Chapter 3

Ring Theory

3.1 Rings, ideals and homomorphisms

Definition 3.1. A ring R is an abelian group with a multiplication operation

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

which is associative, and satisfies the distributive laws

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

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 3.2. 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 3.3. Let R be ring. If $ab = ba$ for any a, b in R, then R is said to be commutative.

A ring was defined above 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.

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 (we will define Hamilton quaternions below), 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. 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.

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

Definition 3.4. 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}. A field is a commutative ring in which every non-zero element is invertible.

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

Definition 3.5. The characteristic of a ring R, denoted by $\mathrm{char}R$, is the smallest positive integer such that

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

If there is no such positive integer, we say that the ring has characteristic 0.

We can also extract smaller rings from a given ring.

Definition 3.6. 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 3.1.

1. $\mathbb{Z}$ is an integral domain but not a field.
2. The integers modulo n form a commutative ring, which is an integral domain if and only if n is prime.
3. For $n \geq 2$, the $n \times n$ matrices $\mathcal{M}_n(\mathbb{R})$ with coefficients in $\mathbb{R}$ are a non-commutative ring, but not an integral domain.
4. The set
$$\mathbb{Z}[i] = \{a + bi,\ a, b \in \mathbb{Z}\},\ i^2 = -1,$$
is a commutative ring. It is also an integral domain, but not a field.

	commutative	non-commutative
has zero divisor	integers mod n, n not a prime	matrices over a field
has no zero divisor	$\mathbb{Z}$	$\{a+bi+cj+dk,\ a,b,c,d \in \mathbb{Z}\}$
non-zero element invertible	$\mathbb{R}$	$\mathbb{H}$

5. 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

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

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

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

As an another example, let us do the classification of rings containing 4 elements.

Example 3.2. Let R be a ring with 4 elements, thus it must contain the two elements $0 \neq 1$, and be an abelian group of order 4. In a group of order 4, elements have order 2 or 4, thus either 1 has order 4, in which case we obtain the integers modulo 4, or 1 has order 2. If 1 has order 2, then $\mathrm{char}(R) = 2$. Now $1+1 = 0$, and we must have another element $u \neq 0, 1$ in R. By the closure property under addition, $u+1$ must be in R. Note that $2u = 0$ and thus $u = -u$. Then by the closure property under multiplication, u^2, $u(u+1) = u^2 + u$ and $(u+1)^2 = u^2 + 2u + 1 = u^2 + 1$ must belong to R. Also this ring is commutative since $u(u+1) = (u+1)u$. Since we are parameterizing the ring by u, we only need to compute u^2 to determine the whole ring multiplication table. The possible

values taken by u^2 are $0, 1, u, u+1$. This gives us the following possibilities:

u^2	u^2+u	u^2+1
0	u	1
1	$u+1$	0
u	0	$u+1$
$u+1$	1	u

This gives us 4 possible multiplication tables:

1.

	1	u	$u+1$
1	1	u	$u+1$
u	u	0	u
$u+1$	$u+1$	u	1

2.

	1	u	$u+1$
1	1	u	$u+1$
u	u	1	$u+1$
$u+1$	$u+1$	$u+1$	0

3.

	1	u	$u+1$
1	1	u	$u+1$
u	u	u	0
$u+1$	$u+1$	0	$u+1$

4.

	1	u	$u+1$
1	1	u	$u+1$
u	u	$u+1$	1
$u+1$	$u+1$	1	u

First, we observe that the first and the second table give the same multiplication. Indeed, take the second table, permute columns 2 and 3, and rows 2 and 3, then switch the labels of u and $u+1$, to get the same multiplication as in the first table.

The Klein group $C_2 \times C_2$ is an instance of the third table, which is seen by setting $u = (0, 1)$. The 4rth table, it is the multiplication table of a group. One can check the closure, the existence of an identity, and that of an inverse for every element. Since we see every element of R but zero, R is a field.

An instance of the first table would be a matrix ring obtained by setting

$$u = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

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.

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.

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 3.7. 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.

We do need to mention that $f(1_R) = 1_S$, otherwise, since a ring is not a group under multiplication, strange things can happen. For example, if $\mathbb{Z}_6$ denotes the integers mod 6, the map $f : \mathbb{Z}_6 \to \mathbb{Z}_6$, $n \mapsto 3n$ satisfies that $f(m+n) = 3(m+n) = 3m + 3n = f(m) + f(n)$, and $f(n)f(m) = 3m3n = 3mn = f(mn)$ but $f(1) \neq 1$ and f is not a ring homomorphism. Notice the difference with group homomorphism: from $f(a+b) = f(a) + f(b)$, we deduce that $f(a+0) = f(a) + f(0)$, that is $f(a) = f(a) + f(0)$. Now because $f(a)$ is invertible, it must be that $f(0) = 0$! Once we reach $f(a) = f(a)f(1)$, because $f(a)$ does not have to be invertible, we cannot conclude!

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 (in an obscure journal, rediscovered in 1847) 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. Here is what it gives in modern terminology:

Definition 3.8. 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.

Example 3.3. The even integers $2\mathbb{Z} = \{2n,\ n \in \mathbb{Z}\}$ form an ideal of $\mathbb{Z}$. The set of polynomials in $\mathbb{R}[X]$ with constant coefficient zero form an ideal of $\mathbb{R}[X]$.

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 3.9. 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:

- $\mathrm{Ker} f$ is an additive subgroup of R.
- Take $a \in \mathrm{Ker} f$ 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 $\mathrm{Ker} f$.
- Then $\mathrm{Ker} f$ has to be proper (that is, $\mathrm{Ker} f \neq R$), since $f(1) = 1$ by definition.

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

Lemma 3.1. *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 $\mathrm{Ker} f$ is a two-sided ideal of R, then either $\mathrm{Ker} f = \{0\}$ or $\mathrm{Ker} f = R$. But $\mathrm{Ker} f \neq R$ since $f(1) = 1$ by definition (in words, $\mathrm{Ker} f$ 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.

3.2 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 3.10. The set of cosets of the two-sided ideal $\mathcal{I}$ given by

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

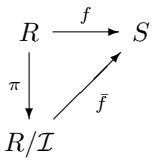
is a ring with identity $1_R + \mathcal{I}$ and zero element $0_R + \mathcal{I}$ called a quotient ring.

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 3.4. We have that $m\mathbb{Z}$ is an ideal of $\mathbb{Z}$, and we can consider the quotient ring $\mathbb{Z}/m\mathbb{Z}$ which is the ring of integers modulo m.

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 two-sided 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 3.2. (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 3.3. (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 3.5. This example uses a polynomial ring, we will study polynomial rings in more details later. Consider the map $f : \mathbb{R}[X] \to \mathbb{C}$, $f(p(X)) = p(i)$, that is, f takes a polynomial $p(X)$ with real coefficients, and evaluate this polynomial in i ($i^2 = -1$). This map is surjective (for $z = a + ib \in \mathbb{C}$, take the polynomial $p(X) = a + bX$) and its kernel is formed by polynomials which, when evaluated in i, are giving 0, meaning that i is a root of the polynomial, or equivalently that $(X^2 + 1)$ is a factor of the polynomial. Thus $\mathrm{Ker}(f) = (X^2 + 1)\mathbb{R}[X] = \{p(X) = (X^2 + 1)q(X),\ q(X) \in \mathbb{R}[X]\}$. Using the first isomorphism for rings, we have

$$\mathbb{R}[X]/(X^2 + 1)\mathbb{R}[X] \simeq \mathbb{C}.$$

We note that we have a second and a third isomorphism theorem for rings. The second one says that the quotient rings $(S + I)/I$ and $S/(S \cap I)$ are isomorphic. The third says that if J is an ideal of R, and I is an ideal of R such that $J \subset I \subset R$, then $(R/J)/(I/J)$ is isomorphic to R/I (note that I/J is an ideal of R/J).

3.3 Maximal and prime ideals

Here are a few special ideals.

Definition 3.11. 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 3.12. An ideal generated by a single element a is called a principal ideal, denoted by $\langle a \rangle$.

Definition 3.13. 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, (a version also exists for groups, sometimes it is also called a 4rth isomorphism theorem) since we will need it for characterizing maximal ideals.

Theorem 3.4. (Correspondence Theorem for rings). *If $\mathcal{I}$ is a two-sided 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 (right/left/two-sided) ideals of R containing $\mathcal{I}$ and the set of all (right/left/two-sided) ideals of $R/\mathcal{I}$.

Proof. Let us thus define two sets, S_1 is the set of ideals of R containing $\mathcal{I}$, and S_2 is the set of ideals of $R/\mathcal{I}$. We define two maps:

$$f : S_1 \to S_2,\ J \mapsto f(J) = \{a + \mathcal{I},\ a \in J\} \subset R/\mathcal{I},$$

and

$$g : S_2 \to S_1,\ \mathcal{J} \mapsto g(\mathcal{J}) = \{a,\ a + \mathcal{I} \in \mathcal{J}\} \subset R.$$

We have that $f(J)$ and $g(\mathcal{J})$ are ideals of $R/\mathcal{I}$ and R respectively. Indeed:

- Consider first $f(J)$. It is a set of cosets, where each coset is such that its representative is chosen in J. It is thus a subset of $R/\mathcal{I}$. To prove that it is an additive subgroup, we take $a + \mathcal{I}$ and $a' + \mathcal{I}$ both in $f(J)$, and we check whether $(a + \mathcal{I}) - (a' + \mathcal{I})$ is in $f(J)$. We know that the difference of two cosets is again a coset in a quotient ring, and that in particular $(a + \mathcal{I}) - (a' + \mathcal{I}) = (a - a') + \mathcal{I}$. Now both $a, a' \in J$, and J itself is an ideal, so $a - a' \in J$. Then we need to check the property of closure under multiplication. Let $(r + \mathcal{I})$ be an element of $R/\mathcal{I}$, then $(r + \mathcal{I})(a + \mathcal{I}) = ra + \mathcal{I}$, this is how we multiply two cosets. Then for J a left ideal, $ra \in J$ and $f(J)$ is a left ideal.

- Consider next $g(\mathcal{J})$. Take $a, b \in g(\mathcal{J})$, we need to check that $a - b$ is such that $a - b + \mathcal{I} \in \mathcal{J}$. But $a - b + \mathcal{I} = (a + \mathcal{I}) - (b + \mathcal{I}) \in \mathcal{J}$ since $\mathcal{J}$ is an ideal. Then take a in $g(\mathcal{J})$ and $r \in R$, we need to check that $ra + \mathcal{I}$ is in $\mathcal{J}$. But again, $ra + \mathcal{I} = (r + \mathcal{I})(a + \mathcal{I})$ which is in $\mathcal{J}$ if $\mathcal{J}$ is a left ideal, showing that $g(\mathcal{J})$ is a left ideal.

We will prove that f and g are inverse of each other, and therefore we have a bijection between the two sets.

If $\mathcal{J} \in S_2$, then $f(g(\mathcal{J})) = \{a + \mathcal{I},\ a \in g(\mathcal{J})\} = \{a + \mathcal{I},\ a + \mathcal{I} \in \mathcal{J}\} = \mathcal{J}$.

If $J \in S_1$, then $g(f(J)) = \{a,\ a + \mathcal{I} \in f(J)\} = \{a,\ a + \mathcal{I} = b + \mathcal{I},\ b \in J\} = \{a,\ a \in b + \mathcal{I},\ b \in J\}$.

The last set contains J, but we need to show that it is actually J. We have

$$a \in b + \mathcal{I} \Rightarrow (a - b) \in \mathcal{I} \subset J \Rightarrow a = b + J \Rightarrow a \in J.$$

This concludes the proof. □

Here is a characterization of maximal ideals in commutative rings.

Theorem 3.5. *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 $a \notin M$. 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 (since R/M is a field). 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. □

To define a prime ideal, we get some inspiration from prime numbers. If p is a prime number, then we have that $p|ab$ implies $p|a$ or $p|b$.

Definition 3.14. 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

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

Example 3.6. For the ring $R = \mathbb{Z}$, the ideal $\mathcal{I} = 5\mathbb{Z}$ is principal and prime. To see that $\mathcal{I}$ is prime, suppose $ab \in 5\mathbb{Z}$. Then ab is a multiple of 5, that is $ab = 5c$ for some $c \in \mathbb{Z}$. But since 5 is prime, and it divides ab, it must be that 5 divides a or 5 divides b, meaning that either $a \in 5\mathbb{Z}$ or $b \in 5\mathbb{Z}$.

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

Theorem 3.6. *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. □

Example 3.7. For the ring $R = \mathbb{Z}$, we get another proof that the ideal $\mathcal{I} = 5\mathbb{Z}$ is prime. We have that $\mathbb{Z}/5\mathbb{Z}$ is the ring of integers modulo 5, which is an integral domain.

Corollary 3.7. *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 3.8. *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 3.8. 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 $\langle X \rangle = \text{Ker}\varphi$ which is prime by the above corollary.

3.4 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.

Example 3.9. Take $f(X) = X^2 - 2$ and $g(X) = 2X - 1$. It is not possible to divide $f(X)$ by $g(X)$ in $\mathbb{Z}[X]$. If it were, then

$$f(X) = X^2 - 2 = (q_0 + q_1 X)(2X - 1) + r_0$$

and the coefficient of X^2 is 1 on the left hand side, and $2q_1$ on the right hand side. Now in $\mathbb{Z}$, there is no solution to the equation $2q_1 = 1$. Of course, this is possible in $\mathbb{Q}$, by taking $q_1 = 1/2$!

This gives the following:

Theorem 3.9. (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 3.10. *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. If f has no root in $R[X]$, then we are done. Let us thus assume that f has a root a_1 in R, 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 $\deg q_1 = n - n_1$ since R is an integral domain. Now if a_1 is the only root of f in R, then $n_1 \leq n$ and we are done. If not, consider similarly $a_2 \neq a_1$ 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

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

where $c(X)$ is a polynomial with no root in R, possibly constant, and

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

Since R is an integral domain, the only possible roots of f are $a_1, \ldots, a_k$, $k \leq n$, and the number of roots counting multiplicity is less than n. □

Example 3.10. 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.

3.5 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 3.15. The elements a, b are called associate if $a = ub$ for some unit u.

Definition 3.16. 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 3.17. If R is an integral domain, then an irreducible element of $R[X]$ is called an irreducible polynomial.

Remark. 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.

Definition 3.18. Let a be a non-zero element which is not a unit. Then a is called prime if 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 3.11. *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 3.11. 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 3.19. 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 3.12. *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):

$$a \cdot ud_1 \cdots d_r = vb_1 \cdots b_s \cdot wc_1 \ldots c_t.$$

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 introduce a notion which is actually stronger than being a unique factorization domain (though we will skip the proof that a PID is actually a UFD).

Definition 3.20. A principal ideal domain (PID) is an integral domain in which every ideal is principal.

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 3.21. Let R be an integral domain. We say that R is a Euclidean domain if 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 3.13. *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 3.12. Consider the ring

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

with

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

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

Note that $\mathbb{Z}[\sqrt{d}]$ is an integral domain. Take $\alpha, \beta \neq 0$ in $\mathbb{Z}[\sqrt{d}]$. Now we would like to perform the division of α by β to get something of the form

$$\alpha = \beta q + r,\ q, r \in \mathbb{Z}[\sqrt{d}].$$

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, $q, r \in \mathbb{Z}[\sqrt{d}]$), however, we can compute the division in $\mathbb{Q}(\sqrt{d})$:

$$\alpha/\beta = q',$$

with $q' = x + \sqrt{d}y$ with x, y rational. Let us now approximate x, y by integers x_0, y_0, namely take x_0, y_0 such that

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

Take

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

where clearly $q \in \mathbb{Z}[\sqrt{d}]$, then

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

which at the same time shows that $r \in \mathbb{Z}[\sqrt{d}]$. We are left to show that $\Psi(r) < \Psi(\beta)$. We have

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

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

ring	ED	PID	UFD	ID
$\mathbb{Z}$	yes	yes	yes	yes
$F[X]$, F a field	yes	yes	yes	yes
$\mathbb{Z}[i]$	yes	yes	yes	yes
$\mathbb{Z}[\sqrt{\pm 2}]$	yes	yes	yes	yes
$\mathbb{Z}[\sqrt{3}]$	yes	yes	yes	yes
$\mathbb{Z}[(1+i\sqrt{19})/2]$	no	yes	yes	yes
$\mathbb{Z}[X]$	no	no	yes	yes
$\mathbb{Z}[\sqrt{-3}]$	no	no	no	yes

Table 3.1: Examples of rings and their properties.

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.

Remark. All the inclusions are strict, since it can be checked that $\mathbb{Z}[\sqrt{-3}]$ is an integral domain but is not a UFD (we saw that 2 is irreducible but not prime), $\mathbb{Z}[X]$ is a UFD which is not PID (it is enough to show that the ideal $\langle 2, X\rangle$ is not principal), while $\mathbb{Z}[(1+i\sqrt{19})/2]$ is a PID which is not a Euclidean domain.

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).** 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, principal ideal domain, Euclidean domain. Connections between prime and irreducible. Hierarchy among UFD, PID and Euclidean domains.