1. Rings

A **ring** is \((R, +, \cdot)\), consisting of a set \(R\) and two binary operations
\[ + : R \times R \rightarrow R, \quad \cdot : R \times R \rightarrow R, \]
called “addition” and “multiplication”,
- \((R, +)\) is an additive group,
- \((R, \cdot)\) is a monoid, and
- multiplication distributes over addition:
\[ a(b + c) = (ab) + (ac), \quad (a + b)c = (ac) + (bc). \]

The identity for + is conventionally called 0, and the inverse of \(a\) under + is called \(-a\). The identity for \(\cdot\) is called 1.

1.1. **Remark.** There is an “associativity” issue when you have two different binary operations: in principle, you can interpret “\(a + b \cdot c\)” as meaning either:
\[ (a + b) \cdot c \quad \text{or} \quad a + (b \cdot c). \]

We all know the solution: multiplication is always assumed to have precedence over addition, so we read this as \(a + (b \cdot c)\).

1.2. **Remark.** A **ring without identity** is one in which \((R, \cdot)\) is merely assumed to be a semi-group, i.e., there is no multiplicative identity. (Some people call this a **rng**.) In many introductory textbooks, “ring” is defined to mean what we are calling a “ring without identity”. Which is odd, since such textbooks have few or no examples of rings without identity, and for the most part, “ring” means the definition I am using for most mathematicians.

A ring is said to be **commutative** if multiplication is commutative: \(ab = ba\) for all \(a, b \in R\).

Basic examples of rings include:
- The real numbers \(\mathbb{R}\) (which is commutative).
- The integers \(\mathbb{Z}\) (commutative).
- The integers modulo \(n\): \(\mathbb{Z}/n\) (commutative).
- The ring \(R = M_{n \times n}(\mathbb{R})\) of \(n \times n\)-matrices over \(\mathbb{R}\). Not commutative if \(n \geq 2\).

1.3. **Example.** For any set \(X\) and ring \(R\), let \(S := \mathcal{F}(X, R) = \{f : X \rightarrow R\}\) be the set of all functions. Then \(S\) is a ring, with operations given by “pointwise” addition and multiplication:
\[ (f + g)(x) := f(x) + g(x), \quad (fg)(x) := f(x)g(x). \]

(Exercise: check that this is a ring.) For instance, the set \(\mathcal{F}(\mathbb{R}, \mathbb{R})\) of real valued functions on \(\mathbb{R}\) is a ring.

If \(S\) is a ring, a **multiplicative inverse** of an element \(a \in S\) is an element \(b \in S\) such that
\[ ab = 1 = ba. \]

Clearly not all elements of a ring can have a multiplicative inverse. For instance, \(0 \in \mathbb{R}\) has none.

---

*Date:* November 13, 2017.
1.4. **Exercise.** If \( a \in S \) has a multiplicative inverse, then this inverse is unique. (Same as the proof for inverses in groups.)

We write \( S^\times \subseteq S \) for the set of elements which have multiplicative inverses.

1.5. **Exercise.** \((S^\times, \cdot)\) is a group.

Examples: \( \mathbb{R}^\times = \mathbb{R} \setminus \{0\} \), \( \mathbb{Z}^\times = \{\pm 1\} \), \( (\mathbb{Z}/n)^\times = \Phi(n) \), \( M_{n \times n}(\mathbb{R})^\times = \text{GL}_n(\mathbb{R}) \).

We say that a ring \( S \) is a **division ring** if \( S^\times = S \setminus \{0\} \), i.e., every non-zero element has a multiplicative inverse, and 0 does not have one.

A commutative division ring is called a **field**. For instance, \( \mathbb{R} \) is a field. Also \( \mathbb{Z}/p \) is a field when \( p \) is prime.

Let’s carefully construct some more examples of rings.

2. **Complex numbers and quaternions**

We define a ring \( \mathbb{C} \) as follows.

- The set of \( \mathbb{C} \) is \( \mathbb{R} \times \mathbb{R} = \{(a,b) \mid a, b \in \mathbb{R}\} \).
- Addition is defined by \((a,b) + (a',b') := (a + a', b + b').\)
- Multiplication is defined by \((a,b)(a',b') := (aa' - bb', ab' + ba').\)

2.1. **Proposition.** \( \mathbb{C} \) is a field.

**Proof sketch.** The operations of + and \( \cdot \) are certainly well-defined. To show that \( \mathbb{C} \) is a ring we need to check the following.

(1) \((\mathbb{C}, +)\) is an abelian group. In fact we have already done this: as a group \((\mathbb{C}, +)\) is the product group of \((\mathbb{R}, +)\) with itself. (Or: it is the same as the additive group of the vector space \( \mathbb{R}^2 \).)

(2) \((\mathbb{C}, \cdot)\) is a monoid. We need to check:
   - Multiplication is associative. This is a little tedious, but here we go:
     \[
     (a,b)(a',b')(a'',b'') = (a'a'' - bb'', ab'b'' - ba'a'' + ba'a'')
     = (aa'a'' - bb'a'' - ab'b'' - ba'a'' + ab'a'' + ba'a''),
     \]
     \[
     (a,b)(a',b')(a'',b'') = (a,b)(a'a'' - bb'' + b'a'')
     = (aa'a'' - ab'b'' - ba'a'' - bb'a'', aa'a'' + ab'a'' + ba'a'' - bb'b''),
     \]
     which are the same.
   - There is a multiplicative unit, which I’ll call 1. In fact, if \( 1 = (1, 0) \), we check that \((1,0)(a,b) = (a,b) = (a,b)(1,0)\).
(3) The distributive law:
\[(a, b)((a', b') + (a'', b'')) = (a, b)(a' + a'', b' + b'')\]
\[= (aa' + aa'' - bb' - bb'', ab' + ab'' + ba' + ba'')\]
\[= (aa' - bb', ab' + ba') + (aa'' - bb'', ab'' + ba'')\]
\[= (a, b)(a', b') + (a, b)(a'', b''),\]
\[(a, b) + (a', b') = (a + a', b + b'),\]
\[(a, b)(a', b') = (aa' - bb', ab' + ba') = (a'a - b'b, b'a + a'b) = (a', b')(a, b).
\]

It is clear from the formula that multiplication is commutative:
\[(a, b)(a', b') = (aa' - bb', ab' + ba') = (a'a - b'b, b'a + a'b) = (a', b')(a, b).
\]

To show that \(\mathbb{C}\) is a field, we have to produce a multiplicative inverse for each \((a, b)\) which is not equal to \((0, 0)\). (Clearly \((0, 0)\) can have no multiplicative inverse, since \((0, 0)(a, b) = (0, 0) \neq (1, 0)\).)

In fact, given \(x = (a, b)\), define
\[y := \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right)\]
This is well defined since \(a^2 + b^2 > 0\) if \((a, b) \neq (0, 0)\). Now check that \(xy = (1, 0)\). (Exercise.)

In practice we identify a real number \(a \in \mathbb{R}\) with the element \((a, 0) \in \mathbb{C}\), and we write \(i := (0, 1)\), so that we write
\[a + bi\]
instead of \((a, b)\).

Of course \(i^2 = -1\). This is the field of complex numbers.

William Hamilton (early 1800s) was one of the first people to realize that complex numbers could be described in this way, as ordered pairs of real numbers plus two binary operations. He worked for many years to do some thing similar for ordered triples of real numbers, but could not. Eventually he discovered that he could if he used ordered 4-tuples. This is the ring of **quaternions**, denoted \(\mathbb{H}\). (Setting this up is left as exercises.)

The quaternions are a non-associative ring with the property that the non-zero elements are exactly the ones with multiplicative inverses. Thus \(\mathbb{H}\) is a division ring.

3. Basic facts

Here are some basic facts about rings.

- The additive identity 0 is unique. (Because it is the additive identity in a group.)
- The multiplicative identity 1 is unique. (By the same proof we used for a group.)
- For any \(a \in S\), we have that
\[0a = 0 = a0.
\]
Proof: \(0 + 0 = 0\), so by the distributive law
\[0a = (0 + 0)a = 0a + 0a.
\]
By cancellation in the additive group we get \(0 = 0a\). A similar proof shows \(a0 = 0\).
- Let \(-1\) be the additive inverse of 1. Then we always have the formula
\[(-1)a = -a = a(-1),
\]
where \(\text{"-a"}\) is the additive inverse of \(a\). Proof:
\[a + (-1)a = (1)a + (-1)a = (1 + (-1))a = 0a = 0,
\]
where we use the distributive law in the second step.
• It is possible for a ring to have $1 = -1$, which implies $a = -a$ for all $a \in S$. For instance $S = \mathbb{Z}/2$.

• It is possible for a ring to have $1 = 0$. The set $S = \{0\}$ with one element has exactly one binary operation, which we use as the definition of both $+$ and $\cdot$, so $0 + 0 = 00 = 0$. Check that this is a ring, with $1 = 0$.

Exercise: if $S$ is a ring with $1 = 0$, then $S$ has only one element.

It is not generally the case that $a \neq 0$ and $b \neq 0$ implies $ab \neq 0$.

3.1. Example. In $S = M_{2 \times 2}(\mathbb{R})$, there exist non-zero matrices $A, B$ such that $AB = 0$.

3.2. Example. In $\mathbb{Z}/4$, we have $[2][2] = [0]$.

We say that a commutative ring $S$ is a domain if the set $S \setminus \{0\}$ is a monoid with respect to multiplication. Equivalently, $S$ is a domain if $1 \neq 0$ and $ab = 0$ implies either $a = 0$ or $b = 0$.

Examples of domains include all fields, and also the integers $\mathbb{Z}$.

4. SUBRINGS

A subring of a ring $R$ is a subset $S \subseteq R$ such that (i) the $+$ and $\cdot$ operations restrict to $S$, and make $S$ a ring in its own right, and (ii) $R$ and $S$ have the same multiplicative identity.

Here is the subring criterion.

4.1. Proposition. A subset $S \subseteq R$ of a ring is a subring if and only if

1. $x, y \in S$ implies $x + y \in S$, i.e., $S$ is closed under addition,
2. $x \in S$ implies $-x \in S$,
3. $x, y \in S$ implies $xy \in S$,
4. $1 \in S$, where $1$ denotes the multiplicative identity in $R$.

Proof. Note that by (4) $S$ is non-empty. Therefore together with (1) and (2) we see that $(S, +)$ is a subgroup of $(R, +)$. Property (3) implies that multiplication is a binary operation on $S$. It is straightforward to check the remaining properties (that multiplication is associative, that $1$ is a multiplicative identity, the distributive law) on $S$, because they hold in $R$. □

4.2. Example. The integers $\mathbb{Z}$ are a subring of $\mathbb{R}$.

4.3. Example. The rational numbers $\mathbb{Q}$ are a subring of $\mathbb{R}$.

4.4. Example. The inclusion $2\mathbb{Z} \subset \mathbb{Z}$ is not a subring. Although closed under the operations the subset does not have a multiplicative identity.

4.5. Example. Let $S \subset M_{2 \times 2}(\mathbb{R})$ be the subset consisting of $2 \times 2$ real matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad a, b \in \mathbb{R}.$$ 

We can check that $S$ is a subring of the ring of matrices. (Verify this.)

4.6. Example. Let $T = \{ \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \mid a \in \mathbb{R} \}$. This is a subset of the ring $S = M_{2 \times 2}(\mathbb{R})$, which is closed under $+$ and $\cdot$, and in fact as such it is a ring in its own right: it’s multiplicative identity is $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

However, we will not consider it as a subring, because the multiplicative identity of $T$ is not the same as the one for $S = M_{2 \times 2}(\mathbb{R})$, which is the identity matrix. (Note: other sources may differ here, and will consider $T$ a subring. I won’t however.)

Compare with groups, where if $H \subseteq G$ is a subset closed under multiplication, and $H$ has an identity element for its product, then the identity element of $H$ must be the same as that for $G$. (The proof of this used the existence of inverses in groups.) Rings are just different.

4.7. Example. The set $C(\mathbb{R}, \mathbb{R})$ of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$ is a subring of $\mathcal{F}(\mathbb{R}, \mathbb{R})$, commutative with identity. This is because of the fact that sums and products of continuous functions are continuous; it has identity because constant functions are continuous.
5. Polynomials

Let $S$ be any ring. A **sequence** in $S$ is a function
\[ a : \mathbb{Z}_{\geq 0} \rightarrow S. \]
I’ll use the notation $a_n \in S$ for the value of this function at $n$, i.e., I’m thinking of $a$ as an infinite sequence.

We define a new ring $P(S)$ as follows.
- Elements of $P(S)$ are sequences $a : \mathbb{Z}_{\geq 0} \rightarrow S$ for which there exists $N \in \mathbb{Z}_{\geq 0}$ such that $a_k = 0$ for all $k > N$. I.e., after finitely many terms the sequence is constant at 0.
- Addition is defined by the “pointwise addition” rule:
  \[ (a + b)_n := a_n + b_n. \]
- Multiplication is defined by the rule:
  \[ (ab)_n := \sum_{i=0}^{n} a_ib_{n-i} = a_nb_0 + a_{n-1}b_1 + \cdots + a_1b_{n-1} + a_0b_n. \]

E.g., $(ab)_0 = a_0b_0$, $(ab)_1 = a_1b_0 + a_0b_1$, $(ab)_2 = a_2b_0 + a_1b_1 + a_0b_2$, etc.

We need to make sure these operations are well-defined, because of the requirement that sequences in $P(S)$ are eventually 0. For instance, given $a, b \in P(S)$ let $N$ be such that $a_k = b_k = 0$ for all $k > N$. Then clearly $(a + b)_k = 0$ for $k > N$, while $(ab)_k = 0$ for $k > 2N$.

**Exercise (Tedious).** With this structure $P(S)$ is a ring. I’ll just note some features of this:
- The additive identity is the zero sequence: $0_n = 0$ for all $n$.
- The multiplicative identity is the sequence $1$ defined by $1_0 = 1$, $1_k = 0$ for $k > 0$.
- If $S$ is commutative, so is $P(S)$.
- Associativity of multiplication is the hardest part to prove, but is is not too bad if you are good at multiple summations:
  \[ ((ab)c)_n = \sum_{i=0}^{n} (ab)_ic_{n-i} = \sum_{i=0}^{n} \sum_{j=0}^{i} a_jb_{i-j}c_{n-i}, \]
  \[ (a(bc))_n = \sum_{k=0}^{n} a_k(bc)_{n-k} = \sum_{k=0}^{n} \sum_{\ell=0}^{n-k} a_kb_{\ell}c_{n-k-\ell}. \]

These work out to the same thing, once you reindex the sums (so that $k = j$ and $\ell = i - j$, whence $n - k - \ell = n - i$). You will actually understand this better by working it out for small values of $n$, like 1 or 2 or 3.

Let $X \in P(S)$ denote the sequence defined by
\[ X_1 = 1, \quad X_k = 0 \text{ if } k \neq 1. \]
If we multiply $X$ by itself a bunch of times, we get $X^n$, which is the sequence with $(X^n)_n = 1$ and $(X^n)_k = 0$ if $k \neq 0$. Given $c \in S$, we use the same symbol $c$ to denote the sequence: $c_0 = c$, $c_k = 0$ if $k > 0$. With this notation we can write any $a \in P(S)$ as the expression
\[ a_0 + a_1X + a_2X^2 + \cdots + a_nX^n, \]
assuming $a_k = 0$ for $k > n$. We often choose to denote such an expression as “$f(X)$”, instead of as “$a$”. 

In other words, $P(S)$ is the ring of **polynomials** in one unknown with coefficients in $S$.

**Warning.** Polynomials are not defined as functions, and they are not the same thing as functions. I’ll talk about this later. 

Another notation for $P(S)$ is $S[X]$. (This is convenient when we want to name the “variable”.)

5.2. **Exercise** (On PS 9). If $D$ is a domain, then $D[X]$ is a domain.

This is also important because we can iterate the construction. Thus we may consider $P(P(S))$, aka $(S[X])[Y]$. Elements $f$ in this ring are expressions

$$f = g_0 + g_1Y + g_2Y^2 + \cdots + g_nY^n,$$

where each $g_k \in S[X]$, so are expressions

$$g_k = a_{0k} + a_{1k}X + a_{2k}X^2 + \cdots a_{2m}X^m$$

with $a_{ij} \in S$. Using the distributive law, we can always rewrite this as

$$f = \sum_{i=0}^m \sum_{j=0}^n a_{ij}X^mY^n.$$ 

We write $S[X,Y]$ for $(S[X])[Y]$, and call it the ring of polynomials in two variables. As we will see soon, the order of the variables isn’t really important: $(S[X])[Y]$ and $(S[Y])[X]$ are the “same” ring (really, they are canonically isomorphic).

You can go on to define $S[X,Y,Z]$, etc.

6. **Center of a ring**

Given a ring $S$ let

$$\text{Cent}(S) := \{a \in S \mid ab = ba \text{ for all } b \in S\},$$

called the **center** of $S$.

6.1. **Exercise** (On PS 9?). The set $R = \text{Cent}(S)$ is a subring of $S$. As a ring $R$ is commutative.

Note that if $S$ is commutative, then $\text{Center}(S) = S$.

6.2. **Example.** The center of the quaternion algebra $\mathbb{H}$ is the subset $\mathbb{R}\mathbf{1} = \{\lambda \mathbf{1} \mid \lambda \in \mathbb{R}\}$ of scalar quaternions. It’s straightforward to check that scalar quaternions are in the center. To see these are the only ones, check commutativity with $i$, $j$, and $k$. For instance, commutativity with $i$ gives

$$(a\mathbf{1} + bi + cj + dk)i = -b\mathbf{1} + ai + dk - cj,$$

$$i(a\mathbf{1} + bi + cj + dk) = -b\mathbf{1} + ai - dk + ck,$$

which means that if $x = a\mathbf{1} + bi + cj + dk$ is in the center then $c = 0 = d$. Checking commutativity with $j$ gives $b = 0$.

6.3. **Exercise** (On PS 9). Let $S = M_{n \times n}(F)$ where $F$ is a field (e.g., $F = \mathbb{R}$). Then

$$\text{Center}(S) = \{\lambda \mathbf{I} \mid \lambda \in F\},$$

the set of diagonal matrices. Thus $\text{Center}(S) \approx F$. 
7. HOMOMORPHISMS AND ISOMORPHISMS OF RINGS

Let $R$ and $S$ be rings. A homomorphism $\phi: R \to S$ is a function such that

- $\phi(a + b) = \phi(a) + \phi(b)$ for all $a, b \in R$,
- $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in \mathbb{R}$, and
- $\phi(1) = 1$.

Note: in groups we did not need a condition (3), because it was true anyway. This is because it was implied by the property that a homomorphism preserve products, which was because elements of groups always have inverses.

Easy fact: if $\phi$ is a homomorphism, then we also have:

$$\phi(0) = 0, \quad \phi(-a) = -\phi(a),$$

since $\phi$ is also a homomorphism of groups $(R, +) \to (S, +)$. We also have the following.

7.1. Proposition. If $a$ has a multiplicative inverse, then so does $\phi(a)$, with $\phi(a)^{-1} = \phi(a^{-1})$.

Proof. Just verify that $\phi(a)\phi(a^{-1}) = \phi(aa^{-1}) = \phi(1) = 1$ and $\phi(a^{-1})\phi(a) = \phi(a^{-1}a) = \phi(1) = 1$. \qed

7.2. Example. Let $R$ be a ring. There is a unique homomorphism of abelian groups $\phi: \mathbb{Z} \to R$ which sends the generator 1 of $\mathbb{Z}$ to $1 \in R$. Thus, $\phi(0) = 0$ (since 0s are identity elements for additive groups); we have $\phi(-1) = -1$ since homomorphisms take inverses to inverses; if $m > 0$, we have

$$\phi(m) = \phi(\underbrace{1 + \cdots + 1}_{m \text{ times}}) = \underbrace{\phi(1) + \cdots + \phi(1)}_{m \text{-times}} = m \cdot 1.$$  

Similarly, $\phi(-m) = -m \cdot 1$.

7.3. Remark. In practice, for any ring $S$ we usually just write the integer $m$ to also denote the element in $S$ given by $m1$ as above. This can be a little confusing. For instance, it is possible to have a non-zero integer $n$ whose image in $S$ is 0.

7.4. Example. Consider the projection map $\phi: \mathbb{Z} \to \mathbb{Z}/n$, defined by $\phi(x) := [x]_n$. This is a ring homomorphism.

An isomorphism of rings is a homomorphism which is a bijection. You can show that the inverse map is also a bijection.

7.5. Proposition. If $\phi: R \to S$ is a homomorphism of rings, then $\phi(R)$ is a subring of $S$. If $\phi$ is injective then it defines an isomorphism between $R$ and $\phi(R)$.

8. AUTOMORPHISMS OF RINGS

An automorphism of a ring is an isomorphism $\phi: R \to R$ from the ring to itself. The set $\text{Aut}(R)$ of automorphisms is a group under composition.

Warning. This group of ring automorphisms of $R$ is not the same thing as the group of group automorphisms of $(R, +)$, even though we use the same notation for both. I may write $\text{Aut}_{\text{group}}$ and $\text{Aut}_{\text{ring}}$ to distinguish them.

8.1. Example. The complex-conjugation function $\phi: \mathbb{C} \to \mathbb{C}$ defined by $\phi(a + bi) = a - bi$ is an automorphism of $\mathbb{C}$. Note that $\phi\phi = \text{id}$, so that the subgroup $G = \langle \phi \rangle \leq \text{Aut}(\mathbb{C})$ has order 2 and acts on $\mathbb{C}$ through ring automorphisms.

(Note: confusingly, complex “conjugation” has nothing to do with the “conjugation” automorphism described below.)

8.2. Example. $\text{Aut}(\mathbb{Z}) \approx \{\text{id}\}$. To see this, note that by definition any homomorphism $\phi: \mathbb{Z} \to \mathbb{Z}$ must send $\phi(1) = 1$. But then $\phi(n) = \phi(1 + \cdots + 1) = \phi(1) + \cdots \phi(1) = n$ for $n > 0$. Also $\phi(-1) = -\phi(1) = -1$, so we can extend the result to show $\phi(n) = n$. 

8.3. Example. \( \text{Aut}(\mathbb{Q}) \approx \{\text{id}\} \). Given \( \phi : \mathbb{Q} \to \mathbb{Q} \), the above argument shows \( \phi(n) = n \) for any integer \( n \). If \( r = ab^{-1} \in \mathbb{Q} \) with \( a, b \in \mathbb{Z} \) and \( b \neq 0 \), then \( a = rb \) so \( \phi(a) = \phi(r)\phi(b) \) which becomes \( a = \phi(r)b \), and thus \( \phi(r) = ab^{-1} = r \).

8.4. Example (Don’t need to know this, it’s just interesting). \( \text{Aut}(\mathbb{R}) \approx \{\text{id}\} \). This uses the following properties of the real numbers:

1. In \( \mathbb{R} \), the non-negative numbers are exactly the ones which have a square root. Therefore, for \( a, b \in \mathbb{R} \), we have that

\[
a \leq b \quad \text{if and only if} \quad \exists x \in \mathbb{R} \text{ such that } b - a = x^2.
\]

2. Every real number is determined by the rational numbers that are smaller than it. That is, for \( a \in \mathbb{R} \) let \( C_a = \{ q \in \mathbb{Q} \mid q \leq a \} \). Then \( a = b \) if and only if \( C_a = C_b \) are the same set.

Suppose \( \phi : \mathbb{R} \to \mathbb{R} \) is an automorphism.

- By the same proof as above, \( \phi(q) = q \) if \( q \in \mathbb{Q} \) is a rational number.
- If \( a \leq b \), then \( \phi(a) \leq \phi(b) \). The proof exactly uses (1): if \( a \leq b \), then \( \exists x \in \mathbb{R} \) such that \( b - a = x^2 \). Then

\[
\phi(b) - \phi(a) = \phi(b-a) = \phi(x^2) = \phi(x)^2,
\]

so there exists \( x' = \phi(x) \) such that \( \phi(b) - \phi(a) = x' \), i.e., \( \phi(a) \leq \phi(b) \).
- Therefore, \( C_{\phi(a)} = C_a \), since \( q \in \mathbb{Q} \) satisfies \( q \leq a \) if and only if \( q = \phi(q) \leq \phi(a) \). Hence from (2) we must have \( \phi(a) = a \).

8.5. Example. \( \text{Aut}(\mathbb{C}) \) is uncountably infinite. Most elements in this group (i.e., except identity and complex conjugation) can’t be written down explicitly: their construction relies on the Axiom of Choice.

Any unit \( u \in R^\times \) defines an automorphism \( \text{conj}_u : R \to R \) by \( \text{conj}_u(x) = uxu^{-1} \). Such automorphisms are called **inner automorphisms** of the ring.

8.6. Example. \( \text{conj}_i : \mathbb{H} \to \mathbb{H} \) is given by \( \text{conj}_i(a1 + bi + cj + dk) = i(a1 + bi + cj + dk)(-i) = a1 + bi - cj - dk \) is an automorphism.

Note that \( \text{conj}_u \text{conj}_v = \text{conj}_{uv} \). We get a homomorphism of groups

\[
\text{conj} : R^\times \to \text{Aut}(R), \quad u \mapsto \text{conj}_u.
\]

Exercise: \( \text{Ker}(\text{conj}) = \text{Center}(R)^\times \), so we get an isomorphism from \( R^\times / \text{Center}(R)^\times \) to the subgroup of inner automorphisms inside \( \text{Aut}(R) \).

9. **Polynomials as functions**

Let \( S \) be a commutative ring. We have two new commutative rings:

- The ring \( S[X] \) of polynomials with coefficients in \( S \).
- The ring \( \mathcal{F}(S, S) \) of functions from \( S \) to \( S \), with pointwise addition and multiplication:

\[
(f + g)(s) := f(s) + g(s), \quad (fg)(s) := f(s)g(s).
\]

Define

\[
\psi : S[X] \to \mathcal{F}(S, S)
\]

by the rule

\[
\psi(p)(s) := \sum_{k=0}^{n} a_k s^k \quad \text{if} \quad p = \sum_{k=0}^{n} a_k X^k, \quad a_k \in S
\]

That is, \( \psi(p) \) is the function associated to the polynomial \( p \).

The function \( \psi \) need not be injective.
9.1. Example. Let $S = \mathbb{Z}/p$ where $p$ is a prime number. Let $p = x^p - x$ in $S[x]$. Then I claim that $\psi(p) = 0$. This amounts to showing that $c^p - c = 0$ for all $c \in \mathbb{Z}/p$, which is Fermat’s little theorem.

Proof of Fermat’s little theorem: either $c = 0$ or $c \neq 0$. If $c = 0$ then obviously $0^p - 0 = 0$. If $c \neq 0$ then $c$ has a multiplicative inverse, i.e., is in the group $\Psi(p) = (\mathbb{Z}/p)^\times$. Since this group has order $p - 1$ we must have $c^{p-1} = 1$, and thus $c^p = c$.

Here’s another proof that $\psi$ is not injective: $\mathbb{Z}/p[X]$ is an infinite set (countably infinite), but $\mathcal{F}(\mathbb{Z}/p, \mathbb{Z}/p)$ is a finite set (size $p^p$).

10. Universal property of $S[X]$

The following proposition tells you how to construct homomorphisms out of a polynomial ring $S[X]$. It is an example of what is called a “universal property” for $S[X]$.

10.1. Proposition. Let $S$ and $T$ be rings. Suppose given

(1) a ring homomorphism $\phi: S \rightarrow T$, and
(2) an element $c \in T$, such that
(3) $\phi(s)c = c\phi(s)$ for all $s \in S$.

Then there exists a unique ring homomorphism

$$\psi: S[X] \rightarrow T$$

such that (i) $\psi(X) = c$ and (ii) $\psi(s) = \phi(s)$ for all $s \in S$.

Note that if $T$ is commutative, then (3) is automatically true.

Proof. Existence. We define $\psi$ by the following rule. If $f \in S[X]$ is given by $f = \sum_{i=0}^{n} a_i X^i$ with $a_i \in S$, then set

$$\psi(f) := \sum_{i=0}^{n} \phi(a_i)c^i.$$

Verify directly that this is a ring homomorphism: i.e., that it preserves addition, multiplication, and multiplicative identity.

I’ll do the case of multiplication, which is the only part that needs hypothesis (3). Let $f = \sum_i a_i X^i$ and $g = \sum_j b_j X^j$. We have

$$\psi(fg) = \psi\left(\left(\sum a_i X^i\right)\left(\sum b_j X^j\right)\right)$$

$$= \psi\left(\sum_{n} \left(\sum_{i=0}^{n} a_i b_{n-i}\right) X^n\right)$$

$$= \sum_{n} \phi\left(\sum_{i=0}^{n} a_i b_{n-i}\right)c^n$$

$$= \sum_{n} \left(\sum_{i=0}^{n} \phi(a_i)\phi(b_{n-i})\right)c^n$$

$\phi$ is a homomorphism,
while
\[
\psi(f)\psi(g) = \psi(\sum_i a_iX^i)\psi(\sum_j b_jX^j) = (\sum_i \phi(a_i)c^i)(\sum_j \phi(b_j)c^j) = \sum_{i,j} \phi(a_i)c^i\phi(b_j)c^j = \sum_{i,j} \phi(a_i)c^iX^j
\]
definition of \(\psi\),
\[
= \sum_i \sum_{n=0}^{\infty} \phi(a_i)c^i\phi(b_j)c^j = \sum_{n=0}^{\infty} \phi(a_i)c^iX^n
\]
distributive law,
\[
= \sum_i \sum_{n=0}^{\infty} \phi(a_i)c^i\phi(b_j)c^j = \sum_{n=0}^{\infty} \phi(a_i)c^i
\]
condition (3).

**Uniqueness.** Conversely, given any ring homomorphism \(\psi: S[X] \to T\) such that \(\psi(s) = \phi(s)\) for \(s \in S\), and \(\psi(X) = c\), the properties of ring homomorphisms force the formula that we used as the construction of \(\psi\): ring homomorphisms recover the formula:
\[
\psi(\sum a_iX^i) = \sum \psi(a_iX^i) = \sum \psi(a_i)\psi(X)^i = \sum \phi(a_i)c^i.
\]
\(\square\)

10.2. **Example.** Fix a commutative ring \(S\), and let \(T = \mathcal{F}(S, S)\), which is also commutative. Let \(\phi: S \to T\) be the map that sends \(a \in S\) to the constant function \((x \mapsto a)\). Let \(c = \text{id}\), the identity function of \(S\). Then the proposition gives the function \(\psi: S[X] \to T\), which sends a polynomial to its function. This function is a homomorphism of rings.

10.3. **Example.** Let \(F\) be a field, and let \(S = M_{n \times n}(F)\). Fix a matrix \(A \in S\). Let \(\phi: F \to M_{n \times n}(F)\) be the homomorphism defined by \(\phi(c) = cA\); note that the image of \(\phi\) is in the center, and so every \(\phi(c)\) commutes with \(A\).

Then the proposition gives a homomorphism \(\psi_A: F[X] \to S\), which sends
\[
c_0 + c_1X + \cdots + c_nX^n \quad \mapsto \quad c_0I + c_1A + c_2A^2 + \cdots + c_nA^n.
\]
That is, we can “plug a matrix into a polynomial”.

10.4. **Example.** Let \(S\) be a commutative ring (e.g., a field), and let \(\phi: S \to S\) be the identity map. Then for any \(c \in S\) we get a homomorphism \(\epsilon_c: S[X] \to S\) defined by
\[
\epsilon_c(\sum a_iX^i) := \sum a_ic^i.
\]
This is the **evaluation at** \(c\) function.

11. **Ideals**

An **ideal** of a ring \(R\) is a subset \(I \subseteq R\) such that
(1) \(I\) is a subgroup of \((R, +)\),
(2) if \(r, r' \in R\) and \(x \in I\), then \(rxr' \in I\). (You can write this condition as \(RIR \subseteq I\).)

Note that an ideal, under our definition of ring, is not usually a subring. (This differs from books where rings are allowed to not have a multiplicative identity.)

Note that since \(1 \in R\), the condition implies \(rx, xr \in I\) if \(x \in I\), \(r \in R\).

**Warning.** The notion of ideal we have defined is sometimes called a **two-sided ideal**, to distinguish it from the notions of left-ideal and right-ideal.

11.1. **Example.** In any ring \(R\), the subsets \(R\) and \(\{0\}\) are ideals of \(R\).

11.2. **Example.** In \(R = \mathbb{Z}\), the subsets \(\mathbb{Z}n = \{nx \mid x \in \mathbb{Z}\}\) are ideals for every \(n\).

**Observation.** We can replace (1) with the weaker condition:
(1') if $x, y \in I$ then $x + y \in I$.
This is because $-1, 0 \in R$, so $x \in I$ and (2) implies $-x = (-1)x \in I$ and $0 = 0x \in I$.

The kernel of a ring homomorphism $\phi : R \to S$ is the set

$$\text{Ker} \phi := \{ r \in R \mid \phi(r) = 0 \}.$$ 

In other words, it is the same as the kernel of $\phi$ thought of as a homomorphism of abelian groups.

11.3. Proposition. The kernel $\text{Ker} \phi \subset R$ of a ring homomorphism $\phi : R \to S$ is an ideal of $R$.

Proof. Straightforward. □

This implies that a ring homomorphism $\phi$ is injective iff $\text{Ker} \phi = \{0\}$ (because $\phi$ is also a homomorphism of abelian groups).

11.4. Exercise (On PS ?). Let $F$ be a field, $c \in F$ an element. Let $I := \{ f \in F[X] \mid f(c) = 0 \}$.
Show that $I$ is an ideal of $F[X]$.

12. Ideals generated by subsets

Let $R$ be a ring and $S \subseteq R$ a subset. The ideal generated by $S$ is

$$(S) := \bigcap_{S \subseteq J \subseteq R} J,$$

the intersection of all ideals of $R$ which contain the subset $S$. That this is an ideal is because of the following.

12.1. Proposition. If $\{J_i\}$ is any collection of ideals in $R$, then $J = \bigcap_i J_i$ is an ideal.

We also have an explicit description of $(S)$.

12.2. Proposition. We have that

$$(S) = \{0\} \cup \{ a_1s_1b_1 + \cdots + a_k s_k b_k \mid \text{for all } k \geq 1, a_i, b_i \in R, s_i \in S \}.$$ 

Proof. First check that the RHS is an ideal. Show (1') it is closed under addition (immediate), and (2) if $r, r' \in R$ and $x \in (S)$ then $rxr' \in S$: if $x = a_1s_1b_1 + \cdots + a_k s_k r_k$ then

$$rxr' = (ra_1)s_1(b_1 r') + \cdots + (ra_k)s_k(b_k r').$$

Second, show that if $J \subseteq R$ is any ideal and $S \subseteq J$, then the RHS above is a subset of $J$. Also straightforward. □

Note. The ideal $(S)$ contains the subgroup $\langle S \rangle$ of $(R, +)$ generated by $S$, but is usually bigger than it.

When $R$ is a commutative ring, this simplifies a bit.

12.3. Proposition. If $R$ is commutative, then

$$(S) = \{0\} \cup \{ a_1 s_1 + \cdots + a_k s_k \mid \text{for all } k \geq 1, a_i \in R, s_i \in S \}.$$ 

Proof. asb = $(ab)s$. □

A principal ideal is an ideal which can be generated by a single element. For $x \in R$ we write

$$(x) := \{ \{x\} \}.$$ 

Thus

$$(x) = \{ a_1 x b_1 + \cdots + a_n x b_n \mid a_i, b_i \in R \}.$$ 

When $R$ is commutative, we can always rearrange:

$$a_1 x b_1 + \cdots + a_n x b_n = (a_1 b_1)x + \cdots + (a_n b_n)x = (a_1 b_1 + \cdots + a_n b_n)x.$$ 

Thus, for commutative $R$, principal ideals have the form

$$(x) = Rx = \{ rx \mid r \in R \}.$$
Sometimes I’ll write \(Rx\) (or \(xR\)) for principal ideals in commutative rings.

We are going to say a lot about principal ideals in commutative rings. Here is an important fact: principal ideals in commutative rings correspond to elements “up to units”.

12.4. Exercise (On PS 9?). Let \(R\) be a commutative ring, and let \(a, b \in R\). Then \((a) = (b)\) if and only if there exists a unit \(u \in R^\times\) such that \(b = ua, a = u^{-1}b\).

13. Simple rings

(I’ll just mention this notion briefly in class; it’s not really important for us.)

A ring \(R\) is simple\(^1\) if \(1 \neq 0\) and its only ideals are \(\{0\}\) and \(R\).

13.1. Example. Let \(D\) be a division ring. Then there are exactly two ideals, namely \(\{0\}\) and \(D\). To see this, suppose \(I \subseteq D\) is an ideal. Obviously \(0 \in I\). If there is an \(a \in I\) with \(a \neq 0\), then because \(D\) is a division ring \(a\) has a multiplicative inverse, so

\[
a^{-1}a = 1 \in I,
\]

which implies \(b = b1 \in I\) for all \(b \in D\).

13.2. Proposition. A commutative ring is simple if and only if it is a field.

Proof. We know fields are simple. Suppose \(R\) is a simple commutative ring, so \(\{0\}\) and \(R\) are the only ideals. We show that any non-zero element of \(R\) has a multiplicative inverse.

Consider any non-zero \(a \in R\). Let \(I = (a)\) be the ideal generated by \(a\). Since \(I \neq \{0\}\) we must have \(I = R\). Because \(I\) is a principal ideal in a commutative ring, we have \(I = \{ra \mid r \in R\}\). Since \(1 \in I\) there exists \(b \in R\) such that \(ba = 1\), so \(a\) has a multiplicative inverse. \(\square\)

This fails for non-commutative rings: a simple ring does not have to be a division ring.

13.3. Example. Let \(S = M_{n \times n}(F)\) for some field \(F\). Then \(S\) is a simple ring. But \(S\) is not a division ring if \(n \geq 2\).

Here is a sketch proof. Let \(I \subseteq S\) be an ideal with a non-zero element \(A \in I\). Since \(A \neq 0\) it has a non-zero entry, say \(a_{uv} \neq 0\). Write \(E(i, j)\) for the “elementary matrix” (which has 1 in position \((i, j)\) and 0 for all other entries).

Note that that every matrix satisfies \(B = \sum b_{ij}E(i, j)\). Thus, to show \(I = S\) it suffices to show all \(E(i, j) \in S\). To do this, check that

\[
E(i, u)AE(v, j) = a_{uv}E(i, j),
\]

and therefore \(E(i, j) = a_{uv}^{-1}(a_{uv}E(i, j)) \in S\).

13.4. Exercise. If \(R\) and \(S\) are rings, then the ideals of the product ring \(R \times S\) are precisely the subsets \(I \times J \subseteq R \times S\), where \(I \subseteq R\) and \(J \subseteq S\) are ideals. In particular, the product of two non-trivial rings is not simple.

14. Ideals in polynomial rings

We recall some exercises from the homework. Let \(R\) be a domain, and let \(P = R[X]\) the polynomial ring. We defined the degree function

\[
\deg: R[X] \to \mathbb{Z}_{\geq 0} \cup \{-\infty\}
\]

which had the properties

- \(\deg(f) = -\infty\) if and only if \(f = 0\).
- \(\deg(f) = 1\) if and only if \(f\) is non-zero and “constant”.

\(^1\)Warning. Many references also require simple rings to satisfy an additional condition, which is sometimes called “left Artinian”. If so, they might call my notion a quasi-simple ring. Don’t worry about this; we’ll just use the simple version of the definition of simple lol.
Exercise. Proposition. Let $f, g \in F[X]$ with $g \neq 0$. Then there exists a unique pair $q, r \in F[X]$ such that

- $f = gq + r$
- $\deg(r) < \deg(g)$.

In other words, “$f \div g$” has quotient $q$ and remainder $r$, where $\deg(r) < \deg(g)$.

Before proving this, let’s think about some consequences.

14.2. Exercise (On PS 10). Let $F$ be a field, $c \in F$ an element, and $f \in F[X]$. Show that if $f(c) = 0$, then $f = (X - c)g$ for some polynomial $g$. (Hint: use the “division algorithm” for $f \div (X - c)$.)

Use this to show that a polynomial of degree $n$ over $F$ has at most $n$ roots in $F$. (Hint: induction on $n$.)

A monic polynomial is a polynomial with coefficient of the leading degree term equal to 1. That is, $f = X^n + a_{n-1}X^{n-1} + \cdots + a_0$. Note that the zero polynomial cannot be a monic polynomial; however, the constant polynomial 1 is monic.

We can classify all ideals in $F[X]$. The proof is very much like the classification of subgroups/ideals in $\mathbb{Z}$.

14.3. Proposition. Let $F$ be a field. Then all ideals in $F[X]$ are principal. In particular, every ideal $I \leq F[X]$ is of the form $I = (f) = R[X]f$ for exactly one polynomial $f$ which is either 0 or a monic polynomial.

Proof. The subset $(0) = \{0\}$ is an ideal, and 0 is the only single element which generates it.

Suppose $I \leq K[X]$ such that $I \neq \{0\}$. Since $I$ contains non-zero elements, it must contain a non-zero element $f$ of minimal degree $d \geq 0$ (by the Well-Ordering Principle of $\mathbb{Z}_{\geq 0}$). We are going to show that every $h \in I$ is of the form $fh$ for some $q \in F[X]$, i.e., that $I = (f)$.

The proof is the division algorithm. Given $h \in I$ consider the division algorithm for $h \div f$: there exist $q, r \in F[X]$ with $\deg(r) < \deg(f)$ such that $h = fq + r$. But by hypothesis $\deg(f)$ is minimal for non-zero elements, so $r = 0$, so $h = fq \in (f)$.

Given $I = (f)$, write $f = cX^d +$ (lower degree) with $c \neq 0$. Then $f' = c^{-1}f$ is monic and $I = (f) = (f')$. This is the only monic polynomial of degree $d$ in $I$, since if there are two such $f', f''$, then $\deg(f' - f'') < d$ so $f' = f''$. □

14.4. Exercise. Given a field $F$ and $c \in F$, let $I = \{f \in F[X] \mid f(c) = 0\}$. What is the monic polynomial which generates $I$?

Proof of the division algorithm. Uniqueness. I’ll do uniqueness first. Suppose there are two solutions, i.e.,

\[ f = gq_1 + r_1 = gq_2 + r_2, \quad \deg(r_1), \deg(r_2) < \deg(g). \]

Then taking differences gives

\[ g(q_2 - q_1) = r_1 - r_2, \quad \deg(r_1 - r_2) \leq \max\{\deg(r_1), \deg(r_2)\} < \deg(g). \]

Set $q = q_2 - q_1$ and $r = r_1 - r_2$, so this becomes $gq = r$. If $r \neq 0$, then also $q \neq 0$, and we would have $\deg(r) = \deg(g) + \deg(q)$. Since $\deg(q) \geq 0$ this contradicts $\deg(r) < \deg(g)$, so this is impossible. Thus we must have $r = 0$ and thus $q = 0$, whence $r_1 = r_2$ and $q_1 = q_2$.

Existence. Suppose $\deg(g) = n \geq 0$. 

In general, prove this by induction on \( \deg(f) = m \): assume the proposition has been proved for \( f \)'s with degree \(< m \).

If \( \deg(f) = -\infty \) then \( f = 0 \), so just take \( q = r = 0 \).

If \( \deg(f) < \deg(g) \), set \( q = 0 \) and \( r = f \), so that

\[
 f = 0g + r .
\]

Now suppose \( \deg(f) \geq \deg(g) \). That is, \( \deg(f) = m \geq n \). Write

\[
 f = a_mX^m + \cdots , \quad g = b_nX^n + \cdots , \quad a_m, b_n \in F \setminus \{0\},
\]

and let

\[
 u = \frac{a_mX^m}{b_nX^n} = a_mb_n^{-1}X^{m-n},
\]

(the fraction in quotes isn’t generally well-defined because \( b_nX^n \) may not have a multiplicative inverse, but the expression of the right is defined since \( b_n \neq 0 \) and \( m \geq n \)). Thus

\[
 (b_nX^n)u = a_mX^m.
\]

Now let \( f' := f - gu \). Looking at the terms of highest degree (= \( m \)) we have

\[
 f' = f - gu = (a_mX^m + \cdots ) - (b_nX^n + \cdots )(a_mb_n^{-1}X^{m-n} + \cdots ) = 0X^m + \text{(lower deg)} .
\]

Therefore \( f' \) has degree strictly less than \( m \), so \( \deg(f') < \deg(f) \). By the induction on degree we know that there exist \( q' \) and \( r' \) such that

\[
 f' = gq' + r' , \quad \deg(r') < \deg(g) .
\]

Then

\[
 f = f' + gu = gq' + r' + gu = g(q' + u) + r'
\]

so \( q' + u \) and \( r' = r \) is a solution.

\[
 \square
\]

15. QUOTIENT RINGS

Given an ideal \( I \) of a ring \( R \), there is a quotient ring \( R/I \). Elements of \( R/I \) are cosets

\[
 a + I = \{ a + x \mid x \in I \} \subset R
\]

of the additive group \((R, +)\) with respect to the subgroup \((I, +)\). Thus, it is automatic that \( R/I \) is an abelian group, with

\[
 (a + I) + (b + I) = (a + b) + I .
\]

Note. Sometimes I will use the notation \( \overline{a} \in R/I \) for the coset \( a + I \).

15.1. Proposition. If \( I \subseteq R \) is an ideal, then \( R/I \) is a ring, with addition as above and multiplication defined by

\[
 (a + I)(b + I) := ab + I .
\]

If \( R \) has multiplicative identity \( 1 \), then \( R/I \) has multiplicative identity \( 1 + I \).

Furthermore, the obvious projection map \( \pi \colon R \to R/I \) defined by \( \pi(a) := a + I \) is a ring homomorphism.

Proof. Check that the formula for product is well defined. Suppose \( a + I = a' + I \) and \( b + I = b' + I \), which implies

\[
 a' = a + x, \quad b' = b + y, \quad x, y \in I .
\]

Then

\[
 a'b' = (a + x)(b + y) = ab + (ay + xb + xy) \in ab + I ,
\]

since \( ay + xb + xy \in I \). Therefore \( a'b' + I = ab + I \).

Checking the various axioms for \( R/I \) to be a ring is straightforward, using that they are true for \( R \).
15.2. Example. \( \mathbb{Z}/n = \mathbb{Z}/(n) \), the quotient of integers by the ideal \((n) = \mathbb{Z}n \). We already know this is a quotient group (under addition. In fact, it is a quotient ring.

16. Quotient rings of polynomials

An important example are quotients of polynomial rings by principal ideals.

16.1. Example. Let \( S = \mathbb{Q}[X]/J \) where \( J = (f) = (X^2 - 2) \). Any element of \( S \) has the form

\[
g + J = (a_nX^n + \cdots + a_1X + a_0) + J, \quad a_n, \ldots, a_n \in \mathbb{Q},
\]

but not uniquely. We can use the division algorithm to find a “canonical form”: a canonical form for \( g + J \) is an expression \( r + J \) such that \( r + J = g + J \) and \( \deg(r) < \deg(f) \). The division algorithm for “\( g \div f \)” tells us that there is a unique canonical form, given by the remainder.

For instance, doing \((X^3 + 4X^2 + 5X - 3) \div f\) gives

\[
X^3 + 4X^2 + 5X - 3 = (X + 4)(X^2 - 2) + (7X + 5) = Xf + (7X + 5)
\]

so

\[
(X^3 - 2X^2 + 5X - 3) + J = (7X + 5) + J.
\]

Another way to think of this: we replaced every “\( X^2 \)” with “2”.

In general, every element of \( S \) is represented uniquely as

\[
a + bX + J, \quad a, b \in \mathbb{Q}.
\]

In particular, the set \( \{ a + J \mid a \in \mathbb{Q} \} \) is a subring of \( S \) which is isomorphic to \( \mathbb{Q} \).

Addition of canonical forms gives a canonical form:

\[
((a + bX) + J) + ((a' + b'X) + J) = ((a + a') + (b + b')X) + J.
\]

Multiplication doesn’t automatically give a canonical form, so we have to work:

\[
((a + bX) + J)((a' + b'X) + J) = (a + bX)(a' + b'X) + J
\]

\[
= (aa' + (ab' + a'b)X + bb'X^2) + J
\]

\[
= (aa' + (ab' + a'b)X + bb'(X^2 - 2)) + J
\]

\[
= ((aa' + 2) + (ab' + a'b)X) + J.
\]

In practice we use the following notational tricks to deal with this:

- Given \( a \in \mathbb{Q} \), we use the same symbol “\( a \)” to represent the element \( a + J \in S \).
- We pick a symbol like \( X \) or \( \omega \) to represent the coset \( X + J \in S \).
- Then any element \( u \in S \) can be written uniquely as

\[
u = a + b\omega, \quad a, b \in \mathbb{Q}.
\]

- The symbol \( \omega \) satisfies the “reduction rule” \( \omega^2 = 2 \). We use this rule whenever we need to put expressions in “canonical form”:

\[
\omega^3 + 4\omega^2 + 5\omega - 3 = 2\omega + 4(2) + 5\omega - 3 = 7\omega + 5.
\]

16.2. Example (Continued). Let \( T \subseteq \mathbb{R} \) denote the following subset of \( \mathbb{R} \):

\[
T := \{ a + b\sqrt{2} \mid a, b \in \mathbb{Q} \}.
\]

Exercise: \( T \) is a subring of \( \mathbb{R} \).

We can define a function \( \phi : S \to T \) by

\[
\phi(g + J) := g(\sqrt{2}).
\]

That this is well defined relies on the following observation: if \( g + J = g' + J \), then \( g' = g + hf \), and therefore

\[
g'(\sqrt{2}) = g(\sqrt{2}) + h(\sqrt{2})f(\sqrt{2}) = g(\sqrt{2}) + h(\sqrt{2})0 = g\sqrt{2},
\]
since $\sqrt{2}$ is a root of the polynomial $f = X^2 - 2$.

As we will soon see, $\phi$ is an isomorphism of rings.

16.3. Example. Let $S = \mathbb{R}[X]/(X^2 + 1)$. If we define $i := X$, then $i^2 = -1$, and an element of $S$ has a unique representation $a + bi$ with $a, b \in \mathbb{R}$. Of course, $S \approx \mathbb{C}$.

Here is the general principle about canonical form for elements in a quotient of a polynomial ring on one variable:

16.4. Proposition. Given $S = F[X]/J$ with $J = (f)$ for a monic polynomial $f = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, then every element of $S$ can be written uniquely as
\[ c_0 + c_1X + \cdots + c_{n-1}X^{n-1}, \quad c_0, \ldots, c_{n-1} \in F, \]
where we identify elements $c \in F$ with elements $c + J \in S$, and $X = X + J$.

Furthermore, any expression $c_0 + c_1X + \cdots + c_mX^m$ can be reduced to a canonical form by repeated applications of replacing $X^n$ with $-(a_0 + a_1X + \cdots + a_{n-1}X^{n-1})$.

Proof. Division algorithm. The process for finding canonical forms for elements in $S$ is exactly the division algorithm in $R[X]$ for $\div f$. □

It is not necessarily the case that $F[X]/(f)$ will be a field.

16.5. Example. Let $S = \mathbb{Q}[X]/(X^2 - 4)$. Then the elements $\alpha = X - 2$ and $\beta = X + 2$ are non-zero in $S$, but satisfy $\alpha \beta = 0$, because $\alpha \beta = (X - 2)(X + 2) = X^2 - 4 = 0$.

You can think of it this way: the ring $S$ is obtained from $\mathbb{Q}$ by "formally adjoining" an element $X$ which is a square root of 4, since $X^2 = 4$. But $\mathbb{Q}$ already has a square root of 4 (in fact two: $\pm 2$), so $S$ has "too many" square roots of 4 if it is going to be a field. (In a field, there are at most $d$ roots of a polynomial of degree $d$.)

16.6. Example. Let $S = \mathbb{Q}[X]/J$ with $J = (f) = (X^3 - 2)$. Write $\omega := [X] = X + J \in S$, and identify elements $a \in \mathbb{Q}$ with $a + J \in S$. Then every element of $S$ can be written uniquely as
\[ a + b\omega + c\omega^2, \quad a, b, c \in \mathbb{Q}. \]

You can carry out calculations using the "reduction" formula $\omega^3 = 2$.

Exercise: $S$ is a field. (This is not obvious, and we will return to this.)

17. Homomorphism theorems

17.1. Theorem (Homomorphism theorem for rings). Let $\phi: R \to S$ be a surjective homomorphism of rings with $I = \ker \phi$. Let $\pi: R \to R/I$ be the quotient homomorphism. Then there exists a unique ring isomorphism $\bar{\phi}: R/I \to S$ such that $\bar{\phi}(\pi(r)) = \phi(r)$ for all $r \in R$.

Proof. If you understand the proof of the homomorphism theorem of groups, then you understand the proof of the homomorphism theorem for rings. □

If $\phi: R \to S$ isn’t surjective, its image $S' := \phi(R) \subseteq S$ is a subring, and you can apply the homomorphism theorem to $R \to S'$, so $S' = \phi(R) \approx R/I$.

17.2. Example. Let $\phi: \mathbb{R}[x] \to \mathbb{C}$ be the homomorphism defined by evaluation at $i \in \mathbb{C}$, and using the usual inclusion $\mathbb{R} \subseteq \mathbb{C}$. Thus $\phi(g) := g(i)$. This is surjective because $a + bi = \phi(a + bx)$.

$\ker \phi$ = set of all polynomials $g$ such that $g(i) = 0$. Clearly $x^2 + 1 \in \ker(\phi)$, therefore $(x^2 + 1) \subseteq \ker(\phi)$.

Given an arbitrary $g \in \mathbb{R}[x]$, by polynomial division we have $g = (x^2 + 1)q + (a + bx)$ for some $q \in \mathbb{R}[x]$. We have $\phi(g) = \phi(x^2 + 1)\phi(q) + (a + bx) = a + bi$.

This is 0 iff $a = 0 = b$, so we see that $\ker(\phi) = (x^2 + 1)$.

The homomorphism theorem gives a unital isomorphism $\mathbb{R}[x]/(x^2 + 1) \approx \mathbb{C}$ of rings.
17.3. Example. Let \( \psi : \mathbb{R}[X] \to \mathbb{C} \) be the homomorphism defined by evaluation at \( \omega = e^{2\pi i/3} = (-1 + i\sqrt{3})/2 \).

This is surjective. Both \( \mathbb{R}[X] \) and \( \mathbb{C} \) are, in particular, real vector spaces over \( \mathbb{R} \), and \( \psi \) is, in particular, an \( \mathbb{R} \)-linear map. The set \( \{\psi(1) = 1, \psi(x) = \omega\} \) in \( \mathbb{C} \) is linearly independent over \( \mathbb{R} \). Since \( \dim_{\mathbb{R}} \mathbb{C} = 2 \), this means \( \psi \) is surjective.

Write \( \text{Ker}(\psi) = (f) \). The homomorphism theorem gives an isomorphism \( \mathbb{R}[X]/(f) \cong _{\mathbb{C}} \) of rings. By counting dimensions, we see that \( \deg f = 2 \). The set \( (f) \) is the collection of all polynomials over \( \mathbb{R} \) which have \( \omega \) as a root.

There is a degree 2 polynomial with \( \omega \) as a root, for instance \( f = X^2 + X + 1 \). We get an isomorphism \( \mathbb{R}[x]/(X^2 + X + 1) \cong _{\mathbb{C}} \) of rings.

18. Domains

Let \( R \) be a commutative ring with 1.

An **domain** (sometimes called an **integral domain**) is a commutative ring \( R \) with 1 such that \( 1 \neq 0 \), and such that \( xy = 0 \) implies either \( x = 0 \) or \( y = 0 \).

As we have noted, this is the same as: a commutative ring \( R \) such that \( R \setminus \{0\} \) is non-empty and closed under multiplication.

18.1. Example. Examples of domains are \( \mathbb{Z} \), fields \( \mathbb{F} \).

Also, the polynomial ring \( D[x] \) with \( D \) a domain (because \( \deg(fg) = \deg(f) \deg(g) \)).

18.2. Proposition. Any subring of a domain is a domain.

Proof. Easy. \( \square \)

For instance, the **Gaussian integers**

\[ \mathbb{Z}[i] := \{ a + bi \mid a, b \in \mathbb{Z} \} \subset \mathbb{C}. \]

18.3. Example. Let \( f_1, f_2 \in F[x] \) be two non-constant polynomials, and let \( g = f_1f_2 \) and \( I = (g) \).

The ring \( F[x]/I \) is not a domain, since \( \overline{f_1} = f_1 + I, \overline{f_2} = f_2 + I \) are non-zero, but \( \overline{f_1f_2} = 0 \).

For instance, \( \mathbb{Q}[X]/(g) \) with \( g = X^2 - 4 = (X - 2)(X + 2) \).

Domains have cancellation of non-zero elements.

18.4. Proposition (Cancellation). If \( R \) is a domain, and \( a, b, c \in R \) such that \( a \neq 0 \), then \( ab = ac \) implies \( b = c \).

Proof. \( a(b - c) = 0 \) implies either \( a = 0 \) or \( b - c = 0 \). \( \square \)

19. Fields of fractions

Every domain is a subring of a field, called its **field of fractions**.

Given a domain \( R \), consider the set

\[ S := \{ (a, b) \in R \times R \mid b \neq 0 \}. \]

Define a relation on \( S \) by

\[ (a, b) \sim (a', b') \iff ab' = ba'. \]

(Remember that in \( \mathbb{Q} \), we have \( a/b = a'/b' \) if and only if \( ab' = ba' \).)

19.1. Lemma. This is an equivalence relation on \( S \).

Proof.

- Reflexive. To see that \( (a, b) \sim (a, b) \), note that \( ab = ba \).
19.2. Example. If \( a, b \in R \setminus \{0\} \), then \( \frac{ac}{bc} = \frac{a}{b} \).

Define operations + and \( \cdot \) on \( F \) by
\[
\frac{a}{b} + \frac{c}{d} := \frac{(ad + bc)}{bd}, \quad \frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}.
\]
Check that these are compatible with the equivalence relation (Exercise!), and so are well-defined operations on \( F \).

19.3. Proposition. If \( R \) is a domain, then \( F = \text{Frac}(R) \) is a field. The function \( \phi : R \to F \) given by \( \phi(a) := (a/1) \) defines an isomorphism of rings from \( R \) to the subring \( \phi(R) = \{ (a/1) \mid a \in \mathbb{R} \} \) of \( F \).

Proof. This is just straightforward, and is partially an exercise. Note: the 0 element is \((0/1)\), the 1 element is \((1/1)\).

I’ll do multiplicative inverses (or leave as exercise?) If \( x \in F \) is not equal to 0, write \( x = (a/b) \). Because \( x \neq 0 \), we have \( (a/b) \neq (0/1) \), i.e., \( a1 \neq b0 \), i.e., \( a \neq 0 \). So set \( y = (b/a) \in F \) (this makes sense exactly because \( a \neq 0 \), and check that \( (a/b)(b/a) = (ab/ab) = (1/1) = 1 \). □

Typically we identify \( R \) with its image in \( F = \text{Frac}(R) \), so we can say that the domain \( R \) is a subring of its fraction field.

19.4. Example. If \( R = \mathbb{Z} \), then \( \text{Frac}(\mathbb{Z}) = \mathbb{Q} \).

19.5. Example. If \( R = \mathbb{F}[X] \), then \( \mathbb{F}(X) := \text{Frac}(\mathbb{F}[X]) \) is called the field of rational functions in one variable. Elements are \( f/g \) where \( f, g \) are polynomials.

Note that polynomial long division of the form \( f \div g \) for \( g \neq 0 \) gives \( f = gq + r \) with \( \deg(r) < \deg(g) \), which in \( \mathbb{F}(X) \) gives
\[
\frac{f}{g} = q + \frac{r}{g}.
\]

19.6. Example. \( \text{Frac}(\mathbb{Z}[i]) = \mathbb{Q}[i] \). To see this, note that every element in \( \mathbb{Q}[i] \) is a ratio of elements in \( \mathbb{Z}[i] \). In fact, if \( x \in \mathbb{Q}[i] \) we can write it as \( x = (a/b) + (c/d)i \) with \( a, b, c, d \in \mathbb{Z} \). Then
\[
x = (bd)^{-1}(ad + bci), \quad bd, ad + bci \in \mathbb{Z}[i].
\]

The claim follows from the following proposition.

19.7. Proposition. Let \( F \) be a field, and \( F \subseteq R \) a subring (which is thus a domain). If every element of \( F \) has the form \( ab^{-1} \) with \( a, b \in R, b \neq 0 \), then \( F \) is isomorphic to the fraction field of \( R \).
**Proof.** (Don’t spend much time on this.)

1. Define a function \( \phi : \text{Frac}(R) \to F \) by \( \phi((a/b)) := ab^{-1} \). First check that this is well-defined: if \( (a/b) = (a'/b') \) represent the same element in \( \text{Frac}(R) \), then \( ab = ba' \) in \( R \), and thus \( a' = ab^{-1}b^{-1} \) in \( F \), whence \( a'b^{-1} = (ab^{-1})b^{-1} = ab^{-1} \).

2. Check that \( \phi \) is a ring homomorphism. This is a straightforward exercise:

\[
\phi((a/b) + (a'/b')) = \phi((ab' + ba')/ab) = (ab' + ba')(bb')^{-1} = ab^{-1} + a'b^{-1} = \phi((a/b)) + \phi((a'/b')),
\]

\[
\phi((a/b)(a'/b')) = \phi((aa')(bb')) = aa'(bb')^{-1} = (ab^{-1})(a'b^{-1}) = \phi((a/b))\phi((a'/b')),
\]

\[
\phi((1/1)) = 1.
\]

3. The function \( \phi \) is surjective, because of the hypothesis that every element of \( F \) has the form \( ab^{-1} \) with \( a, b \in R \). To see that \( \phi \) is surjective, consider Ker \( \phi \) which is an ideal in \( F \). Since \( F \) has only two ideals, and \( 1 \not\in \text{Ker} \phi \), we have Ker \( \phi = \{0\} \) so \( \phi \) is injective.

\[ \square \]

### 20. Factorization in Domains

Let \( R \) be a domain. Remember that this means that the set \( R \setminus \{0\} \) of non-zero elements forms a monoid under the operation of multiplication.

Let \( R^\times \subseteq R \setminus \{0\} \) be the group of units.

Given \( a, b \in R \), we say \( a \) divides \( b \) and \( a \) is a factor of \( b \) if there exists \( c \in R \) such that \( b = ac \). (I.e., if “\( b/a \in \text{Frac}(R) \)” is actually contained in the subring \( R \).) We write \( a|b \).

We say that \( a, b \in R \) are the same up to units if there exists a unit \( u \in R^\times \) such that \( b = ua \). Note that “same up to units” is an equivalence relation on \( R \).

Note: I’m going to write \( Ra := \{ ra \mid r \in R \} \). Because \( R \) is commutative, this is also the principal ideal \((a)\).

**Important observation.** \( a|b \) iff \( b \in Ra \) iff \( Rb \subseteq Ra \) (inclusion of principal ideals). This is often the best way to think of this.

**Another important observation.** \( a, b \in R \) are the same up to units iff \( Ra = Rb \).

In other words: “divisibility” and “same up to units” are properties that only depend on the principal ideals generated by elements.

Here is a categorization of elements in a domain \( R \).

- The zero element \( 0 \in R \).
- Units \( u \in R^\times \).
- \( a \in R \) is reducible if (i) \( a \in R \setminus (R^\times \cup \{0\}) \) (i.e., \( a \neq 0 \) and \( a \) is not a unit) such that (ii) there exist \( b, c \in R \setminus (R^\times \cup \{0\}) \) such that \( a = bc \).

That is, \( a \) is reducible if it is a non-zero non-unit which can be factored into two non-units.
- \( a \in R \) is irreducible if \( a \in R \setminus (R^\times \cup \{0\}) \) and \( a \) is not reducible. That is, \( a \) is a non-zero non-unit such that \( a = bc \) implies either \( b \) or \( c \) is a unit.

This is a partition of the domain \( R \): an element is in exactly one of these four classes.

**20.1. Exercise.** If \( a \) and \( b \) are the same up to units, then both \( a \) and \( b \) are in the same one of these four classes.

**20.2. Example.** For \( R = \mathbb{Z} \), we have

- \( 0 \).
- Units(\( \mathbb{Z} \)) = \{ 1, -1 \}.
- Irreducible elements = \{ \pm p \mid p \in \mathbb{N} \text{ a prime number} \}.
- Reducible elements: the rest, i.e., the composite integers (both positive and negative).

**20.3. Example.** For \( R = K \) a field, we have

- \( 0 \).
20.4. Example. For $R = F[X]$ with $F$ a field, we have

- Units$(F[X]) = \{ f(x) = c \in F[X] \mid c \in F \setminus \{ 0 \} \}$. I.e., units are the non-zero constant polynomials.
- Irreducible elements = set of irreducible polynomials, i.e., polynomials of degree $\geq 1$ which do not factor into product of polys of smaller degree.
- Reducible elements = set of polys of degree $\geq 1$ which do factor into polys of smaller degree.

For instance, if $F = \mathbb{R}$ (real numbers), then the four classes in $\mathbb{R}[X]$ are

- $0$.
- Units = $c \in \mathbb{R} \setminus \{ 0 \}$.
- Irreducible = $\{ aX + b \mid a, b \in \mathbb{R}, a \neq 0 \} \cup \{ aX^2 + bX + c \mid a, b, c \in \mathbb{R}, a \neq 0, b^2 - 4ac < 0 \}$.
- Reducible = Everything else.

This is a fact you are probably aware of, though the proof is fairly deep. It is true because (i) every real polynomial splits into degree 1 factors over $\mathbb{C}$ (i.e., the “Fundamental Theorem of Algebra”, which is already a deep theorem), and (ii) any non-real roots of a real polynomial come in conjugate pairs.

20.5. Example. Let $R = \mathbb{Z}[i] = \{ a + bi \mid a, b \in \mathbb{Z} \} \subset \mathbb{C}$ be the Gaussian integers. It is useful to consider the norm map

$N: \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}, \quad N(a + bi) := (a + bi)(a - bi) = |a + bi|^2 = a^2 + b^2$ if $a, b \in \mathbb{Z},$

which has the properties:

$N(1) = 1, \quad N(xy) = N(x)N(y)$ for $x, y \in \mathbb{Z}[i]$.

- Note that $N(x) = 0$ if and only if $x = 0$.
- $R^x = \{ 1, -1, i, -i \}$. To see this, first note that if $u \in R$ is a unit, so $u$ has a multiplicative inverse, then $1 = N(1) = N(uu^{-1}) = N(u)N(u^{-1})$. Since the outputs of $N$ are non-negative integers we must have $N(u) = 1$.

Thus if $u = a + bi$ is a unit for $a, b \in \mathbb{Z}$, then $a^2 + b^2 = 1$. We see that the only possible values for $u$ are $\pm 1, \pm i$, and it is obvious that these actually are units.
- Some irreducible elements of $\mathbb{Z}$ are also irreducible in $\mathbb{Z}[i]$. For instance, $3$ is irreducible. Proof. Suppose $3 = xy$ for some $x = a + bi, y = c + di \in \mathbb{Z}[i]$ with $a, b, c, d \in \mathbb{Z}$. Then

$9 = N(3) = N(x)N(y) = (a^2 + b^2)(c^2 + d^2)$.

Since $N(x), N(y) \in \mathbb{Z}_{\geq 0}$, we must have $N(x), N(y) \in \{ 1, 3, 9 \}$. If we want a factorization into non-units, we must have $N(x) = N(y) = 3$. But there are no solutions in integers to the equation $a^2 + b^2 = 3$, so this is impossible.
• On the other hand, 2 is reducible in \( \mathbb{Z}[i] \), since \( 2 = (1 + i)(1 - i) \), and neither \( 1 + i \) nor \( 1 - i \) are units since \( \mathcal{N}(1 + i) = 2 \). In fact, \( 1 + i \) and \( 1 - i \) are irreducible in \( \mathbb{Z}[i] \).

Similarly 13 is reducible, since \( (3 + 2i)(3 - 2i) = 3^2 + 2^2 = 13 \). So the primes 2, 5, 13, 17, 29, . . . which are sums of two squares of integers are always reducible in \( \mathbb{Z}[i] \). However, it can be shown that the primes 3, 7, 11, 19, 23, . . . which are not sums of two squares of integers are not reducible in \( \mathbb{Z}[i] \).

Note: \( 1 + i \) and \( 1 - i \) are actually the same up-to-units, since \( 1 + i = i(1 - i) \). However, \( 3 + 2i \) and \( 3 - 2i \) are no the same up-to-units.

Each of the four classes of elements can be related to a property of principal ideals. (Skip this in class for now.)

20.6. Proposition. Let \( R \) be a domain and \( a \in R \). Then

1. \( a = 0 \) if and only if \( Ra = \{0\} \) is the trivial ideal.
2. \( a \) is a unit if and only if \( Ra = R \).
3. \( a \) is reducible if and only if (i) \( Ra \neq 0 \) and (ii) there exists \( Rb \) such that \( Ra \subseteq Rb \subseteq R \).
4. \( a \) is irreducible if and only if (i) \( Ra \neq 0 \) and \( Ra \neq R \), and (ii) if \( Ra \subseteq Rb \subseteq R \), then either \( Rb = Ra \) or \( Rb = R \).

Proof. Easy exercise. (PS ?.)

Property (4) is a “maximality” property: \( a \) is irreducible if and only if \( a \neq 0 \) and \( Ra \) is maximal among proper principal ideals of \( R \).

21. Unique factorization domains

A unique factorization domain (UFD) is a domain \( R \) such that every non-zero non-unit factors into irreducibles, and the factorization is unique up to order and to multiplication by units.

That is, for a UFD \( R \), if \( a \in R \) is not zero and not a unit, we can write

\[
a = p_1 \cdots p_m, \quad m \geq 1,
\]

for some sequence of irreducible elements \( p_i \) in \( R \). Furthermore, if

\[
a = p_1 \cdots p_m = q_1 \cdots q_n, \quad m, n \geq 1,
\]

with all \( p_i, q_j \) irreducible, then \( m = n \) and (after possibly reordering the \( q_j \)s), we have that \( p_i \) and \( q_i \) are the same up to units.

21.1. Example. The integers are a UFD. We actually proved this at the start of the semester.

We are going to have a big theorem: every polynomial ring with coefficients in a UFD is a UFD.

In particular, \( F[X] \) with \( F \) a field is a UFD. This case has an easier proof, and is also the consequence of a theorem: every PID is a UFD. This will also show that the Gaussian integers are a UFD.

But not every domain is a UFD.

21.2. Example. Let \( R = \mathbb{Z} [\sqrt{5}] \subseteq \mathbb{C} \), the subring of the complex numbers consisting of elements

\[
a + b\sqrt{5} \quad \text{such that } a, b \in \mathbb{Z}.
\]

By \( \sqrt{5} \) I mean the element \( \sqrt{5} i \in \mathbb{C} \), which has the property that its square is \( -5 \). (In what follows everything works the same way if we take it to be \( -\sqrt{5} i \). I’m always going to write the element as “\( \sqrt{5} \)” : just remember that it is a fixed square root of \( -5 \). Another way to think about this: \( R \approx \mathbb{Z}[X] / (X^2 + 5) \), and \( \sqrt{5} \) corresponds to the element \( \overline{X} \).)

\( R \) is a domain since it is a subring of the field \( \mathbb{C} \). There is a norm function \( N: R \to \mathbb{Z}_{\geq 0} \) defined by

\[
N(a + b\sqrt{5}) = (a + b\sqrt{5})(a - b\sqrt{5}) = |a + b\sqrt{5}|^2 = a^2 + 5b^2,
\]
which is multiplicative: $N(1) = 1$ and $N(xy) = N(x)N(y)$.

- The units of $R$ are $R^\times = \{\pm 1\}$. (Exercise: Solve $N(x) = 1$.)
- The element 6 is reducible in $R$:

$$6 = (2)(3) = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

- The elements 2, 3, $1 + \sqrt{-5}$, $1 - \sqrt{-5}$ are all irreducible in $R$. (Exercise. Use $N$.)
- The element 2 does not divide either $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$, and thus is not the same up-to-units-in-$R$ as either $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$. Likewise for 3. (Exercise. Use $N$.)

Thus, 6 admits two non-equivalent factorizations into irreducibles in $\mathbb{Z}[\sqrt{-5}]$. So it is not a UFD!

### 22. Primes in domains

An element $p \in R$ in a domain is **prime** if (i) it is non-zero and not a unit, and (ii) if $p|ab$ then either $p|a$ or $p|b$.

We can formulate this as a property of principal ideals: $p$ is prime if and only if (i) $Rp \neq \{0\}$ and $Rp \neq R$, and (ii) $ab \in Rp$ implies either $a \in Rp$ or $b \in Rp$.

**22.1. Exercise.** $p$ is prime if and only if $R/(p)$ is a domain.

Primes are a subset of irreducibles.

**22.2. Proposition.** Prime elements are irreducible.

**Proof.** Suppose $p$ is prime and $p = ab$; I’ll show that at least one of $a$ or $b$ is a unit. Then $p|ab$ so $p$ divides one of the factors, say $p|a$. Thus $a = pc$ for some $c$, so $p = ab = pcbc$, so $1 = cb$, so $b$ is a unit.

The converse of the proposition is not true: irreducibles do not need to be prime.

**22.3. Example.** Let $R = \mathbb{Z}[\sqrt{-5}]$, and let $p = 2$. Then 2 is irreducible in $R$, it is not prime in $R$. For instance, $2|6$, but $6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$ and you can show that 2 does not divide either of $1 \pm \sqrt{-5}$. (Exercise. Use $N$.)

**Warning.** The usual definition of “prime number” in $\mathbb{N}$ (no positive factors other than 1 and self), looks almost the same as what we have called an “irreducible element” of $\mathbb{Z}$ (non-zero non-unit with no factorization into two non-units). It is very easy see that irreducible integers are exactly prime numbers or negatives of prime numbers.

On the other hand, the definition of “prime element” in a domain is not at all like the definition of prime number. It is a fact that the positive prime elements of $\mathbb{Z}$ are exactly the usual prime numbers, but this is a theorem that needs to be proved, and it is not entirely trivial. (In fact, we proved this in the first week of the course, and we’ll prove it again: it relies on the fact that $\mathbb{Z}$ is a PID.)

I personally find this confusing, and it confused me a lot when I first learned this stuff. Just remember that “irreducible” is the concept that generalizes “prime integer” to arbitrary domains; it’s also the notion that’s familiar to you in the context of factoring polynomials. “Prime element” is a different, more subtle notion. **End warning.**

However, in a UFD things are good. (Maybe skip this?)

**22.4. Proposition.** If $R$ is a UFD, then all irreducibles are prime.

**Proof.** First I’m going to show that if $p$ is irreducible and $p|d$, then $p$ must be (up to units) one of the terms in the irreducible factorization of $d$.

If $p|d$ then $d = pc$ for some $c$. Choose irreducible factorizations $c = p_1 \cdots p_m$ and $d = q_1 \cdots q_n$. Then

$$q_1 \cdots q_n = pp_1 \cdots p_m.$$
But uniqueness of such factorizations means \( p \) must be the same as one of the \( q_i \)'s up to units, so \( q_i = pu \) for some unit \( u \).

Now if \( p|ab \), choose irreducible factorizations \( a = p_1 \cdots p_m \) and \( b = q_1 \cdots q_n \), so \( ab = p_1 \cdots p_m q_1 \cdots q_n \) is also an irreducible factorization. The element \( p \) is the same (up to units) as either a \( p_i \) or \( q_j \), so either \( p|a \) or \( p|b \).

\[ \square \]

## 23. Principal ideal domains

A **principal ideal domain (PID)** is a domain such that every ideal is principal.

Examples, which we have already proved:

- Any field \( K \). (These are boring examples of PIDs.)
- \( \mathbb{Z} \).
- \( F[X] \) where \( F \) is a field.

The fact that both \( \mathbb{Z} \) and \( F[X] \) are PIDs means we can prove some facts about both using the same proof: in particular, facts about factorization into irreducibles.

Not every domain is a PID.

### 23.1. Example

The ring \( \mathbb{Z}[X] \) of polynomials with coefficients in \( \mathbb{Z} \) is a domain, but contains non-principal ideals such as \( I = (2, X) \). (On PS 10. Note that one needs to show that \( I \) cannot be generated by a single element.)

Similarly, the ring \( F[X, Y] \) where \( F \) is a field is not a PID, e.g., \( I = (X, Y) \). (Proof similar to \( (2, X) \subseteq \mathbb{Z}[X] \).

The proofs that \( \mathbb{Z} \) and \( F[X] \) are PIDs both used forms of the “division algorithm”. We can use a similar idea on the Gaussian integers.

### 23.2. Example

The Gaussian integers \( \mathbb{Z}[i] \) are a PID. This is proved using a version of the division algorithm based on complex norms.

**Division algorithm in \( \mathbb{Z}[i] \) for \( x \div y \).** Given \( x, y \in \mathbb{Z}[i] \) with \( y \neq 0 \), there exist \( q, r \in \mathbb{Z}[i] \) such that

\[ x = qy + r, \quad |r| < |y|, \]

where \( |y| \) is the usual argument (=length) of a complex number: \( |a + bi| = \sqrt{a^2 + b^2} \). (Note that \( q \) and \( r \) are not assumed unique.)

**Proof of division algorithm.** Fix \( x \) and \( y \), and consider the ideal generated by \( y \) as a subset of the complex numbers.

\( I := (y) := \{ qy \mid q \in \mathbb{Z}[i] \} = \{ ay + byI \mid a, b \in \mathbb{Z} \} \).

It really helps to draw a picture of this set: elements of \( I \) are points on a square grid in the plane, whose vertices include the vectors \( 0, y, \) and \( yi \) (note that \( y \) and \( yi \) are perpendicular and have the same length.) The length of the sides of the squares is \( L = |y| = |yi| \).

Consider an arbitrary complex number \( x \in \mathbb{C} \). The point \( x \) will land on a vertex of \( G \), or on a side of one of the squares, or at worst inside a square. The smallest distance from \( x \) to some point of \( G \) is \( \leq L/\sqrt{2} \). (Because the center of a square of side \( L \) is that far from any vertex.)

In particular, if \( x \in \mathbb{Z}[i] \), choose \( q \in \mathbb{Z}[i] \) such that \( |x - qy| \) is minimized, and set \( r = x - qy \in \mathbb{Z}[i] \).

Then we win, because

\[ |r| = |x - qy| \leq L/\sqrt{2} < L = |y|. \]

The proof works exactly because the “radius” of a square of side \( L \) is \( L/\sqrt{2} \), which is strictly less than \( L \) (because \( 1/\sqrt{2} < 1 \)).

**Proof that \( \mathbb{Z}[i] \) is a PID using the division algorithm.** Let \( J \subseteq \mathbb{Z}[i] \) be an ideal. If \( J \neq \{0\} \), there exists a \( y \in J \setminus \{0\} \) for which \( |y| \) is as small as possible. The claim is that \( J = (y) \). The proof is the division algorithm: given \( x \in J \), there exist \( q, r \) such that \( x = qy + r \) with \( |r| < |y| \). Since \( r = x - qy \in J \) we must have \( |r| = 0 \) so \( r = 0 \) so \( x = qy \) so \( x \in (y) \).
23.3. Example. The domain $R = \mathbb{Z}[\sqrt{-5}]$ is not a PID. The ideal $J = (2, 1 + \sqrt{-5})$ is not principal: there is no $x \in R$ such that $J = (x)$. Exercise. (Hint: Show that if $(2, 1 + \sqrt{-5}) = (x)$, then $N(x)$ must divide $N(2) = 4$ and $N(1 + \sqrt{-5}) = 6$ (in $\mathbb{Z}$). Since $N(x) = 2$ is impossible, this forces $N(x) = 1$, i.e., $x$ is a unit so $1 \in J$. Complete the proof by showing $1 \notin J$.)

Note that if you try to carry out for $R$ the argument that worked for Gaussian integers, you fail: there is no reasonable “division algorithm” for $\mathbb{Z}[\sqrt{-5}]$. What goes wrong is that $I = (y) \subset \mathbb{C}$ is a “rectangular grid” whose sides are of length $L$ and $\sqrt{5}L$ (where $L = |y|$). The “radius” of this rectangle is $(\sqrt{6}/2)L$, which is actually greater than $L$.

23.4. Proposition. If $R$ is a PID, then $a \in R$ is prime iff it is irreducible.

Proof. We’ve shown prime implies irreducible in any domain.

Conversely, suppose $p$ is irreducible; we show $p$ is prime. (The proof is exactly the same as the one we gave for $\mathbb{Z}$ near the start of the course.) Suppose $ab \in pR$ and $a \notin pR$; we want to show $b \in pR$.

We will do this by showing that when $a \notin pR$ there exist $s, t \in R$ such that $1 = sp + ta$. Given this, we can write

$$b = 1b = (sp + ta)b = spb + t(ab).$$

Since $ab \in pR$, both terms on the right are in $pR$, and so $b \in pR$.

Now we show: if $a \notin pR$ then there exist $s, t \in R$ such that $1 = sp + ta$. To prove this, consider the ideal $I := (p, a) = pR + aR$. Because $R$ is a PID, this ideal is also principal: we have $I = cR$ for some $c \in R$. We know that $pR \subseteq cR \subseteq R$ (since $a \notin pR$ but $a \in I = cR$). Since $p$ is irreducible, we must have $cR = R$. (I.e., we have $p \in cR$ so $p = cd$ so either $c$ or $d$ is a unit, but if $d$ is a unit then $pR = cR$ which is not allowed.)

Therefore $(p, a) = cR = R$, so $1 = sp + ta$ for some $s, t \in R.$

We can also write this proof this way: we want to show $ab \in pR$ implies either $a \in pR$ or $b \in pR$. Set $I := (p, a)$. We have $pR \subseteq I \subseteq R$. By the PID property and $p$ irreducible we must have either $I = pR$ or $I = R$. If $I = pR$ then $a \in pR$. If $I = R$ then $bI = R$, whence $b \in bR = bI = (pb, ab) \subseteq pR$ since $pb, ab \in pR$.

24. Application: Fermat’s Theorem on Sums of Squares

Question. Which integers $m$ are the sum of two squares of integers? That is, given $m$, does there exist $a, b \in \mathbb{Z}$ such that $m = a^2 + b^2$?

The answer to this question is related to the Gaussian integers. This is because $m = a^2 + b^2$ implies that you can factor $m$ in $\mathbb{Z}[i]$ as $m = (a + bi)(a - bi) = a^2 + b^2$.

Because of the relation to factorization, it turns out it is easiest to answer the question for integers which are prime.

Question (special case). Which prime integers $p$ are the sum of two squares?

- For $p = 2$, we have $2 = 1^2 + 1^2$.
- For odd $p$ with $p \equiv 3 \mod 4$ this is not possible, because $a^2, b^2 \in [0]_4 \cup [1]_4$ but $p \in [3]_4$.
  (In fact, this is not possible for any integer in $[3]_4$.)
- The remaining case is odd primes $p$ such that $p \equiv 1 \mod 4$. In this case, such primes are sums of two squares.

24.1. Theorem (Fermat, Euler). If $p = 4n + 1$ is a prime integer, then $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.

For instance: $5 = 2^2 + 1^2$, $13 = 3^2 + 2^2$, $17 = 4^2 + 1$, $29 = 5^2 + 2^2$, $37 = 6^2 + 1^2$, $41 = 5^2 + 4^2$, $53 = 7^2 + 2^2$, $61 = 6^2 + 5^2$, etc. This is a non-obvious theorem, since even if you know it’s true, it is not obvious how to solve for $a, b$ such that $p = a^2 + b^2$ (except by brute force trying all possibilities).
We are going to prove this using the fact that the Gaussian integers \( \mathbb{Z}[i] \) are a PID, and so irreducibles in \( \mathbb{Z}[i] \) are also prime. (This proof is due to Dedekind (1877).) These ideas will prove the following proposition.

24.2. Proposition. If \( p = 4n + 1 \) is a prime integer, then it is reducible in the Gaussian integers \( \mathbb{Z}[i] \).

Given this, the proof of the Theorem is not hard.

Proof of Theorem using the Proposition. By the proposition, \( p = xy \) for two not unit \( x,y \in \mathbb{Z}[i] \). Taking norms gives \( N(x)N(y) = N(p) = p^2 \). Since \( x,y \) are non-units their norms cannot be 1, so we must have \( N(x) = N(y) = p \). Therefore if \( x = a + bi \) with \( a,b \in \mathbb{Z} \) then \( p = N(x) = a^2 + b^2 \), as promised.

To prove the proposition, we need a lemma, which I will put in the homework. It turns out to be equivalent to a fact about finite fields.

24.3. Lemma (Lagrange). If \( p = 4n + 1 \) is a prime integer, then there exists \( m \in \mathbb{Z} \) such that \( p | (m^2 + 1) \). (Divisibility in \( \mathbb{Z} \).

Proof. Another way to say this: for such primes \( p = 4n + 1 \), we want an integer \( m \) such that \( m^2 \equiv -1 \mod p \). Or said another way: If \( p \) is a prime congruent to 1 mod 4, then \( \mathbb{Z}/p \) contains a square root of \(-1\). Left as a (non-obvious) exercise. [On PS ?]

Before giving the proof of the proposition, let’s note something about ideals in \( \mathbb{Z}[i] \). If \( p \in \mathbb{Z} \) is an integer, then the principal that it generates in \( \mathbb{Z}[i] \) has the form

\[
(p) = \mathbb{Z}[i]p = \{ a + bi \mid a,b \in \mathbb{Z}p \}.
\]

This is just because for \( c + di \in \mathbb{Z}[i] \) (with \( c,d \in \mathbb{Z} \)) we have \((c + di)p = (cp) + (dp)i\), so if \( a + bi \in (p) \) (with \( a,b \in \mathbb{Z} \)) then both integers \( a \) and \( b \) are divisible by \( p \) as integers.

Proof of Proposition. Let \( p = 4n + 1 \) be a prime integer. By the lemma, there exists \( m \in \mathbb{Z} \) such that \( p | (m^2 + 1) \) (divisibility in \( \mathbb{Z} \)). This implies that \( p \) also divides \( m^2 + 1 \) in the Gaussian integers \( \mathbb{Z}[i] \). (Given \( m^2 + 1 = pk \) with \( k \in \mathbb{Z} \), this is also a factorization of \( m^2 + 1 \) in \( \mathbb{Z}[i] \).)

But in \( \mathbb{Z}[i] \) we can factor \( m^2 + 1 = (m + i)(m - i) \). We suppose \( p \) irreducible and derive a contradiction.

Because \( \mathbb{Z}[i] \) is a PID, irreducibles are primes, so we must have \( p|m + i \) or \( p|m - i \) in \( \mathbb{Z}[i] \). But either of these is impossible: e.g., as noted above, \( p|m + i \) implies \( p|1 \) as integers.

Thus, \( p \) must be irreducible in \( \mathbb{Z}[i] \). \( \square \)

The original question has a similar answer. (Maybe skip this in class?)

24.4. Proposition. A positive integer \( m \) is a sum of two squares of integers if and only if its prime factorization (in \( \mathbb{Z} \)) \( m = p_1^{k_1} \cdots p_r^{k_r} \) is such that: if \( p_i \equiv 3 \mod 4 \), then \( k_i \) is even.

In other words, a positive \( m \) is not a sum of two squares if it is divisible an odd number of times by some prime \( p \) with \( p \equiv 3 \mod 4 \). Thus, the integers \( \leq 50 \) which are not a sum of two squares are:

3, 6, 7, 11, 12, 14, 15, 19, 21, 22, 23, 24, 27, 28, 30, 31, 33, 35, 38, 39, 42, 43, 44, 46, 47, 48.
On the other hand, we have

\begin{align*}
1 &= 1^2 + 0^2, & 9 &= 3^2 + 0^2, & 18 &= 3^2 + 3^2, & 32 &= 4^2 + 4^2, & 41 &= 5^2 + 4^2, \\
2 &= 1^2 + 1^2, & 10 &= 3^2 + 1^2, & 20 &= 4^2 + 2^2, & 34 &= 5^2 + 3^2, & 45 &= 6^2 + 3^2, \\
4 &= 2^2 + 0^2, & 13 &= 3^2 + 2^2, & 25 &= 5^2 + 0^2, & 36 &= 6^2 + 0^2, & 49 &= 7^2 + 0^2, \\
5 &= 2^2 + 1^2, & 16 &= 4^2 + 0^2, & 26 &= 5^2 + 1^2, & 37 &= 6^2 + 1^2, & 50 &= 7^2 + 1^2, \\
8 &= 2^2 + 2^2, & 17 &= 4^2 + 1^2, & 29 &= 5^2 + 2^2, & 40 &= 6^2 + 2^2,
\end{align*}

Proof. Let

\[ S = \{ m \in \mathbb{Z}_{>0} \mid m = a^2 + b^2 \text{ for some } a, b \in \mathbb{Z} \} \]

and

\[ T = \{ m \in \mathbb{Z}_{>0} \mid m = p_1^{k_1} \cdots p_r^{k_r}, \text{ } p_i \text{ distinct primes, } k_i \text{ even if } p_i \equiv 3 \mod 4 \}. \]

We want to show \( S = T \).

First note that \( S \) is closed under multiplication, a fact we can prove using the Gaussian integers:

\[
(a^2 + b^2)(c^2 + d^2) = (a + bi)(a - bi)(c + di)(c - di) \\
= ((ac - bd) + (bc + ad)i)((ac - bd) + (bc + ad)i) \\
= (ac - bd)^2 + (bc + ad)^2.
\]

Also, any square \( m^2 \) is a sum of two squares: \( m^2 = m^2 + 0^2 \).

Note also that \( T \) is closed under multiplication: if \( m \) and \( n \) each have an even number of factors of a prime \( p \), then so does their product \( mn \).

Therefore if \( m \in T \), i.e., \( m = p_1^{k_1} \cdots p_r^{k_r} \in S \) with the \( p_i \) distinct primes and \( k_i \) even if \( p_i \equiv 3 \mod 4 \), then \( m \in S \). Because: (i) \( p_i \in S \) if \( p_i \not\equiv 3 \mod 4 \) by what we have shown, and (ii) \( p_i^2 \in S \) for any prime. We have shown \( T \subseteq S \).

Now suppose \( m \in S \), and show that \( m \in T \). We prove this by induction on \( m \). Note that the statement is true for \( m = 1, 2 \), which are in both \( S \) and \( T \).

Let \( m = a^2 + b^2 \), and suppose we know that any integer smaller than \( m \) which is a sum of squares is in \( T \). If \( m \) has no prime factors (in \( \mathbb{Z} \)) which are \( \equiv 3 \mod 4 \) then \( m \in T \), and there is nothing else to do. So suppose there is a \( p \equiv 3 \mod 4 \) which divides \( m \) in \( \mathbb{Z} \), and therefore divides it in \( \mathbb{Z}[i] \).

Then since \( m = (a + bi)(a - bi) \) we have \( p|(a + bi)(a - bi) \) in \( \mathbb{Z}[i] \). We know that \( p \) is irreducible in \( \mathbb{Z}[i] \), and thus prime since \( \mathbb{Z}[i] \) is a PID. Therefore \( p \) must divide one of the factors, either \( p|a + bi \) or \( p|a - bi \). In either case, this implies that \( p|a \) and \( p|b \) in \( \mathbb{Z} \), since \( p \) is actually an integer as we have noted above. Thus \( a + bi = p(c + di) \) for some \( c, d \in \mathbb{Z} \), so

\[
m = (a + bi)(a - bi) = p^2(c + di)(c - di) = p^2(c^2 + d^2).
\]

Since \( n = c^2 + d^2 \) is a sum of two squares and less than \( m \), by induction it is in \( T \). We know \( p^2 \in T \), and therefore \( m = np^2 \) is also in \( T \). \( \square \)