MA314 (Part 2) 2012-2013 : Lecture Notes
Introduction to Abstract Algebra : Rings and Fields

Dr Rachel Quinlan
School of Mathematics, Statistics and Applied Mathematics, NUI Galway

March 22, 2013

Contents

MA314 (PART 2) 2012-13 : RINGS AND FIELDS

LECTURER Dr Rachel Quinlan
Room C105, Ground Floor Áras de Brún
(091) 493796 (Extension 3796 from inside NUI Galway)
rachel.quinlan@nuigalway.ie
http://www.maths.nuigalway.ie/~rquinlan
Students are welcome in my office whenever I am there

LECTURES Tuesday 1.00 Larmor Theatre, Friday 12.00 AM150

SYLLABUS

This is a short (10 lectures) introduction to the algebraic theory of rings and fields, which are components of the wider subject of *abstract algebra*. The philosophy of this subject is that we focus on similarities in arithmetic structure between sets (of numbers, matrices, functions or polynomials for example) which might look initially quite different but are connected by the property of being equipped with operations of addition and multiplication. The set of integers and the set of 2 by 2 matrices with real numbers as entries are examples of rings. These sets are obviously not the same, but they have some similarities - and some differences - in terms of their algebraic structure. Although people have been studying specific examples of rings for thousands of years, the emergence of ring theory as a branch of mathematics in its own right is a very recent development. Much of the activity that led to the modern formulation of ring theory took place in the first half of the 20th century. Ring theory is powerful in terms of its scope and generality, but it can be simply described as the study of systems in which addition and multiplication are possible. Fields are special examples of rings, in which multiplication is commutative and in which it is possible to divide by any non-zero element. Examples of fields include the field $\mathbb{Q}$ of rational numbers and the field $\mathbb{C}$ of complex numbers.
The lecture notes contain a list of more specific topics, itemized over the ten lectures.

LEARNING OUTCOMES

By the end of this course you will be able to :

- Explain the meaning of the terms *ring* and *field* and provide and identify examples (and non-examples) of rings and fields.
- Discuss the important features of given examples of rings and fields.
- Implement the Euclidean algorithm.
- Construct examples of finite rings and fields and perform calculations in them.

ASSESSMENT

End-of-Semester Examinations: One two-hour examination in MA314. More later on the format and on what to expect.
Continuous Assessment: There will be two written assignments over the course of the 10 lectures.

RESOURCES

Online resources for this course will be maintained on the MA314 Blackboard page (click on "Part 2 : Rings and Fields" in the main menu). These will include lecture notes which constitute the "text" for the course. You are expected to study the lecture notes, which are more detailed than the discussion in lectures. As well as the "official" assignments, there are exercises throughout the lecture notes. You are advised to think about these and to maintain your own expanded set of notes including your responses to these exercises.
For supplementary reading, you may find the following books helpful :

- *A First Course in Abstract Algebra*, John B. Fraleigh (512.02)
- *Modern Algebra*, John R. Durbin (512.02 DUR)

A WORD OF ADVICE

At this advanced level, Mathematics is more about understanding and explaining concepts and their logical connections than about doing calculuations or "working out answers". This means that it involves learning a formal and technical language, which takes some practice. Don't be discouraged if it takes you some time to get used to this in the context of abstract algebra - having reached this point you are capable of it.

1 Rings and Fields - Examples

We begin with a list of familiar *algebraic structures* that will be useful for reference and as a source of examples as we proceed. All of these objects are probably well known to readers of these notes, but it is important to make sure that you are also familiar with all the notation used here and that you can use this notation in a precise and careful way, both in reading and in writing.
NOTE For these lectures you will need to be familiar with the notations $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ and their meanings. You will need to know how these number systems relate to and differ from each other and be able to recall this information easily. If you are rusty on this at all, have a look for example at the Wikipedia page on "Number" - `en.wikipedia.org/wiki/Number`

1. The set of integers, $\mathbb{Z}$.
$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, 3, \dots\}$$
The origin of the symbol $\mathbb{Z}$ comes from *Zahlen*, the German word for numbers.

2. The set of rational numbers, $\mathbb{Q}$.
$$\mathbb{Q} = \left\{ \frac{a}{b} : a \in \mathbb{Z},\ b \in \mathbb{Z}, b \neq 0 \right\}$$
$\mathbb{Q}$ is the set of all numbers that can be expressed as the quotient of two integers, in other words as a fraction with integers as numerator and denominator (the denominator must be non-zero obviously). The notation $\mathbb{Q}$ comes from "quotient". Note that $\mathbb{Z} \subset \mathbb{Q}$ (i.e. every integer is a rational number; $\mathbb{Z}$ is a subset of $\mathbb{Q}$).

3. The set $2\mathbb{Z}$ of even integers.
$$2\mathbb{Z} = \{2k : k \in \mathbb{Z}\} = \{\dots, -4, -2, 0, 2, 4, 6, \dots\}$$

4. The set $M_3(\mathbb{R})$ of real 3×3 matrices.
Elements of this set are 3×3 matrices whose entries are real numbers ($\mathbb{R}$ is the set of real numbers - the real numbers include all points on the number line). Note that if we can add or multiply any two of these 3×3 matrices. In each case the result will again be a 3×3 matrix with real entries.

5. $\mathbb{Q}[x]$ - the set of polynomials in x over $\mathbb{Q}$
A *polynomial* in x over $\mathbb{Q}$ is an expression of the form
$$a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0,$$
where d is a non-negative integer, and $a_0, a_1, \dots, a_d$ are rational numbers (elements of $\mathbb{Q}$). So a polynomial (in x over $\mathbb{Q}$) involves a constant term and a finite number of positive integer powers of x which appear with rational coefficients. The following are all examples of elements of $\mathbb{Q}[x]$:
$$x + 2,\ \ x^2 - 5,\ \ x^4 + \frac{1}{2}x^3 + x,\ \ \frac{21}{5},\ \ 0$$
Note that rational numbers themselves are included in $\mathbb{Q}[x]$, as those elements in which the positive powers of x all have zero coefficient.
We can define $\mathbb{Z}[x]$, $\mathbb{R}[x]$ etc. in the same way - the $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$ or whatever indicates where the coefficients live.

6. $D_2(\mathbb{Z})$ - the set of *diagonal* 2×2 matrices with integer entries.
$$D_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a, b \in \mathbb{Z} \right\}.$$
Note that the sum of two diagonal matrices is diagonal, and the product of two diagonal matrices is diagonal.

7. $UT_3(\mathbb{R})$ - the set of *upper triangular* 3×3 matrices with real entries.

$$UT_3(\mathbb{R}) = \left\{ \begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix} : a, b, c, d, e, f \in \mathbb{R} \right\}.$$

   Again, verify that the sum and product of two upper triangular matrices are again upper triangular.

8. $\mathbb{C}$ - the set of complex numbers.

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R},\ i^2 = -1\}$$

   The set of complex numbers is obtained from the set of real numbers by adjoining an "imaginary" square root of -1, denoted by i. Complex numbers can be added together and multiplied to produce new complex numbers.

9. $\mathbb{Q}(i)$ - the set of *Gaussian rational numbers*
   $\mathbb{Q}(i)$ is the subset of $\mathbb{C}$ consisting of all those numbers $a+bi$ whose real part a and imaginary part b are both rational. So for example $1 + 3i$ belongs to $\mathbb{Q}(i)$ but $1 - \sqrt{2}i$ does not. Note that $\mathbb{Q}(i)$ is a subset of $\mathbb{C}$ but not of $\mathbb{R}$.

   EXERCISE: Show that the sum and product of two Gaussian rational numbers is always a Gaussian rational number. (This is not completely obvious at first glance - not every complex number belongs to $\mathbb{Q}(i)$, so you have to explain why taking the sum or product of two elements of $\mathbb{Q}(i)$ couldn't take you outside $\mathbb{Q}(i)$).

These sets are all different, and their elements are mathematical objects of different kinds (integers, real numbers, complex numbers, matrices, polynomials etc.). However, in terms of their *algebraic structure*, they have some overall similarties. All of them are naturally equipped with operations of addition, subtraction and multiplication. This means

given any pair of elements of any of these sets, it is possible to add them together, multiply them, or subtract one from the other, TO GET ANOTHER ELEMENT OF THE SAME SET

This structural similarity between all these sets (and many others) is essentially the theme of the mathematical theory of rings. A few points to note :

1. It is important to note that when we talk about these sets being "equipped" with addition, multiplication and subtraction, we mean not only that it makes sense to add, subtract or multiply any two elements of the set, but that when we do so the result is still within the same set (we could express this by saying that each of our sets is *closed* under addition, multiplication and subtraction).

   In this sense the set $\mathbb{N}$ of natural numbers (or positive integers) is not equipped with a subtraction operation. While it is true that it is always possible to subtract one natural number from another, doing this might take us outside the set of natural numbers. For example 5 and 8 are both natural numbers but the difference $5 - 8 = -3$ is not. So $\mathbb{N}$ is NOT closed under subtraction.

   Similarly, if we were to consider the set $\{\ldots, -3, -1, 1, 3, 5, \ldots\}$ of odd integers in this context, it would not be closed under addition (or subtraction), since the sum of two odd integers is not an odd integer. On the other hand, the sum, difference and product of two even integers is always an even integer, so the set of even integers can make it into our list.

   It is not automatically obvious that all of the sets in our list are closed under their own addition, subtraction and multiplication operations. For example it is not completely obvious that the product of two upper triangular 3×3 matrices is again upper triangular. These things need to be checked and you are advised to check them.

2. The words "addition" and "multiplication" are being used here in a generic way. For example multiplication of $3\times$ matrices is not the same thing as multiplication of real numbers or of polynomials, but the same word "multiplcation" is used for all of these - the specific mechanism of what "addition" or "multiplication" involves depends on the context.

Informal and Incomplete Definition: *A* ring *is an algebraic structure in which it is possible to add, multiply and subtract elements as in the examples above.*

The reason why this definition is incomplete is because it does not specify what properties the addition and multiplication (and subtraction) operations should have or how they should interact with each other.

The multiplication operations in our examples do not all behave in the same way. There are two features of multiplication that are of interest right now.

- First, multiplication may or may not be *commutative*. This means that the order in which two elements are multiplied might or might not make a difference. For example, multiplication of integers is commutative, because $ab = ba$ whenever a and b are integers. Multiplication of 3×3 matrices is not commutative, because AB may not be equal to BA for 3×3 matrices A and B.

  EXERCISE: Determine in which of our examples the multiplication is commutative. (The hardest example of these is probably $UT_3(\mathbb{R})$.)

- In some (but) of our examples, it is possible to divide by any non-zero elements (and stay within the set). This is not the case in the ring $\mathbb{Z}$ of integers, because while we may divide one integer by another, the result of this (e.g. $\frac{3}{5}$) might take us outside $\mathbb{Z}$. We can divide in $\mathbb{Q}$ by any non-zero element, basically because the reciprocal of a non-zero rational number is another non-zero rational number. The same is true in the set of complex numbers $\mathbb{C}$ - recall that if $z = a + bi$ is a (non-zero) complex number, then

$$\frac{1}{z} = \frac{1}{a+bi} = \frac{1}{a+bi} \times \frac{a-bi}{a-bi} = \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

  So the reciprocal of a non-zero complex number is a non-zero complex number, and in the ring of complex numbers we have a division operation (these means that we can divide by any non-zero complex number).
  We do not have a division operation in the polynomial ring $\mathbb{Q}[x]$ - i.e. it is not in general possible to divide one polynomial by another and get a polynomial as the result.
  We do not have a notion of division for matrices either (although some matrices have inverses and some do not).

EXERCISE: Show that the reciprocal of every non-zero element of the ring $\mathbb{Q}(i)$ is also an element of $\mathbb{Q}(i)$.
This means : if a and b are rational numbers, show that the real and imaginary parts of $\frac{1}{a+bi}$ are also rational numbers.

Informal Definition: *A ring in which the multiplication is commutative and in which it is possible to divide by any non-zero element (and stay inside the ring) is called a* field.

EXERCISE: Which of our examples are fields?
To answer this : you can rule out any example in which the multiplication is not commutative straight away. For the commutative examples, what you need to do is figure out whether every non-zero element has a reciprocal that belongs to the ring.

2 Axiomatic Definitions

This section contains the "official and complete" definitions of the terms ring and field. These consist of axioms specifying the properties that the division and multiplication operations must have and how they interact with each other.

These definitions make no reference to any particular set or to the nature of the objects that are elements of the ring - whether they are numbers, functions, matrices or whatever. When thinking about the definitions, try to bear this in mind and try also to bear in mind that "addition" and "multiplication" are just words describing two ways of putting pairs of elements together to get new elements in the set. It can't be assumed that they always correspond to familiar versions of multiplication and addition - that is why in the definition these operations are characterized by their properties. This in a way is the whole point of abstract algebra - to characterize algebraic structures in terms of their properties rather than in terms of the details of specific examples.

Definition (Formal definition of a ring) A ring R is a non-empty set equipped with operations of addition $(+)$ and multiplication $(\times)$ satisfying the following axioms.
(The first five axioms are concerned with addition and are labelled A1 to A5 for this reason.)

A1 R is *closed* under addition. This means that $r + s \in R$ whenever $r \in R$ and $s \in R$.

A2 The addition operation is *commutative*. This means that $r + s = s + r$ for all pairs of elements r and s.

A3 The addition operation is *associative*. This means that $(r + s) + t = r + (s + t)$ for all elements r, s and t of R.

A4 R contains an *identity element* or *neutral element* 0_R for addition (this is often referred to as the *zero element* because for addition it resembles the number 0). The property that the zero element must satisfy is

$$r + 0_R = r\ (=)_R + r) \text{ for all} r \in R.$$

So adding the zero element to another element has no effect on that element (hence the term "neutral" although "identity" is more usual).

A5 For every element r of R, there exists an element $-r$ in R for which

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

So $-r$ is the "negative" or "additive inverse" of r.

NOTE: It is the existence of additive inverses that makes it possible to define a subtraction operation - subtracting a particular element from something means adding the negative of that element.

For a set (with addition and multiplication) to be considered a ring, its addition must satisfy all of axioms A1 to A5. For example the set of odd integers is not a ring because it fails axiom A1 (and A4). The set $\{0, 1, 2, 3, \dots\}$ of non-negative integers is not a ring because it fails axiom A5.

For the multiplication operation, only one special property (apart from closure) is specified.

M1 If r and s are elements of R, then $r \times s$ and $s \times r$ are also elements of R.

M2 The multiplication in R is associative. This means that for all elements r, s, t of R we have

$$(r \times s) \times t = r \times (s \times t).$$

Axiom M2 means that if we want to multiply elements r, s and t (in that order), we don't have to specify whether we first take $r \times s$ and then multiply that on the right by t, or whether we first take $s \times t$ and then multiply that on the left by r.

The last two axioms specify how the addition and multiplication operations interact - as we might expect we have *distributivity* of multiplication over addition. This means :

D1 For any elements r, s and t of R,

$$r \times (s + t) = (r \times s) + (r \times t)$$

D2 For any elements r, s and t of R,

$$(r + s) \times t = (r \times t) + (s \times t)$$

So a ring is a non-empty set equipped with addition and multiplication operations that satisfy all of these axioms.

REMARK: Composing this list of axioms may seem at first glance like a lot of unnecessary fussing to do about systems of arithmetic with which we are all very familiar. If you are puzzled about the rationale behind this, one thing to bear in mind is that the point is not just to give an abstract, formal description of things that are familiar, but also to characterize addition and multiplication by their abstract properties so that our definition (and whatever theory might be developed from it) applies also to situations where the addition and multiplication work in less familiar and less obviously "natural" ways.

Another way to think about it : if someone asked you what "addition" means, what would you say? It is hard to imagine trying to explain what addition is without having to first start with some things that can be added together, like integers for example. The description in Axioms A1 to A5 above gives us a versatile way of describing the concept of *an addition operation* (for rings) that does not rely on any particular collection of things to be added together.

Now that we know what a ring is, we turn our attention to fields. A *field* is a ring whose multiplication satisfies three extra conditions.

Definiton (Formal definition of a field).
A field F is a ring whose multiplication operation satisfies the following additional properties (the label "FM" stands for "field multiplication").

FM1 The multiplication is commutative, i.e.

$$x \times y = y \times x \text{ for all } x, y \in F.$$

FM2 There is an element 1_F (different from the zero element) which is an *identity element* or *neutral element* with respect to multiplication. Multiplication by this element has no effect. This means that for all elements x of F

$$x \times 1_F = x \ (= 1_F \times x).$$

FM3 If x is a non-zero element of F (this means that x is not the additive identity element), then F contains an element which we denote x^{-1}, which is a multiplicative inverse of x. This means that multiplying x by x^{-1} gives the multiplicative identity element.

$$x \times x^{-1} = 1_F \ \ (-x^{-1} \times x).$$

NOTE ON AXIOMS FM2 AND FM3: The multiplicative inverse acts like a reciprocal and it is this that enables division (by non-zero elements) in a field. Dividing by an element x means multiplying by x^{-1}. The effect of multiplying by x^{-1} "cancels" the effect of multiplying by x. Multiplying an element both by x and by x^{-1} has no effect - it amounts to multiplying by the identity element.

TERMINOLOGY

1. A ring whose multiplicaton is commutative is referred to as a *commutative ring*.
2. A ring that has an identity element for multiplication is referred to as a *unital ring* or as a *ring with identity*.
3. In a ring with identity, we can consider which elements have multiplicative inverses. These elements are called the *units* of the ring.
4. A ring in which every non-zero element is a unit is called a *division ring*. A division ring satisfies all the axioms of a field except for the commutativity of multiplication. We will meet an example of a non-commutative division ring later.

EXERCISES

1. Which of the examples from Lecture 1 are commutative?
2. Which of the examples form Lecture 1 are rings with identity? For each such example, state what the multiplicative identity element is.
3. In those examples from Lecture 1 that are have an identity element for multiplication, state which elements are units. (This is trickier than it might appear at first glance).
4. What are the units in the ring $M_2(\mathbb{Z})$ of 2×2 matrices with integer entries?

The following lemma explains why Axiom FM3 only deals with *non-zero* elements. It says that it is a consequence of the ring axioms that multiplying any element by the zero element will always result in the zero element - so the zero element in a ring never has a chance of having a multiplicative inverse. This lemma is also a nice example of the sort of reasoning from axiomatic foundations that is central to abstract algebra.

Lemma 2.1. *Suppose that* r *is any element of a ring* R*. Then*

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

Proof. We deal with the product $0_R \times r$. Because adding 0_R to any element has no effect, we know that $0_R + 0_R = 0_R$. Then

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

Now we can use the distributive axiom D2 to rewrite $(0_R + 0_R) \times r$ as $(0_R \times r) + (0_R \times r)$ and say

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

So we are saying that the element $0_R \times r$ (which we would like to show is equal to 0_R) remains the same when added to itself. Now we can use subtraction (or Axiom A5) to simplify the above equation.

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

as required. The proof for $r \times 0_R$ is similar, but uses Axiom D1 instead of D2. □

Note that proving this statement (that multiplying by zero always results in zero) uses quite a few of the ring axioms. In particular it uses the distributive axioms. These describe how addition and multiplication interact. What is special about the zero element is concerned with addition, but what we are trying to prove is something about how this element behaves under multiplication. So it is not surprising to see the axioms describing how addition and multiplication interact entering the picture.

3 Polynomials and Roots

This section focusses on polynomial rings, especially on the polynomial ring $\mathbb{Q}[x]$. This ring has many structural properties in common with the ring $\mathbb{Z}$ of integers. For example both of these rings are commutative and each contains an identity element for multiplication.

We finished the last section by observing that in a ring, multiplying by the zero element always results in the zero element. One could ask the converse question : if the product of two elements in ring is zero, must one (or both) of the two elements be zero?

The answer to this question is not always yes.

Exercise: Give an example of a pair of non-zero elements of $D_2(\mathbb{Z})$ whose product is zero.

However, for many rings the answer *is* yes and we make frequent use of this fact. For example, when we solve the quadratic equation $x^2 + 5x + 6 = 0$ by writing

$$x^2 + 5x + 6 = 0 \Longrightarrow (x+2)(x+3) = 0 \Longrightarrow x + 2 = 0 \text{ or} x + 3 = 0,$$

we are making use of the fact that the product of two real numbers is 0 if and only if at least one of the numbers is 0.

Definition An *integral domain* R is a commutative ring with identity in which the product $a \times b$ is equal to the zero element 0_R only if either $a = 0_R$ or $b = 0_R$.

Example The ring $\mathbb{Z}$ of integers is an integral domain. It is from the term "integer" that the terminology "integral domain" arose. The field $\mathbb{Q}$ of rational numbers is also an integral domain. In fact all fields are integral domains.

The ring $\mathbb{Q}[x]$ is another example of an integral domain. Certainly it is a commutative ring with identity. What is required to see that it is an integral domain is a demonstration that the product of two non-zero polynomials in $\mathbb{Q}[x]$ cannot be zero. So suppose that $p(x)$ and $q(x)$ are non-zero polynomials in $\mathbb{Q}[x]$ and let their respective *leading terms* be ax^t and bx^s. This means that a and b are non-zero rational numbers, and that no higher power of x than x^t appears with non-zero coefficient in $p(x)$, and no higher power of x than x^s appears with non-zero coefficient in $q(x)$. So

$$p(x) = ax^t + \text{possible lower order terms in } x, \quad q(x) = bx^s + \text{possible lower order terms in } x$$

Then the leading term of the product $p(x)q(x)$ is abx^{t+s}: we know that $ab \neq 0$ since a and b are non-zero rational numbers. Note that we are using the fact that only the leading terms of $p(x)$ and $q(x)$ can contribute to a term of order $t + s$ in the product, so x^{t+s} is the highest powerof x that can appear in $p(x)q(x)$, and it appears there with coefficient ab. Finally, since we have identified at least one non-zero term in $p(x)q(x)$, we can assert that $p(x)q(x) \neq 0$ and hence that $\mathbb{Q}[x]$ is an integral domain.

Notes

1. It is perfectly possible for either t or s (or both) to be equal to zero in the argument above - this just means that either $p(x)$ or $q(x)$ (or both) is a non-zero constant polynomial. This "extreme case" causes no problems in the proof above.

2. A piece of terminology: the highest power of x to appear with non-zero coefficient in a polynomial $p(x)$ is called the *degree* of $p(x)$. The degree of a non-zero constant polynomial is 0. (The degree of the zero polynomial is not defined). If $p(x)$ and $q(x)$ are two non-zero polynomials in $\mathbb{Q}[x]$, then the degree of the product $p(x)q(x)$ is equal to the sum of the degrees of $p(x)$ and $q(x)$.

Exercises

1. Prove that every field is an integral domain.

2. Give an example of a commutative ring with identity that is *not* an integral domain. (Hint: if you can't think of one, look at the examples from Lecture 1)

One of the reasons for considering the algebra of polynomials is that it provides a means of constructing examples of fields, other than "obvious" examples such as $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. To see the mechanisms of this, we first need to consider some features of factorization and divisibility in $\mathbb{Q}[x]$. We start by mentioning another arithmetic similarity between $\mathbb{Z}$ and $\mathbb{Q}[x]$.

Theorem 3.1 (Division algorithm in $\mathbb{Z}$)**.** *Let n, m be integers with $m \neq 0$. Then n can be written in a unique way in the form*

$$n = qm + r,$$

where q and r are integers and $0 < r < m - 1$.

What this is saying is that we can *divide* n by m, we get a quotient q and a remainder r. So qm is the greatest multiple of m that is less than or equal to n, and r is the "left over" remainder. For example if $n = 26$ and $m = 3$, we divide 26 by 3 and get a quotient (q) of 8 and a remainder (r) of 2, i.e.

$$26 = 8 \cdot 3 + 2.$$

There is an analogous statement for $\mathbb{Q}[x]$.

Theorem 3.2. *[Division algorithm in $\mathbb{Q}[x]$] Let $f(x)$, $g(x)$ be polynomials in $\mathbb{Q}[x]$ with $g(x) \neq 0$. Then $f(x)$ can be written in a unique way in the form*

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

where $q(x)$ and $r(x)$ are polynomials and $r(x) = 0$ or $\deg r(x) < \deg g(x)$.

This means that (for example) we should be able to divide $x^5+x^3-2x^2-5x+1$ by x^2+2x+2 and get a quotient (of degree 3) and a remainder of degree less than 2. We can calculate this quotient and remainder by long division. We find

$$x^5 + x^3 - 2x^2 - 5x + 1 = (x^3 - 2x^2 + 3x - 4)(x^2 + 2x + 2) + (-3x + 7),$$

so the *remainder* for this division is $-3x + 7$.

The reason for making a fuss about this in the case of polynomial rings is because of a connection between divisibility and roots of polynomials. In algebra, it is helpful to think of a polynomial in $\mathbb{Q}[x]$ simultaneously as a formal expression involving powers of x and as a function that can be evaluated at different rational numbers. For example, if $p(x) = x^3 + 3x + 2$, we could write $p(2) = 2^3 + 3(2) + 2 = 16$, $p(-1) = (-1)^3 + 3(-1) + 2 = -2$, etc.

Definition 3.3. *An element a of $\mathbb{Q}$ (or of $\mathbb{R}$ or $\mathbb{C}$) is called a* root *of the polynomial $p(x)$ if $p(a) = 0$.*

The problem of finding solutions of polynomial equations is an ancient one and it is prompted much of the development of modern number theory, ring theory and field theory. In the case of quadratic, cubic and quartic polynomials, explicit formulae exist for expressing the solutions in terms of various square, cube and 4th roots of expressions involving the coefficients. So such general formulae exist for polynomials of degree 5 or more. There is a close link between roots and factorization, given by the following *remainder theorem*.

Theorem 3.4. *Let $p(x)$ be a polynomial in $\mathbb{Q}[x]$ and let $a \in \mathbb{Q}$. Then $p(a)$ is equal to the remainder on dividing $p(x)$ by $x - a$.*

Proof. By the division algorithm there exist elements $q(x)$ and r of $\mathbb{Q}[x]$ for which

$$p(x) = q(x)(x - a) + r,$$

and $r = 0$ or $\deg(r) < 1$. This means that $r \in \mathbb{Q}$. Now

$$p(a) = q(a)(a - a) + r = r,$$

as required. □

Note: Theorem 3.4 says for example that the remainder on dividing $x^3 - 4x^2 + x + 4$ by $x - 2$ should be $(2)^3 - 4(2)^2 + 2 + 4 = -10$. Check that this is true.

Closely related to the remainder theorem is the following *factor theorem*.

Theorem 3.5. *Let $p(x)$ be a polynomial of degree at least 1 in $\mathbb{Q}[x]$, and let $\alpha \in \mathbb{Q}$. Then α is a root of $p(x)$ if and only if $x - \alpha$ is a factor of $p(x)$ in $\mathbb{Q}[x]$.*

Proof. By Theorem 3.5, $p(\alpha)$ is the remainder on dividing $p(x)$ by $(x - \alpha)$. Thus $p(\alpha) = 0$ if and only if this remainder is 0, which happens if and only if $p(x) = (x - \alpha)q(x)$ for some $q(x) \in \mathbb{Q}[x]$. □

So there is a connection between roots and linear factors of polynomials. A polynomial in $\mathbb{Q}[x]$ has a root in $\mathbb{Q}$ if and only if it has a linear factor in $\mathbb{Q}$. The general problem of factorizing a given polynomial in $\mathbb{Q}[x]$ is both computationally and theoretically difficult (although it is not too bad for quadratics and cubics). We have the concept of an *irreducible polynomial* in $\mathbb{Q}[x]$, which is analogous to the idea of a *prime number* in $\mathbb{Z}$.

Definition 3.6. *A polynomial of degree at least 1 in $\mathbb{Q}[x]$ is called* irreducible *if it cannot be written as the product of two polynomials in $\mathbb{Q}[x]$ both having degree at least 1.*

An analogous definition applies to other fields. Irreducible polynomials in $\mathbb{Q}[x]$ are bascially analogous to prime numbers in $\mathbb{Z}$. The problem of deciding whether a given polynomial is irreducible is difficult in general, although there are some shortcuts that work for certain classes of examples.

Problem Decide whether each of the following polynomials is irreducible in the indicated ring.

1. $x^3 + 3x + 2$ in $\mathbb{Q}[x]$
2. $x^3 + 3x + 2$ in $\mathbb{R}[x]$
3. $x^4 - x^3 + x^2 - 5x - 2$ in $\mathbb{Q}[x]$

Solution

1. This is cubic. So either it is irreducible or it is the product of a quadratic and a linear factor (in this case the quadratic factor might factorize further of course). So if the cubic polynomial is reducible, it has a linear factor and hence it has a root in $\mathbb{Q}$.

   Fact: If a polynomial with integer coefficients factorizes in $\mathbb{Q}[x]$ then it also factorizes in $\mathbb{Z}[x]$ (with factors of the same degrees).

   (Unfortunately we won't get a chance to discuss an explanation for this but it is a consequence of a statement known as Gauss's Lemma.) In our example our polynomial is $p(x) = x^3 + 3x + 2$. A possible factorization in $\mathbb{Z}[x]$ would have the form

   $$(x + a)(x^2 + bx + c),$$

where $ac = 2$. In this factorization $-a$ would be a root. So the only candidates for roots are integer factors of 2, i.e. $1, 2, -1, -2$. Check these to see if any is a root :

$$\begin{aligned} p(1) &= 6 \\ p(2) &= 16 \\ p(-1) &= -2 \\ p(-2) &= -12 \end{aligned}$$

We conclude that $p(x)$ has no root in $\mathbb{Z}$ and hence (since it is cubic) that is has no factorization in $\mathbb{Z}[x]$ or in $\mathbb{Q}[x]$ and is irreducible in $\mathbb{Q}[x]$.

2. $p(x) = x^3 + 3x + 2$ in $\mathbb{R}[x]$
   Think of $p(x)$ as a continuous function of x and imagine its graph. When $|x|$ is large, the leading term x^3 dominates. This term is negative when x is negative and positive when x is positive. So there are values of x for which $p(x)$ is negative and values for which it is positive. It follows then from the Intermediate Value Theorem that $p(x)$ must have a root in $\mathbb{R}$ and that $p(x)$ is reducible in $\mathbb{R}[x]$.

   Note: The same argument would apply to any polynomial of odd degree in $\mathbb{R}[x]$.

3. $f(x) = x^4 - x^3 + x^2 - 5x - 2$ in $\mathbb{Q}[x]$
   As in the first example above, search for roots amongst the integer factors of the constant term -2. Note that
   $$f(2) = 16 - 8 + 4 - 10 - 2 = 0.$$
   Hence 2 is a root of $f(x)$, so $x - 2$ is a factor and $f(x)$ is *reducible* in $\mathbb{Q}[x]$.

   Important Note: In this example $f(x)$ is quartic (degree 4). If we had found no roots of $f(x)$ amongst the integer factors of -2, *we could not have concluded anything about the irreducibility or reducibility of* $f(x)$ *in* $\mathbb{Q}[x]$. This would not have ruled out the possibility that $f(x)$ could be the product of two quadratic factors. It is only in the case of a cubic or quadratic polynomial in $\mathbb{Q}[x]$ that the absence of a root in $\mathbb{Q}$ can allow us to conclude that the polynomial is irreducible.

We conclude this section by stating one of the most important theorems in all of mathematics.

Theorem 3.7 (Fundamental Theorem of Algebra)**.** *Every polynomial in* $\mathbb{C}[x]$ *has a root in* $\mathbb{C}$ *and factorizes in* $\mathbb{C}[x]$ *as the product of linear factors.*

So the only irreducible polynomials in $\mathbb{C}[x]$ are the linear ones.

4 The Euclidean Algorithm

In the ring of integers $\mathbb{Z}$ and in the ring $\mathbb{Q}[x]$ of rational polynomials, we have the idea that one element may be a divisor of another. Our interest in this section is in $\mathbb{Z}$, although everything in it could equally well be applied for example to $\mathbb{Q}[x]$.

Definition 4.1. *For integers* a *and* b*, we say that* a *is a divisor of* b*, or* a *is a factor of* b*, or* a *divides* b*, if* $b = ka$ *for some* $k \in \mathbb{Z}$*. We write* $a|b$ *to say that* a *is a divisor of* b*.*

Note on terminology: Do not confuse the term "a divides b" with "a is divisible by b". To say that a divides b means saying that a is a factor of b. For example we could say that 7 divides 35 (this is a true statement).
Please also familiarise yourself with the mathematical symbol "|", which is a very useful compact notation for "divides". The statement 7|35 means "7 divides 35" or "7 is a factor of 35" and that is how it should be read. The symbol is a vertical bar not a slash or a hyphen, and it absolutely should not be confused with the forward slash that is used in fractions like 2/3, it has no connection to that.
We could also write $7 \not| 36$, to state that 7 is not a factor of 36.

Given a pair of non-zero integers (or polynomials in $\mathbb{Q}[x]$) we have the notion of their *greatest common divisor* or gcd (sometimes referred to as highest common factor).

Definition 4.2. *If our two integers are* a *and* b*, then* $\gcd(a, b)$ *is the unique integer* d *with the following properties.*

- $d > 0$
- $d|a$ *and* $d|b$ *(*d *is a* common factor *of* a *and* b*.*
- *If* c *is a common factor of* a *and* b*, the* $c|d$*.*

The greatest common divisor of 30 and 45 is 15 - we write $\gcd(45, 30) = 15$ (some authors just write $(45, 30) = 15$).

Note: The third item in the definition above says that $\gcd(a, b)$ has the property that it is a multiplie of *every* common divisor of a and b. It might seem more obvious to define the gcd just as the largest amongst all the common divisors of a and b. For integers this amounts to the same thing. The reason for choosing the form in the definition is that it can also be used for polynomials and for other rings whose elements are not necesarily ordered by "greater than" and "less than" relations.

We will show that every pair (a, b) of (non-zero) integers has a unique gcd. We note that we can assume that both a and b are positive, since $-a$ and a have the same integer divisors, as do $-b$ and b.

Given positive integers a and b with $a > b$, we can calculate $\gcd(a, b)$ as in the following example.

Example 4.3. *Calculate* $\gcd(770, 528)$.

Step 1 Write $a = bq_1 + r_1$ where $0 \leqslant r_1 < b$. Then $r_1 = a - bq_1$, so every common divisor of a and b is a divisor of r_1, and hence a common divisor of b and r_1. On the other hand since $a = bq_1 + r_1$, every common divisor of b and r_1 is a divisor of a, and hence a common divisor of a and b. Thus the pairs (a, b) and (b, r_1) have the same sets of common divisors and

$$\gcd(a, b) = \gcd(b, r_1).$$

In our example

$$770 = 528(1) + 242, \quad r_1 = 242.$$

Step 2 Now write $b = r_1q_2+r_2$ where $0 \leqslant r_2 < r_1$. By the above reasoning $\gcd(b, r_1) = \gcd(r_1, r_2) = \gcd(a, b)$.

$$528 = 242(2) + 44, \quad r_2 = 44.$$

Step 3 Now write $r_1 = r_2q_3 + r_3$ with $0 \leqslant r_3 < r_2$. Continuing like this we create a sequence

$$b > r_1 > r_2 > \cdots > 0.$$

This is a strictly decreasing sequence of non-negative integers, so it reaches 0 after a finite number of steps. Each pair of successive terms in the sequence has the same gcd and this is $\gcd(a, b)$. So $\gcd(a, b)$ is the last non-zero term in the sequence. We have

$$242 = 44(5) + 22, \quad r_3 = 22.$$

$$44 = 22(2) + 0, \quad r_4 = 0.$$

We conclude that $\gcd(770, 528) = 22$.

Note $22 = \gcd(22, 44) = \gcd(44, 242) = \gcd(242, 528) = \gcd(528, 770)$.

The procedure that we have just used to calculate $\gcd(770, 528)$ is called the *Euclidean Algorithm*.

Example 4.4. *Calculate* $\gcd(1704, 1344)$ *using the Euclidean Algorithm.*

SOLUTION:

1. $1704 = 1344(1) + 360, \quad r_1 = 360$
2. $1344 = 360(3) + 264, \quad r_2 = 264$
3. $360 = 264(1) + 96, \quad r_3 = 96$
4. $264 = 96(2) + 72, \quad r_4 = 72$
5. $96 = 72(1) + 24, \quad r_5 = 24$
6. $72 = 24(3) + 0.$

So $\gcd(1704, 1344) = 24$.

The next theorem, which is one of the main themes of this chapter, captures an important property of the greatest common divisor.

Theorem 4.5. *Let* a *and* b *be integers and let* $d = \gcd(a, b)$. *Then there exist integers* m *and* n *for which*

$$d = ma + nb.$$

REMARKS

1. What this theorem says about Example 4.4 is that there exist integers m and n for which

   $$24 = 1704m + 1344n.$$

   To understand what the theorem is about, think about what integers you might expect to get by adding a multiple (positive or negative) of 1704 to a multiple (positive or negative) of 1344. Convince yourself that any number that could possibly arise that way would have to be a multiple of 24. What is not obvious yet is that 24 itself arises this way - that is the content of the theorem.

2. More generally, it is easy enough to see that any integer that can be written in the form $ma + nb$ for integers m and n must be divisible by all common divisors of a and b, and hence by $\gcd(a, b)$. It is perhaps less obvious that $\gcd(a, b)$ can be written in this form.

3. Theorem 4.5 can be proved by going backwards through the steps involved in calculating $\gcd(a, b)$ using the Euclidean algorithm. Rather than giving a formal proof of this theorem we will demonstrate how it works by writing 24 in the form $1704m + 1344n$ for integers m and n.

Step 1 Look at Step 5 in the calculation of $\gcd(1704, 1344)$. This says

$$24 = 96 + 72(-1).$$

Step 2 Now use Step 4 to replace 72 with a combination of 96 and 264.

$$24 = 96 + 72(-1) = 96 + (264 + 96(-2))(-1) = 96(3) + 264(-1).$$

Step 3 Use Step 3 to replace 96 with a combination of 360 and 264.

$$24 = 96(3) + 264(-1) = (360 + 264(-1))(3) + 264(-1) = 360(3) + 264(-4).$$

Step 4 Use Step 2 to write 264 as a combination of 1344 and 360. Then

$$24 = 360(3) + 264(-4) = 360(3) + (1344 + 360(-3))(-4) = 360(15) + 1344(-4).$$

Step 5 Finally use Step 1 to write 360 as a combination of 1344 and 1704. Then

$$24 = 360(15) + 1344(-4) = (1704 + 1344(-1))(15) + 1344(-4) = 1704(15) + 1344(-19).$$

So we have succeeded in writing 24 in the form $1704m+1344n$, where $m = 15$ and $n = -19$.

Definition 4.6. *Let* a *and* b *be non-zero integers. The* a *and* b *are* coprime *or* relatively prime *(to each other) if* $\gcd(a, b) = 1$.

Equivalently a and b are relatively prime if they have no common divisors except 1 and -1.

From Theorem 4.5 we can say that a and b are relatively prime if and only if there exist integers m and n for which

$$1 = ma + nb.$$

Example 4.7. *Find integers* m *and* n *for which*

$$1 = 98m + 85n.$$

SOLUTION: First apply the Euclidean algorithm to 98 and 85.

1. $98 = 85(1) + 13$
2. $85 = 13(6) + 7$
3. $13 = 7(1) + 6$
4. $7 = 6(1) + 1$
5. $6 = 1(6) + 0$

Now reverse the steps :

4. $1 = 7 + 6(-1)$
3. $1 = 7 + (13 + 7(-1))(-1) = 7(2) + 13(-1)$
2. $1 = (85 + 13(-6))(2) + 13(-1) = 85(2) + 13(-13)$

1. $1 = 85(2) + (98 + 85(-1))(-13) = 85(15) + 98(-13)$

So we have $1 = 85(15) + 98(-13)$; $m = -13,\ n = 15$.

REMARKS:

1. The integers m and n in these problems are not unique. For example in the above problem we could obtain another solution as follows :

$$1 = 85(15) + 98(-13) = 85(15) + 85(-98) + 98(85) + 98(-13) = 85(-83) + 98(72).$$

   So $m = 72, n = -83$ would be another solution.
   Exercise - think about how all possible solutions are related.

2. Let a and b be non-zero integers. An integer d can be written as $ma + nb$ for integers m and n if and only if $\gcd(a, b)$ divides d.

5 The finite rings $\mathbb{Z}/n\mathbb{Z}$

We will consider some examples of *finite* rings and fields. All of our examples so far have infinitely many elements. For every positive integer n it is possible to define a ring $\mathbb{Z}/n\mathbb{Z}$ (or $\mathbb{Z}_n$) that has exactly n elements. The definitions and basic properties of these rings are discussed in this section. The main theoretical ingredients that are needed are the Euclidean algorithm and the notion of relative primality.

The elements of the ring $\mathbb{Z}/n\mathbb{Z}$ (also denoted $\mathbb{Z}_n$) are

$$\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{n-1}.$$

Addition and multiplication are defined as follows :

- to add $\overline{a}$ and $\overline{b}$, just add a and b in $\mathbb{Z}$ - if the result of this exceeds $n-1$, take its remainder r on division by n; then $\overline{a} + \overline{b} = \overline{r}$ in $\mathbb{Z}/n\mathbb{Z}$.
- to multiply $\overline{a}$ and $\overline{b}$, just multiply a and b in $\mathbb{Z}$ - if the result of this exceeds $n-1$, take its remainder r on division by n; then $\overline{a} \times \overline{b} = \overline{r}$ in $\mathbb{Z}/n\mathbb{Z}$.

NOTE: For any integer x, we think of $\overline{x}$ as the element of $\mathbb{Z}_n$ that is represented by the remainder on dividing x by n. So for example, in $\mathbb{Z}_5$, the elements $\overline{16}, \overline{21}, \overline{41}$ etc., are all the same and they are all the same as the element $\overline{1}$. The elements of $\mathbb{Z}_n$ are really "remainders on division by n".

The following are the addition and multiplication tables of $\mathbb{Z}_5$ and $\mathbb{Z}_6$. In the multiplication table, the zero element is omitted in each case, just because it would only add a row and column of zeroes to the table. We will not go through the process of verifying that the addition and multiplication of $\mathbb{Z}_n$ satisfy all the ring axioms, although really we should.

REMARK ON NOTATION: In the definition above, the bars on the elements $\overline{1}, \overline{2}$ etc. are included to indicate that these are not the usual numbers $1, 2, \ldots$ but are equipped with different arithmetic operations (that depend on the n in question). If we wanted to be really careful and precise, we would include the "n" in the notation for the elements, and use something like $[1]_5$, $[2]_5, \ldots$ for the elements of $\mathbb{Z}_5$. In practice however, this level of specification becomes quite cumbersome, and if the context is clear people often drop the bar as well, as we do in the tables below.

$\mathbb{Z}_5, +$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\mathbb{Z}_5\backslash\{0\}, \times$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

$\mathbb{Z}_6, +$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\mathbb{Z}_6\backslash\{0\}, \times$	1	2	3	4	5
1	1	2	3	4	5
2	2	4	0	2	4
3	3	0	3	0	3
4	4	2	0	4	2
5	5	4	3	2	1

OBSERVATIONS

1. In $\mathbb{Z}_5$, every non-zero element has an inverse for multiplication. The elements 2 and 3 are inverses of each other, and 1 and 4 are their own inverses. So $\mathbb{Z}_5$ is an example of a *field*. It is often referred to as "the field of 5 elements" and sometimes written $\mathbb{F}_5$.

2. On the other hand, $\mathbb{Z}_6$ is not a field. Amongst the non-zero elements of $\mathbb{Z}_6$, only 1 and 5 have inverses for multiplication (and each of these is its own inverse). Amongst the elements 1,2,3,4,5, another thing that is special about 1 and 5 is that they are relatively prime to 6 (which 2,3 and 4 are not).

Theorem 5.1. *The element $\overline{a}$ has an inverse for multiplication in $\mathbb{Z}_n$ if and only if* $\gcd(a, n) = 1$.

Proof. The element $\overline{a}$ has an inverse in $\mathbb{Z} - n$ if and only if there exists a $b \in \mathbb{Z}$ for which $\overline{a}\overline{b} = \overline{1}$ in $\mathbb{Z}_n$. This means exactly that the remainder on dividing ab by n in $\mathbb{Z}$ is 1, which means that

$$ab = nt + 1, \text{ for some } t \in \mathbb{Z}.$$

From the Euclidean algorithm we know that integers t and b for which $1 = nt - ab$ exist if and only if $\gcd(a, n) = 1$. □

Corollary 5.2. *$\mathbb{Z}_n$ is a field if and only if n is prime.*

Proof. Every integer in the range $1, 2, \ldots, n-1$ is relatively prime to n if and only if n is prime. □

Problem: Find the inverse of 14 in $\mathbb{Z}_{33}$.

Solution: First note that $\gcd(14, 33) = 1$, so 14 *has* an inverse in $\mathbb{Z}_{33}$. In order to calculate this inverse, use the Euclidean algorithm to write 1 in the form $14s + 33t$ for integers s and t. First apply the Euclidean algorithm directly.

$$\begin{aligned} 33 &= 14(2) + 5, \ \ r_1 = 5 \\ 14 &= 5(2) + 4, \ \ r_2 = 4 \\ 5 &= 4(1) + 1 \end{aligned}$$

Then reverse the steps

$$\begin{aligned} 1 &= 5 + 4(-1) \\ &= 5 + [(14 + 5(-2)](-1) = 5(3) + 14(-1) \\ &= [33 + 14(-2)](3) + 14(-1) = 33(3) + 14(-7). \end{aligned}$$

Thus $1 = 33(3) + 14(-7)$ which means that $14(-7) = 1 + 33(-3)$ and $14(-7)$ represents the multiplicative identity element of $\mathbb{Z}_{33}$. So the inverse of 14 in $\mathbb{Z}_{33}$ is the element of $\mathbb{Z}_{33}$ represented by -7. To express this in the usual range $0, 1, \ldots, 32$, just add 33 to get 26. Thus the inverse of 14 in $\mathbb{Z}_{33}$ is 26.
CHECK that $14 \times 26 = 1$ in $\mathbb{Z}_{33}$, by checking that 14×26 has remainder 1 on division by 33.

6 More finite fields - extensions of $\mathbb{Z}/p\mathbb{Z}$

The theme of this final section is the use of irreducible polynomials in $\mathbb{Z}_p[x]$ to construct further examples of finite fields, in which the number of elements is a prime power, i.e. a number of the form p^k for a prime p and positive integer k. We know from the last section that the ring $\mathbb{Z}/n\mathbb{Z}$ (or $\mathbb{Z}_n$) is a field if and only if n is prime. For a prime p, we will write $\mathbb{F}_p$ for the field $\mathbb{Z}/p\mathbb{Z}$ of p elements. So we have $\mathbb{F}_2$, $\mathbb{F}_3$, $\mathbb{F}_5$, etc. We will show in this section how to use an irreducible polynomial of degree k in $\mathbb{F}_p[x]$ to construct a field with p^k elements. It's not obvious in advance why an irreducible polynomial of degree k must exist in $\mathbb{F}_p[x]$ for a given p and k - to prove this would take another course in field theory. But we will be able to produce examples of irreducible quadratic and cubic polynomials over finite fields that are small enough to deal with by hand.

First we mention a few familiar examples of using irreducible polynomials to construct fields.

1. The polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ - we know this because it has no real root, a root of this polynomial would be a square root of -1. The field $\mathbb{C}$ of complex numbers is constructed by introducing an "imaginary" square root i of -1 and taking all numbers of the form $a + bi$ with $a, b \in \mathbb{R}$. That this produces a field can be checked directly as we discussed in Lecture 1.

2. Similarly, the polynomial $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$ (since there is no square root of 2 in $\mathbb{Q}$). We can use this irreducible polynomial to construct a field $\mathbb{Q}(\sqrt{2})$ consisting of all numbers of the form $a + b\sqrt{2}$ where $a, b \in \mathbb{Q}$. That this is a field is easy enough to check - the only issue really is to check that every non-zero number of the form $a + b\sqrt{2}$ (with $a, b \in \mathbb{Q}$) has an inverse of the same form. To do this, first notice that $(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2$ is a rational number. Then the inverse of $a + b\sqrt{2}$ is given by

$$\frac{1}{a^2 - 2b^2}(a - b\sqrt{2}) = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}.$$

   So for example the inverse of $1 + \sqrt{2}$ is $-1 + \sqrt{2}$ (check).
   Note that this description of how to write down the inverse of an element of $\mathbb{Q}(\sqrt{2})$ is entirely analogous to how we calculate the inverse of a complex number.

3. The polynomial $x^3 - 2$ is also irreducible in $\mathbb{Q}[x]$. If we introduce a root $\sqrt[3]{2}$ of this polynomial, we can construct a field $\mathbb{Q}(\sqrt[3]{2})$ that consists of all numbers of the form

$$a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2, \text{ where } a, b, c \in \mathbb{Q}.$$

   Showing directly that this set of numbers is a field takes a bit of work and patience, but it can be done.

We now show how to use exactly the same ideas to extend the field $\mathbb{F}_p$ to a larger field. As an example we will construct a field $\mathbb{F}_9$ with 9 elements ($9 = 3^2$).

STEP 1: Find an irreducible quadratic polynomial in $\mathbb{F}_3[x]$. This is relatively easy - it is just a matter of trying out some quadratics with coefficients in $\mathbb{F}_3$ and finding one with no roots in $\mathbb{F}_3$. Checking for roots in $\mathbb{F}_3$ is easy because there are only three candidates. Try

$$f(x) = x^2 + 1.$$

Note $f(0) = 1$, $f(1) = 1^2 + 1 = 2$, $f(2) = 2^2 + 1 = 2$ in $\mathbb{F}_3$. Since $f(x)$ has no root in $\mathbb{F}_3$ it is irreducible in $\mathbb{F}_3[x]$.

STEP 2: Introduce a root α of $f(x)$ (exactly as we introduce an imaginary square root of -1 in order to construct $\mathbb{C}$ from $\mathbb{R}$). Then

$$\alpha^2 + 1 = 0 \Longrightarrow \alpha^2 = 2.$$

The elements of the field $\mathbb{F}_9$ will have the form $a + b\alpha$ where $a, b \in \mathbb{F}_3$. So the elements of $\mathbb{F}_9$ are

$$0, 1, 2, \alpha, 2\alpha, 1+\alpha, 2+\alpha, 1+2\alpha, 2+2\alpha.$$

Checking that what we have constructed is a field is relatively easy, since we have only nine elements we can write out the full addition and multiplication tables. When multiplying elements of $\mathbb{F}_9$ we use the fact that $\alpha^2 = 2$ (just as we use $i^2 = -1$ when multiplying complex numbers).

STEP 3: The addition and multiplication tables of $\mathbb{F}_9$.

$\mathbb{F}_9, +$	0	1	2	α	2α	$1+\alpha$	$2+\alpha$	$1+2\alpha$	$2+2\alpha$
0	0	1	2	α	2α	$1+\alpha$	$2+\alpha$	$1+2\alpha$	$2+2\alpha$
1	1	2	0	$1+\alpha$	$1+2\alpha$	$2+\alpha$	α	$2+2\alpha$	2α
2	2	0	1	$2+\alpha$	$2+2\alpha$	α	$1+\alpha$	2α	$1+2\alpha$
α	α	$1+\alpha$	$2+\alpha$	2α	0	$1+2\alpha$	$2+2\alpha$	1	2
2α	2α	$1+2\alpha$	$2+2\alpha$	0	α	1	2	$1+\alpha$	$2+\alpha$
$1+\alpha$	$1+\alpha$	$2+\alpha$	α	$1+2\alpha$	1	$2+2\alpha$	2α	2	0
$2+\alpha$	$2+\alpha$	α	$1+\alpha$	$2+2\alpha$	2	2α	$1+2\alpha$	0	1
$1+2\alpha$	$1+2\alpha$	$2+2\alpha$	2α	1	$1+\alpha$	2	0	$2+\alpha$	α
$2+2\alpha$	$2+2\alpha$	2α	$1+2\alpha$	2	$2+\alpha$	0	1	α	$1+\alpha$

When writing out the mutiplication table, bear in mind that whenver we encounter a term involving α^2 we should make use of the fact that $\alpha^2 = 2$. Thus for example

$$(1+\alpha)(1+2\alpha) = 1 + 3\alpha + 2\alpha^2 = 1 + 0 + 2(2) = 1 + 1 = 2.$$

$\mathbb{F}_9\backslash\{0\}, \times$	1	2	α	2α	$1+\alpha$	$2+\alpha$	$1+2\alpha$	$2+2\alpha$
1	1	2	α	2α	$1+\alpha$	$2+\alpha$	$1+2\alpha$	$2+2\alpha$
2	2	1	2α	α	$2+2\alpha$	$1+2\alpha$	$2+\alpha$	$1+\alpha$
α	α	2α	2	1	$2+\alpha$	$2+2\alpha$	$1+\alpha$	$1+2\alpha$
2α	2α	α	1	2	$1+2\alpha$	$1+\alpha$	$2+2\alpha$	$2+\alpha$
$1+\alpha$	$1+\alpha$	$2+2\alpha$	$2+\alpha$	$1+2\alpha$	2α	1	2	α
$2+\alpha$	$2+\alpha$	$1+2\alpha$	$2+2\alpha$	$1+\alpha$	1	α	2α	2
$1+2\alpha$	$1+2\alpha$	$2+\alpha$	$1+\alpha$	$2+2\alpha$	2	2α	α	1
$2+2\alpha$	$2+2\alpha$	$1+\alpha$	$1+2\alpha$	$2+\alpha$	α	2	1	2α

Note that we can see from this table that every non-zero element of $\mathbb{F}_9$ has an inverse in $\mathbb{F}_9$ and we can read off what these inverses are. For example the inverse of $1 + 2\alpha$ is $2 + 2\alpha$. What is the inverse of $2 + \alpha$?

This detailed example is indicative of how a finite field of order p^k can generally be constructed from an irreducible polynomial of degree k with coefficients in $\mathbb{F}_p$. To write out the full multiplication table by hand becomes unfeasible even for $p = 5$ and $k = 2$ (this would be a field of 25 elements). However, it is a good exercise to use an irreducible quadratic and an irreducible cubic in $\mathbb{F}_2[x]$ to construct (respectively) a field $\mathbb{F}_4$ with 4 elements and a field $\mathbb{F}_8$ with 8 elements (see the current problem sheet).