MA416 Rings: Lecture Notes

Dr Rachel Quinlan
School of Mathematical and Statistical Sciences
NUI Galway

Contents

1 What is a Ring? 2
- 1.1 Some Examples . . . 2
- 1.2 The Axioms of a Ring . . . 5
- 1.3 Units in Rings . . . 9
- 1.4 Integral Domains and Zero–Divisors . . . 12

2 Factorization in Polynomial Rings 14
- 2.1 Polynomial Rings . . . 14
- 2.2 Divisibility and Irreducibility . . . 17

3 Ideals, Homomorphisms and Factor Rings 25
- 3.1 Ring Homomorphisms and Ideals . . . 25
- 3.2 Principal Ideal Domains . . . 30
- 3.3 Factor Rings . . . 31
- 3.4 Maximal and Prime Ideals . . . 36

4 Unique Factorization Domains (UFDs) 39
- 4.1 Unique Factorization Domains (UFDs) . . . 39
- 4.2 Every PID is a UFD . . . 43

Chapter 1

What is a Ring?

1.1 Some Examples

Consider the following sets :

1. $\mathbb{Z} = \{\ldots, -1, 0, 1, 2, \ldots\}$ – the set of *integers*.
2. $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z},\ b \neq 0\}$ – the set of *rational numbers*.
3. $M_2(\mathbb{R})$ – the set of 2×2 matrices with real numbers as entries.
4. $2\mathbb{Z} = \{\ldots, -2, 0, 2, 4, \ldots\}$ – the set of *even integers*.
5. $C(\mathbb{R})$ – the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$.
6. $\mathbb{Q}[x] = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \mid a_n, \ldots, a_0 \in \mathbb{Q}\}$ – the set of polynomials with rational coefficients.
7. $\mathbb{Z}/6\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}\}$ – the set of congruence classes in $\mathbb{Z}$ modulo 6.

Remember the very general definition of an *algebraic structure* as a set equipped with a *binary operation*, that is a scheme for combining any pair of elements of the set to produce a new element *of the same set*. All of the sets in our list above have binary operations defined on them in natural and probably mostly familiar ways. Of course it is possible for a set to have more than one "natural" binary operation defined on it. Algebra, in its broadest sense, is the study of algebraic structures.

What do all the six sets described above have in common as algebraic structures?

Each of them is equipped with two binary operations called addition and multiplication. In $\mathbb{Z}$, $\mathbb{Q}$ and $2\mathbb{Z}$ we have the usual addition and multiplication of integers and rational numbers. In $M_2(\mathbb{R})$ we have matrix addition and matrix multiplication. In $C(\mathbb{R})$ we have addition and multiplication defined by

$$(\underbrace{f+g}_{+\text{ in } C(\mathbb{R})})(x) = \underbrace{f(x)+g(x)}_{+\text{in}\mathbb{R}}, \text{ for all } x \in \mathbb{R} \text{ and all } f, g \in C(\mathbb{R})$$

$$(\underbrace{f \times g}_{\times \text{ in } C(\mathbb{R})})(x) = \underbrace{f(x) \times g(x)}_{\times \text{in} \mathbb{R}}, \text{ for all } x \in \mathbb{R} \text{ and all } f, g \in C(\mathbb{R}).$$

In $\mathbb{Q}[x]$ we have the usual addition and multiplication of polynomials, e.g.

$$(x^2 + 2x + 4) + (x^3 - 3x + 2) = x^3 + x^2 - x + 6,$$

$$(x^2 - 2x + 1)(x + 5) = x^3 + 5x^2 - 2x^2 - 10x + x + 5 = x^3 + 3x^2 - 9x + 5.$$

In $\mathbb{Z}/6\mathbb{Z}$ the addition and multiplication are defined modulo 6, e.g. $\bar{4} + \bar{5} = \bar{3}$; $\bar{4} \times \bar{5} = \bar{2}$, etc.

Note: In each case the set under consideration is *closed* under the relevant operations of addition and multiplication; this means that in each case the product and sum of a pair of elements in a particular set also belong to that set. For example the set of odd integers is *not* closed under addition, since the sum of two odd integers is not odd.

ADDITION IN OUR EXAMPLES

- All the above examples contain an identity element for addition, which we refer to as the zero element and write as 0. This element has the property that adding it to another element has no effect. The zero elements in our examples are

  1. The integer 0
  2. The rational number 0
  3. The zero matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
  4. The integer 0
  5. The function $f_0 : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = 0$, $\forall\, x \in \mathbb{R}$
  6. The zero polynomial 0
  7. The congruence class $\bar{0}$ modulo 6

- In each of our sets, every element has an additive inverse or "negative". Two elements are additive inverses each other if their sum is the zero element. The fact that every element of a set has an additive inverse means that subtraction can be defined in the set.

- In all of our sets, addition is commutative, i.e. $a + b = b + a$ for all pairs a and b of elements.

MULTIPLICATION IN OUR EXAMPLES

- The multiplication is commutative in all these examples except for $M_2(\mathbb{R})$. For 2×2 matrices A and B, the products AB and BA need not be equal.

- Except for $2\mathbb{Z}$ each of these examples contains an identity element for multiplication, i.e. an element e for which $e \times a = a \times e = a$ for all elements a of the set; multiplying by e has no effect. The multiplicative identities are

  1. The integer 1
  2. The rational number 1
  3. The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
  4. No identity element for multiplication
  5. The function $f_1 : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f(x) = 1$ for all $x \in \mathbb{R}$
  6. The polynomial 1
  7. The congruence class $\bar{1}$ modulo 6

- Two elements are multiplicative inverses of each other if their product is the multiplicative identity element. In $\mathbb{Q}$, every element except 0 has a multiplicative inverse, namely its reciprocal. All the other examples contain non-zero elements without multiplicative inverses.

The seven algebraic structures mentioned in this section are all examples of *rings*.

1.2 The Axioms of a Ring

NOTE: In this section and throughout these lecture notes, please do not confuse the symbol R, which is used for a general ring, with the symbol $\mathbb{R}$ which is used for the set of real numbers.

Definition 1.2.1 *A* ring *is a non-empty set* R *equipped with two binary operations called addition* $(+)$ *and multiplication* $(\times)$*, satisfying the following properties :*

The first four are concerned with the operation that is called addition.

A1 Addition is associative.
$(r + s) + t = r + (s + t)$ for all $r, s, t \in R$.

A2 Addition is commutative. $r + s = s + r$ for all $r, s \in R$.

A3 R contains an identity element for addition, denoted by 0_R and called the *zero element* of R.
$r + 0_R = 0_R + r = r$ for all $r \in R$.

A4 Every element of R has an inverse with respect to addition. (The additive inverse of r is often denoted by $-r$).
For every $r \in R$, there exists an element $-r \in R$ for which $r + (-r) = 0_R$.

NOTE : Axioms A1 to A4 could be summarized by saying that R is an abelian group under addition. (If this remark is not helpful for you, disregard it for now).

The multiplication operation is required only to satisfy one special condition :

M1 Multiplication is associative.
$(r \times s) \times t = r \times (s \times t)$ for all $r, s, t \in R$.

The last two axioms are concerned with the manner in which the two operations must interact.

D1 $r \times (s + t) = (r \times s) + (r \times t)$ for all $r, s, t \in R$.

D2 $(r + s) \times t = (r \times t) + (s \times t)$ for all $r, s, t \in R$.
-Distributive laws for multiplication over addition.

REMARKS

1. A ring is called *commutative* if its multiplication is commutative.

2. A ring R is called *unital* or referred to as a *ring with identity* if it contains an identity element for multiplication. In this case we will denote the multiplicative identity by 1_R or just 1. We have already met one example of a ring without identity, namely the ring $2\mathbb{Z}$ of *even integers*.

3. The term "ring" was introduced by David Hilbert in the late 19th century, when he referred to a "Zahlring" or "number ring".

Our first theorem about rings is the following consequence of the ring axioms.

Theorem 1.2.2 *Let* R *be a ring. Then for all elements* r *of* R *we have*

$$0_R \times r = 0_R \text{ and } r \times 0_R = 0_R.$$

i.e. multiplying any element of R *by the zero element results in the zero element as the product.*

Proof : Let $r \in R$. We have

$$\begin{aligned}(0_R \times r) + (0_R \times r) &= (0_R + 0_R) \times r \\ &= 0_R \times r.\end{aligned}$$

Adding the additive inverse of the element $0_R \times r$ to both sides of this equation gives

$$0_R \times r = 0_R.$$

A similar argument shows that $r \times 0_R = 0_R$. □

THREE REMARKS

1. The problem of deducing the truth of a statement like Theorem 1.2.2 from the axioms of a ring might be somewhat daunting. The proof may not be too hard to follow, but could you have come up with it yourself? If you were trying to, and you didn't know where to start, there are certain observations you could make that might help. There are seven axioms for rings - which might be likely to be helpful in proving the two (left and right) statements of Theorem 1.2.2? Well, the statement is about multiplication and about the zero element. According to the ring axioms, what is special about the zero element has to do with addition not multiplication. So it might seem likely that the statement in the theorem is essentially connected to the interaction of the addition and multiplication - the two axioms that deal with that are the *distributive laws*, so maybe we should not be so surprised that these have a crucial role in the proof.

2. The next two remarks are about the philosophy of abstract algebra and the mechanisms by which the subject progresses. The definition of a ring consists of a list of technical properties, but the motivation for this definition is the ubiquity of objects having these properties, like the ones in Section 1.1. When making a definition like that of a ring (or group or vector space), the goal is to arrive at a set of axioms that exactly captures the crucial unifying properties of those objects that you wish to study. In familiar number systems like the integers, the rational numbers and the real numbers, we are all used to the fact with which Theorem 1.2.2 is concerned, namely that "multiplying by zero gives zero". The same fact is easily observed to hold in the polynomial ring $\mathbb{Q}[x]$ and in the ring of matrices $M_2(\mathbb{R})$. We might well speculate that in any ring, it is probably the case that multiplying by the zero element always results in the zero element. But before we can assume that this property holds in *every ring* and incorporate it into our mental scheme for thinking about rings we must *deduce this property as a consequence of the ring axioms*.

   If we were unable to do this, but we only wanted to study rings with the property described in Theorem 1.2.2, we could an extra axiom to our definition of a ring insisting on this "multiplication by zero" property. However the fact that this property *does* turn out to follow from the standard ring axioms means that it does not need to be included in the definition.

3. On looking at Definition 1.2.1, you may wonder why these seven axioms in particular are chosen to comprise the definition of a ring. Does it look like an arbitrary selection of rules? Why do we insist that the addition have an identity element and that every element have an inverse for addition, but where the multiplication is concerned only ask that it be associative? What happens if we add more axioms about how the multiplication should behave, or drop some of the axioms about addition? The answer is that people do these things and they lead to different areas of study within abstract algebra. Relaxing the addition axioms in various ways leads to different types of algebraic structures such as *near–rings* and *semirings*. If you drop the requirement that multiplication must be associative then you are studying *non-associative rings* – people do study all of these variants and some of them have important connections to other areas of mathematics. You can even relax the distributive laws and people do this too. However *rings* themselves as defined in Definition 1.2.1 are of paramount importance in mathematics.

   On the other hand, if you want more instead of fewer axioms, you can insist that multiplication be commutative as well as associative, then you are studying *commutative rings*. In fact much of this course will be concerned with commutative rings. If you further insist that you want an identity element for multiplication and that every (non-zero) element have an in-

verse for multiplication, then you are studying *fields*. Fields are examples of rings, and field theory itself is a vast area of mathematical activity. A crucial practice in studying abstract algebra is to be absolutely clear on the precise axioms that determine the class of objects that you are studying.

1.3 Units in Rings

As we have already mentioned, the axioms of a ring are not very restrictive concerning how the operation of multiplication should behave - all we ask is that it should be associative. We do not even insist that every ring should contain an identity element for multiplication (although incidentally some authors in ring theory do). If a ring does contain an identity element for multiplication, then we can enter a discussion about whether or not something like *division* is possible in the ring; we can try to identify pairs of elements that are related to each other in the way that a rational number is related to its reciprocal or in the way that a non-singular matrix is related to its inverse.

Definition 1.3.1 *Let* R *be a ring with identity element* 1_R *for multiplication. An element* $r \in R$ *is called a* unit *in* R *if there exists* $s \in R$ *for which*

$$r \times s = 1_R \text{ and } s \times r = 1_R.$$

In this case r *and* s *are (multiplicative) inverses of each other.*

Example 1.3.2

1. In $\mathbb{Q}$ every element except 0 is a unit; the inverse of a non-zero rational number is its reciprocal.
2. In $\mathbb{Z}$ the only units are 1 and -1 : no other integer can be multiplied by an integer to give 1.
3. In $M_2(\mathbb{R})$, the units are the 2×2 matrices with non-zero determinant, and the identity element is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
4. In $\mathbb{Z}/6\mathbb{Z}$ the only units are $\bar{1}$ and $\bar{5}$; each of these is its own inverse.
5. Question for discussion in the seminar : *what are the units in* $M_2(\mathbb{Z})$*, the ring of* 2×2 *matrices with integer entries?*

NOTATION: We will denote the set of units in a ring R with identity by $\mathcal{U}(R)$.

REMARKS

1. If R is a unital ring having two or more elements then it follows from Theorem 1.2.2 that the zero element of R and the multiplicative identity in R cannot be the same element.
2. If R has two or more elements then 0_R cannot be a unit in R, again by Theorem 1.2.2.

3. It is possible for a ring to have only one element; for example the subset of $\mathbb{Z}$ containing only 0 is a ring. (This is called the zero ring and as an example of a ring it is not very instructive)

4. 1_R is always a unit in R since it is its own inverse.

The next theorem is concerned with a special property of the subset of a ring consisting of the units. Suppose that R is a unital ring. Then from the above comments it follows that $\mathcal{U}(R)$ is a subset of R that includes the (multiplicative) identity element but not the zero element. Is $\mathcal{U}(R)$ just a set, or does it have algebraic structure of its own? The full ring R has addition and multiplication defined on it. If we take two units of R we can add them in R; will the result be a unit? If we take two units of R and multiply them (in R), will the result be a unit? If the answer to this second question is yes, then the set of units of R is itself an algebraic structure with respect to the multiplication of R, and we can study its properties.
Algebraists are always on the lookout for substructures of the objects that they are studying, which are themselves algebraic structures with respect to the operation(s) of the larger object. The general thinking behind this practice is that small things are usually easier to understand than big things, and that we have some chance of understanding (at least partically) a large complicated algebraic structure if we can identify smaller parts of it that are themselves algebraic structures.

Theorem 1.3.3 *Let R be a ring with identity element 1_R. Then $\mathcal{U}(R)$ is a group under the multiplication of R. ($\mathcal{U}(R)$ is called the unit group of R).*

Note : The statement that $\mathcal{U}(R)$ is a group under multiplication means that :

- $\mathcal{U}(R)$ is *closed* under multiplication - whenever elements a and b belong to $\mathcal{U}(R)$, so does their product ab.
- $\mathcal{U}(R)$ contains an identity element for multiplication.
- $\mathcal{U}(R)$ contains a multiplicative inverse for each of its elements.

Proof of Theorem 1.3.3: We need to show

1. $\mathcal{U}(R)$ is closed under the multiplication of R; i.e. that rs is a unit in R whenever r and s are units in R. So assume that r and s belong to $\mathcal{U}(R)$ and let r^{-1} and s^{-1} denote their respective inverses in R. Then

$$\begin{aligned}(rs)(s^{-1}r^{-1}) &= r(ss^{-1})r^{-1} \\ &= r1_R r^{-1} \\ &= rr^{-1} \\ &= 1_R.\end{aligned}$$

Similarly $(s^{-1}r^{-1})(rs) = 1_R$ and so $s^{-1}r^{-1}$ is an inverse in R for rs, and $rs \in \mathcal{U}(R)$.

2. $\mathcal{U}(R)$ contains an identity element for multiplication. This is true since $1_R \in \mathcal{U}(R)$.

3. $\mathcal{U}(R)$ contains an inverse for each of its elements.
To see this, suppose $r \in \mathcal{U}(R)$, and let r^{-1} be the inverse of r in R. Then $r^{-1}r = 1_R$ and $rr^{-1} = 1_R$, so r is the inverse of r^{-1}, and r^{-1} is in $\mathcal{U}(R)$.

This proves the theorem. □

EXAMPLES

1. $\mathcal{U}(\mathbb{Z}) = \{-1, 1\}$ is a cyclic group of order 2.

2. The unit group of the matrix ring $M_n(\mathbb{R})$ is the general linear group $GL(n, \mathbb{R})$ of $n \times n$ invertible matrices over $\mathbb{R}$.

3. The unit group of $\mathbb{Q}$ is denoted $\mathbb{Q}^\times$ and consists of all non-zero rational numbers.

QUESTION FOR DISCUSSION IN THE SEMINAR: *In general, is there anything to be said about the behaviour of* $\mathcal{U}(R)$ *with respect to addition in* R*?*

Suppose that R is a ring with identity. Then we know that the unit group of R cannot include the zero element of R, but any non-zero element of R could potentially be a unit. A particularly nice thing to happen is for *every* non-zero element of R to be a unit. Rings in which this occurs are worthy of special study.

Definition 1.3.4 *A ring with identity is called a* field *if it is commutative and every non-zero element is a unit (so we can divide by every non-zero element).*

Examples of fields include $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ and $\mathbb{Z}/5\mathbb{Z}$ (check).

A ring with identity in which every non-zero element is a unit is called a *division ring*. Commutative division rings are fields. Examples of non-commutative division rings are not easy to find, but we will see at least one in this course.

1.4 Integral Domains and Zero–Divisors

We saw in Theorem 1.2.2 that whenever an element of a ring is multiplied by zero, the result is zero. When working in the set of real numbers we often use the converse of this - a product ab can be zero in $\mathbb{R}$ only if at least one of a and b is equal to zero.
QUESTION FOR THE SEMINAR: *When/how do we use this?*

QUESTION: Is it true in every ring that the product of two elements can be zero only if at least one of the elements is zero? To think about this question, look at some examples.

Example 1.4.1 *1. In* $M_2(\mathbb{Q})$

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

i.e. the product of two non-zero matrices in $M_2(\mathbb{Q})$ *can be the zero matrix.*

2. In $\mathbb{Z}/6\mathbb{Z}$, $\bar{2} \times \bar{3} = \bar{0}$

So the answer to the question is no in general. However, it *is* of interest to study the class of rings in which the property described in the question holds.

Definition 1.4.2 *Let* R *be a ring with zero element* 0_R. *An element* a *of* R *is called a* (left) zero–divisor *in* R *if* $a \neq 0_R$ *and there exists an element* $b \neq 0_R$ *of* R *for which* $ab = 0_R$. *(In this case* b *is a right zero–divisor).*

NOTE: If R is commutative then $ab = ba$ and we just talk about zero–divisors (not left and right zero–divisors).

Definition 1.4.3 *A commutative ring with identity that contains no zero-divisors is called an* integral domain *(or just a domain).*

In an integral domain, the product of two elements can be zero only if one of the elements is zero.

EXAMPLES

1. $\mathbb{Z}$ is an integral domain. Somehow it is the "primary" example - it is from the ring of integers that the term "integral domain" is derived. The adjective "integral" in this context is related to "integer" (nothing to do with integrals in the calculus sense!).

2. Every *field* is an integral domain. For let F be a field and suppose that a, b are elements of F for which $ab = 0_F$. Assume $a \neq 0$. Then a has a multiplicative inverse in F and

$$\begin{aligned} ab &= 0_F \\ \Longrightarrow a^{-1}(ab) &= a^{-1}0_F \\ \Longrightarrow (a^{-1}a)b &= 0_F \text{ by Theorem 1.2.2} \\ \Longrightarrow 1_F b &= 0_F \\ \Longrightarrow b &= 0_F. \end{aligned}$$

   REMARK: It follows from the above argument that no unit can be a (left or right) zero-divisor in any ring.

   EXERCISE: Write down a proof of the statement of the above remark.

3. An example of a commutative ring with identity that is not an integral domain is $\mathbb{Z}/6\mathbb{Z}$ (or $\mathbb{Z}/n\mathbb{Z}$ for any composite natural number n).

QUESTIONS FOR THE SEMINAR:

1. *For which natural numbers n is $\mathbb{Z}/n\mathbb{Z}$ a field?*

2. *For which natural numbers n is $\mathbb{Z}/n\mathbb{Z}$ an integral domain?*

3. *For a natural number n, which elements of $\mathbb{Z}/n\mathbb{Z}$ are units?*

4. *Is it true for every natural number n that every non-zero element of $\mathbb{Z}/n\mathbb{Z}$ is either a unit or a zero-divisor? Can we prove this?*

5. *Suppose that R is a commutative ring with identity that is not an integral domain. Must it be true that every non-zero element of R is either a zero-divisor or a unit?*

Chapter 2

Factorization in Polynomial Rings

2.1 Polynomial Rings

If R is any ring, we can define the ring $R[x]$ of polynomials with coefficients in R. If F is a field, then the polynomial ring $F[x]$ is a particular interest. Polynomial rings over fields have some resemblance to the ring $\mathbb{Z}$ of integers in terms of their divisibility properties. Integers can sometimes be factorized in nontrivial ways and sometimes not, and every integer $\geqslant 2$ can be written in a (more or less) unique manner as a product of primes, which are "elementary components" of integers with respect to multiplication. The theme of this chapter is to explore analagous properties of polynomial rings over fields. Note that notions like factorization and prime are lost when we move from the integers to the rational numbers.
QUESTION FOR THE SEMINAR : Why is this?

Definition 2.1.1 *Let R be a ring. A* polynomial *in x with coefficients in R is an expression of the form*

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $n \geqslant 0$ is an integer and $a_i \in R$ for $i = 0, \ldots, n$.
The set of all such expressions is denoted by $R[x]$.

NOTE: The symbol x is an *indeterminate*. The expressions

$$a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \text{ and } b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0$$

are (by definition) equal in $R[x]$ if and only if $a_i = b_i$ for all $i \geqslant 0$. (Here we assign $a_j = 0$ for $j > m$ and $b_j = 0$ for $j > n$, in order for the statement " $a_i = b_i$ for all $i \geqslant 0$" to make sense.)

The set $R[x]$ is a ring under polynomial addition and multiplication, which are defined as follows. Let

$$\begin{aligned} f(x) &= a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0 \\ g(x) &= b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0 \end{aligned}$$

be elements of $R[x]$.

- The sum $f(x) + g(x)$ is the polynomial in which the constant term is $a_0 + b_0$ and the coefficient of x_i is $a_i + b_i$ for $i \geqslant 1$.
- The product $f(x)g(x)$ has constant term a_0b_0. For $i > m + n$ the coefficient of x^i is 0 and for $i \leqslant m + n$ the coefficient of x^i is

$$\sum_{j=0}^{i} a_j b_{i-j}.$$

  (For example the coefficient of x^2 is $a_0b_2 + a_1b_1 + a_2b_0$).

NOTES:

1. $R[x]$ is commutative if and only if R is commutative.
2. If R contains an identity element 1_R for multiplication, then 1_R is also an identity element for multiplication in $R[x]$.
3. Those polynomials in $R[x]$ in which the coefficient of x^i is zero whenever $i \geqslant 1$ (i.e. those in which the indeterminate x does not actually appear) are called the *constant* polynomials. They are just the elements of R.

   Of course the set of constant polynomials is itself a ring under the operations of $R[x]$ (which for the constant polynomials are just the addition and multiplication of R). We say that R is a *subring* of $R[x]$).

The remarks above show that the properties of $R[x]$ are influenced by the properties of R. We will shortly assume that R is an integral domain, and later that R is a field.

Definition 2.1.2 *Let R be a ring. The* degree *of a polynomial $f(x)$ in $R[X]$ is defined to be the maximum i for which x^i appears with non-zero coefficient in $f(x)$, if any such i exists. The degree of a non-zero constant polynomial is zero. The degree of the zero polynomial is not defined.*

So associated to every non-zero polynomial we have its *degree*, which is a non-negative integer. The next result describes how the degree behaves with regard to multiplication in polynomial rings over integral domains.

Lemma 2.1.3 *Let R be an integral domain and let $f(x)$ and $g(x)$ be non-zero elements of $R[x]$ of degrees m and n respectively. Then the polynomial $f(x)g(x)$ has degree $m + n$.*

Proof: Write

$$\begin{aligned} f(x) &= a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0,\ a_m \neq 0 \\ g(x) &= b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0,\ b_n \neq 0 \end{aligned}$$

Then the highest power of x to possibly appear in the product $f(x)g(x)$ is x^{m+n} which has coefficient $a_m b_n$. Note that this element is not zero in R since it is the product of two non-zero elements in a ring without zero-divisors. □

Corollary 2.1.4 *If R is an integral domain then $R[x]$ is also an integral domain.*

Proof: Exercise for the seminar.

Corollary 2.1.5 *Let R be an integral domain. Then the unit group of $R[x]$ is just the unit group of R.*

NOTE: This is saying that the only elements of $R[x]$ that are units in $R[x]$ are those constant polynomials which are units in R.

Proof: The identity element of $R[x]$ is the constant polynomial 1_R, which is also the identity element of R. Since $R \subset R[x]$ and $1_R \in R$, it is clear that $\mathcal{U}(R) \subseteq \mathcal{U}(R[x])$. On the other hand suppose that $f(x)$ is a non–constant polynomial in $R[x]$, so $\deg(f(x)) \geqslant 1$. If $g(x)$ is a non-zero element of $F[x]$, then

$$\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x)) \geqslant 1,$$

so $f(x)g(x) \neq 1_R$. Thus $f(x)$ has no inverse in $R[x]$ and $f(x)$ does not belong to $\mathcal{U}(R[x])$.

Example: If F is a field then $\mathcal{U}(F[x]) = F^\times$, the multiplicative group of non-zero elements of F.

QUESTION FOR THE SEMINAR: Suppose that R is not an integral domain. Then could it happen that a non-constant polynomial could be a unit in $R[x]$?

2.2 Divisibility and Irreducibility

RECALL: The division algorithm in $\mathbb{Z}$: if m is a positive integer and n is any integer, then there exist unique integers q and r (respectively called the quotient and remainder on dividing n by m) with $0 \leqslant r < m$ and

$$n = mq + r.$$

We will discuss in the seminar how the division algorithm for $\mathbb{Z}$ can be proved (although it is not very difficult to persuade yourself that it is true). In this section we will see that for a field F, the polynomial ring $F[x]$ has many properties in common with the ring $\mathbb{Z}$ of integers. The first of these is a version of the division algorithm.

Definition 2.2.1 *Let* $f(x)$, $g(x)$ *be polynomials in* $F[x]$. *We say that* $g(x)$ divides $f(x)$ *in* $F[x]$ *if* $f(x) = g(x)q(x)$ *for some* $q(x) \in F[x]$ *(i.e. if* $f(x)$ *is a multiple of* $g(x)$ *in* $F[x]$*).*

Theorem 2.2.2 (Division Algorithm in $F[x]$). *Let* F *be a field and let* $f(x)$ *and* $g(x)$ *be non-zero polynomials in* $F[x]$ *with* $g(x) \neq 0$. *respectively. Then there exist* unique *polynomials* $q(x)$ *and* $r(x)$ *in* $F[x]$ *with* $r(x) = 0$ *or* $\deg(r(x)) < \deg(g(x))$ *and*

$$f(x) = g(x)q(x) + r(x).$$

NOTES

1. In this situation $q(x)$ and $r(x)$ are called the quotient and remainder upon dividing $f(x)$ by $g(x)$.

2. There are two separate assertions to be proved - the existence of such a $q(x)$ and $r(x)$, and their uniqueness.

Proof: (Existence) Define S to be the set of all polynomials in $F[x]$ of the form $f(x) - g(x)h(x)$ where $s(x) \in F[x]$. So S is the set of all those polynomials in $F[x]$ *that differ from* $f(x)$ *by a multiple of* $g(x)$. Our goal for the existence part of the proof is show that either the zero polynomial belongs to S, or S contains some element whose degree is less than that of $g(x)$.

1. If $0 \in S$ then $f(x) - g(x)h(x) = 0$ for some $h(x) \in F[x]$, so $f(x) = g(x)h(x)$ and we can take $q(x) = h(x)$ and $r(x) = 0$.

2. If $0 \notin S$, let $r(x)$ be an element of minimal degree in S.

Let m denote the degree of $g(x)$ and write

$$g(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_1 x + a_0,\ a_m \neq 0.$$

Let $t = \deg(r(x))$ and write

$$r(x) = b_t x^t + b_{t-1}x^{t-1} + \cdots + b_1 x + b_0,\ b_t \neq 0.$$

We claim that $t < m$. We know since $r(x) \in S$ that there exists a polynomial $h(x) \in F[x]$ for which

$$r(x) = f(x) - g(x)h(x).$$

Thus

$$b_t x^t + b_{t-1}x^{t-1} + \cdots + b_1 x + b_0 = f(x) - g(x)h(x).$$

If $t \geqslant m$ then $t - m \geqslant 0$. Also $a_m \neq 0$ in F, so a_m has an inverse $1/a_m$ in F and the element b_t/a_m belongs to F. Now subtract the polynomial $g(x)(b_t/a_m)x^{t-m}$ (which has leading term $b_t x^t$) from both sides of the above equation to get

$$b_t x^t + \cdots + b_1 x + b_0 - g(x)(b_t/a_m)x^{t-m} = f(x) - g(x)h(x) - g(x)(b_t/a_m)x^{t-m}.$$

The left side of the above equation is $r_1(x)$, a polynomial of degree less than t in $F[x]$. The right hand side is $f(x) - g(x)h_1(x)$ where $h_1(x) = h(x) + (b_t/a_m)x^{t-m}$. Thus $r_1(x)$ belongs to S, contrary to the choice of $r(x)$ as an element of minimal degree in S. We conclude that $t < m$ and

$$f(x) = g(x)h(x) + r(x)$$

is a description of $f(x)$ of the required type. This proves the existence.

QUESTIONS FOR THE SEMINAR:

1. How do we know that $r_1(x)$ above has degree less than t?
2. Why can we conclude that $t < m$ at the third last line above?
3. Where does the proof use the fact that F is a field?

Uniqueness: Suppose that

$$\begin{aligned} f(x) &= g(x)q_1(x) + r_1(x),\ \deg(r_1(x)) < m \\ \text{and } f(x) &= g(x)q_2(x) + r_2(x),\ \deg(r_2(x)) < m. \end{aligned}$$

Then

$$0 = g(x)(q_1(x) - q_2(x)) + (r_1(x) - r_2(x)) \Longrightarrow g(x)(q_1(x) - q_2(x)) = r_2(x) - r_1(x).$$

Now $g(x)(q_1(x) - q_2(x))$ is either zero or a polynomial of degree at least m, and $r_2(x) - r_1(x)$ is either zero or a polynomial of degree less than m. Hence these two can be equal only if they are both zero, which means $q_1(x) = q_2(x)$ and $r_1(x) = r_2(x)$. This completes the proof. □

QUESTION FOR THE SEMINAR: Why can we say that if $g(x)(q_1(x) - q_2(x)) = 0$ then it must follow that $q_1(x) = q_2(x)$?

Let $f(x) \in R[x]$ for some ring R; suppose

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

If $\alpha \in R$ then we let $f(\alpha)$ denote the element

$$a_n \alpha^n + a_{n-1} \alpha^{n-1} + \cdots + a_1 \alpha + a_0$$

of R. Thus associated to the polynomial $f(x)$ we have a function from R to R sending α to $f(\alpha)$. Forming the element $f(\alpha)$ is called *evaluating* the polynomial $f(x)$ at α.

Definition 2.2.3 *In the above context,* $\alpha \in R$ *is a* root *of* $f(x)$ *if* $f(\alpha) = 0$.

Theorem 2.2.4 *(The Factor Theorem) Let* $f(x)$ *be a polynomial of degree* $n \geqslant 1$ *in* $F[x]$ *and let* $\alpha \in F$. *Then* α *is a root of* $f(x)$ *if and only if* $x - \alpha$ *divides* $f(x)$ *in* $F[x]$.

Proof: By the division algorithm (Theorem 2.2.2), we can write

$$f(x) = q(x)(x - \alpha) + r(x),$$

where $q(x) \in F[x]$ and either $r(x) = 0$ or $r(x)$ has degree zero and is thus a non-zero element of F. So $r(x) \in F$; we can write $r(x) = \beta$. Now

$$\begin{aligned} f(\alpha) &= q(\alpha)(\alpha - \alpha) + \beta \\ &= 0 + \beta \\ &= \beta. \end{aligned}$$

Thus $f(\alpha) = 0$ if and only if $\beta = 0$, i.e. if and only if $r(x) = 0$ and $f(x) = q(x)(x-\alpha)$ which means $x - \alpha$ divides $f(x)$. □

QUESTION FOR THE SEMINAR:
This actually proves more than the statement of the theorem - explain.

Now that we have some language for discussing divisibility in polynomial rings, we can also think about factorization. In $\mathbb{Z}$, we are used to calling an integer *prime* if it does not have any interesting factorizations. In polynomial rings, we call a polynomial *irreducible* if it does not have any interesting factorizations.

QUESTION FOR THE SEMINAR:
What does "interesting" mean in this context?

Definition 2.2.5 *Let* F *be a field and let* $f(x)$ *be a non-constant polynomial in* $F[x]$*. Then* $f(x)$ *is* irreducible *in* $F[x]$ *(or irreducible over* F*) if* $f(x)$ *cannot be expressed as the product of two factors both of degree at least 1 in* $F[x]$*. Otherwise* $f(x)$ *is* reducible *over* F*.*

NOTES:

1. Any polynomial $f(x) \in F[x]$ can be factorized (in an uninteresting way) by choosing $a \in F^\times$ and writing

$$f(x) = a(a^{-1}f(x).$$

   This is not considered to be a proper factorization of $f(x)$.

2. Every polynomial of degree 1 is irreducible.

3. It is possible for a polynomial that is irreducible over a particular field to be reducible over a larger field. For example $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$. However it is not irreducible in $\mathbb{R}[x]$, since here $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$. Therefore when discussing irreducibility, it is important to specify what field we are talking about (sometimes this is clear from the context).

4. The only irreducible polynomials in $\mathbb{C}[x]$ are the linear (i.e. degree 1) polynomials. This is basically the Fundamental Theorem of Algebra, which states that every non-constant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.

Let $f(x)$ be a polynomial of degree $\geqslant 2$ in $F[x]$. If $f(x)$ has a root α in F then $f(x)$ is not irreducible in $F[x]$ since it has $x - \alpha$ as a proper factor. This statement has a partial converse.

Theorem 2.2.6 *Let* $f(x)$ *be a quadratic or cubic polynomial in* $f(x)$*. Then* $f(x)$ *is irreducible in* $F[x]$ *if and only if* $f(x)$ *has no root in* F*.*

Proof: Since $f(x)$ is quadratic or cubic any proper factorization of $f(x)$ in $F[x]$ involves at least one linear (i.e. degree 1) factor. Suppose that $r(x) = ax + b$ is a linear factor of $f(x)$ in $F[x]$. Then we have $f(x) = r(x)g(x)$ for some $g(x)$ in $F[x]$. Since F is a field we can rewrite this as

$$f(x) = (x + b/a)(ag(x)).$$

Thus $x - (-b/a)$ divides $f(x)$ in $F[x]$ and by Theorem 2.2.4 $-b/a$ is a root of $f(x)$ in F. □

QUESTION FOR THE SEMINAR: Theorem 2.2.6 certainly does not hold for polynomials of degree 4 or higher. That is, for a polynomial of degree 4 or more, having no roots in a particular field does not mean being irreducible over that field. Give an example to demonstrate this.

In general, deciding whether a given polynomial is reducible over a field or not is a difficult problem. We will look at this problem in the case where the field of coefficients is $\mathbb{Q}$. The problem of deciding reducibility in $\mathbb{Q}[x]$ is basically the same as that of deciding reducibility in $\mathbb{Z}[x]$, as the following discussion will show.

Lemma 2.2.7 *For a field* F, *let* $a \in F^\times$ *and let* $f(x) \in F[x]$. *Then* $f(x)$ *is reducible in* $F[x]$ *if and only if* $af(x)$ *is reducible in* $F[x]$.

Proof: Exercise for the seminar.

Note that any polynomial in $\mathbb{Q}[x]$ can be multiplied by a non-zero integer to produce a polynomial in $\mathbb{Z}[x]$. Then by Lemma 2.2.7 the problem of deciding reducibility in $\mathbb{Q}[x]$ is the same as that of deciding reducibility over $\mathbb{Q}$ for polynomials in $\mathbb{Z}[x]$.
Suppose that $f(x)$ is a polynomial with coefficients in $\mathbb{Z}$. Surprisingly, $f(x)$ has a proper factorization with factors in $\mathbb{Q}[x]$ if and only if $f(x)$ has a proper factorization with factors (of the same degree) that belong to $\mathbb{Z}[x]$. This fact is a consequence of Gauss's lemma which is discussed below. It means that a polynomial with integer coefficients is irreducible over $\mathbb{Q}$ provided that it is irreducible over $\mathbb{Z}$. This is good news because irreducibility over $\mathbb{Z}$ is in principle easier to decide.

QUESTION FOR THE SEMINAR: Why is irreducibility over $\mathbb{Z}$ is in principle easier to decide than irreducibility over $\mathbb{Q}$, for a polynomial with integer coefficients?

Definition 2.2.8 *A polynomial in* $\mathbb{Z}[x]$ *is called* primitive *if the greatest common divisor of all its coefficients is 1.*

EXAMPLE
$3x^4 + 6x^2 - 2x - 2$ is primitive.
$3x^4 + 6x^2 = 18x$ is not primitive, since 3 divides each of the coefficients.

Theorem 2.2.9 *(Gauss's Lemma) : Let* $f(x)$ *and* $g(x)$ *be primitive polynomials in* $\mathbb{Z}[x]$. *Then their product is again primitive.*

Proof: We need to show that no prime divides all the coefficients of $f(x)g(x)$. We can write

$$\begin{aligned} f(x) &= a_s x^s + a_{s-1}x^{s-1} + \cdots + a_1 x + a_0,\ a_s \neq 0, \\ f(x) &= b_t x^t + b_{t-1}x^{t-1} + \cdots + b_1 x + b_0,\ b_t \neq 0. \end{aligned}$$

Let p be a prime. Since $f(x)$ and $g(x)$ are primitive we can choose k and m to be the least integers for which p does not divide a_k and p does not divide b_m. Now look at the coefficient of x^{k+m} in $f(x)g(x)$. This is

$$a_{k+m}b_0 + \cdots + a_{k+1}b_{m-1} + a_k b_m + a_{k-1}b_{m+1} + \cdots + a_0 b_{k+m}.$$

Since $p|b_i$ for $i < m$ and $p|a_i$ for $i < k$, every term in the above expression is a multiple of p except for $a_k b_m$ which is definitely not. Thus p does not divide the coefficient of x^{k+m} in $f(x)g(x)$, p does not divide all the coefficients in $f(x)g(x)$ and $f(x)g(x)$ is primitive. □

Corollary 2.2.10 *Suppose $f(x)$ is a polynomial of degree $\geqslant 2$ in $\mathbb{Z}[x]$. Then $f(x)$ has a proper factorization in $\mathbb{Q}[x]$ if and only if it has a proper factorization in $\mathbb{Z}[x]$, with factors of the same degrees.*

This means : if $f(x)$ can be properly factorized in $\mathbb{Q}[x]$ it can also be properly factorized in $\mathbb{Z}[x]$; if it can be written as the product of two polynomials of degree $\geqslant 1$ with rational coefficients, it can be written as the product of two such polynomials with *integer* coefficients.
Proof: $\Longleftarrow$: This direction is obvious, since any factorization in $\mathbb{Z}[x]$ is a factorization in $\mathbb{Q}[x]$.
$\Longrightarrow$: First assume that $f(x)$ is primitive in $\mathbb{Z}[x]$.
Suppose that $f(x) = g_1(x)h_1(x)$ where $g_1(x)$ and $h_1(x)$ are polynomials of degree $k \geqslant 1$ and $m \geqslant 1$ in $\mathbb{Q}[x]$. Then we can find integers a_1 and b_1 for which $a_1 g_1(x)$ and $b_1 h_1(x)$ are elements of $\mathbb{Z}[x]$, both of degree at least 1. Let d_1 and d_2 denote the greatest common divisors of the coefficients in $a_1 g_1(x)$ and $b_1 h_1(x)$ respectively. Then $(a_1/d_1)g_1(x)$ and $(b_1/d_2)h_1(x)$ are primitive polynomials in $\mathbb{Z}[x]$. Call these polynomials $g(x)$ and $h(x)$ respectively, and let a and b denote the rational numbers a_1/d_1 and b_1/d_2. Now

$$f(x) = g_1(x)h_1(x) \Longrightarrow abf(x) = ag_1(x)bh_1(x) = g(x)h(x).$$

Since $g(x)h(x) \in \mathbb{Z}[x]$ and $f(x)$ is primitive it follows that ab is an integer. Furthermore since $g(x)h(x)$ is primitive by Theorem 2.2.9, $abf(x)$ is primitive. This means $ab = 1$ or -1. Now either $ab = 1$ and $f(x) = g(x)h(x)$ or $ab = -1$ and $f(x) = (-g(x))h(x)$. Thus $f(x)$ factorizes in $\mathbb{Z}[x]$.
Finally, if $f(x)$ is not primitive we can write $f(x) = df_1(x)$ where d is the gcd of the coefficients in $f(x)$ and $f_1(x)$ is primitive. By Lemma 2.2.7 $f(x)$ is irreducible in $\mathbb{Q}[x]$ if and only if $f_1(x)$ is. By the above, $f_1(x)$ factorizes in $\mathbb{Q}[x]$ if and only if it factorizes in $\mathbb{Z}[x]$. Finally, $f(x)$ clearly factorizes in $\mathbb{Z}[x]$ if $f_1[x]$ does. □

Theorem 2.2.9 and Corollary 2.2.10 make the reducibility question in $\mathbb{Q}[x]$ much easier.

Theorem 2.2.11 *Let $f(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial of degree $n \geqslant 2$ in $\mathbb{Z}[x]$, with $a_0 \neq 0$. If $f(x)$ has a root in $\mathbb{Q}$ this root has the form b/a where a and b are integers (positive or negative) for which $b|a_0$ and $a|a_n$.*

Proof: By Theorem 2.2.4, $f(x)$ has a root in $\mathbb{Q}$ only if $f(x)$ has a linear factor in $\mathbb{Q}[x]$. By Corollary 2.2.10 this happens only if

$$f(x) = (ax + b)(g(x))$$

where $a, b \in \mathbb{Z},\ a \neq 0$ and $g(x) \in \mathbb{Z}[x]$. Then if

$$g(x) = c_{n-1}x^{n-1} + \cdots + c_1x + c_0,$$

we have $ac_{n-1} = a_n$ and $b_0c_0 = a_0$. Thus $a|a_n,\ b|a_0$ and $-b/a$ is a root of $f(x)$ in $\mathbb{Q}$. □

Example: Let $f(x) = \frac{3}{5}x^3 + 2x - 1$ in $\mathbb{Q}[x]$. Determine if $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Solution: By Lemma 2.2.7 $f(x)$ is irreducible in $\mathbb{Q}[x]$ if and only if $5f(x) = 3x^3 + 10x - 5$ is irreducible. By Theorem 2.2.6 this would mean having no root in $\mathbb{Q}$. By Theorem 2.2.11 possible roots of $5f(x)$ in $\mathbb{Q}$ are

$$1, -1, 5, -5, \frac{1}{3}, -\frac{1}{3}, \frac{5}{3}, -\frac{5}{3}.$$

It is easily checked that none of these is a root. Since $f(x)$ is cubic it follows that $f(x)$ is irreducible in $\mathbb{Q}[x]$.

NOTE: A polynomial is called *monic* if its leading coefficient is 1. If $f(x)$ is a monic polynomial in $\mathbb{Z}[x]$ then any rational roots of $f(x)$ are integer divisors of the constant term (provided that this is not zero).

EXAMPLE: Decide if the polynomial $f(x) = x^5+3x^4-3x^3-8x^2+3x-2$ is irreducible in $\mathbb{Q}[x]$.

Solution : Possible rational roots of $f(x)$ are integer divisors of the constant term -2 - i.e. $1, -1, 2, -2$. Inspection of these possibilities reveals that -2 is a root. Thus $f(x)$ is reducible in $\mathbb{Q}[x]$.

NOTE: Since $f(x)$ has degree 5, a discovery that $f(x)$ had no rational roots would not have told us anything about the irreducibility or not of $f(x)$ over $\mathbb{Q}$.
There is one known criterion for irreducibility over $\mathbb{Q}$ that applies to polynomials of high degree, but it only applies to polynomials with a special property.

Theorem 2.2.12 *(The Eisenstein irreducibility Criterion) Let* $f(x) = a_nx^n+\cdots+a_1x+a_0$ *be a polynomial in* $\mathbb{Z}[x]$ *where* $a_n \neq 0$, *and* $n \geqslant 2$. *Suppose that there exists a prime number* p *for which*

- p *divides all of* $a_0, a_1, \ldots, a_{n-1}$
- p *does not divide* a_n
- p^2 *does not divide* a_0.

Then $f(x)$ *is irreducible in* $\mathbb{Q}[x]$.

For example the Eisenstein test says that $2x^4 - 3x^3 + 6x^2 - 12x + 3$ is irreducible in $\mathbb{Q}[x]$ since the prime 3 divides all the coefficients except the leading one, and 9 does not divide the constant term.

Proof of Theorem 2.2.12: Assume (in the hope of contradiction) that $f(x)$ is reducible and write

$$f(x) = (\underbrace{b_s x^s + \cdots + b_1 x + b_0}_{g(x)})(\underbrace{c_t x^t + \cdots + c_1 x + c_0}_{h(x)})$$

where $g(x), h(x) \in \mathbb{Z}[x]$, $b_s \neq 0$, $c_t \neq 0$, $s \geqslant 1$, $t \geqslant 1$ and $s + t = n$.
Now $b_0 c_0 = a_0$ which means p divides exactly one of b_0 and c_0, as p^2 does not divide a_0. Suppose $p|b_0$ and $p \not| c_0$. Now $a_1 = b_1 c_0 + b_0 c_1$, which means $p|b_1$ since p divides a_1 and b_0 but not c_0. Similarly looking at a_2 shows that p must divide b_2. However p does not divide all the b_i - it does not divide b_s, otherwise it would divide $a_n = b_s c_t$.
Now let k be the least for which $p \not| b_k$. Then $k \leqslant s \Longrightarrow k < n$ and

$$a_k = b_k c_0 + \underbrace{b_{k-1} c_1 + \cdots + b_0 c_k}_{\text{all multiplesof } p}$$

Now $p \not| b_k c_0$ since $p \not| b_k$ and $p \not| c_0$. Since the remaining terms in the above description of a_k are all multiples of p, it follows that $p \not| a_k$, contrary to hypothesis. We conclude that any polynomial in $\mathbb{Z}[x]$ satisfying the hypotheses of the theorem is irreducible in $\mathbb{Q}[x]$. □

NOTE: Theorem 2.2.12 says nothing at all about polynomials in $\mathbb{Z}[x]$ for which no prime satisfies the requirements in the statement.

Chapter 3

Ideals, Homomorphisms and Factor Rings

3.1 Ring Homomorphisms and Ideals

In this section we develop some more of the abstract theory of rings. In particular we will describe those functions between rings that preserve the ring structure, and we will look at another way of forming new rings from existing ones.

Definition 3.1.1 *Let* R *be a ring. A non-empty subset* S *of* R *is a* subring *of* R *if it is itself a ring under the addition and multiplication of* R.

This means that S is closed under the addition and multiplication of R, that it contains the zero element of R, and that it contains the negative of each of its elements.

EXAMPLES

1. $\mathbb{Z}$ is a subring of $\mathbb{Q}$.
   $\mathbb{Q}$ is a subring of $\mathbb{R}$.
   $\mathbb{R}$ is a subring of $\mathbb{C}$.

2. The ring $M_n(F)$ of $n \times n$ matrices over a field F has the following subrings :

   - $D_n(F)$ - the ring of *diagonal* $n \times n$ matrices over F.
   - $U_n(F)$ - the ring of *upper triangular* $n \times n$ matrices over F.

3. For any field F, F is a subring of the polynomial ring $M_n(F)$. So also is $F[x^2]$, the subset of $F[x]$ consisting of those polynomials in which the coefficient of x^i is zero whenever i is odd.

4. Every (non-zero) ring R has at least two subrings - the full ring R and the zero subring $\{0_R\}$

QUESTIONS FOR THE SEMINAR:

1. Give two more examples of subrings of $M_n(\mathbb{Q})$.

2. Suppose that S is a subring of a ring R. Is it possible that S could have an identity element for multiplication that is different from the identity element of R?
   Could this happen if R is an integral domain?

Definition 3.1.2 *Let* R *and* S *be rings. A function* $\phi : R \longrightarrow S$ *is a* ring homomorphism *if for all* $x, y \in R$ *we have*

$$\phi(x + y) = \phi(x) + \phi(y)$$

and

$$\phi(xy) = \phi(x)\phi(y).$$

EXAMPLES

1. Choose a positive integer n and define $\phi_n : \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}$ to be the function that sends $k \in \mathbb{Z}$ to the congruence class modulo n to which k belongs. Then ϕ_n is a ring homomorphism.

2. Let F be a field. If $a \in F$ we can define a homomorphism

   $$\phi_a : F[x] \longrightarrow F$$

   given by $\phi_a(f(x)) = f(a)$ for $f(x) \in F[x]$.

QUESTION FOR THE SEMINAR: Determine whether each of the following is a ring homomorphism :

1. The function det : $M_2(\mathbb{Q}) \longrightarrow \mathbb{Q}$ that associates to every matrix its determinant.

2. The function $g : \mathbb{Z} \longrightarrow \mathbb{Z}$ defined by $g(n) = 2n$, for $n \in \mathbb{Z}$.

3. The function $\phi : \mathbb{Q}[x] \longrightarrow \mathbb{Q}$ defined for $f(x) \in \mathbb{Q}[x]$ by

   $$\phi(f(x)) = \text{the sum of the coefficients of } f(x).$$

Definition 3.1.3 *Suppose that* $\phi : R \longrightarrow S$ *is a homomorphism of rings. The* kernel *of* ϕ *is the subset of* R *defined by*

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

The image *of* ϕ *is the subset of* S *defined by*

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

Lemma 3.1.4 $\mathrm{Im}\phi$ *is a subring of* S.

Proof: First we need to show that $\mathrm{Im}\phi$ is closed under the addition and multiplication of S. So suppose that $s_1,\ s_2$ are elements of $\mathrm{Im}\phi$ and let $r_1,\ r_2$ be elements of R for which $s_1 = \phi(r_1)$ and $s_2 = \phi(r_2)$. Then

$$\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2) = s_1 + s_2$$

and so $s_1 + s_2 \in \mathrm{Im}\phi$. Also

$$\phi(r_1 r_2) = \phi(r_1)\phi(r_2) = s_1 s_2$$

and so $s_1 s_2 \in \mathrm{Im}\phi$.
Next we show that $0_S \in \mathrm{Im}\phi$. To see this observe that

$$\phi(0_R) + \phi(0_r) = \phi(0_R + 0_R) = \phi(0_R).$$

Subtracting the element $\phi(0_R)$ of S from both sides gives

$$\phi(0_R) = 0_S.$$

Thus $0_S \in \mathrm{Im}\phi$ - in fact we have proved something more than this, namely that 0_S is the image of 0_R.
Finally we show that $\mathrm{Im}\phi$ contains the additive inverse in S of each of its elements. Let $s \in \mathrm{Im}\phi$ and let r be an element of R for which $\phi(r) = s$. Then

$$\phi(-r) + \phi(r) = \phi(0_R) = 0_S.$$

Thus $\phi(-r)$ is the additive inverse of s in S, i.e. $-s = \phi(-r)$ and $\mathrm{Im}\phi$ contains the negative of each of its elements. □

Lemma 3.1.5 $\ker\phi$ *is a subring of* R.

Proof: Let $r_1, r_2 \in \ker\phi$. Then $\phi(r_1) = \phi(r_2) = 0_S$. We have

$$\begin{aligned} \phi(r_1 + r_2) &= \phi(r_1) + \phi(r_2) = 0_S + 0_S = 0_S, \\ \text{and } \phi(r_1 r_2) &= \phi(r_1)\phi(r_2) = 0_S 0_S = 0_S. \end{aligned}$$

Thus $\ker\phi$ is closed under addition and multiplication in R.
To see that $0_R \in \ker\phi$ we note that $\phi(0_R) = 0_S$ by the proof of Lemma 3.1.4 above.
Finally if $r \in \ker\phi$ then

$$0_S = \phi(-r + r) = \phi(-r) + \phi(r) = \phi(-r) + 0_S$$

and so $\phi(-r) = 0$ and $-r \in \ker\phi$. Thus $\ker\phi$ is a subring of R. □

In fact $\ker\phi$ is not just a subring of R - it has an extra property. Suppose $r \in \ker\phi$ and let x be any element of R. Then xr and rx belong to $\ker\phi$, since

$$\begin{aligned} \phi(xr) &= \phi(x)\phi(r) = \phi(x)0_S = 0_S, \\ \phi(rx) &= \phi(r)\phi(x) = 0_S\phi(x) = 0_S. \end{aligned}$$

So not only is $\ker\phi$ closed under its own multiplication, it is also closed under the operation of multiplying an element of $\ker\phi$ by any element of R.

Definition 3.1.6 *Let* R *be a ring.*
A left ideal *of* R *is a subring* I_L *of* R *with the additional property that* $xa \in I_L$ *whenever* $a \in I_L$ *and* $x \in R$.
A right ideal *of* R *is a subring* I_R *of* R *with the additional property that* $ax \in I_R$ *whenever* $a \in I_R$ *and* $x \in R$.
A two-sided ideal *of* R *is a subring* I *of* R *with the additional property that both* xa *and* ax *are in* I *whenever* $a \in I$ *and* $x \in R$.

QUESTION FOR THE SEMINAR: Find some examples of left, right, or two-sided ideals in each of the following rings :

$$\mathbb{Z},\ \mathbb{Q},\ \mathbb{Q}[x],\ \mathbb{Z}[x],\ M_2(\mathbb{Q}).$$

NOTES

1. If R is commutative then every left or right ideal of R is a two-sided ideal. We do not talk about two-sided ideals in this case, just ideals.

2. (Two-sided) ideals play a role in ring theory similar to that played by normal subgroups in group theory.

EXAMPLES

1. Let R be a ring. We have already seen that the kernel of any ring homomorphism with domain R is a (two-sided) ideal of R.

2. The subrings

$$\begin{aligned} 2\mathbb{Z} &= \{\ldots, -2, 0, 2, 4, \ldots\} \\ 3\mathbb{Z} &= \{\ldots, -3, 0, 3, 6, \ldots\} \end{aligned}$$

   are ideals of $\mathbb{Z}$. In general if $n \in \mathbb{Z}$ we will denote by $n\mathbb{Z}$ or $\langle n \rangle$ the subring of $\mathbb{Z}$ consisting of all the integer multiples of n. In each case $\langle n \rangle$ is an ideal of $\mathbb{Z}$, since a multiple of n can be multiplied by *any* integer and the result is always a multiple of n.

   Note that $\langle n \rangle$ is the kernel of the homomorphism $\phi_n : \mathbb{Z} \longrightarrow \mathbb{Z}/n\mathbb{Z}$ that sends $k \in \mathbb{Z}$ to the class of k modulo n.

3. Fix a polynomial $f(x) \in \mathbb{Q}[x]$. We denote by $\langle f(x) \rangle$ the subring of $\mathbb{Q}[x]$ consisting of all those polynomials of the form $g(x)f(x)$ for an element $g(x)$ of $\mathbb{Q}[x]$. Then $\langle f(x) \rangle$ is an ideal of $\mathbb{Q}[x]$, called the principal ideal generated by $f(x)$.

4. Let R be any ring and let $a \in R$. We define

$$Ra = \{ra : r \in R\}.$$

   Then Ra is a left ideal of R called the principal left ideal generated by a. Similarly $aR = \{ar : r \in R\}$ is the principal right ideal generated by a.
   If R is commutative then $aR = Ra$ for all $a \in R$, and this ideal is called the *principal ideal* generated by a. It is denoted by $\langle a \rangle$. In $\mathbb{Z}$, $n\mathbb{Z}$ is the principal ideal generated by n.
   In general an ideal in a commutative ring is called *principal* if it is the principal ideal generated by some element.

5. Every non-zero ring R has at least two ideals, namely the full ring R and the zero ideal $\{0_R\}$.

Lemma 3.1.7 *Let R be a ring, and let I be an ideal of R. If I contains a unit u of R, then $I = R$.*

Proof: Let u^{-1} denote the inverse of u in R. Then $u \in I$ implies $u^{-1}u = 1_R$ belongs to I. Now let $r \in R$. Then $r1_R = r$ belongs to I, so $R \subseteq I$ and $R = I$. □

Corollary 3.1.8 *If F is a field, then the only ideals in F are the zero ideal (consisting only of the zero element) and F itself.*

3.2 Principal Ideal Domains

Definition 3.2.1 *A* principal ideal domain *(PID) is an integral domain in which every ideal is principal.*

Lemma 3.2.2 *$\mathbb{Z}$ is a PID.*

NOTE: Showing that $\mathbb{Z}$ is a PID means showing that if I is an ideal of $\mathbb{Z}$, then there is some integer n for which I consists of all the integer multiples of n.

Proof: Suppose that $I \subseteq \mathbb{Z}$ is an ideal. If $I = \{0\}$ then I is the principal ideal generated by 0 and I is principal. If $I \neq \{0\}$ then I contains both positive and negative elements. Let m be the least positive element of I. We will show that $I = \langle m \rangle$.
Certainly $\langle m \rangle \subseteq I$ as I must contain all integer mulitples of m. On the other hand suppose $a \in I$. Then we can write

$$a = mq + r$$

where $q \in \mathbb{Z}$ and $0 \leqslant r < m$. Then $r = a - qm$. Since $a \in I$ and $-qm \in I$, this means $r \in I$. It follows that $r = 0$, otherwise we have a contradiction to the choice of m. Thus $a = qm$ and $a \in \langle m \rangle$. We conclude $I = \langle m \rangle$. □

Note: In fact every subring of $\mathbb{Z}$ is an ideal - think about this.

Lemma 3.2.3 *Let F be a field. Then the polynomial ring $F[x]$ is a PID.*

NOTE: Recall that $F[x]$ has one important property in common with $\mathbb{Z}$, namely a division algorithm. This is the key to showing that $F[x]$ is a PID.
Proof: Let $I \subseteq F[x]$ be an ideal. If $I = \{0\}$ then $I = \langle 0 \rangle$ and I is principal. If $I \neq \{0\}$, let $f(x)$ be a polynomial of minimal degree m in I. Then $\langle f(x) \rangle \subseteq I$ since every polynomial multiple of $f(x)$ is in I.
We will show that $I = \langle f(x) \rangle$. To see this suppose $g(x) \in I$. Then

$$g(x) = f(x)q(x) + r(x)$$

where $q(x), r(x) \in F[x]$ and $r(x) = 0$ or $\deg(r(x)) < m$. Now

$$r(x) = g(x) - f(x)q(x)$$

and so $r(x) \in I$. It follows that $r(x) = 0$ otherwise $r(x)$ is a polynomial in I of degree strictly less than m, contrary to the choice of $f(x)$.
Thus $g(x) = f(x)q(x)$, $g(x) \in \langle f(x) \rangle$ and $I = \langle f(x) \rangle$. □

QUESTION FOR THE SEMINAR: If R is a ring (not a field) it is not always true that $R[x]$ is a PID.
Find an example of a non-principal ideal in $\mathbb{Z}[x]$.

3.3 Factor Rings

Suppose that R is a ring and that I is a (two-sided) ideal of R. Then we can use R and I to create a new ring, called "the factor ring of R modulo I". This ring is denoted R/I (read "R mod I"), and its elements are certain subsets of R associated to I. The most well known examples are the rings $\mathbb{Z}/n\mathbb{Z}$, created from the ring $\mathbb{Z}$ of integers and its ideals.

Definition 3.3.1 *Let R be a ring and let I be a (two-sided) ideal of R. If $a \in R$, the* coset *of I in R determined by a is defined by*

$$a + I = \{a + r : r \in I\}.$$

Thus $a + I$ is a subset of R; it consists of all those elements of R that differ from a by an element of I. Note that $a + I$ does not generally have algebraic structure in its own right, it is typically not closed under the addition or multiplication of R. We will show that the set of cosets of I in R is itself a ring, with addition and multiplication defined in terms of the operations of R.

NOTES

1. $a + I$ is a coset of the subgroup $(I, +)$ of the additive group of R.

2. Suppose $R = \mathbb{Z}$ and $I = \langle 5 \rangle = 5\mathbb{Z}$. Then

$$2 + I = \{2 + 5n,\ n \in \mathbb{Z}\} = \{\ldots, -3, 2, 7, 12, \ldots\}.$$

   This is the congruence class of 2 modulo 5. So in $\mathbb{Z}$, the cosets of $n\mathbb{Z}$ in Z are the congruence classes modulo n - there is a finite number n of them and each has exactly one representative in the range $0, \ldots, n-1$ (this is guaranteed by the division algorithm in $\mathbb{Z}$).

3. Let F be a field and let I be an ideal in $F[x]$. Then $I = \langle f(x) \rangle$ for some polynomial $f(x)$, by Lemma 3.2.3. If $g(x) \in F[x]$ then the coset $g(x) + I$ contains all those polynomials that differ from $g(x)$ by a multiple of $f(x)$.

   If F is infinite then the number of cosets of I in $F[x]$ is infinite but each has exactly one representative of degree less than that of $f(x)$.

   QUESTION FOR THE SEMINAR: Why is this?

   If F is finite (e.g. $F = \mathbb{Z}/p\mathbb{Z}$ for some prime p), then the number of cosets of I in $F[x]$is finite.

Lemma 3.3.2 *Let a and b be elements of a ring R in which I is a two-sided ideal. Then*

(i) If $a - b \in I$, $a + I = b + I$.

(ii) If $a - b \notin I$, *the cosets* $a + I$ *and* $b + I$ *are disjoint subsets of* R.

Proof: (i): Suppose $a - b \in I$ and let $x \in a + I$. Then $x = a + m$ for some $m \in I$ and we can write

$$x = a - b + b + m = b + (a - b) + m.$$

Since $a - b \in I$ and $m \in I$ this means $(a - b) + m \in I$ and so $x \in b + I$. Thus $a + I \subseteq b + I$.
Now $a - b$ belongs to I and so $b - a = -(a - b)$ does also. It then follows from the above argument that $b + I \subseteq a + I$. Thus $a + I = b + I$.

(ii) Suppose $a - b \notin I$ and let $c \in (a + I) \cap (b + I)$. Then

$$c = a + m_1 = b + m_2$$

where $m_1, m_2 \in I$. It follows that $a - b = m_2 - m_1$ which is a contradiction since $a - b \notin I$. □

Lemma 3.3.2 shows that the different cosets of I in R are disjoint subsets of R. We note that their union is all of R since every element a of R belongs to *some* coset of I in R : $a \in a + I$. The set of cosets of I in R is denoted R/I. We can define addition and multiplication in R/I as follows.
Let $a + I,\ b + I$ be cosets of I in R. We define their *sum* by

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

Claim: This addition is well-defined.

QUESTION FOR THE SEMINAR: What is this claim saying? Why is there doubt about the definition of addition given above?

What the claim is concerned with is the following : if $a + I = a_1 + I$ and $b + I = b_1 + I$, how do we know that $(a + b) + I = (a_1 + b_1) + I$? How do we know that the coset sum $(a + I) + (b + I)$ as defined above does not depend on the choice a and b of representatives of these cosets to be added in R?

PROOF OF CLAIM: Suppose

$$a + I = a_1 + I \text{ and } b + I = b_1 + I$$

for elements a_1, b_1 of R. Then $a - a_1 \in I$ and $b - b_1 \in I$, by Lemma 3.3.2. Hence $(a - a_1) + (b - b_1) = (a + b) - (a_1 + b_1)$ belongs to I. Thus

$$(a + b) + I = (a_1 + b_1) + I,$$

by Lemma 3.3.2 again.

Multiplication in R/I is defined by

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

for cosets $a + I$ and $b + I$ of I in R.

Claim: Multiplication is well-defined in R/I
(i.e. the coset $ab + I$ does not depend on the choice of representatives of $a + I$ and $b + I$).

PROOF OF CLAIM: Suppose that

$$a + I = a_1 + I \text{ and } b + I = b_1 + I$$

for elements a_1, b_1 of R. Then $a - a_1 \in I$ and $b - b_1 \in I$, by Lemma 3.3.2. We need to show that

$$ab + I = a_1b_1 + I.$$

By Lemma 3.3.2, this means showing that $ab - a_1b_1 \in I$. To see this observe that

$$\begin{aligned} ab - a_1b_1 &= ab - a_1b + a_1b - a_1b_1 \\ &= (a - a_1)b + a_1(b - b_1). \end{aligned}$$

Now since I is a two-sided ideal we know that $(a - a_1)b \in I$ and $a(b - b_1) \in I$. Thus

$$(a - a_1)b + a_1(b - b_1) = ab - a_1b_1 \in I,$$

and this proves the claim. □

That addition and multiplication in R/I satisfy the ring axioms follows easily from the fact that these axioms are satisfied in R. The ring R/I, with addition and multiplication defined as above, is called the *factor ring* "R modulo "I".

NOTES:

1. The zero element of R/I is the coset $0_R + I = I$.

2. It is clear that R/I has some properties in common with R. For example

   - R/I is commutative if R is commutative.
   - If R contains an identity element 1_R for multiplication, then $1_R + I$ is an identity element for multiplication in R/I
   - If u is a unit in R with inverse u^{-1}, then $u + I$ is a unit in R/I, with inverse $u^{-1} + I$.

3. However, R/I can be structurally quite different from R. For example, R/I can contain zero-divisors, even if R does not. It is also possible for R/I to be a field if R is not.
   QUESTION FOR THE SEMINAR: Find examples of both of these phenomena.

In the next section we will look at conditions on I under which R/I is an integral domain or a field, for a commutative ring R.
Our final goal in this section is to prove the *Fundamental Homomorphism Theorem* for rings, which states that if $\phi : R \longrightarrow S$ is a ring homomorphism, then the image of ϕ is basically a copy of the factor ring $R/\ker\phi$.

Definition 3.3.3 *Let $\phi : R \longrightarrow S$ be a ring homomorphism. Then ϕ is called an isomorphism if*

1. *ϕ is surjective (onto); i.e.* $\mathrm{Im}\phi = S$, *and*
2. *ϕ is injective (one-to-one) i.e. $\phi(r_1) \neq \phi(r_2)$ whenever $r_1 \neq r_2$ in R.*

NOTE: ϕ is injective if and only if $\ker\phi$ is the zero ideal of R.
To see this first suppose ϕ is injective. Then $\ker\phi = \{0_R\}$, otherwise if $r \in \ker\phi$ for some $r \neq 0$ we would have $\phi(r) = \phi(0_R)$, contrary to the injectivity of ϕ.
On the other hand suppose $\ker\phi = \{0_R\}$. Then if there exist elements r_1 and r_2 of R with $\phi(r_1) = \phi(r_2)$ we must have $\phi(r_1 - r_2) = \phi(r_1) - \phi(r_2) = 0_S$. This means $r_1 - r_2 \in \ker\phi$, so $r_1 - r_2 = 0_R$ and ϕ is injective.

The characterisation of injectivity in the above note can be very useful.

If $\phi : R \longrightarrow S$ is an isomorphism, then S is an "exact copy" of R. This means that S and R are structurally identical, and only differ in the way their elements are labelled. We say that R and S are *isomorphic* and write $R \cong S$.

Theorem 3.3.4 *(The Fundamental Homomorphism Theorem) Let $\phi : R \longrightarrow S$ be a homomorphism of rings. Then the image of ϕ is isomorphic to the factor ring $R/\ker\phi$.*

Proof: Let I denote the kernel of ϕ, so I is a two-sided ideal of R. Define a function $\bar{\phi} : R/I \longrightarrow \mathrm{Im}\phi$ by

$$\bar{\phi}(a + I) = \phi(a) \text{ for } a \in R.$$

1. $\bar{\phi}$ is well-defined (i.e. the image of $a + I$ does not depend on a choice of coset representative). Suppose that $a + I = a_1 + I$ for some $a, a_1 \in R$. Then $a - a_1 \in I$ by Lemma 3.3.2. Hence $\phi(a - a_1) = 0_S = \phi(a) - \phi(a_1)$. Thus $\phi(a) = \phi(a_1)$ as required.

2. $\bar{\phi}$ is a ring homomorphism.
   Suppose $a + I, b + I$ are elements of R/I. Then

$$\begin{aligned}
\bar{\phi}\left((a + I) + (b + I)\right) &= \bar{\phi}\left((a + b) + I\right) \\
&= \phi(a + b) \\
&= \phi(a) + \phi(b) \\
&= \bar{\phi}(a + I) + \bar{\phi}(b + I).
\end{aligned}$$

So ϕ is additive.

Also

$$\begin{aligned}\bar{\phi}\left((a+I)(b+I)\right) &= \bar{\phi}(ab+I)\\ &= \phi(ab)\\ &= \phi(a)\phi(b)\\ &= \bar{\phi}(a+I)\bar{\phi}(b+I).\end{aligned}$$

So $\bar{\phi}$ is multiplicative - $\bar{\phi}$ is a ring homomorphism.

3. $\bar{\phi}$ is injective.
   Suppose $a+I \in \ker\bar{\phi}$. Then $\bar{\phi}(a+I) = 0_S$ so $\phi(a) = 0_S$. This means $a \in \ker\phi$, so $a \in I$. Then $a+I = I = 0_R + I$, $a+I$ is the zero element of R/I. Thus $\ker\bar{\phi}$ contains only the zero element of R/I.

4. $\bar{\phi}$ is surjective.
   Let $s \in \mathrm{Im}\phi$. Then $s = \phi(r)$ for some $r \in R$. Thus $s = \bar{\phi}(r+I)$ and every element of $\mathrm{Im}\phi$ is the image under $\bar{\phi}$ of some coset of I in R.

Thus $\bar{\phi} : R/\ker\phi \longrightarrow \mathrm{Im}\phi$ is a ring isomorphism, and $\mathrm{Im}\phi$ is isomorphic to the factor ring $R/\ker\phi$. □

3.4 Maximal and Prime Ideals

The goal of this section is to characterize those ideals of commutative rings with identity which correspond to factor rings that are either integral domains or fields.

Definition 3.4.1 *Let* R *be a ring. A two-sided ideal* I *of* R *is called* maximal *if* $I \neq R$ *and no proper ideal of* R *properly contains* I.

EXAMPLES

1. In $\mathbb{Z}$, the ideal $\langle 6\rangle = 6\mathbb{Z}$ is not maximal since $\langle 3\rangle$ is a proper ideal of $\mathbb{Z}$ properly containing $\langle 6\rangle$ (by a *proper* ideal we mean one which is not equal to the whole ring).

2. In $\mathbb{Z}$, the ideal $\langle 5\rangle$ is maximal. For suppose that I is an ideal of $\mathbb{Z}$ properly containing $\langle 5\rangle$. Then there exists some $m \in I$ with $m \notin \langle 5\rangle$, i.e. 5 does not divide m. Then $\gcd(5, m) = 1$ since 5 is prime, and we can write

$$1 = 5s + mt$$

   for integers s and t. Since $5s \in I$ and $mt \in I$, this means $1 \in I$. Then $I = \mathbb{Z}$, and $\langle 5\rangle$ is a maximal ideal in $\mathbb{Z}$.

3. The maximal ideals in $\mathbb{Z}$ are precisely the ideals of the form $\langle p\rangle$, where p is prime.

The following is a generalization of the statement that $\mathbb{Z}/n\mathbb{Z}$ is a field precisely when n is prime.

Theorem 3.4.2 *Let* R *be a commutative ring with identity, and let* M *be an ideal of* R. *Then the factor ring* R/M *is a field if and only if* M *is a maximal ideal of* R.

COMMENT ON PROOF: There are two things to be shown here. We must show that if R/M is a field (i.e. if every non-zero element of R/M is a unit), then M is a maximal ideal of R. A useful strategy for doing this is to suppose that I is an ideal of R properly containing M, and try to show that I must be equal to R.
We must also show that if M *is* a maximal ideal of R, then every non-zero element of R/M is a unit. A strategy for doing this is as follows : if $a \in R$ does not belong to M (so $a+M$ is not the zero element in R/M), then the fact that M is maximal as an ideal of R means that the only ideal of R that contains both M and the element a is R itself. In particular the only ideal of R that contains both M and the element a contains the identity element of R.

Proof of Theorem 4.2.6: ($\Longleftarrow$) Suppose that R/M is a field and let I be an ideal of R properly containing M. Let $a \in I$, $a \notin M$. Then $a + M$ is not the zero element of R/M, and so $(a + M)(b + M) = 1 + M$, for some $b \in R$. Then $ab - 1 \in M$; let

$m = ab - 1$. Now $1 = ab - m$ and so $1 \in I$ since $a \in I$ and $m \in I$. It follows that $I = R$ and so M is a maximal ideal of R.

($\Longrightarrow$): Suppose that M is a maximal ideal of R and let $a+M$ be a non-zero element of R/M. We need to show the existence of $b + m \in R/M$ with $(a + M)(b + M) = 1 + M$. This means $ab + M = 1 + M$, or $ab - 1 \in M$.
So we need to show that there exists $b \in R$ for which $ab - 1 \in M$. Let M' denote the set of elements of R of the form

$$ar + s, \text{ for some } r \in R \text{ and } s \in M.$$

Then M' is an ideal of R (check), and M' properly contains M since $a \in M'$ and $a \notin M$. Then $M' = R$ since M is a maximal ideal of R. In particular then $1 \in M'$ and $1 = ab + m$ for some $b \in R$ and $m \in M$. Then $ab - 1 \in M$ and

$$(a + M)(b + M) = 1 + M \text{ in } R/M.$$

So $a + M$ has an inverse in R/M as required. □

We will now characterize those ideals I of R for which R/I is an integral domain.

Definition 3.4.3 *Let* R *be a commutative ring. An ideal* I *of* R *is called* prime *if* $I \neq R$ *and whenever* $ab \in I$ *for elements* a *and* b *of* R*, either* $a \in I$ *or* $b \in I$*.*

EXAMPLE: The ideal $\langle 6 \rangle$ is not a prime ideal in $\mathbb{Z}$, since $2 \times 3 \in \langle 6 \rangle$ although neither 2 nor 3 belongs to $\langle 6 \rangle$. However the ideal $\langle 5 \rangle$ *is* prime in $\mathbb{Z}$, since the product of two integers is a multiple of 5 only if at least one of the two is a multiple of 5.
The prime ideals of $\mathbb{Z}$ are precisely the maximal ideals; they have the form $\langle p \rangle$ for a prime p.

Theorem 3.4.4 *Let* R *be a commutative ring with identity, and let* I *be an ideal of* R*. Then the factor ring* R/I *is an integral domain if and only if* I *is a prime ideal of* R*.*

Proof: R/I is certainly a commutative ring with identity, so we need to show that R/I contains zero-divisors if and only if I is not a prime ideal of R. So let $a+I$, $b+I$ be non-zero elements of R/I. This means neither a nor b belongs to I. We have $(a + I)(b + I) = 0 + I$ in R/I if and only if $ab \in I$. This happens for some pair a and b if and only if I is not prime. □

Corollary 3.4.5 *Let* R *be a commutative ring with identity. Then every maximal ideal of* R *is prime.*

Proof: Let M be a maximal ideal of R. Then R/M is a field so in particular it is an integral domain. Thus M is a prime ideal of R. □

QUESTION FOR THE SEMINAR: Try to prove Corollary 3.4.5 using only the definitions of prime and maximal ideals.

It is not true that every prime ideal of a commutative ring with identity is maximal. For example

1. We have already seen that the zero ideal of $\mathbb{Z}$ is prime but not maximal.

2. In $\mathbb{Z}[x]$, let I denote the ideal consisting of all elements whose constant term is 0 (I is the principal ideal generated by x). The I is a prime ideal of $\mathbb{Z}[x]$ but it is not maximal, since it is contained for example in the ideal of $\mathbb{Z}[x]$ consisting of all those polynomials whose constant term is even.

Theorem 3.4.6 *Let F be a field and let I be an ideal of the polynomial ring $F[x]$. Then*

1. *I is maximal if and only if $I = \langle p(x)\rangle$ for some irreducible polynomial $p(x)$ in $F[x]$.*

2. *I is prime if and only if $I = \{0\}$ or $I = \langle p(x)\rangle$ for an irreducible $p(x) \in F[x]$.*

Proof: By Lemma 3.2.3 I is principal, $I = \langle p(x)\rangle$ for some $p(x) \in F[x]$.

1. ($\Longleftarrow$) Assume $p(x)$ is irreducible and let I_1 be an ideal of $F[x]$ containing I. Then $I_1 = \langle f(x)\rangle$ for some $f(x) \in F[x]$. Since $p(x) \in I_1$ we have $p(x) = f(x)q(x)$ for some $q(x) \in F[x]$. Since $p(x)$ is irreducible this means that either $f(x)$ has degree zero (i.e. is a non-zero element of F) or $q(x)$ has degree zero. If $f(x)$ has degree zero then $f(x)$ is a unit in $F[x]$ and $I_1 = F[x]$. If $q(x)$ has degree zero then $p(x) = af(x)$ for some nonzero $a \in F$, and $f(x) = a^{-1}p(x)$; then $f(x) \in I$ and $I_1 = I$. Thus either $I_1 = I$ or $I_1 = F[x]$, so I is a maximal ideal of $F[x]$.
($\Longrightarrow$): Suppose $I = \langle p(x)\rangle$ is a maximal ideal of $F[x]$. Then $p(x) \neq 0$. If $p(x) = g(x)h(x)$ is a proper factorization of $p(x)$ then $g(x)$ and $h(x)$ both have degree at least 1 and $\langle g(x)\rangle$ and $\langle h(x)\rangle$ are proper ideals of $F[x]$ properly containing I. This contradicts the maximality of I, so we conclude that $p(x)$ is irreducible. This proves 1.

2. Certainly the zero ideal of $F[x]$ and the principal ideals generated by irreducible polynomials are prime. Every other ideal has the form $\langle f(x)\rangle$ for a reducible $f(x)$. If $I = \langle f(x)\rangle$ and $f(x) = g(x)h(x)$ where $g(x)$ and $h(x)$ both have degree less than that of $f(x)$ then neither $g(x)$ nor $h(x)$ belongs to I but their product does. Thus I is not prime.

Chapter 4

Unique Factorization Domains (UFDs)

4.1 Unique Factorization Domains (UFDs)

Throughout this section R will denote an integral domain (i.e. a commutative ring with identity containing no zero-divisors). Recall that a *unit* of R is an element that has an inverse with respect to multiplication. If a is any element of R and u is a unit, we can write

$$a = u(u^{-1}a).$$

This is not considered to be a proper factorization of a. For example we do not consider $5 = 1(5)$ or $5 = (-1)(-5)$ to be proper factorizations of 5 in $\mathbb{Z}$. We do not consider

$$x^2 + 2 = 2\left(\frac{1}{2}x^2 + 1\right)$$

to be a proper factorization of $x^2 + 2$ in $\mathbb{Q}[x]$.

Definition 4.1.1 *An element* a *in an integral domain* R *is called* irreducible *if it is not zero or a unit, and if whenever* a *is written as the product of two elements of* R*, one of these is a unit.*

An element p *of an integral domain* R *is called* prime *if* p *is not zero or a unit, and whenever* p *divides* ab *for elements* a, b *of* R*, either* p *divides* a *or* p *divides* b.

Note

1. Elements r and s are called *associates* of each other if $s = ur$ for a unit u of R. So $a \in R$ is irreducible if it can only be factorized as the product of a unit and one of its own associates.

2. If R is an integral domain, every prime element of R is irreducible. To see this let $p \in R$ be prime and suppose that $p = rs$ is a factorisation of p in R. Then since p divides rs, either p divides r or $p|s$. There is no loss of generality in assuming p divides r. Then $r = pa$ for some element a of R, and $p = rs$ so $p = pas$. Then $p - pas = 0$ so $p(1 - as) = 0$ in R. Thus $as = 1$ since R is an integral domain and $p \neq 0$. Then s is a unit and $p = rs$ is not a proper factorisation of p. Hence p is irreducible in R.

   It is *not* true that every irreducible element of an integral domain must be prime, as we will shortly see.

Examples:

1. In $\mathbb{Z}$ the units are 1 and -1 and each non-zero non-unit element has two associates, namely itself and its negative. SO 5 and -5 are associates, 6 and -6 are associates, and so on. The irreducible elements of $\mathbb{Z}$ are p and $-p$, for p prime.

2. In $\mathbb{Q}[x]$, the units are the non-zero constant polynomials. The associates of a non-zero non-constant polynomial $f(x)$ are the polynomials of the form $af(x)$ where $a \in \mathbb{Q}^\times$. So $x^2 + 2$ is associate to $3x^2 + 6$, $\frac{1}{2}x^2 + 1$, etc.

3. In $\mathbb{Z}$ the irreducible elements are the integers p and $-p$ where p is a prime numbers. The prime elements of $\mathbb{Z}$ are exactly the irreducible elements - the prime numbers and their negatives.

Definition 4.1.2 *An integral domain* R *is a* unique factorization domain *if the following conditions hold for each element* a *of* R *that is neither zero nor a unit.*

1. a *can be written as the product of a finite number of irreducible elements of* R.

2. *This can be done in an essentially unique way. If* $a = p_1 p_2 \dots p_r$ *and* $a = q_1 q_2 \dots q_s$ *are two expressions for* a *as a product of irreducible elements, then* $s = r$ *and* $q_1, \dots, q_s$ *can be reordered so that for each* i, q_i *is an associate of* p_i.

Example 4.1.3 *$\mathbb{Z}$ is a UFD.*

(This is the Fundamental Theorem of Arithmetic).

Example 4.1.4 *Let $\mathbb{Z}[\sqrt{-5}$ denote the set of complex numbers of the form $a + b\sqrt{-5}$ where a and b are integers (and $\sqrt{-5}$ denotes the complex number $\sqrt{5}i$. We will show that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD (it is easily shown to be a ring under the usual addition and multiplication of complex numbers).*

Claim: $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.
The proof if this claim will involve a number of steps.

1. We define a function $\phi : \mathbb{Z}[\sqrt{-5}] \longrightarrow \mathbb{Z}_{\geqslant 0}$ by $\phi(\alpha) = \alpha\bar{\alpha}$ where $\bar{\alpha}$ denotes the complex conjugate of α. Thus

$$\phi(a + b\sqrt{-5}) = (a + b\sqrt{-5})(a - b\sqrt{-5}) = a^2 + 5b^2.$$

   Let $\alpha,\ \beta \in \mathbb{Z}\sqrt{-5}$. Then

$$\phi(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\beta\bar{\alpha}\bar{\beta} = \alpha\bar{\alpha}\beta\bar{\beta} = \phi(\alpha)\phi(\beta).$$

   So ϕ is multiplicative.

2. Suppose α is a unit of $\mathbb{Z}[\sqrt{-5}]$ and let β be its inverse. Then $\phi(\alpha\beta) = \phi(1) = 1 = \phi(\alpha)\phi(\beta)$. Since $\phi(\alpha)$ and $\phi(\beta)$ are positive integers this means $\phi(\alpha) = 1$ and $\phi(\beta) = 1$. So $\phi(\alpha) = 1$ whenever α is a unit.

   On the other hand $\phi(a + b\sqrt{-5}) = 1$ implies $a^2 + 5b^2 = 1$ for integers a and b which means $b = 0$ and $a = \pm 1$. So the only units of $\mathbb{Z}[\sqrt{-5}]$ are 1 and -1.

3. Suppose $\phi(\alpha) = 9$ for some $\alpha \in \mathbb{Z}[\sqrt{-5}]$. If α is not irreducible in $\mathbb{Z}[\sqrt{-5}]$ then it factorizes as $\alpha_1\alpha_2$ where α_1 and α_2 are non-units. Then we must have

$$\phi(\alpha_1) = \phi(\alpha_2) = 3.$$

   Now this would means $3 = c^2 + 5d^2$ for integers c and d which is impossible. So if $\phi(\alpha) = 9$ then α is irreducible in $\mathbb{Z}[\sqrt{-5}]$.

4. Now $9 = 3 \times 3$ and $9 = (2 + \sqrt{-5})(2 - \sqrt{-5})$ in $\mathbb{Z}[\sqrt{-5}]$. The elements $3,\ 2 + \sqrt{-5}$ and $2 - \sqrt{-5}$ are all irreducible in $\mathbb{Z}[\sqrt{-5}]$ by item 3. above. Furthermore 3 is not an associate of either $2 + \sqrt{-5}$ or $2 - \sqrt{-5}$ as the only units in $\mathbb{Z}[\sqrt{-5}]$ are 1 and -1. We conclude that the factorizations of 9 above are genuinely different, and $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Note that 3 is an example of an element of $\mathbb{Z}[\sqrt{-5}]$ that is irreducible but not prime.
Remark: The ring $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ *is* a UFD.

Theorem 4.1.5 *Let F be a field. Then the polynomial ring $F[x]$ is a UFD.*

Proof: We need to show that every non-zero non-unit in $F[x]$ can be written as a product of irreducible polynomials in a manner that is unique up to order and associates.
So let $f(x)$ be a polynomial of degree $n \geqslant 1$ in $F[x]$. If $f(x)$ is irreducible there is nothing to do. If not then $f(x) = g(x)h(x)$ where $g(x)$ and $h(x)$ both have degree

less than n. If $g(x)$ or $h(x)$ is reducible further factorization is possible; the process ends after at most n steps with an expression for $f(x)$ as a product of irreducibles. To see the uniqueness, suppose that

$$\begin{aligned} f(x) &= p_1(x)p_2(x)\dots p_r(x) \text{ and} \\ f(x) &= q_1(x)q_2(x)\dots q_s(x) \end{aligned}$$

are two such expressions, with $s \geqslant r$. Then $q_1(x)q_2(x)\dots q_s(x)$ belongs to the ideal $\langle p_1(x)\rangle$ of $F[x]$. Since this ideal is prime (as $p_1(x)$ is irreducible) this means that either $q_1(x) \in \langle p_1(x)\rangle$ or $q_2(x)\dots q_s(x) \in \langle p_1(x)\rangle$. Repeating this step leads to the conclusion that at least one of the $q_i(x)$ belongs to $\langle p_1(x)\rangle$. After reordering the $q_i(x)$ if necessary we have $q_1(x) \in \langle p_1(x)\rangle$. Since $q_1(x)$ is irreducible this means $q_1(x) = u_1p_1(x)$ for some unit u_1. Then

$$p_1(x)p_2(x)\dots p_r(x) = u_1p_1(x)q_2(x)\dots q_s(x).$$

Since $F[x]$ is an integral domain we can cancel $p_1(x)$ from both sides to obtain

$$p_2(x)\dots p_r(x) = u_1q_2(x)\dots q_s(x).$$

After repeating this step a further $r-1$ times we have

$$1 = u_1u_2\dots u_rq_{r+1}(x)\dots q_s(x),$$

where $u_1,\dots,u_r$ are units in $F[x]$ (i.e. non-zero elements of F). This means $s = r$, since the polynomial on the right in the above expression must have degree zero. We conclude that $q_1(x),\dots,q_s(x)$ are associates (in some order) of $p_1(x),\dots,p_r(x)$. This completes the proof. □

4.2 Every PID is a UFD

Recall that an ideal I of a commutative ring with identity R is *principal* if $I = \langle a \rangle$ for some $a \in R$, i.e.

$$I = \{ra : r \in R\}.$$

An integral domain R is a *principal ideal domain* if all the ideals of R are principal. Examples of PIDs include $\mathbb{Z}$ and $F[x]$ for a field F.

Definition 4.2.1 *A commutative ring* R *satisfies the* ascending chain condition *(ACC) on ideals if there is no infinite sequence of ideals in* R *in which each term properly contains the previous one. Thus if*

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$$

is a chain of ideals in R*, then there is some* m *for which* $I_k = I_m$ *for all* $k \geqslant m$*.*

Note: Commutative rings satisfying the ACC are called *Noetherian*.

To understand what the ACC means it may be helpful to look at an example of a ring in which it does not hold.

Example 4.2.2 *Let* $C(\mathbb{R})$ *denote the ring of continuous functions from* $\mathbb{R}$ *to* $\mathbb{R}$ *with addition and multiplication defined by*

$$(f + g)(x) = f(x) + g(x); \quad (fg)(x) = f(x)g(x), \quad \text{for } f, g \in C(\mathbb{R}), x \in \mathbb{R}.$$

For $n = 1, 2, 3, \dots$, define I_n to be the subset of $C(\mathbb{R})$ consisting of those functions that map every element of the interval $\left[-\frac{1}{n}, \frac{1}{n}\right]$ to 0.
Then I_n is an ideal of $C(\mathbb{R})$ for each n and

$$I_1 \subset I_2 \subset I_3 \subset \dots$$

is an infinite strictly ascending chain of ideals in $C(\mathbb{R})$ (i.e. every term is this chain is strictly contained in the next one). So the ACC is not satisfied in $C(\mathbb{R})$.

Example 4.2.3 *The ACC is satisfied in* $\mathbb{Z}$*.*

Proof: Let $I_1 \subseteq I_2 \subseteq \dots$ be an ascending chain of ideals in $\mathbb{Z}$. Choose k with $I_k \neq \{0\}$. Then $I_k = \langle n \rangle$ for some positive integer n. Now for an ideal $\langle m \rangle$ of $\mathbb{Z}$ we have $n \in \langle m \rangle$ if and only if $m|n$. Since n has only a finite number of divisors in $\mathbb{Z}$, this means only finitely many different ideals can appear after I_k in the chain.

Theorem 4.2.4 *Let* R *be a PID. Then the ACC is satisfied in* R*.*

Proof: Let $I_1 \subseteq I_2 \subseteq \dots$ be an ascending chain of ideals in $\mathbb{R}$. Let $I = \cup_{i=0}^{\infty} I_i$. Then

1. I is closed under addition and multiplication, for suppose a and b are elements of I. Then there are ideals I_j and I_k in the chain with $a \in I_j$ and $b \in I_k$. If $m \geqslant \max(j,k)$ then both a and b belong to I_m and so do $a+b$ and ab. So $a+b \in I$ and $ab \in I$.

2. $0 \in I$ since $0 \in I_i$ for each i.

3. Suppose $a \in I$. Then $a \in I_j$ for some j, and $-a \in I_j$. So $-a \in I$. Thus I is a subring of R.

4. Furthermore I is an ideal of R. To see this let $a \in I$. Then $a \in I_j$ for some j. If r is any element of R then $ra \in I_j$ and $ra \in I$. So whenever $a \in I$ we have $ra \in I$ for all $r \in R$. Thus I is an ideal of R.

Now since R is a PID we have $I = \langle c \rangle$ for some $c \in \mathbb{R}$. Since $c \in I$ there exists n with $c \in I_n$. Then $I_n = \langle c \rangle$ and $I_r = \langle c \rangle$ for all $r \geqslant n$. So the chain of ideals stabilizes after a finite number of steps, and the ACC holds in R.

Theorem 4.2.5 *Let R be a PID. Then every element of R that is neither zero nor a unit is the product of a finite number of irreducibles.*

Proof: Let $a \in R,\ a \neq 0,\ a \notin \mathcal{U}(R)$ (i.e. a not a unit).

1. First we show that a has an irreducible factor. If a is irreducible, this is certainly true. If not then we can write $a = a_1 b_1$ where neither a_1 nor b_1 is a unit. Then $a \in \langle a_1 \rangle$, and $\langle a \rangle \subset \langle a_1 \rangle$. This inclusion is strict for $\langle a \rangle = \langle a_1 \rangle$ would imply $a_1 = ac$ and $a = acb_1$ for some $c \in R$. Since R is an integral domain this would imply that b_1 is a unit, contrary to the fact that the above factorization of a is proper.

   If a_1 is not irreducible then we can write $a_1 = a_2 b_2$ for non-units a_2 and b_2 and we obtain

   $$\langle a \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle,$$

   where each of the inclusions is strict. If a_2 is not irreducible we can extend the above chain, but since the ACC is satisfied in R the chain must end after a finite number of steps at an ideal $\langle a_r \rangle$ generated by an irreducible element a_r. So a has a_r as an irreducible factor.

2. Now we show that a is the product of a finite number of irreducible elements of R. If a is not irreducible then by the above we can write $a = p_1 c_1$ where p_1 is irreducible and c_1 is not a unit. Thus $\langle a \rangle$ is strictly contained in the ideal $\langle c_1 \rangle$. If c_1 is not irreducible then $c_1 = p_2 c_2$ where p_2 is irreducible and c_2 is not a unit. We can build a strictly ascending chain of ideals :

   $$\langle a \rangle \subset \langle c_1 \rangle \subset \langle c_2 \rangle \ldots$$

This chain must end after a finite number of steps at an ideal $\langle c_r \rangle$ with c_r irreducible. Then

$$a = p_1 p_2 \dots p_r c_r$$

is an expression for a as the product of a finite number of irreducibles in R.

□

So in order to show that every PID is a UFD, it remains to show uniqueness of factorizations of the above type.

Lemma 4.2.6 *Let* I *be an ideal of a PID* R*. Then* I *is maximal if and only if* $I = \langle p \rangle$ *for an irreducible element* p *of* R*.*

Proof: Suppose I is maximal and write $I = \langle p \rangle$ for some $p \in R$. If p is reducible then $p = ab$ for non-units a and b of R, and $\langle p \rangle \subseteq \langle a \rangle$. Furthermore $\langle p \rangle \neq \langle a \rangle$ since $a \in \langle p \rangle$ would imply $a = pc$ and $p = pcb$ which would mean that b is a unit in R. Also $\langle a \rangle \neq R$ since a is not a unit of R. Thus reducibility of p would contradict the maximality of I.
On the other hand suppose p is irreducible and let I_1 be an ideal of R containing $I = \langle p \rangle$. Then $I_1 = \langle q \rangle$ for some $q \in R$ and $p \in I_1$ means $p = rq$ for some $r \in R$. Then either q is a unit or r is a unit. In the first case $I_1 = R$ and in the second case $q = r^{-1}p$ and $q \in \langle p \rangle$ implies $\langle q \rangle = \langle p \rangle$ and $I_1 = I$. Thus I is a maximal ideal of R.
□

Note: The notation $a|b$ (a divides b) in an integral domain R means $b = ac$ for some $c \in R$.

Lemma 4.2.7 *Let* R *be a PID and let* p *be an irreducible in* R*. Then* p *is a prime in* R*.*

Proof: Let a and b be elements of R for which $p|ab$. By Lemma 4.2.6 $I = \langle p \rangle$ is a maximal ideal of R. Thus I is a prime ideal of R by Corollary 3.4.5. Now $ab \in I$ implies either $a \in I$ or $b \in I$. Thus either $p|a$ or $p|b$ in R. □.
So in a PID the notions of prime and irreducible coincide.

Theorem 4.2.8 *Every PID is a UFD.*

Proof: Let R be a PID and suppose that a non-zero non-unit element a of R can be written in two different ways as a product of irreducibles. Suppose

$$a = p_1 p_2 \dots p_r \text{ and } a = q_1 q_2 \dots q_s$$

where each p_i and q_j is irreducible in R, and $s \geqslant r$. Then p_1 divides the product $q_1 \dots q_s$, and so p_1 divides q_j for some j, as p_1 is prime. After reordering the q_j if

necessary we can suppose $p_1|q_1$. Then $q_1 = u_1 p_1$ for some unit u_1 of R, since q_1 and p_1 are both irreducible. Thus

$$p_1 p_2 \dots p_r = u_1 p_1 q_2 \dots q_s$$

and

$$p_2 \dots p_r = u_1 q_2 \dots q_s.$$

Continuing this process we reach

$$1 = u_1 u_2 \dots u_r q_{r+1} \dots q_s.$$

Since none of the q_j is a unit, this means $r = s$ and $p_1, p_2, \dots, p_r$ are associates of $q_1, q_2, \dots, q_r$ in some order. Thus R is a unique factorization domain. □

Note: It is not true that every UFD is a PID.
For example $\mathbb{Z}[x]$ is not a PID (e.g. the set of polynomials in $\mathbb{Z}[x]$ whose constant term is even is a non-principal ideal) but $\mathbb{Z}[x]$ *is* a UFD.
To see this note that irreducible elements in $\mathbb{Z}[x]$ are either integers of the form $\pm p$ for a prime p, or primitive irreducible polynomials of degree $\geqslant 1$. (Recall that a polynomial in $\mathbb{Z}[x]$ is primitive if the gcd of its coefficients is 1.) Let $f(x)$ be a non-zero non-unit in $\mathbb{Z}[x]$.
If $f(x) \in \mathbb{Z}$, then $f(x)$ has a unique factorization as a product of primes. If not then $f(x) = dh(x)$, where d is the gcd of the coefficients in $f(x)$ and $h(x) \in \mathbb{Z}[x]$ is primitive. So $h(x)$ is the product of a finite number of primitive irreducible polynomials in $\mathbb{Z}[x]$, and f(x) is the product of a finite number of irreducible elements of $\mathbb{Z}[x]$. Now suppose that

$$f(x) = p_1 \dots p_k f_1(x) \dots f_r(x) = q_1 \dots q_l g_1(x) \dots g_s(x),$$

where $p_1, \dots, p_k, q_1, \dots, q_l$ are irreducibles in $\mathbb{Z}$ and $f_1(x), \dots, f_r(x), g{-}1(x), \dots, g_s(x)$ are primitive irreducible polynomials in $\mathbb{Z}[x]$. Then $p_1 \dots p_k = \pm$(the gcd of the coefficients in $f(x)$), and $p_1 \dots p_k = \pm q_1 \dots q_l$. Thus $l = k$ and $p_1, \dots, p_k$ are associates in some order of $q_1, \dots, q_k$. Now

$$f_1(x) \dots f_r(x) = \pm g_1(x) \dots g_s(x).$$

Then each $f_i(x)$ and $g_j(x)$ is irreducible not only in $\mathbb{Z}[x]$ but in $\mathbb{Q}[x]$ and since $\mathbb{Q}[x]$ is a UFD this means that $s = r$ and $f_1(x), \dots, f_r(x)$ are associates (in some order) of $g_1(x), \dots, g_r(x)$. After reordering the $g_j(x)$ we can suppose that for $i = 1, \dots, r$ $f_i(x) = u_i(g_i(x)$ where u_i is a non-zero rational number. However since $f_i(x)$ and $g_i(x)$ are both primitive polynomials in $\mathbb{Z}[x]$, we must have $u_i = \pm 1$ for each i, so $f_i(x)$ and $g_i(x)$ are associates not only in $\mathbb{Q}[x]$ but in $\mathbb{Z}[x]$.
Thus $\mathbb{Z}[x]$ is a UFD.