{S}(f,g) = \operatorname{Res}(f,g) = 0$. Applying our preceding lemma, we get the claim. ■

**N.B.** nonzero not non-trivial solution.

As promised, this is where we even get the idea of defining the resultant. It tells us when two polynomials $f, g$ has an (at least) linear common factor.

## 7.2 Main results with discriminants

The following proposition tells us about the resultant of two polynomials when they have a specific form. This specific form is quite natural if you think about the zeros of the polynomials.

**Proposition 7.1.** Let $R$ be a Noetherian UFD and let $f, g \in R[X]$ be defined by

$$f(X) = a\prod_{i=1}^{n}(X - x_i), \quad g(X) = b\prod_{j=1}^{m}(X - y_j),$$

where the $x_i$ and $y_j$ are further indeterminates. Then

$$\operatorname{Res}(f,g) = a^m b^n \prod_{i,j}(x_i - y_j).$$

**Proof.** By the preceding lemmas, $\mathrm{Res}(f,g)=0$ if and only if $f,g$ have a common factor, if and only if, by unique factorization, there exists $i,j$ such that $x_i=y_j$. So $\mathrm{Res}(f,g)$ is divisible by $\prod_{i,j}(x_i-y_j)$. Also, it is clear that $\mathrm{Res}(f,g)$ is divisible by $a^m b^n$ simply from definition. So $\mathrm{Res}(f,g)$ is divisible by $a^m b^n\prod_{i,j}(x_i-y_j)=S$. Next, observe that simply of definition of $g(X)$, we have

$$\prod_{i=1}^{n} g(x_i)=b^n\prod_{i,j}(x_i-y_j).$$

Therefore, we have

$$S=a^m\prod_{i=1}^{n} g(x_i). \qquad (*)$$

Similarly, from the definition of $f(X)$, we have

$$f(y_j)=a\prod_{i=1}^{n}(y_j-x_i)=a(-1)^n\prod_{i=1}^{n}(x_i-y_j).$$

So therefore, we have

$$\prod_{j=1}^{m} f(y_j)=(-1)^{mn}a^m\prod_{i,j}(x_i-y_j). \qquad (\dagger)$$

Now $(*)$ implies that $S$ is homogeneous of degree $n$ as a function of the variables $b_0,\dots,b_m$. And $(\dagger)$ shows that $S$ is homogeneous of degree $m$ as a function of the variables $a_0,\dots,a_n$. Moreover, $\mathrm{Res}(f,g)$ has the same homogeneity by inspection. Therefore, $\mathrm{Res}(f,g)$ and $S$ have the same degree, and $S\mid\mathrm{Res}(f,g)$. Thus,

$$\mathrm{Res}(f,g)=\lambda S,$$

for some $\lambda\in R$. In both $S$ and $\mathrm{Res}(f,g)$, the coefficient of $a^m b^n$ is 1, and so $\lambda=1$. ■

Observe that if $f(X)=a_0X^n+\dots+a_n$ and $x_1,\dots,x_n$ is a complete set of its zeros, then we have an equality $f(X)=a_0\prod_{i=1}^{n}(X-x_i)$. This was the specific polynomial form we had in the preceding proposition.

**Definition** (*Polynomial discriminant*)**.** Let $f(X)=a_0X^n+\dots+a_n$ be a polynomial over a ring $R$ with zeros $x_1,\dots,x_n$. Then the **discriminant** of $f$ is defined to be

$$D(f)=a_0^{2n-2}\prod_{1\leqslant i<j\leqslant n}(x_j-x_i)^2.$$

So we can see that $D(f)=0$ if and only if $f$ has repeated zeros i.e. a zero of multiplicity greater than 1.

Define the **derivative** of a polynomial the way we define it in real analysis. So if

$$f(X)=\sum_{i=0}^{n} a_iX^i,$$

is a polynomial over some ring $R$, then its derivative is

$$f'(X)=\sum_{i=0}^{n} ia_iX^{i-1}.$$

By abuse of notation, we can think of it as $f'(X) = \frac{d}{dX} f(X)$.

**Lemma 7.3.** Let $R$ be a Noetherian UFD and let $f \in R[X]$ be the polynomial $f = a_0 X^n + \cdots + a_n$. Then

$$\operatorname{Res}(f, f') = a_0 (-1)^{\binom{n}{2}} D(f).$$

**Proof.** Suppose $f$ has zeros $x_1, \ldots, x_n \in R$. From $(*)$ in the proof of the preceding proposition, we know that we can write

$$\operatorname{Res}(f, f') = a_0^{n-1} \prod_{i=1}^{n} f'(x_i).$$

Now

$$\frac{f'}{f} = (\log f)' = \left( \log a_0 + \sum_{i=1}^{n} (X - x_i) \right)' = \sum_{i=1}^{n} \frac{1}{X - x_i},$$

since $f(X) = a_0 \prod_{i=1}^{n} (X - x_i)$. This implies that

$$f'(X) = \sum_{i=1}^{n} \frac{f(X)}{X - x_i} \implies f'(x_i) = a_0 \prod_{j \neq i} (x_j - x_i).$$

Therefore, we have

$$\begin{aligned} \operatorname{Res}(f, f') &= a_0^{n-1} a_0^n \prod_{i=1}^{n} \left( \prod_{j \neq i} (x_j - x_i) \right) \\ &= a_0^{2n-1} (-1)^{\binom{n}{2}} \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)^2 = a_0 (-1)^{\binom{n}{2}} D(f). \end{aligned}$$

■

Let's calculate some discriminants.

**Example.** Let $f = X^3 + pX + q$ with derivative $f' = 3X^2 + p$. Then

$$\operatorname{Res}(f, f') = (-1)^{\binom{3}{2}} D(f) = -D(f).$$

But $\operatorname{Res}(f, f') = 4p^3 + 27q^2$. So we have

$$D(f) = -4p^3 - 27q^2.$$

This discriminant should be familiar to you. It is the discriminant (in the usual high school sense) for cubic polynomials.

**Example.** Consider the trinomial $f = X^n + pX + q$ with derivative $f' = nX^{n-1} + p$. Then we can easily[49] calculate that

$$D(f) = (-1)^{\binom{n}{2}} \left[ (-1)^{n-1} (n-1)^{n-1} p^n + n^n q^{n-1} \right].$$

We have two things name discriminants, here is where they are related.

**Theorem 7.1.** Let $\alpha \in \mathbb{C}$ and let $K = \mathbb{Q}[\alpha]$. Let $f \in \mathbb{Z}[X]$ be monic and irreducible such that $f(\alpha) = 0$. Then

[49] *this is sarcasm btw.*

(1). $\mathbb{Z}[\alpha] \leqslant \mathcal{O}_K$,

(2). $\mathbb{Z}[\alpha] \cong \mathbb{Z}[X]/(f)$,

(3). $D_{\mathbb{Z}[\alpha]} = D(f)$.

Moreover, if $\mathbb{Z}[\alpha]$ has index $r$ in $\mathcal{O}_K$, then $D_{\mathbb{Z}[\alpha]} = r^2 D_{\mathcal{O}_K}$.

Statement (1) shouldn't be surprising. $\alpha \in \mathcal{O}_K$, so $\mathbb{Z}[\alpha] \leqslant \mathcal{O}_K$. Statement (2) is obvious by using the evaluation homomorphism. What's interesting and not obvious to prove before is (3).

**Proof.** Suppose $\deg f = n$ and suppose $x_1, \ldots, x_n$ are the conjugates of $\alpha$. Then $\left\{1, \alpha, \ldots, \alpha^{n-1}\right\}$ is a $\mathbb{Z}$-basis of $A$. This implies that

$$D_A = \det \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \\ \vdots & & \vdots \\ x_1^{n-1} & \cdots & x_n^{n-1} \end{pmatrix}^2 = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)^2 = D(f),$$

where the second equality is due to the Vandermonde's identity. ■

The hypothesis on $f$ in the theorem really just means that $f$ is the minimal polynomial of $\alpha$ (since we assumed it is irreducible). That was Nick's version. The fact that

$$\mathbb{Z}[\alpha] \leqslant \mathcal{O}_{\mathbb{Q}[\alpha]} \leqslant \mathbb{Q}[\alpha],$$

is quite trivial. Using notation that we build in this notes, here is our simpler version.

**Theorem 7.2.** Let $\alpha \in \mathbb{C}$ and let $m_\alpha$ be its minimal polynomial over $\mathbb{Z}$. Then

$$D_{\mathbb{Z}[\alpha]} = D(m_\alpha).$$