Chapter 2

Ring Theory

In the first section below, a ring will be defined as an abstract structure with a commutative addition, and a multiplication which may or may not be commutative. This distinction yields two quite different theories: the theory of respectively commutative or non-commutative rings. These notes are mainly concerned about commutative rings.

Non-commutative rings have been an object of systematic study only quite recently, during the 20th century. Commutative rings on the contrary have appeared though in a hidden way much before, and as many theories, it all goes back to Fermat's Last Theorem.

In 1847, the mathematician Lamé announced a solution of Fermat's Last Theorem, but Liouville noticed that the proof depended on a unique decomposition into primes, which he thought was unlikely to be true. Though Cauchy supported Lamé, Kummer was the one who finally published an example in 1844 to show that the uniqueness of prime decompositions failed. Two years later, he restored the uniqueness by introducing what he called "ideal complex numbers" (today, simply "ideals") and used it to prove Fermat's Last Theorem for all $n < 100$ except $n = 37$, 59, 67 and 74.

It is Dedekind who extracted the important properties of "ideal numbers", defined an "ideal" by its modern properties: namely that of being a subgroup which is closed under multiplication by any ring element. He further introduced prime ideals as a generalization of prime numbers. Note that today we still use the terminology "Dedekind rings" to describe rings which have in particular a good behavior with respect to factorization of prime ideals. In 1882, an important paper by Dedekind and Weber developed the theory of rings of polynomials. At this stage, both rings of polynomials and rings of numbers (rings appearing in the context of Fermat's Last Theorem, such as what we call now the Gaussian integers) were being studied. But it was separately, and no one made connection between these two topics. Dedekind also introduced the term "field" (Körper) for a commutative ring in which every non-zero element has a

multiplicative inverse but the word "ring" is due to Hilbert, who, motivated by studying invariant theory, studied ideals in polynomial rings proving his famous "Basis Theorem" in 1893.

It will take another 30 years and the work of Emmy Noether and Krull to see the development of axioms for rings. Emmy Noether, about 1921, is the one who made the important step of bringing the two theories of rings of polynomials and rings of numbers under a single theory of abstract commutative rings.

In contrast to commutative ring theory, which grew from number theory, non-commutative ring theory developed from an idea of Hamilton, who attempted to generalize the complex numbers as a two dimensional algebra over the reals to a three dimensional algebra. Hamilton, who introduced the idea of a vector space, found inspiration in 1843, when he understood that the generalization was not to three dimensions but to four dimensions and that the price to pay was to give up the commutativity of multiplication. The quaternion algebra, as Hamilton called it, launched non-commutative ring theory.

Other natural non-commutative objects that arise are matrices. They were introduced by Cayley in 1850, together with their laws of addition and multiplication and, in 1870, Pierce noted that the now familiar ring axioms held for square matrices.

An early contributor to the theory of non-commutative rings was the Scottish mathematician Wedderburn, who in 1905, proved "Wedderburn's Theorem", namely that every finite division ring is commutative and so is a field.

It is only around the 1930's that the theories of commutative and non-commutative rings came together and that their ideas began to influence each other.

2.1 Rings, ideals and homomorphisms

Definition 2.1. A ring R is an abelian group with a multiplication operation

$$(a, b) \mapsto ab$$

which is associative, and satisfies the distributive laws

$$a(b + c) = ab + ab, \ (a + b)c = ac + bc$$

with identity element 1.

There is a group structure with the addition operation, but not necessarily with the multiplication operation. Thus an element of a ring may or may not be invertible with respect to the multiplication operation. Here is the terminology used.

Definition 2.2. Let a, b be in a ring R. If $a \neq 0$ and $b \neq 0$ but $ab = 0$, then we say that a and b are zero divisors. If $ab = ba = 1$, we say that a is a unit or that a is invertible.

While the addition operation is commutative, it may or not be the case with the multiplication operation.

Definition 2.3. Let R be ring. If $ab = ba$ for any a, b in R, then R is said to be commutative.

Here are the definitions of two particular kinds of rings where the multiplication operation behaves well.

Definition 2.4. An integral domain is a commutative ring with no zero divisor. A division ring or skew field is a ring in which every non-zero element a has an inverse a^{-1}.

Let us give two more definitions and then we will discuss several examples.

Definition 2.5. The characteristic of a ring R, denoted by char$R$, is the smallest positive integer such that

$$n \cdot 1 = \underbrace{1 + 1 + \ldots + 1}_{n\text{times}} = 0.$$

We can also extract smaller rings from a given ring.

Definition 2.6. A subring of a ring R is a subset S of R that forms a ring under the operations of addition and multiplication defined in R.

Examples 2.1.

1. $\mathbb{Z}$ is an integral domain but not a field.
2. The integers modulo n form a ring, which is an integral domain if and only if n is prime.
3. The $n \times n$ matrices $\mathcal{M}_n(\mathbb{R})$ with coefficients in $\mathbb{R}$ are a ring, but not an integral domain if $n \geq 2$.
4. Let us construct the smallest and also most famous example of division ring. Take $1, i, j, k$ to be basis vectors for a 4-dimensional vector space over $\mathbb{R}$, and define multiplication by

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

   Then

   $$\mathbb{H} = \{a + bi + cj + dk,\ \ a, b, c, d \in \mathbb{R}\}$$

   forms a division ring, called the Hamilton's quaternions. So far, we have only seen the ring structure. Let us now discuss the fact that every non-zero element is invertible. Define the conjugate of an element $h = a + bi + cj + dk \in \mathbb{H}$ to be $\bar{h} = a - bi - cj - dk$ (yes, exactly the same way you did it for complex numbers). It is an easy computation (and a good exercise if you are not used to the non-commutative world) to check that

   $$q\bar{q} = a^2 + b^2 + c^2 + d^2.$$

Now take q^{-1} to be

$$q^{-1} = \frac{q}{q\bar{q}}.$$

Clearly $qq^{-1} = q^{-1}q = 1$ and the denominator cannot possibly be 0, but if $a = b = c = d = 0$.

5. If R is a ring, then the set $R[X]$ of polynomials with coefficients in R is a ring.

Similarly to what we did with groups, we now define a map from a ring to another which has the property of carrying one ring structure to the other.

Definition 2.7. Let R, S be two rings. A map $f : R \to S$ satisfying

1. $f(a + b) = f(a) + f(b)$ (this is thus a group homomorphism)
2. $f(ab) = f(a)f(b)$
3. $f(1_R) = 1_S$

for $a, b \in R$ is called ring homomorphism.

The notion of "ideal number" was introduced by the mathematician Kummer, as being some special "numbers" (well, nowadays we call them groups) having the property of unique factorization, even when considered over more general rings than $\mathbb{Z}$ (a bit of algebraic number theory would be good to make this more precise). Today only the name "ideal" is left, and here is what it gives in modern terminology:

Definition 2.8. Let $\mathcal{I}$ be a subset of a ring R. Then an additive subgroup of R having the property that

$$ra \in \mathcal{I} \text{ for } a \in \mathcal{I},\ r \in R$$

is called a left ideal of R. If instead we have

$$ar \in \mathcal{I} \text{ for } a \in \mathcal{I},\ r \in R$$

we say that we have a right ideal of R. If an ideal happens to be both a right and a left ideal, then we call it a two-sided ideal of R, or simply an ideal of R.

Of course, for any ring R, both R and $\{0\}$ are ideals. We thus introduce some terminology to precise whether we consider these two trivial ideals.

Definition 2.9. We say that an ideal $\mathcal{I}$ of R is proper if $\mathcal{I} \neq R$. We say that is it non-trivial if $\mathcal{I} \neq R$ and $\mathcal{I} \neq 0$.

If $f : R \to S$ is a ring homomorphism, we define the kernel of f in the most natural way:

$$\mathrm{Ker} f = \{r \in R,\ f(r) = 0\}.$$

Since a ring homomorphism is in particular a group homomorphism, we already know that f is injective if and only if $\mathrm{Ker} f = \{0\}$. It is easy to check that $\mathrm{Ker} f$ is a proper two-sided ideal:

- Ker$f$ is an additive subgroup of R.
- Take $a \in \text{Ker}f$ and $r \in R$. Then

$$f(ra) = f(r)f(a) = 0 \text{ and } f(ar) = f(a)f(r) = 0$$

  showing that ra and ar are in Ker$f$.
- Then Ker$f$ has to be proper (that is, Ker$f \neq R$), since $f(1) = 1$ by definition.

We can thus deduce the following (extremely useful) result.

Lemma 2.1. *Suppose $f : R \to S$ is a ring homomorphism and the only two-sided ideals of R are $\{0\}$ and R. Then f is injective.*

Proof. Since Ker$f$ is a two-sided ideal of R, then either Ker$f = \{0\}$ or Ker$f = R$. But Ker$f \neq R$ since $f(1) = 1$ by definition (in words, Ker$f$ is a proper ideal). □

At this point, it may be worth already noticing the analogy between on the one hand rings and their two-sided ideals, and on the other hand groups and their normal subgroups.

- Two-sided ideals are stable when the ring acts on them by multiplication, either on the right or on the left, and thus

$$rar^{-1} \in \mathcal{I},\ a \in \mathcal{I},\ r \in R,$$

  while normal subgroups are stable when the groups on them by conjugation

$$ghg^{-1} \in H,\ h \in H,\ g \in G\ (H \leq G).$$

- Groups with only trivial normal subgroups are called simple. We will not see it formally here, but rings with only trivial two-sided ideals as in the above lemma are called simple rings.
- The kernel of a group homomorphism is a normal subgroup, while the kernel of a ring homomorphism is an ideal.
- Normal subgroups allowed us to define quotient groups. We will see now that two-sided ideals will allow to define quotient rings.

2.2 Quotient rings

Let $\mathcal{I}$ be a proper two-sided ideal of R. Since $\mathcal{I}$ is an additive subgroup of R by definition, it makes sense to speak of cosets $r + \mathcal{I}$ of $\mathcal{I}$, $r \in R$. Furthermore, a ring has a structure of abelian group for addition, so $\mathcal{I}$ satisfies the definition of a normal subgroup. From group theory, we thus know that it makes sense to speak of the quotient group

$$R/\mathcal{I} = \{r + \mathcal{I},\ r \in R\},$$

group which is actually abelian (inherited from R being an abelian group for the addition).

We now endow $R/\mathcal{I}$ with a multiplication operation as follows. Define

$$(r+\mathcal{I})(s+\mathcal{I}) = rs+\mathcal{I}.$$

Let us make sure that this is well-defined, namely that it does not depend on the choice of the representative in each coset. Suppose that

$$r+\mathcal{I} = r'+\mathcal{I},\ s+\mathcal{I} = s'+\mathcal{I},$$

so that $a = r'-r \in \mathcal{I}$ and $b = s'-s \in \mathcal{I}$. Now

$$r's' = (a+r)(b+s) = ab+as+rb+rs \in rs+\mathcal{I}$$

since ab, as and rb belongs to $\mathcal{I}$ using that $a, b \in \mathcal{I}$ and the definition of ideal. This tells us $r's'$ is also in the coset $rs+\mathcal{I}$ and thus multiplication does not depend on the choice of representatives. Note though that this is true only because we assumed a two-sided ideal $\mathcal{I}$, otherwise we could not have concluded, since we had to deduce that both as and rb are in $\mathcal{I}$.

Definition 2.10. The set of cosets of the two-sided ideal $\mathcal{I}$ given by

$$R/\mathcal{I} = \{r+\mathcal{I},\ r \in R\}$$

is a ring with identity $1_R+\mathcal{I}$ and zero element $0_R+\mathcal{I}$ called a quotient ring.

Note that we need the assumption that $\mathcal{I}$ is a proper ideal of R to claim that $R/\mathcal{I}$ contains both an identity and a zero element (if $R = \mathcal{I}$, then $R/\mathcal{I}$ has only one element).

Example 2.2. Consider the ring of matrices $\mathcal{M}_2(\mathbb{F}_2[i])$, where $\mathbb{F}_2$ denotes the integers modulo 2, and i is such that $i^2 = -1 \equiv 1 \mod 2$. This is thus the ring of 2×2 matrices with coefficients in

$$\mathbb{F}_2[i] = \{a+ib,\ a,b \in \{0,1\}\}.$$

Let $\mathcal{I}$ be the subset of matrices with coefficients taking values 0 and $1+i$ only. It is a two-sided ideal of $\mathcal{M}_2(\mathbb{F}_2[i])$. Indeed, take a matrix $U \in \mathcal{I}$, a matrix $M \in \mathcal{M}_2(\mathbb{F}_2[i])$, and compute UM and MU. An immediate computation shows that all coefficients are of the form $a(1+i)$ with $a \in \mathbb{F}_2[i]$, that is all coefficients are in $\{0, 1+i\}$. Clearly $\mathcal{I}$ is an additive group.

We then have a quotient ring

$$\mathcal{M}_2(\mathbb{F}_2[i])/\mathcal{I}.$$

We have seen that $\mathrm{Ker} f$ is a proper ideal when f is a ring homomorphism. We now prove the converse.

Proposition 2.2. *Every proper ideal $\mathcal{I}$ is the kernel of a ring homomorphism.*

Proof. Consider the canonical projection π that we know from group theory. Namely

$$\pi : R \to R/\mathcal{I},\ r \mapsto \pi(r) = r + \mathcal{I}.$$

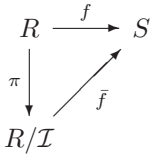
We already know that π is group homomorphism, and that its kernel is $\mathcal{I}$. We are only left to prove that π is a ring homomorphism:

- $\pi(rs) = rs + \mathcal{I} = (r + \mathcal{I})(s + \mathcal{I}) = \pi(r)\pi(s)$.
- $\pi(1_R) = 1_R + \mathcal{I}$ which is indeed the identity element of $R/\mathcal{I}$.

□

We are now ready to state a factor theorem and a 1st isomorphism theorem for rings, the same way we did for groups. It may help to keep in mind the analogy between two-sided ideals and normal subgroups mentioned above.

Assume that we have a ring R which contains a proper ideal $\mathcal{I}$, another ring S, and $f : R \to S$ a ring homomorphism. Let π be the canonical projection from R to the quotient group $R/\mathcal{I}$:



We would like to find a ring homomorphism $\bar{f} : R/\mathcal{I} \to S$ that makes the diagram commute, namely

$$f(a) = \bar{f}(\pi(a))$$

for all $a \in R$.

Theorem 2.3. (Factor Theorem for Rings). *Any ring homomorphism f whose kernel K contains $\mathcal{I}$ can be factored through $R/\mathcal{I}$. In other words, there is a unique ring homomorphism $\bar{f} : R/\mathcal{I} \to S$ such that $\bar{f} \circ \pi = f$. Furthermore*

1. *$\bar{f}$ is an epimorphism if and only if f is.*
2. *$\bar{f}$ is a monomorphism if and only if $K = \mathcal{I}$.*
3. *$\bar{f}$ is an isomorphism if and only if f is an epimorphism and $K = \mathcal{I}$.*

Proof. Since we have already done the proof for groups with many details, here we will just mention a few important points in the proof.

Let $a + \mathcal{I} \in R/\mathcal{I}$ such that $\pi(a) = a + \mathcal{I}$ for $a \in R$. We define

$$\bar{f}(a + \mathcal{I}) = f(a).$$

This is the most natural way to do it, however, we need to make sure that this is indeed well-defined, in the sense that it should not depend on the choice of the representative taken in the coset. Let us thus take another representative,

say $b \in a + \mathcal{I}$. Since a and b are in the same coset, they satisfy $a - b \in \mathcal{I} \subset K$, where $K = \mathrm{Ker}(f)$ by assumption. Since $a - b \in K$, we have $f(a - b) = 0$ and thus $f(a) = f(b)$.

Now that $\bar{f}$ is well defined, it is an easy computation to check that $\bar{f}$ inherits the property of ring homomorphism from f.

The rest of the proof works exactly the same as for groups. □

The first isomorphism theorem for rings is similar to the one for groups.

Theorem 2.4. (1st Isomorphism Theorem for Rings). *If $f : R \to S$ is a ring homomorphism with kernel K, then the image of f is isomorphic to R/K:*

$$\mathrm{Im}(f) \simeq R/\mathrm{Ker}(f).$$

Proof. We know from the Factor Theorem that

$$\bar{f} : R/\mathrm{Ker}(f) \to S$$

is an isomorphism if and only if f is an epimorphism, and clearly f is an epimorphism on its image, which concludes the proof. □

Example 2.3. Let us finish Example 2.2. We showed there that $\mathcal{M}_2(\mathbb{F}_2[i])/\mathcal{I}$ is a quotient ring, where $\mathcal{I}$ is the ideal formed of matrices with coefficients in $\{0, 1 + i\}$. Consider the ring homomorphism:

$$f : \mathcal{M}_2(\mathbb{F}_2[i]) \to \mathcal{M}_2(\mathbb{F}_2), M = (m_{k,l}) \mapsto f(M) = (m_{k,l} \mod 1 + i)$$

that is f looks at the coefficients of $M \mod 1 + i$. Its kernel is $\mathcal{I}$ and it is surjective. By the first isomorphism for rings, we have

$$\mathcal{M}_2(\mathbb{F}_2[i])/\mathcal{I} \simeq \mathcal{M}_2(\mathbb{F}_2).$$

2.3 The Chinese Remainder Theorem

We will prove a "general" Chinese Remainder Theorem, rephrased in terms of rings and ideals.

For that let us start by introducing some new definitions about ideals, that will collect some of the manipulations one can do on ideals. Let us start with the sum.

Definition 2.11. Let $\mathcal{I}$ and $\mathcal{J}$ be two ideals of a ring R. The sum of $\mathcal{I}$ and $\mathcal{J}$ is the ideal

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

If $\mathcal{I}$ and $\mathcal{J}$ are right (resp. left) ideals, so is their sum.

Note that the intersection $\mathcal{I} \cap \mathcal{J}$ of two (resp. right, left, two-sided) ideals of R is again a (resp. right, left, two-sided) ideal of R. We can define a notion of being co-prime for ideals as follows.

Definition 2.12. The ideals $\mathcal{I}$ and $\mathcal{J}$ of R a commutative ring are relatively prime if

$$\mathcal{I} + \mathcal{J} = R.$$

Finally, let us extend the notion of "modulo" to ideals.

Definition 2.13. If $a, b \in R$ and $\mathcal{I}$ is an ideal of R, we say that a is congruent to b modulo $\mathcal{I}$ if

$$a - b \in \mathcal{I}.$$

A last definition this time about rings is needed before we can state the theorem.

Definition 2.14. If $R_1, \ldots, R_n$ are rings, the direct product of the R_i is defined as the ring of n-tuples $(a_1, \ldots, a_n)$, $a_i \in R_i$, with componentwise addition and multiplication. The zero element is $(0, \ldots, 0)$ and the identity is $(1, \ldots, 1)$ where 1 means 1_{R_i} for each i.

This definition is an immediate generalization of the direct product we studied for groups.

Theorem 2.5. (Chinese Remainder Theorem). *Let R be a commutative ring, and let $\mathcal{I}_1, \ldots, \mathcal{I}_n$ be ideals in R, such that*

$$\mathcal{I}_i + \mathcal{I}_j = R, \ i \neq j.$$

1. *If $a_1, \ldots, a_n$ are elements of R, there exists an element $a \in R$ such that*

$$a \equiv a_i \mod \mathcal{I}_i, \ i = 1, \ldots, n.$$

2. *If b is another element of R such that $b \equiv a_i \mod \mathcal{I}_i$, $i = 1, \ldots, n$, then*

$$b \equiv a \mod \cap_{i=1}^{n} \mathcal{I}_i.$$

Conversely, if b satisfies the above congruence, then $b \equiv a_i \mod \mathcal{I}_i$, $i = 1, \ldots, n$.

3. *We have that*

$$R/\cap_{i=1}^{n} \mathcal{I}_i \simeq \prod_{i=1}^{n} R/\mathcal{I}_i.$$

Proof. 1. For $j > 1$, we have by assumption that $\mathcal{I}_1 + \mathcal{I}_j = R$, and thus there exist $b_j \in \mathcal{I}_1$ and $d_j \in \mathcal{I}_j$ such that

$$b_j + d_j = 1, \ j = 2, \ldots, n.$$

This yields that

$$\prod_{j=2}^{n} (b_j + d_j) = 1. \tag{2.1}$$

Now if we look at the left hand side of the above equation, we have

$$(b_2+d_2)(b_3+d_3)\cdots(b_n+d_n) = (\underbrace{b_2b_3+b_2d_3+d_2b_3}_{\in\mathcal{I}_1}+d_2d_3)\cdots(b_n+d_n)$$

and all the terms actually belong to $\mathcal{I}_1$, but $c_1 := \prod_{j=2}^n d_j \in \prod_{j=2}^n \mathcal{I}_j$. Thus

$$c_1 \equiv 1 \mod \mathcal{I}_1$$

from (2.1). On the other hand, we also have

$$c_1 \equiv 0 \mod \mathcal{I}_j$$

for $j > 1$ since $c_1 \in \prod_{j=2}^n \mathcal{I}_j$.

More generally, for all i, we can find c_i with

$$c_i \equiv 1 \mod \mathcal{I}_i,\ c_i \equiv 0 \mod \mathcal{I}_j,\ j \neq i.$$

Now take arbitrary elements $a_1, \ldots, a_n \in R$, and set

$$a = a_1c_1 + \ldots + a_nc_n.$$

Let us check that a is the solution we are looking for. Rewrite

$$a - a_i = a - a_ic_i + a_ic_i - a_i = (a - a_ic_i) + a_i(c_i - 1).$$

If we look modulo $\mathcal{I}_i$, we get

$$a - a_i \equiv a - a_ic_i \equiv a_1 + \ldots + a_{i-1}c_{i-1} + a_{i+1}c_{i+1} + \ldots + a_nc_n \equiv 0 \mod \mathcal{I}_i$$

where the first congruence follows from $c_i - 1 \equiv 0 \mod \mathcal{I}_i$ and the third congruence comes from $c_j \equiv 0 \mod \mathcal{I}_j$,$j \neq i$.

2. We have just shown the existence of a solution a modulo $\mathcal{I}_i$ for $i = 1, \ldots, n$. We now discuss the question of unicity, and show that the solution is actually not unique, but any other solution than a is actually congruent to $a \mod \cap_{i=1}^n \mathcal{I}_i$.

   We have for all $i = 1, \ldots, n$ that

   $$b \equiv a_i \mod \mathcal{I}_i \iff b \equiv a \mod \mathcal{I}_i \iff b - a \equiv 0 \mod \mathcal{I}_i$$

   which finally is equivalent to

   $$b - a \in \cap_{i=1}^n \mathcal{I}_i.$$

3. Define the ring homomorphism $f : R \to \prod_{i=1}^n R/\mathcal{I}_i$, sending

   $$a \mapsto f(a) = (a + \mathcal{I}_1, \ldots, a + \mathcal{I}_n).$$

- This map is surjective: for any $(a_1 + \mathcal{I}_1, \ldots, a_n + \mathcal{I}_n) \in \prod_{i=1}^n R/\mathcal{I}_i$, there exists an $a \in R$ such that $f(a) = (a_1 + \mathcal{I}_1, \ldots, a_n + \mathcal{I}_n)$, that is $a_i \equiv a \mod \mathcal{I}_i$, by the first point.
- Its kernel is given by

$$\begin{aligned} \mathrm{Ker} f &= \{a \in R,\ f(a) = (\mathcal{I}_1, \ldots, \mathcal{I}_n)\} \\ &= \{a \in R,\ a \in \mathcal{I}_i,\ i = 1, \ldots, n\} \\ &= \prod_{i=1}^n \mathcal{I}_i. \end{aligned}$$

We conclude using the first isomorphism Theorem for rings.

□

The Chinese remainder Theorem does not hold in the non-commutative case. Consider the ring R of non-commutative real polynomials in X and Y. Denote by $\mathcal{I}$ the principal two-sided ideal generated by X and $\mathcal{J}$ the principal two-sided ideal generated by $XY + 1$. Then $\mathcal{I} + \mathcal{J} = R$ but $\mathcal{I} \cap \mathcal{J} \neq \mathcal{I}\mathcal{J}$.

2.4 Maximal and prime ideals

Here are a few special ideals.

Definition 2.15. The ideal generated by the non-empty set X of R is the smallest ideal of R that contains X. It is denoted by $\langle X \rangle$. It is the collection of all finite sums of the form $\sum_i r_i x_i s_i$.

Definition 2.16. An ideal generated by a single element a is called a principal ideal, denoted by $\langle a \rangle$.

Definition 2.17. A maximal ideal in the ring R is a proper ideal that is not contained in any strictly larger proper ideal.

One can prove that every proper ideal is contained in a maximal ideal, and that consequently every ring has at least one maximal ideal. We skip the proof here, since it heavily relies on set theory, requires many new definitions and the use of Zorn's lemma.

Instead, let us mention that a correspondence Theorem exists for rings, the same way it exists for groups, since we will need it for characterizing maximal ideals.

Theorem 2.6. (Correspondence Theorem for rings). *If $\mathcal{I}$ is an ideal of a ring R, then the canonical map*

$$\pi : R \to R/\mathcal{I}$$

sets up a one-to-one correspondence between

- *the set of all subrings of R containing $\mathcal{I}$ and the set of all subrings of $R/\mathcal{I}$,*
- *the set of all ideals of R containing $\mathcal{I}$ and the set of all ideals of $R/\mathcal{I}$.*

Here is a characterization of maximal ideals in commutative rings.

Theorem 2.7. *Let M be an ideal in the commutative ring R. We have*

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

Proof. Let us start by assuming that M is maximal. Since R/M is a ring, we need to find the multiplicative inverse of $a+M \in R/M$ assuming that $a+M \neq 0$ in R/M, that is for $a \notin M$. Since M is maximal, the ideal $Ra+M$ has to be R itself, since $M \subset Ra+M$. Thus $1 \in Ra+M = R$, that is

$$1 = ra + m,\ r \in R,\ m \in M.$$

Then

$$(r+M)(a+M) = ra + M = (1-m) + M = 1 + M$$

proving that $r+M$ is $(a+M)^{-1}$.

Conversely, let us assume that R/M is a field. First we notice that M must be a proper ideal of R, since if $M = R$, then R/M contains only one element and $1 = 0$.

Let N be an ideal of R such that $M \subset N \subset R$ and $N \neq R$. We have to prove that $M = N$ to conclude that M is maximal.

By the correspondence Theorem for rings, we have a one-to-one correspondence between the set of ideals of R containing M, and the set of ideals of R/M. Since N is such an ideal, its image $\pi(N) \in R/M$ must be an ideal of R/M, and thus must be either $\{0\}$ or R/M. The latter yields that $N = R$, which is a contradiction, letting as only possibility that $\pi(N) = \{0\}$, and thus $N = M$, which completes the proof. □

Definition 2.18. A prime ideal in a commutative ring R is a proper ideal P of R such that for any $a, b \in R$, we have that

$$ab \in P \Rightarrow a \in P \text{ or } b \in P.$$

Here is again a characterization of a prime ideal P of R in terms of its quotient ring R/P.

Theorem 2.8. *If P is an ideal in the commutative ring R*

$$P \text{ is a prime ideal} \iff R/P \text{ is an integral domain.}$$

Proof. Let us start by assuming that P is prime. It is thus proper by definition, and R/P is a ring. We must show that the definition of integral domain holds, namely that

$$(a+P)(b+P) = 0 + P \Rightarrow a + P = P \text{ or } b + P = P.$$

Since

$$(a+P)(b+P) = ab + P = 0 + P,$$

we must have $ab \in P$, and thus since P is prime, either $a \in P$ or $b \in P$, implying respectively that either $a + P = P$ or $b + P = P$.

Conversely, if R/P is an integral domain, then P must be proper (otherwise $1 = 0$). We now need to check the definition of a prime ideal. Let us thus consider $ab \in P$, implying that

$$(a+P)(b+P) = ab + P = 0 + P.$$

Since R/P is an integral domain, either $a + P = P$ or $b + P = P$, that is

$$a \in P \text{ or } b \in P,$$

which concludes the proof. □

Corollary 2.9. *In a commutative ring, a maximal ideal is prime.*

Proof. If M is maximal, then R/M is a field, and thus an integral domain, so that M is prime. □

Corollary 2.10. *Let $f : R \to S$ be an epimorphism of commutative rings.*

1. *If S is a field, then* $\mathrm{Ker}f$ *is a maximal ideal of R.*
2. *If S is an integral domain, then* $\mathrm{Ker}f$ *is a prime ideal of R.*

Proof. By the first isomorphism theorem for rings, we have that

$$S \simeq R/\mathrm{Ker}f.$$

□

Example 2.4. Consider the ring $\mathbb{Z}[X]$ of polynomials with coefficients in $\mathbb{Z}$, and the ideal generated by the indeterminate X, that is $\langle X \rangle$ is the set of polynomials with constant coefficient 0. Clearly $\langle X \rangle$ is a proper ideal. To show that it is prime, consider the following ring homomorphism:

$$\varphi : \mathbb{Z}[X] \to \mathbb{Z},\ f(X) \mapsto \varphi(f(X)) = f(0).$$

We have that $\langle X \rangle = \mathrm{Ker}\varphi$ which is prime by the above corollary.

2.5 Polynomial rings

For this section, we assume that R is a commutative ring. Set $R[X]$ to be the set of polynomials in the indeterminate X with coefficients in R. It is easy to see that $R[X]$ inherits the properties of ring from R.

We define the evaluation map E_x, which evaluates a polynomial $f(X) \in R[X]$ in $x \in R$, as

$$E_x : R[X] \to R,\ f(X) \mapsto f(X)|_{X=x} = f(x).$$

We can check that E_x is a ring homomorphism.

The degree of a polynomial is defined as usual, that is, if $p(X) = a_0 + a_1 X + \ldots + a_n X^n$ with $a_n \neq 0$, then $\deg(p(X)) = \deg p = n$. By convention, we set $\deg(0) = -\infty$.

Euclidean division will play an important role in what will follow. Let us start by noticing that there exists a polynomial division algorithm over $R[X]$, namely: if $f, g \in R[X]$, with g monic, then there exist unique polynomials q and r in $R[X]$ such that

$$f = qg + r,\ \deg r < \deg g.$$

The requirement that g is monic comes from R being a ring and not necessarily a field. If R is a field, g does not have to be monic, since one can always multiply g by the inverse of the leading coefficient, which is not possible if R is not a field.

This gives the following:

Theorem 2.11. (Remainder Theorem). *If $f \in R[X]$, $a \in R$, then there exists a unique polynomial $q(X) \in R[X]$ such that*

$$f(X) = q(X)(X - a) + f(a).$$

Hence $f(a) = 0 \iff X - a \mid f(X)$.

Proof. Since $(X - a)$ is monic, we can do the division

$$f(X) = q(X)(X - a) + r(X).$$

But now since $\deg r < \deg(X - a)$, $r(X)$ must be a constant polynomial, which implies that

$$f(a) = r(X)$$

and thus

$$f(X) = q(X)(X - a) + f(a)$$

as claimed. Furthermore, we clearly have that

$$f(a) = 0 \iff X - a \mid f(X).$$

□

The following result sounds well known, care should be taken not to generalize it to rings which are not integral domain!

Theorem 2.12. *If R is an integral domain, then a non-zero polynomial f in $R[X]$ of degree n has at most n roots in R, counting multiplicity.*

Proof. Take a_1 a root of f, that is $f(a_1) = 0$. Then

$$X - a_1 \mid f(X)$$

by the remainder Theorem above, meaning that

$$f(X) = q_1(X)(X - a_1)^{n_1}$$

where $q_1(a_1) \neq 0$ and $\deg q_1 = n - n_1$. Similarly, consider $a_2 \neq a_1$ another root of f, so that

$$0 = f(a_2) = q_1(a_2)(a_2 - a_1)^{n_1}.$$

Since R is an integral domain, we must have that $q_1(a_2) = 0$, and thus a_2 is a root of $q_1(X)$. We can repeat the process with $q_1(X)$ instead of $f(X)$: since a_2 is a root of $q_1(X)$, we have

$$q_1(X) = q_2(X)(X - a_2)^{n_2}$$

with $q_2(a_2) \neq 0$ and $\deg q_2 = n - n_1 - n_2$. By going on iterating the process, we obtain

$$\begin{aligned} f(X) &= q_1(X)(X - a_1)^{n_1} \\ &= q_2(X)(X - a_2)^{n_2}(X - a_1)^{n_1} \\ &= \dots \\ &= (X - a_1)^{n_1}(X - a_2)^{n_2} \cdots (X - a_k)^{n_k} \cdot c \end{aligned}$$

where c is some constant and

$$n = n_1 + n_2 + \cdots + n_k.$$

Since R is an integral domain, the only possible roots of f are $a_1, \dots, a_k$, $k \leq n$. □

Example 2.5. Take $R = \mathbb{Z}_8$ the ring of integers modulo 8. Consider the polynomial

$$f(X) = X^3.$$

It is easy to check that is has 4 roots: $0, 2, 4, 6$. This comes from the fact that $\mathbb{Z}_8$ is not an integral domain.

2.6 Unique factorization and Euclidean division

In this section, all rings are assumed to be integral domains.

Let us start by defining formally the notions of irreducible and prime. The elements a, b, c, u in the definitions below all belong to an integral domain R.

Definition 2.19. The elements a, b are called associate if $a = ub$ for some unit u.

Definition 2.20. Let a be a non-zero element which is not a unit. Then a is said to be irreducible if $a = bc$ implies that either b or c must be a unit.

Definition 2.21. Let a be a non-zero element which is not a unit. Then a is called prime is whenever $a \mid bc$, then $a \mid b$ or $a \mid c$.

Between prime and irreducible, which notion is the stronger? The answer is in the proposition below.

Proposition 2.13. *If a is prime, then a is irreducible.*

Proof. Suppose that a is prime, and that $a = bc$. We want to prove that either b or c is a unit. By definition of prime, we must have that a divides either b or c. Let us say that a divides b. Thus

$$b = ad \Rightarrow b = bcd \Rightarrow b(1 - cd) = 0 \Rightarrow cd = 1$$

using that R is an integral domain, and thus c is a unit. The same argument works if we assume that a divides c, and we conclude that a is irreducible. □

Example 2.6. Consider the ring

$$R = \mathbb{Z}[\sqrt{-3}] = \{a + ib\sqrt{3},\ a, b \in \mathbb{Z}\}.$$

We want to see that 2 is irreducible but not prime.

- Let us first check that 2 is indeed irreducible. Suppose that

  $$2 = (a + ib\sqrt{3})(c + id\sqrt{3}).$$

  Since 2 is real, it is equal to its conjugate, and thus

  $$2\bar{2} = (a + ib\sqrt{3})(c + id\sqrt{3})(a - ib\sqrt{3})(c - id\sqrt{3})$$

  implies that

  $$4 = (a^2 + 3b^2)(c^2 + 3d^2).$$

  We deduce that $a^2 + 3b^2$ must divide 4, and it cannot possibly be 2, since we have a sum of squares in $\mathbb{Z}$. If $a^2 + 3b^2 = 4$, then $c^2 + 3d^2 = 1$ and $d = 0$, $c = \pm 1$. Vice versa if $c^2 + 3d^2 = 4$ then $a^2 + 3b^2 = 1$, and $b = 0$, $a = \pm 1$. In both cases we get that one of the factors of 2 is unit, namely ± 1.

- We now have to see that 2 is not a prime. Clearly

  $$2 \mid (1 + i\sqrt{3})(1 - i\sqrt{3}) = 4.$$

  But 2 divides neither $1 + i\sqrt{3}$ nor $1 - i\sqrt{3}$.

We can see from the above example that the problem which arises is the lack of unique factorization.

Definition 2.22. A unique factorization domain (UFD) is an integral domain R satisfying that

1. every element $0 \neq a \in R$ can be written as a product of irreducible factors $p_1, \dots p_n$ up to a unit u, namely:
$$a = up_1 \dots p_n.$$
2. The above factorization is unique, that is, if
$$a = up_1 \dots p_n = vq_1 \dots q_m$$
are two factorizations into irreducible factors p_i and q_j with units u, v, then $n = m$ and p_i and q_i are associate for all i.

We now prove that the distinction between irreducible and prime disappear in a unique factorization domain.

Proposition 2.14. *In a unique factorization domain R, we have that a is irreducible if and only if a is prime.*

Proof. We already know that prime implies irreducible. Let us show that now, we also have irreducible implies prime.

Take a to be irreducible and assume that $a \mid bc$. This means that $bc = ad$ for some $d \in R$. Using the property of unique factorization, we decompose d, b and c into products of irreducible terms (resp. d_i, b_i, c_i up to units u, v, w):
$$a \cdot ud_1 \cdots d_r = vb_1 \cdots b_s \cdot wc_1 \dots c_t.$$
Since the factorization is unique, a must be associate to some either b_i or c_i, implying that a divides b or c, which concludes the proof. □

We now want to connect the property of unique factorization to ideals.

Definition 2.23. Let $a_1, a_2, \dots$ be elements of an integral domain R. If the sequence of principal ideals
$$(a_1) \subseteq (a_2) \subseteq (a_3) \subseteq \dots$$
stabilizes, i.e., we have
$$(a_n) = (a_{n+1}) = \dots$$
for some n, then we say that R satisfies the ascending chain condition on principal ideals.

If the same condition holds but for general ideals, not necessarily principal, we call R a Noetherian ring, in honor of the mathematician Emmy Noether.

Theorem 2.15. *Let R be an integral domain.*

1. *If R is a UFD, then R satisfies the ascending chain condition on principal ideals.*

Figure 2.1: Amalie Emmy Noether (1882-1935)

2. *If R satisfies the ascending chain condition on principal ideals, then every non-zero element of R can be factored into irreducible (this says nothing about the unicity of the factorization).*

3. *If R is such that every non-zero element of R can be factored into irreducible, and in addition every irreducible element is prime, then R is a UFD.*

Thus R is a UFD if and only if it satisfies the ascending chain condition on principal ideals and every irreducible element of R is prime.

Proof. 1. Recall that in a UFD, prime and irreducible are equivalent. Consider an ascending chain of principal ideals

$$(a_1) \subseteq (a_2) \subseteq (a_3) \subseteq \dots$$

We have that $a_{i+1} \mid a_i$ for all i. Thus the prime factors of a_{i+1} consist of some (possibly all) prime factors of a_i. Since a_1 has a unique factorization into finitely many prime factors, the prime factors will end up being the same, and the chain will stabilize.

2. Take $0 \neq a_1 \in R$. If a_1 is irreducible, we are done. Let us thus assume that a_1 is not irreducible, that is

$$a_1 = a_2 b_2$$

where a_2 and b_2 are not unit. Since $a_2 \mid a_1$, we have $(a_1) \subseteq (a_2)$, and actually

$$(a_1) \subsetneq (a_2).$$

Indeed, if $(a_1) = (a_2)$, then a_2 would be a multiple of a_1, namely $a_2 = ca_1$ and thus

$$a_1 = a_2 b_2 \Rightarrow a_1 = c a_1 b_2 \Rightarrow a_1(1 - c b_2) = 0$$

implying that $cb_2 = 1$ and thus b_2 is a unit. This contradicts the assumption that a_1 is not irreducible. This computation has shown us that whenever we get a factor which is not irreducible, we can add a new principal ideal to the chain of ideals. Thus, if $a_2 b_2$ is a product of irreducible, we are done. Otherwise, we have that say a_2 is not irreducible, and $a_2 = a_3 b_3$, yielding

$$(a_1) \subsetneq (a_2) \subsetneq (a_3).$$

Since R satisfies the ascending chain condition on principal ideals, this process cannot go on and must stop, showing that we have a factorization into irreducible.

3. We now know that R allows a factorization into irreducible. We want to prove that this factorization is unique, under the assumption that every irreducible is prime. Suppose thus that

$$a = u p_1 p_2 \cdots p_n = v q_1 q_2 \cdots q_m$$

where u, v are units and p_i, q_j are irreducible. p_1 is an irreducible but also a prime by assumption, thus it must divides one of the q_j, say q_1, and we have $q_1 = p_1 d$. Since q_1 is irreducible, d must be a unit, and q_1 and p_1 are associate. We can iterated the process to find that q_i and p_i are associate for all i.

□

We now introduce a notion stronger than being a unique factorization domain.

Definition 2.24. A principal ideal domain (PID) is an integral domain in which every ideal is principal.

Theorem 2.16. *A principal ideal domain R is a unique factorization domain.*

Proof. What we will prove is that if R is a principal ideal domain, then

- R satisfies the ascending chain condition on principle ideals.
- every irreducible in R is also prime.

Having proved these two claims, we can conclude using the above theorem.

Let us first prove that R satisfies the ascending chain condition on principle ideals. Consider the following sequence of principal ideals

$$(a_1) \subseteq (a_2) \subseteq (a_3) \dots$$

and let $\mathcal{I} = \cup_{i=1}^{\infty}(a_i)$. Note that $\mathcal{I}$ is an ideal of R (be careful, a union of ideals is not an ideal in general!). Indeed, we have that $\mathcal{I}$ is closed under addition:

take $a, b \in \mathcal{I}$, then there are ideals $\mathcal{I}_j$ and $\mathcal{I}_k$ in the chain with $a \in \mathcal{I}_j$ and $b \in \mathcal{I}_k$. If $m \geq \max(j, k)$, then both $a, b \in \mathcal{I}_m$ and so do $a + b$. To check that $\mathcal{I}$ is closed under multiplication by an element of R, take again $a \in \mathcal{I}$. Then $a \in \mathcal{I}_j$ for some j. If $r \in R$, then $ra \in \mathcal{I}_j$ implying that $ra \in \mathcal{I}$.

Now by assumption, $\mathcal{I}$ is a principal ideal, generated by, say b: $\mathcal{I} = (b)$. Since b belongs to $\cup_{i=1}^{\infty}(a_i)$, it must belongs to some (a_n). Thus $\mathcal{I} = (b) \subseteq (a_n)$. For $j \geq n$, we have

$$(a_j) \subseteq \mathcal{I} \subseteq (a_n) \subseteq (a_j)$$

which proves that the chain of ideal stabilizes.

We are left to prove that every irreducible element is also prime. Let thus a be an irreducible element. Consider the principal ideal (a) generated by a. Note that (a) is a proper ideal: if $(a) = R$, then $1 \in (a)$ and thus a is a unit, which is a contradiction.

We have that (a) is included in a maximal ideal $\mathcal{I}$ (this can be deduced from either the ascending chain condition or from the theorem (Krull's theorem) that proves that every ideal is contained in a maximal ideal). Since R is a principal ideal domain, we have that $\mathcal{I} = (b)$. Thus

$$(a) \subseteq (b) \Rightarrow b \mid a \Rightarrow a = bd$$

where a is irreducible, b cannot be a unit (since $\mathcal{I}$ is by definition of maximal ideal a proper ideal), and thus d has to be a unit of R. In other words, a and b are associate. Thus

$$(a) = \mathcal{I} = (b).$$

Since $\mathcal{I}$ is a maximal ideal, it is prime implying that a is prime, which concludes the proof. □

Determining whether a ring is a principal ideal domain is in general quite a tough question. It is still an open conjecture (called Gauss's conjecture) to decide whether there are infinitely many real quadratic fields which are principal (we use the terminology "principal" for quadratic fields by abuse of notation, it actually refers to their ring of integers, that is rings of the form either $\mathbb{Z}[\sqrt{d}]$ if $d \equiv 2$ or $3 \mod 4$ or $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ else).

One way mathematicians have found to approach this question is to actually prove a stronger property, namely whether a ring R is Euclidean.

Definition 2.25. Let R be an integral domain. We say that R is a Euclidean domain is there is a function Ψ from $R\backslash\{0\}$ to the non-negative integers such that

$$a = bq + r,\ a, b \in R,\ b \neq 0,\ q, r \in R$$

where either $r = 0$ or $\Psi(r) < \Psi(b)$.

When the division is performed with natural numbers, it is clear what it means that $r < b$. When we work with polynomials instead, we can say that $\deg r < \deg b$. The function Ψ generalizes these notions.

Theorem 2.17. *If R is a Euclidean domain, then R is a principal ideal domain.*

Proof. Let $\mathcal{I}$ be an ideal of R. If $\mathcal{I} = \{0\}$, it is principal and we are done. Let us thus take $\mathcal{I} \neq \{0\}$. Consider the set

$$\{\Psi(b),\ b \in \mathcal{I},\ b \neq 0\}.$$

It is included in the non-negative integers by definition of Ψ, thus it contains a smallest element, say n. Let $0 \neq b \in \mathcal{I}$ such that $\Psi(b) = n$.

We will now prove that $\mathcal{I} = (b)$. Indeed, take $a \in \mathcal{I}$, and compute

$$a = bq + r$$

where $r = 0$ or $\Psi(r) < \Psi(b)$. This yields

$$r = a - bq \in \mathcal{I}$$

and $\Psi(r) < \Psi(b)$ cannot possibly happen by minimality of n, forcing r to be zero. This concludes the proof. □

Example 2.7. Consider the ring

$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d},\ a, b \in \mathbb{Z}\}$$

with

$$\Psi(a + b\sqrt{d}) = |a^2 - bd^2|.$$

We will show that we have a Euclidean domain for $d = -2, -1, 2, 3$.

Note that $\mathbb{Z}[\sqrt{d}]$ is an integral domain. Take $\alpha, \beta \neq 0$ in $\mathbb{Z}[\sqrt{d}]$. Now we perform the division of α by β to get something of the form $x + \sqrt{d}y$. Since $\mathbb{Z}[\sqrt{d}]$ is not a field, there is no reason for this division to give a result in $\mathbb{Z}[\sqrt{d}]$ (that is, $x, y \in \mathbb{Z}$), however, we can compute it in $\mathbb{Q}(\sqrt{d})$, and get a result with x, y rational. Since x, y have no reason to be integers, let us approximate them by integers x_0, y_0, namely take x_0, y_0 such that

$$|x - x_0| < 1/2,\ |y - y_0| < 1/2.$$

Let

$$q = x_0 + y_0\sqrt{d},\ r = \beta((x - x_0) + (y - y_0)\sqrt{d})$$

then

$$\begin{aligned} \beta q + r &= \beta(x_0 + y_0\sqrt{d}) + \beta((x - x_0) + (y - y_0)\sqrt{d}) \\ &= \beta(x + y\sqrt{d}) = \alpha. \end{aligned}$$

We are left to show that $\Psi(\alpha) < \Psi(\beta)$. We have

$$\begin{aligned} \Psi(r) &= \Psi(\beta)\Psi((x - x_0) + (y - y_0)\sqrt{d}) \\ &= \Psi(\beta)[(x - x_0)^2 - d(y - y_0)^2] \\ &< \Psi(\beta)\left(\frac{1}{4} + |d|\frac{1}{4}\right) \end{aligned}$$

showing that $\mathbb{Z}[\sqrt{d}]$ is indeed a Euclidean domain for $d = -2, -1, 2, 3$.

Below is a summary of the ring hierarchy (recall that PID and UFD stand respectively for principal ideal domain and unique factorization domain):

integral domains $\supset$ UFD $\supset$ PID $\supset$ Euclidean domains

Note that though the Euclidean division may sound like an elementary concept, as soon as the ring we consider is fancier than $\mathbb{Z}$, it becomes quickly a difficult problem. We can see that from the fact that being Euclidean is stronger than being a principal ideal domain. All the inclusions are strict, since one may check that $\mathbb{Z}[\sqrt{-3}]$ is an integral domain but is not a UFD, $\mathbb{Z}[X]$ is a UFD which is not PID, while $\mathbb{Z}[(1+i\sqrt{19})/2]$ is a PID which is not a Euclidean domain.

2.7 Irreducible polynomials

Recall the definition of irreducible that we have seen: a non-zero element a which is not a unit is said to be irreducible if $a = bc$ implies that either b or c is a unit. Let us focus on the case where the ring is a ring of polynomials $R[X]$ and R is an integral domain.

Definition 2.26. If R is an integral domain, then an irreducible element of $R[X]$ is called an irreducible polynomial.

In the case of a field F, then units of $F[X]$ are non-zero elements of F. Then we get the more familiar definition that an irreducible element of $F[X]$ is a polynomial of degree at least 1, that cannot be factored into two polynomials of lower degree.

Let us now consider the more general case where R is an integral domain (thus not necessarily a field, it may not even be a unique factorization domain). To study when polynomials over an integral domain R are irreducible, it is often more convenient to place oneselves in a suitable field that contains R, since division in R can be problematic. To do so, we will now introduce the field of fractions, also called quotient field, of R. Since there is not much more difficulty in treating the general case, that is, when R is a commutative ring, we present this construction.

Let S be a subset of R which is closed under multiplication, contains 1 and does not contain 0. This definition includes the set of all non-zero elements of an integral domain, or the set of all non-zero elements of a commutative ring that are not zero divisors. We define the following equivalence relation on $R \times S$:

$$(a, b) \sim (c, d) \iff s(ad - bc) = 0 \text{ for some } s \in S.$$

It is clearly reflexive and symmetric. Let us check the transitivity. Suppose that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. Then

$$s(ad - bc) = 0 \text{ and } t(cf - de) = 0$$

for some $s, t \in S$. We can now multiply the first equation by tf, the second by sb and add them

$$stf(ad - bc) + tsb(cf - de) = 0$$

to get

$$sdt(fa - be) = 0$$

which proves the transitivity.

What we are trying to do here is to mimic the way we deal with $\mathbb{Z}$. If we take non-zero $a, b, c, d \in \mathbb{Z}$, we can write down $a/b = c/d$, or equivalently $ad = bc$, which is also what $(a, b) \sim (c, d)$ satisfies by definition if we take R to be an integral domain. In a sense, (a, b) is some approximation of a/b.

Formally, if $a \in R$ and $b \in S$, we define the fraction a/b to be the equivalence class of the pair (a, b). The set of all equivalence classes is denoted by $S^{-1}R$. To make it into a ring, we define the following laws in a natural way:

- addition:
$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}.$$
- multiplication:
$$\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$
- additive identity:
$$\frac{0}{1} = \frac{0}{s},\ s \in S.$$
- additive inverse:
$$-\frac{a}{b} = \frac{-a}{b}.$$
- multiplicative identity:
$$\frac{1}{1} = \frac{s}{s},\ s \in S.$$

To prove that we really obtain a ring, we need to check that all these laws are well-defined.

Theorem 2.18. *With the above definitions, the set of equivalence classes $S^{-1}R$ is a commutative ring.*

1. *If R is an integral domain, so is $S^{-1}R$.*
2. *If R is an integral domain, and $S = R\backslash\{0\}$, then $S^{-1}R$ is a field.*

Proof. **Addition is well-defined.** If $a_1/b_1 = c_1/d_1$ and $a_2/b_2 = c_2/d_2$, then for some $s, t \in S$, we have

$$s(a_1d_1 - b_1c_1) = 0 \text{ and } t(a_2d_2 - b_2c_2) = 0.$$

We can now multiply the first equation by tb_2d_2 and the second by sb_1d_1 to get

$$tb_2d_2s(a_1d_1 - b_1c_1) = 0 \text{ and } sb_1d_1t(a_2d_2 - b_2c_2) = 0,$$

and adding them yields

$$st[d_2d_1(b_2a_1 + b_1a_2) - b_2b_1(d_2c_1 + d_1c_2)] = 0$$

that is

$$\frac{b_2a_1 + b_1a_2}{b_2b_1} = \frac{d_2c_1 + d_1c_2}{d_2d_1},$$

which can be rewritten as

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{c_1}{d_1} + \frac{c_2}{d_2}$$

and we conclude that addition does not depend on the choice of a representative in an equivalence class.

Multiplication is well-defined. We start as before. If $a_1/b_1 = c_1/d_1$ and $a_2/b_2 = c_2/d_2$, then for some $s, t \in S$, we have

$$s(a_1d_1 - b_1c_1) = 0 \text{ and } t(a_2d_2 - b_2c_2) = 0.$$

Now we multiply instead the first equation by ta_2d_2, the second by sc_1b_1 and we add them:

$$st[a_2d_2a_1d_1 - c_1b_1b_2c_2] = 0.$$

This implies, as desired, that

$$\frac{a_1a_2}{b_1b_2} = \frac{c_1c_2}{d_1d_2}.$$

To be complete, one should check that the properties of a ring are fulfilled, but this follows from the fact that addition and multiplication are carried the usual way.

1. We want to prove that $S^{-1}R$ is an integral domain. We assume that R is an integral domain, and we need to check the definition of an integral domain. Namely, suppose that $(a/b)(c/d) = 0$ in $S^{-1}R$, that is

   $$\frac{a}{b}\frac{c}{d} = \frac{0}{1}.$$

   This means that $(ac, bd) \sim (0, 1)$ and $acs = 0$ for some $s \in S$. Now $acs = 0$ is an equation in R, which is an integral domain, and $s \neq 0$, thus $ac = 0$, so either a or c is 0, and consequently either a/b or c/d is zero.

2. To conclude, we want to prove that $S^{-1}R$ is a field, assuming that R is an integral domain, and $S = R\backslash\{0\}$. We consider a/b a non-zero element of $S^{-1}R$, for which we need to find an inverse. Note that a and b are non-zero, thus they are both in S meaning that both a/b and b/a are in $S^{-1}R$ and b/a is the multiplicative inverse of a/b.

□

Definition 2.27. Let R be a commutative ring. Based on the above, the set of equivalence classes $S^{-1}R$ is a commutative ring, called the ring of fractions of R by S. If R is an integral domain, and $S = R\backslash\{0\}$, then $S^{-1}R$ is called the field of fractions or quotient field of R.

Now that we have defined a suitable field, we are left to prove that we can embed an integral domain R in its quotient field.

Proposition 2.19. *A commutative ring R can be embedded in its ring of fractions $S^{-1}R$, where S is the set of all its non-divisors of zero. In particular, an integral domain can be embedded in its quotient field, which is furthermore the smallest field containing R.*

Proof. Consider the following map:

$$f : R \to S^{-1}R,\ a \mapsto f(a) = a/1.$$

It is not hard to check that f is a ring homomorphism. If S has no zero divisor, we have that the kernel of f is given by the set of a such that $f(a) = a/1 = 0/1$, that is the set of a such that $sa = 0$ for some s. Since s is not a zero divisor, we have $a = 0$ and f is a monomorphism. □

Let us get back to the irreducible polynomials, and consider now the case where D is a unique factorization domain. It is not necessarily a field, but we now know how to embed it in a suitable field, namely its field of fractions, or quotient field. Take the polynomial $f(X) = a + abX$, $a \neq 0$ not a unit. Since we can factor it as

$$f(X) = a(1 + bX)$$

where a is not a unit by assumption, this polynomial is not irreducible. But we do not really have a factorization into two polynomials of lower degree. What happens here is that the constant polynomials are not necessarily units, unlike in the case of fields. To distinguish this case, we introduce the notion of primitive polynomial.

Definition 2.28. Let D be a unique factorization domain and let $f \in D[X]$. We call the greatest common divisor of all the coefficients of f the content of f, denoted by $c(f)$. A polynomial whose content is a unit is called a primitive polynomial.

We can now rule out the above example, and we will prove later that this allows us to say that a primitive polynomial is irreducible if and only if it cannot be factored into two polynomials of lower degree. Be careful however that "primitive polynomial" has a different meaning if it is defined over a field.

The next goal is to prove Gauss lemma, which in particular implies that the product of two primitive polynomials is a primitive polynomial.

We start with a lemma.

Lemma 2.20. *Let D be a unique factorization domain, and consider $f \neq 0, g, h \in D[X]$ such that $pf(X) = g(X)h(X)$ with p a prime. Then either p divides all the coefficients of g or p divides all the coefficients of h.*

Figure 2.2: Carl Friedrich Gauss (1777-1855)

Before starting the proof, let us notice that this lemma is somehow a generalization of the notion of prime. Instead of saying that $p|ab$ implies $p|a$ or $p|b$, we have $p|g(X)h(X)$ implies that $p|g(X)$ or $p|h(X)$ (dividing the whole polynomial means dividing all of its coefficients).

Proof. Denote

$$g(X) = g_0 + g_1X + \ldots + g_sX^s,\ h(X) = h_0 + h_1X + \ldots + h_tX^t.$$

Suppose by contradiction that p does not divide all coefficients of g and does not divide all coefficients of h either. Then let g_u and h_v be the coefficients of minimum index not divisible by p. Then the coefficient of X^{u+v} in $g(X)h(X)$ is

$$g_0h_{u+v} + g_1h_{u+v-1} + \ldots + g_uh_v + \ldots + g_{u+v-1}h_1 + g_{u+v}h_0.$$

By definition of u and v, p divides every term but g_uh_v, thus p cannot possibly divide the entire expression, and thus there exists a coefficient of $g(X)h(X)$ not divisible by p. This contradicts the fact that $p|g(X)h(X)$. □

Proposition 2.21. (Gauss Lemma). *Let f, g be non-constant polynomials in $D[X]$ where D is a unique factorization domain. The content of a product of polynomials is the product of the contents, namely*

$$c(fg) = c(f)c(g),$$

up to associates. In particular, the product of two primitive polynomials is primitive.

Proof. Let us start by noticing that by definition of content, we can rewrite

$$f(X) = c(f)f^*(X),\ g(X) = c(g)f^*(X),$$

where $f^*, g^* \in D[X]$ are primitive. Clearly

$$fg = c(f)c(g)f^*g^*.$$

Since $c(f)c(g)$ divides fg, it divides every coefficient of fg and thus their greatest common divisor:

$$c(f)c(g) \mid c(gf).$$

We now prove the converse, namely that $c(gf)| \ |c(g)c(g)$. To do that, we consider each prime p appearing in the factorization of $c(gf)$ and argue that $p \mid c(f)c(g)$. Let thus p be a prime factor of $c(gf)$. Since $fg = c(fg)(fg)^*$, we have that $c(fg)$ divides fg, that is

$$p \mid c(f)c(g)f^*g^* = (c(f)f^*)(c(g)g^*).$$

By the above lemma, either $p \mid c(f)f^*$ or $p \mid c(g)g^*$, say $p \mid c(f)f^*$, meaning that either $p \mid c(f)$ or $p \mid f^*$. Since f^* is primitive, p cannot possibly divide f^*, and thus

$$p \mid c(f) \Rightarrow p \mid c(f)c(g).$$

If p appears with multiplicity, we iterate the reasoning with the same p. □

We are now ready to connect irreducibility over a unique factorization domain and irreducibility over the corresponding quotient field or field of fractions.

Proposition 2.22. *Let D be a unique factorization domain with quotient field F. If f is a non-constant polynomial in $D[X]$, then f is irreducible over D if and only if f is primitive and irreducible over F.*

Proof. First assume that f is irreducible over D.

f **is primitive.** Indeed, if f were not primitive, then we could write

$$f = c(f)f^*,$$

where $c(f)$ denotes the content of f and f^* is primitive. Since we assume f is not primitive, its content cannot be a unit, which contradicts the irreducibility of f over D, and we conclude that f is primitive.

f **is irreducible over** F**.** Again assume by contradiction that f is not irreducible over F. Now F is a field, thus reducible means f can be factored into a product of two non-constant polynomials in $F[X]$ of smaller degree:

$$f(X) = g(X)h(X), \ \deg g < \deg f, \ \deg h < \deg f.$$

Since g, h are in $F[X]$, and F is the field of fractions of D, we can write

$$g(X) = \frac{a}{b}g^*(X), \ h(X) = \frac{c}{d}h^*(X), \ a, b, c, d \in D$$

and g^*, h^* primitive. Thus

$$f(X) = \frac{ac}{bd}g^*(X)h^*(X)$$

Figure 2.3: Ferdinand Eisenstein (1823-1852)

where g^*h^* is a primitive polynomial by Gauss Lemma. Since we have already proved that f^* is primitive, it must be that $bd = ac$. But this would mean that

$$f(X) = g^*(X)h^*(X)$$

which contradicts the fact that $f(X)$ is irreducible over $D[X]$ and we conclude that f is also irreducible over $F[X]$.

We are left to prove the converse. Let then f be a primitive and irreducible polynomial over F. We do it by contraction, and assume that the primitive polynomial f is not irreducible over D:

$$f(X) = g(X)h(X).$$

Since f is primitive, $\deg g$ and $\deg h$ are at least 1. But then neither g not h can be a unit in $F[X]$ (these are units in F) and thus

$$f = gh$$

contradicts the irreducibility of f over F. □

In other words, we have proven that f irreducible over D is equivalent to f primitive and cannot be factored into two polynomials of lower degree in $F[X]$.

To conclude, we present a practical criterion to decide whether a polynomial in $D[X]$ is irreducible over F.

Proposition 2.23. (Eisenstein's criterion). *Let D be a unique factorization domain, with quotient field F and let*

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

be a polynomial in $D[X]$ with $n \geq 1$ and $a_n \neq 0$.

If p is a prime in D and p divides a_i, $0 \leq i \leq n$ but p does not divide a_n nor does p^2 divide a_0, then f is irreducible over F.

Proof. We first divide f by its content, to get a primitive polynomial. By the above proposition, it is enough to prove that this primitive polynomial is irreducible over D.

Let thus f be a primitive polynomial and assume by contradiction it is reducible, that is

$$f(X) = g(X)h(X)$$

with

$$g(X) = g_0 + \ldots + g_r X^r,\ h(X) = h_0 + \ldots + h_s X^s.$$

Notice that r cannot be zero, for if $r = 0$, then $g_0 = g$ would divide f and thus all a_i implying that g_0 divides the content of f and is thus a unit. But this would contradict the fact that f is reducible. We may from now on assume that

$$r \geq 1,\ s \geq 1.$$

Now by hypothesis, $p \mid a_0 = g_0 h_0$ but p^2 does not divide a_0, meaning that p cannot divide both g_0 and h_0. Let us say that

$$p \mid g_0$$

and p does not divide h_0 (and vice-versa).

By looking at the dominant coefficient $a_n = g_r h_s$, we deduce from the assumption that p does not divide a_n that p cannot possibly divide g_r. Let i be the smallest integer such that p does not divide g_i. Then

$$1 \leq i \leq r < n = r + s.$$

Let us look at the ith coefficient

$$a_i = g_0 h_i + g_1 h_{i-1} + \ldots + g_i h_0$$

and by choice of i, p must divide $g_0, \ldots, g_{i-1}$. Since p divides a_i by assumption, it thus must divide the last term $g_i h_0$, and either $p \mid g_i$ or $p \mid h_0$ by definition of prime. Both are impossible: we have chosen p dividing neither h_0 nor g_i. This concludes the proof. □

The main definitions and results of this chapter are

- **(2.1-2.2).** Definitions of: ring, zero divisor, unit, integral domain, division ring, subring, characteristic, ring homomorphism, ideal, quotient ring. Factor and 1st Isomorphism Theorem for rings.
- **(2.3-2.4).** Operations on ideals, Chinese Remainder Theorem, Correspondence Theorem for rings. Definitions of: principal ideal, maximal ideal, prime ideal, the characterization of the two latter in the commutative case.
- **(2.5).** Polynomial Euclidean division, number of roots of a polynomial.
- **(2.6).** Definitions of: associate, prime, irreducible, unique factorization domain, ascending chain condition, principal ideal domain, Euclidean domain. Connections between prime and irreducible. Hierarchy among UFD, PID and Euclidean domains.
- **(2.7).** Construction of ring of fractions. Definitions of: content of a polynomial, primitive polynomial. Gauss Lemma, Eisenstein's criterion.