Abstract Algebra Refresher

Review, Amplification, Examples

Abstract Algebra Refresher

Review, Amplification, Examples

Thomas R. Shemanske
Dartmouth College

Version: December 30, 2022

Preface

These notes are not intended as a first or second course in abstract algebra, though they assume the reader has seen the material in a basic algebra course, covered for example in [1].

These notes will undertake a review of many basic topics from a typical first course, often taking the opportunity to interleave more advanced concepts with simpler ones when convenient. It will refresh the reader's memory of definitions, structural results, core examples, and provide some computational tools to help the reader come to a deeper appreciation of the ideas first met perhaps a long time ago.

Computations in these notes uses Sage (`sagemath.org`) which is a free, open source, software system for advanced mathematics. Sage can be used either on your own computer, a local server, or on SageMathCloud (`cocalc.com`).

Thomas R. Shemanske
Hanover, NH
2020

Contents

Back Matter

Chapter 1

A quick review of a first course

As stated in the preface, these notes presume the reader has seen the material in a basic abstract algebra course such as in

1.1 What is Algebra?

Well, if you ask a random acquaintance “what is algebra?”, more than likely the answer will be something which *they* learned in high school, though a quick perusal of your textbook may cast a bit of doubt. Still, shouldn’t there be some connection between that subject you learned in high school and the one by that name in College?

In high school, algebra was about solving equations or simple systems of equations, about how to use the quadratic formula to find points of intersection of two conics, and so on. Linear Algebra undertakes the systematic study of solving systems of linear equations which is best answered when cast in terms of algebraic structures called vector spaces, and structure preserving maps between them. Abstract algebra, often called modern algebra, fully embraces the study of sets endowed with one or more binary operations.

One of the main goals of mathematics is to classify things by type, and to characterize when two things are of the same type. For vector spaces being of the same type is called isomorphic, and you learned that two vector spaces are isomorphic if and only if they have the same dimension. That is a remarkably simple characterization. For groups, rings and fields, this characterization is significantly more challenging.

Another important goal of mathematics is to see the manner in which a given object can be built up from simpler ones. For example,

- When is a group G isomorphic to the direct product of simpler groups?
- How can one classify all the groups G which contain a normal subgroup isomorphic to a fixed group H having prescribed quotient G/H?

Abstract algebra creates an extensive toolbox with which to embark on these investigations.

1.2 Partitions and Equivalence Relations

In our bid to classify things by type, we introduce the notion of isomorphism to say two objects are of the same type. That means that when we look at the set of all objects (say groups), we **partition** that set of objects into equivalence classes so that any two objects in one class are isomorphic, but no two objects from different classes can be isomorphic.

The notions of a partition and of a set of equivalence classes are deeply intertwined. Let's review the basics.

Given a nonempty set X, a **partition** of X is simply a collection of non-overlapping subsets whose union is the original set. For example, the pieces of a puzzle form a partition of the image which is their union. The set of all M&M's in a bag can be partitioned by color. We give a formal definition.

Definition 1.2.1 Let X be a non-empty set. A **partition** of X is a collection $P = \{X_i \mid i \in I\}$ of nonempty subsets so that

- $X = \bigcup\limits_{i \in I} X_i$, and
- $X_i \cap X_j = \emptyset$ for all $i \neq j$.

$\Diamond$

The closely related notion is that of an **equivalence relation** on a nonempty set. Formally, we have

Definition 1.2.2 Let X be a non-empty set. A **relation** on X is a subset $R \subseteq X \times X$, that is a collection of ordered pairs. Often instead of saying $(x, y) \in R$, write $x \sim y$ and say x is **related** to y.

An **equivalence relation** on X is a relation which satisfies three properties:

- $x \sim x$ (i.e., $(x, x) \in R$) for all $x \in X$. This is called the **reflexive** property of the relation.
- If $x \sim y$, then $y \sim x$, that is, whenever the ordered pair $(x, y) \in R$, then also $(y, x) \in R$. This is called the **symmetric** property of the relation.
- If $x \sim y$ and $y \sim z$, then $x \sim z$, that is, if $(x, y), (y, z) \in R$, then so it (x, z). This is called the **transitive** property of the relation.

$\Diamond$

Let $\sim$ be an equivalence relation on a set X. For each $x \in X$, the **equivalence class** containing x is given by:

$$[x] = \{y \in X \mid y \sim x\}.$$

Notice that by the reflexive property, $x \in [x]$, and we did not need to fuss in the definition about whether $y \sim x$, or $x \sim y$ since the relation is symmetric. And the transitive property shows that two equivalence classes are either the same or disjoint. It follows that

Proposition 1.2.3 *Let $\sim$ be an equivalence relation on a set X. Then $P = \{[x] \mid x \in X\}$ forms a partition of X. In words, the set of equivalence classes forms a partition of the set.*

Conversely, we have

Proposition 1.2.4 *Given a partition $P = \{X_i \mid i \in I\}$ of a set X, the relation $x \sim y$ if and only if $x, y \in X_j$ for some (unique) $j \in I$ defines an equivalence relation on X in which the equivalence classes are the elements of the original partition.*

Example 1.2.5 The Integers modulo n. One of the most familiar example comes from introducing an equivalence relations on the integers, $\mathbb{Z}$. We fix a positive integer n, and define the relation by

$$j \sim k \text{ iff } j \equiv k \pmod{n}.$$

Then the set of equivalences classes is called the **integers modulo** n. There are multiple notations for this, but the set of classes is generally denoted

$$\mathbb{Z}_n \text{ (or) } \mathbb{Z}/n\mathbb{Z} = \{[k] \mid k \in \mathbb{Z}\}.$$

As you recall from your course, $\mathbb{Z}_n$ consists of n equivalence classes often denoted with representatives given by $[0], [1], \ldots, [n-1]$. □

Example 1.2.6 Similarity classes of matrices. If F is a field, we say that two matrices $A, B \in M_n(F)$ are **similar** if there is an invertible matrix P so that $B = P^{-1}AP$. We often adopt the more general term that A and B are **conjugate.**

Once again it is easy to verify that similarity is an equivalence relation. If you have had an advanced linear algebra course, you would know that each similarity class $[A]$ has a distinguished representative which is the **rational canonical form** of A. □

Now often, as in the case of $\mathbb{Z}_n$, we wish to introduce an algebraic structure on the set of equivalence classes. With $\mathbb{Z}_n$ you know that the operations

$$[a] + [b] = [a+b] \text{ and } [a][b] = [ab]$$

makes $\mathbb{Z}_n$ into a commutative ring with identity.

Insight 1.2.7 And other times, algebraic objects that we have known since grade school suddenly reveal themselves in the guise of equivalence classes. The rational numbers, $\mathbb{Q}$, is such an example. How can something we understand so intrinsically be hiding this underlying structure?

Well, if we went back to grade school we could solve the following arithmetic problem using some curious-looking rules learned long ago:

$$\frac{2}{5} + \frac{3}{7} = \frac{29}{35}.$$

Just what is the rule that tells you how to add $\frac{a}{b} + \frac{c}{d}$? Yes, after a moment's thought you could write down the rule, but where in the world did it come from and why is such a complicated-looking rule really needed?

Indeed, what could we possible mean by writing

$$\frac{1}{2} = \frac{2}{4} = \frac{5}{10}, \text{ etc.?}$$

It is becoming very clear that this symbol $\frac{a}{b}$ or a/b has a nontrivial meaning.

It is easy to believe writing a/b where there is a distinguished numerator and denominator that we are doing nothing more than introducing a fancy notation for an ordered pair (a, b) where the first coordinate is the numerator and the second the denominator. But then we note that while $\frac{1}{2} = \frac{2}{4}$, it is not true that $(1, 2) = (2, 4)$. So it is time to understand what we have been doing for years without a thought about higher-level mathematics.

The short version is that we define a set X as

$$X = \mathbb{Z} \times \mathbb{Z} \setminus \{0\},$$

the set of ordered pairs with arbitrary first coordinate, but nonzero second coordinate. Then we define the relation $(a, b) \sim (c, d)$ iff $ad - bc = 0$. As this is an equivalence relation, we can talk about its equivalence classes, and we denote the equivalence class of (a, b) not as $[(a, b)]$, but as a/b or $\frac{a}{b}$.

Then we are left with the task of giving definitions for addition and multiplication which are well-defined on the equivalence classes, that is independent of the choice of representative of the class. But you know all that.

1.3 Structure-preserving maps and quotient structures

In our attempt to understand the structure of algebraic objects, we frequently try to gain insight by looking at substructures and quotient structures as well as maps between algebraic structures.

1.3.1 Morphisms

Having studied groups, rings, vector spaces and similar objects, you have considered the notion of a **homomorphism,** a structure-preserving map. While the notions of a linear map, group or ring homomorphism differ in the details, they all are defined to preserve whatever structure the algebraic object has. For example,

- A **linear map** $T : V \to W$ between two vector spaces over the same field F requires that for all $v, v' \in V$ and $\lambda \in F$,

  $$T(v + v') = T(v) + T(v') \text{ and } T(\lambda v) = \lambda T(v),$$

  transporting the structure of vector addition and scalar multiplication on V to the corresponding ones on W.

- A **group homomorphism** $\varphi : G \to H$ between groups G, H requires that for all $g, g' \in G$,

  $$\varphi(g * g') = \varphi(g) * \varphi(g')$$

  transporting the binary operation on G to the one on H.

- A **ring homomorphism** $\varphi : R \to S$ between rings R, S requires that for all $r, r' \in R$

  $$\varphi(r + r') = \varphi(r) + \varphi(r') \text{ and } \varphi(r * r') = \varphi(r) * \varphi(r'),$$

  transporting the additive abelian group structure and multiplicative structure on R to the corresponding structures on S.

The subject of **category theory** generalizes the notion of a homomorphism to an extreme providing a definition of a **morphism** $\varphi : X \to Y$ between **objects** X and Y, where the objects need not be restricted to algebraic objects, but could include things like topological or analytic spaces.

In this very general context, one defines the notion of an **isomorphism**, as a morphism $\varphi : X \to Y$ for which there is a morphism $\psi : Y \to X$, so that $\varphi \circ \psi = id_Y$ and $\psi \circ \varphi = id_X$. In particular, this means that φ is a bijective morphism.

Remark 1.3.1 When you studied algebra for the first time, you noted (and may have proven) that given a bijective homomorphism $\varphi : X \to Y$ between algebraic objects (groups, rings, etc), that the inverse map, ψ, which is guaranteed to exist set theoretically, is automatically a homomorphism.

It is important to note that this result is not always true in other categories. For example, in topology, the notion of a morphism is that of a continuous map, so that an isomorphism is a continuous bijection for which there is a continuous inverse. It is possible to have a continuous bijection whose inverse is not continuous. As an easy example consider the map from the half-open interval to the complex unit circle

$$\varphi : [0, 1) \to S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$$

given by $\varphi(x) = e^{2\pi i x}$. The inverse map is not continuous, for if it were, the two spaces would be topologically isomorphic implying that any topological property which holds for one would hold for the other. But the unit circle is compact (closed and bounded in Euclidean space), while the half-open interval is not compact.

The bottom line is that we catch a break in dealing with algebraic homomorphisms, and need only verify that our chosen homomorphism is bijective.

1.3.2 Quotients

Given an algebraic object X and a sub-object Y (e.g., group and subgroup, ring and subring, vector space and subspace), we can ask whether the quotient X/Y is again an algebraic object of the same type.

The notion of a quotient is certainly deeply connected with that of homomorphisms, the most universal theorem being the **first isomorphism theorem**, which for a homomorphism $\varphi : X \to Y$ always has the form

$$X/\ker\varphi \cong \operatorname{Im}\varphi$$

In general the kernel, $\ker\varphi$, always has the right properties to guarantee that the quotient $X/\ker\varphi$ is an algebraic object of the same type. While it is always true if W is a subspace of a vector space V, that V/W is again a vector space, quotients do not always have the desired properties. Recall for groups, one needs a subgroup H to be a normal subgroup of a group G for the set of cosets G/H to be a group, and we need S to be an ideal of a ring R for the quotient R/S to be a ring. Indeed these definitions arise and are natural precisely because they describe the conditions under which a quotient will have the appropriate structure.

1.3.3 Cosets, partitions, and equivalence relations

Meeting the notion of a coset for the first time often seems to cause a bit of confusion, some of which arises from notation (additive or multiplicative), and some of which arises from failing to understand the underlying equivalence relation. Let's try to understand all the issues.

There are couple of considerations. The first is that vector spaces and rings have underlying group structures so that when we form cosets with them, they are first and foremost group cosets. The set of cosets may enjoy additional structure depending on the context, but what sets the notation — and the equivalence relation — is the group structure.

Second in most (though not all) introductions to abstract algebra, groups enter the picture first, and being a set with a single binary operation, it is very often written multiplicatively. This works well in our experience since in most settings (think rings) addition is commutative, but multiplication may not be (e.g., matrices). So introducing notation for a group that may or may not be abelian, multiplication is generally the more intuitive choice. So for a group G, subgroup H, and element $x \in G$, we write xH or Hx depending on whether we are talking about **left** cosets or **right** cosets.

The complication arises when we have a ring R and an ideal I and want to discuss the structure of the set of cosets R/I. Now we need both operations of multiplication and addition, so choosing a multiplicative notation for the groups structure would lead to a notational nightmare. Moreover, the additive groups structure is now abelian, so it is a bit more natural to use additive one for cosets, so we write $x + I$ (or $I + x$ though they turn out to be the same since addition is commutative).

Now that we have some sense of why we choose the notation, let's understand the equivalence relation and the associated partition. Since all the cosets start out with an underlying group, we use group notation (both multiplicative and additive) to describe the cosets and equivalence relation.

Let G be a group, H a subgroup, and $x \in G$. We define left cosets:

$$xH = \{y \in G \mid y = xh \text{ for some } h \in H\} = \{xh \mid h \in H\}$$
$$x + H = \{y \in G \mid y = x + h \text{ for some } h \in H\} = \{x + h \mid h \in H\}$$

We denote the **set of left cosets** by

$$G/H = \{xH \mid x \in G\} \text{ or } \{x + H \mid x \in G\}$$

depending upon whether we are using multiplicative or additive notation.

Now it is obvious that each coset is a subset of G, and what you prove is that the set of left cosets forms a partition of G. Since the identity of the group, e, is contained in every subgroup H, every element $x \in G$ is an element of the coset xH since $x = xe \in xH$. It follows that

$$G = \bigcup_{x \in G} xH,$$

the first property of a partition.

The critical condition is that cosets are either disjoint or identical which makes their collection a partition.

Proposition 1.3.2 *The (left) cosets of H in G are either disjoint or identical.*

Proof. Suppose we have two cosets xH and yH. If they are disjoint that is fine, but if they intersect, we must show that are equal. Let's do this out in full detail so we better understand the rule we shall write down for determining whether or not two cosets are equal.

Let $z \in xH \cap yH$. To show that $xH = yH$ (recall they are sets), we must show $xH \subseteq yH$ and $yH \subseteq xH$. The argument we give will be symmetric, so we need only show one containment, say $xH \subseteq yH$.

The condition $z \in xH \cap yH$ says that we may write

$$z = xh_1 = yh_2 \text{ for some } h_1, h_2 \in H,$$

so $x = yh_2h_1^{-1} \in yH$, and so of course every element $xh \in xH$ is equal to $xh = yh_2h_1^{-1}h \in yH$ which gives the desired conclusion $xH \subseteq yH$. ■

What we must also understand is that while

$$G = \bigcup_{x \in G} xH$$

there is a great deal of redundancy in the (multi)set $\{xH \mid x \in G\}$. An example should help.

Example 1.3.3 Let $G = S_3$, the symmetric group on 3 letters, i.e., the permutations of the set $\{1, 2, 3\}$. Let $H = \{e, (1\ 2)\}$ be the cyclic subgroup generated by the transposition $(1\ 2)$.

Solution. Then the set of left cosets is

$$G/H = \{H, (1\ 3)H, (2\ 3)H\} = \{H, (1\ 2\ 3)H, (1\ 3\ 2)H\}$$

where

$$H = (1\ 2)H, \quad (1\ 3)H = (1\ 2\ 3)H, \text{ and } (2\ 3)H = (1\ 3\ 2)H.$$

Note that the associated partition of G is

$$\mathcal{P}_L = \{P_1 = \{1, (1\ 2)\}, P_2 = \{(1\ 3), (1\ 2\ 3)\}, P_3 = \{(2\ 3), (1\ 3\ 2)\},$$

where the P_i are simply the elements in the respective cosets.

For the record we record the set of right cosets and their associated partition which differs from the one from left cosets (since H is not a normal subgroup of G).

$$H\backslash G = \{H, H(1\ 3), H(2\ 3)\} = \{H, H(1\ 3\ 2), H(1\ 2\ 3)\}$$

The associated partition is of G is

$$\mathcal{P}_R = \{Q_1 = \{1, (1\ 2)\}, Q_2 = \{(1\ 3), (1\ 3\ 2)\}, Q_3 = \{(2\ 3), (1\ 2\ 3)\}.$$

□

To distinguish when two cosets are equal or the same, we have the following criterion.

Proposition 1.3.4 *Let G be a group and H a subgroup. Then two (left) cosets xH and yH are equal if and only if and of the following equivalent conditions hold:*

- $y^{-1}x \in H$
- $x = yh$ *for some* $h \in H$
- $x^{-1}y \in H$
- $y = xh$ *for some* $h \in H$

Proof. It is quite straightforward to show that these four conditions are all equivalent, but the important part is why they are equivalent to $xH = yH$.

The key to that is to remember Proposition 1.3.2, that cosets are disjoint or equal, so to show that $xH = yH$, one needs only show that $xH \cap yH \neq \emptyset$. So it is enough to show that $x \in yH$ or $y \in xH$ which lead naturally to the conditions in the proposition. ■

Remark 1.3.5 If our operation is commutative, cosets are often written additively, so we would have two cosets $x + H$ and $y + H$ equal if and only if and of the following equivalent conditions hold:

- $-y + x \in H$
- $x = y + h$ for some $h \in H$
- $-x + y \in H$
- $y = x + h$ for some $h \in H$

1.3.4 Introducing an algebraic structure on the set of cosets.

Whether we are talking about the quotients of groups, rings, or vector spaces, there is always an underlying group G and a subgroup H. For quotient groups this is obvious; for rings, the group is the additive group of the ring, and for vector spaces, the group is the additive group of the vector space.

To make a set of cosets into a group, we need H to be a **normal** subgroup of G. Recall that

Proposition 1.3.6 *Let G be a group and H a subgroup of G. The following conditions are equivalent and define what it means for H **normal** subgroup of G.*

- *$gHg^{-1} = H$ for all $g \in G$*
- *$gH = Hg$ for all $g \in G$*
- *$gHg^{-1} \subseteq H$ for all $g \in G$*

Proof. The equivalence of the first two is easy to check, and of course the first implies the third, so we are left to show that $gHg^{-1} \subseteq H$ for all $g \in G$ implies that $gHg^{-1} = H$ for all $g \in G$.

Fix a $g \in G$ for which $gHg^{-1} \subseteq H$; we must show the reverse inclusion $H \subseteq gHg^{-1}$. Since $xHx^{-1} \subseteq H$ for all $x \in G$, choose $x = g^{-1}$. Then

$$g^{-1}Hg \subseteq H \text{ which implies } H \subseteq gHg^{-1},$$

which is the desired inclusion. ■

Checkpoint 1.3.7 Why is normality the correct notion to make a set of cosets into a group? Let's make sense of the question. Our set of left cosets is

$$G/H = \{xH \mid x \in G\}.$$

Our job (still working multiplicatively) is to define the product of two cosets in a well-defined manner:

$$xH \cdot yH = zH$$

for some $z \in G$. Now everybody knows that they want the answer to be

$$xH \cdot yH = xyH,$$

but it is not clear that always makes sense.

Let's go back to our example with $G = S_3$ and $H = \{e, (1\ 2)\}$. We observed that the cosets paired up as

$$H = eH = (1\ 2)H, \quad (1\ 3)H = (1\ 2\ 3)H, \text{ and } (2\ 3)H = (1\ 3\ 2)H.$$

So to be well-defined, whatever definition we come up with cannot depend on how we name the sets, so in our case it must be true that (for example)

$$(1\ 2)H \cdot (1\ 3)H = eH \cdot (1\ 2\ 3)H,$$

but if we used our desired rule we would get

$$(1\ 2)H \cdot (1\ 3)H = (1\ 3\ 2)H \text{ while } eH \cdot (1\ 2\ 3)H = (1\ 2\ 3)H$$

which are not equal.

The exercise is to see how the definition of normality arises naturally in trying to rectify this ambiguity.

Remark 1.3.8 Given a vector space V and a subspace W, we know that the additive structure of V is that of an abelian group, which means W is automatically a normal subgroup, which in turn says the set of cosets is an additive abelian group. But as the group operation is additive, we write the cosets additively, so the set of left cosets

$$V/W = \{v + W \mid v \in V\}$$

is a group under the natural operation

$$(v + W) + (v' + W) = (v + v') + W.$$

We then go farther to define a vector space structure on the abelian group V/W by

$$\lambda(v + W) = \lambda v + W,$$

an operation we easily check is well defined.

Similarly, given a ring R and a subring S, we know that the set of cosets

$$R/S = \{r + S \mid r \in R\}$$

is an abelian group under addition, but what about multiplication? Can we make the set of cosets into a ring?

Checkpoint 1.3.9 Why is the property of being an ideal the correct property to make R/S into a ring? For any subring S of R we already have the quotient group $R/S = \{r + S \mid r \in R\}$. It is natural to define multiplication as

$$(r + S) \cdot (r' + S) = rr' + S,$$

but we need to check that things are well-defined.

To do so we need to ensure that for $s, t \in S$,

$$(r + s + S) \cdot (r' + t + S) = rr' + S,$$

but our would-be definition tells us we would need

$$rr' + rt + sr' + st + S = rr' + S.$$

Since S is subring we know that $st \in S$, but we also need that $rt + sr' \in S$ for any $s, t \in S$. Thus we arrive at the condition that S be a **two-sided ideal** of R, namely that in addition to being a subring of R, we must have

$$R \cdot S \subseteq S \text{ and } S \cdot R \subseteq S.$$

1.4 A fundamental isomorphism theorem for groups, rings, vector spaces

If X, Y are algebraic objects of the same type (group, ring, vector space) and $\varphi : X \to Y$ is a homomorphism of the appropriate type, then first and foremost all such φ are group homomorphisms, and that is what sets the stage.

It follows that the kernel of the homomorphism is the kernel of the underlying group homomorphism, so $\ker \varphi = \{x \in X \mid \varphi(x) = e\}$ where e is the identity of the underlying group, in particular, $e = 0$ for rings and vector spaces.

We also recall that the kernel of any group homomorphism is a normal subgroup; the kernel of any ring homomorphism is a two-sided ideal, and the kernel of any linear map a vector subspace. So in all cases $X/\ker \varphi$ is an algebraic object of the same type as X, but always a group.

Theorem 1.4.1 Fundamental theorem for group homomorphisms. *Let G, H be groups, and let $\varphi : G \to H$ be a homomorphism. Let K be any normal subgroup of G with $K \subseteq \ker \varphi$, and let $\pi : G \to G/K$ be the usual projection ($g \mapsto gK$). Then there exists a unique group homomorphism $\varphi_* : G/K \to H$, so that for all $g \in G$, $\varphi(g) = \varphi_*(\pi(g))$. In this case we say that φ **factors through the quotient** G/K.*

Moreover, the image of φ_* is the same as the image of φ, and $\ker \varphi_* = \ker \varphi / K$.

$$\begin{tikzcd}
G \arrow[r, "\varphi"] \arrow[d, "\pi"] & H \\
G/K \arrow[ur, dashed, "\varphi_*"'] &
\end{tikzcd}$$

Figure 1.4.2 *Factoring a homomorphism through a quotient*

Proof. The condition that $\varphi(g) = \varphi_*(\pi(g))$ requires that

$$\varphi_*(gK) = \varphi(\pi(g)) = \varphi(g)$$

from which it is immediate that there is only one possible definition for φ_* (hence uniquely defined), and also that the images of φ and φ_* are the same.

Presuming the map φ_* is well-defined, we see that it is a group homomorphism since

$$\varphi_*(gKg'K) = \varphi_*(gg'K) = \varphi(gg') = \varphi(g)\varphi(g') = \varphi_*(gK)\varphi_*(g'K),$$

where we have used the group operation on G/K and that φ is a group homomorphism.

The most important issue is that the map makes sense, i.e., is well-defined. In this case, we must check that

$$\varphi_*(gK) = \varphi_*(gkK) \text{ for any } k \in K.$$

But

$$\varphi_*(gkK) = \varphi(gk) = \varphi(g)\varphi(k) = \varphi(g) = \varphi_*(gK)$$

since $\varphi(k) = e$ because $K \subseteq \ker\varphi$.

Finally, we compute

$$\ker\varphi_* = \{gK \in G/K \mid \varphi_*(gK) = \varphi(g) = e_H\},$$

but that says

$$\ker\varphi_* = \{gK \in G/K \mid g \in \ker\varphi\} = \ker\varphi/K.$$

■

An immediate corollary of this theorem is the

Theorem 1.4.3 First Isomorphism Theorem. *Let G, H be groups, and let $\varphi : G \to H$ be a homomorphism. Then*

$$G/\ker\varphi \cong \operatorname{Im}\varphi.$$

Proof. Let $K = \ker\varphi$. The fundamental theorem gives us a surjective homomorphism $\varphi_* : G/K \to \operatorname{Im}\varphi$ whose kernel is $K/K = eK = e_{G/K}$, the identity of G/K, so the map φ_* is injective. ■

Remark 1.4.4 The importance of the fundamental theorem cannot be overstated. The main takeaway is that if ever faced with the job of finding a homomorphism $\Psi : G/N \to H$ you do your best to recognize a homomorphism $\psi : G \to H$ whose kernel contains N. In this way, the map you define is not only more natural, but you never have to check that the map $\Psi : G/N \to H$ is well-defined. Attempting to define the map directly forces you to always check that fact.

Example 1.4.5 The canonical example? For a positive integer n, we have the group $\mathbb{Z}_n$ consisting of all the congruence classes of integers modulo n. The set $n\mathbb{Z}$ (all the integer multiples of n) is a (normal) subgroup of $\mathbb{Z}$, so $\mathbb{Z}/n\mathbb{Z}$ is a group. We want to show that

$$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n.$$

Solution. Of course we could write down a map taking $m + n\mathbb{Z} \mapsto [m]_n$, but then we would have to check that it is well-defined, a homomorphism, and eventually an isomorphism.

Instead we invoke the fundamental theorem. There is certainly a natural map $\varphi : \mathbb{Z} \to \mathbb{Z}_n$ which takes an integer m to its residue class $[m]_n$ modulo n. It is a homomorphism by virtue that $\mathbb{Z}_n$ is a group:

$$m + m' \mapsto [m + m']_n = [m]_n + [m']_n$$

and it is certainly surjective. What is the kernel of φ?

$$\ker \varphi = \{m \in \mathbb{Z} \mid [m]_n = [0]_n\} = n\mathbb{Z}.$$

By the first isomorphism theorem, the result is achieved. □

Now what we want to achieve is a fundamental (and first isomorphism) theorem for rings, vector spaces, etc. The key is that all these maps are group homomorphisms at their core and so most of the theorems are already in place.

Theorem 1.4.6 A meta fundamental theorem for homomorphisms. *Let X, Y be algebraic objects of the same type and $\varphi : X \to Y$ an associated homomorphism. Suppose that $Z \subseteq \ker \varphi$ has enough structure so that X/Z is again an algebraic object of the same type as X and Y. Let $\pi : X \to X/Z$ be the standard projection. Then there is a unique homomorphism $\varphi_* : X/Z \to Y$ (of the appropriate type) so that $\varphi = \varphi_* \circ \pi$ whose image is the same as φ, and whose kernel is $\ker \varphi_* = \ker \varphi/Z$.*

Proof. Because all these structures have underlying group structures, that map φ_* is uniquely determined, well-defined and has the correct kernel and image. The only thing missing is to verify that φ_* has the additional properties necessary to be a homomorphism of the correct type (e.g., ring or vector space). But this is easily checked. For example, if the objects were rings, then Z would necessarily be an ideal. The map φ_* takes $x + Z$ to $\varphi(x)$. We check (using the ring structure of X/Z and that φ is a ring homomorphism) that

$$\begin{aligned}\varphi_*((x+Z)(x'+Z)) &= \varphi_*(xx'+Z) = \varphi(xx') \\ &= \varphi(x)\varphi(x') = \varphi_*(x+Z)\varphi_*(x'+Z).\end{aligned}$$

■

Remark 1.4.7 The above meta theorem now tells us that $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$ as rings as well as abelian groups.

There are further isomorphism theorems that you learned, but they require some new constructions which we give in the next section.

1.5 New algebraic objects from old: products and sums

With one of our goals to understand how a given algebraic object can be decomposed into simpler objects, it is necessary to understand how to build larger objects up from smaller ones. In one direction this leads to additional isomorphism theorems alluded to previously, but it also leads to things like direct and semi-direct products which are used to reveal the structure of many groups.

As usual, we begin with the simplest of our algebraic objects under consideration, groups. Suppose that H and K are two subgroups of a group G.

Question 1.5.1 A natural question is what is the smallest subgroup of G which contains both H and K?

Answer. Mathematics has an answer for us, but it may not be satisfying. The answer is that the smallest subgroup of G containing both H and K is

$$\langle H, K\rangle := \bigcap_{J \supseteq H \cup K} J$$

where the intersection is over all subgroups J of G which contain H and K.

It is a very good answer in that it makes it clear such a group exists and is unique, but it gives us no idea how to construct it. □

Let's try again from a slightly different angle. If J is any subgroup of G which contains H and K, then since it is closed, it must contain all the products of the form hk and kh for $k \in K$ and $h \in H$. We could ask a more naive question.

Question 1.5.2 Given subgroups H and K of a group G, when is $HK := \{hk \mid h \in H, k \in K\}$ a subgroup of G, and perhaps when are HK and KH related? □

The answer to the question above is a well-known theorem.

Theorem 1.5.3 *Let H and K be subgroups of a group G. Assume that K is a normal subgroup of G, or more generally that H contained in $N_G(K)$, the normalizer of K. Then $HK = KH$ is a subgroup of G, in particular the smallest* *subgroup of G containing H and K.*

An important fact to remember is the $HK = KH$ does not mean that the elements of H and K commute, rather that for $h \in H, k \in K$, there exists $h' \in H, k' \in K$ so that $hk = k'h'$.

Proof of Theorem 1.5.3. The proof is just some simple manipulations using the concept of the normalizer. It should be clear that the identity is an element of HK.

Given $h, h' \in H, k, k' \in K$, we need to know that $(hk)(h'k') = h''k''$ for some $h'' \in H$ and $k'' \in K$. But

$$(hk)(h'k') = h(h'h'^{-1})k(h'k') = (hh')(h'^{-1}kh')k' = h''k''$$

where $h'' = hh'$ and $k'' = (h'^{-1}kh')k'$, the last using $H \subseteq N_G(K)$.

Given $h \in H, k \in K$, we need to know that $(hk)^{-1} = h'k'$ for some $h' \in H$ and $k' \in K$. But

$$(hk)^{-1} = k^{-1}h^{-1} = (h^{-1}h)k^{-1}h^{-1} = h^{-1}(hk^{-1}h^{-1}) = h'k'.$$

Finally to show $HK = HK$, we need to show the inclusions $HK \subseteq KH$ and $KH \subseteq HK$. With predictable notation we see

$$\begin{aligned} hk &= hk(h^{-1}h) = (hkh^{-1})h = k'h \\ kh &= (hh^{-1})kh = h(h^{-1}kh) = hk'. \end{aligned}$$

■

Example 1.5.4 Let $G = S_3$, the symmetric group on three letters having order 6. Let $K = \langle(1\ 2\ 3)\rangle = \langle(1\ 3\ 2)\rangle$ be the subgroup generated by either 3-cycle. We know this to be a normal subgroup the easiest reason being a consequence of Lagrange's theorem (reviewed in the next chapter). Let H be any subgroup of order two (i.e., generated by any of the three transpositions. Then HK is a subgroup of G; indeed $G = HK$, a fact we shall explore in the next chapter as well. □

Next we turn our attention to rings and their ideals.

Definition 1.5.5 Let R be a ring with ideals I and J. We can define three new ideals from them: their sum and product and intersection. The sum, $I + J$ is the smallest ideal of R which contains both I and J. The product, IJ, is the smallest ideal of R which contains all the the elements of the form ij with $i \in I$ and $j \in J$. And of course the intersection, $I \cap J$ is the largest ideal contained in both.

While these properties define the ideals, the first two are not constructive definitions, but their characterization is not too hard to discern. One just has to ask how to make the generating sets closed under the operations of addition and multiplication by elements of the ring. One finds

$$I + J = \{i + j \mid i \in I, j \in J\}$$

$$IJ = \left\{ \sum_{k=1}^{r} i_k j_k \mid i_k \in I, j_k \in J \right\}$$
$$I \cap J = \{k \mid k \in I \text{ and } k \in J\}$$

so the elements in IJ are finite sums of products of the form ij. ♢

Example 1.5.6 Let $R = \mathbb{Z}$ and let I, J be ideals. We know that all the ideals of $\mathbb{Z}$ are principal ideals, so let's look at three interesting cases.

1. $I = 12\mathbb{Z}$ and $J = 4\mathbb{Z}$
2. $I = 12\mathbb{Z}$ and $J = 15\mathbb{Z}$
3. $I = 12\mathbb{Z}$ and $J = 5\mathbb{Z}$

The resulting ideals $I + J, IJ$, and $I \cap J$ are:

1. $I + J = 4\mathbb{Z}$; $IJ = 48\mathbb{Z}$; $I \cap J = 12\mathbb{Z}$
2. $I + J = 3\mathbb{Z}$; $IJ = 180\mathbb{Z}$; $I \cap J = 60\mathbb{Z}$
3. $I + J = \mathbb{Z}$; $IJ = 60\mathbb{Z}$; $I \cap J = 60\mathbb{Z}$

□

Checkpoint 1.5.7 Let $R = \mathbb{Z}$ and let $I = m\mathbb{Z}$ and $J = n\mathbb{Z}$ be ideals.

- Determine $I + J, IJ$, and $I \cap J$ in terms of m and n. The expressions should be quite familiar to you.
- Based upon the examples above, one might conjecture that

  $$(I + J) \cdot (I \cap J) = IJ.$$

  Note that $(I + J) \cdot (I \cap J)$ is a product of ideals. Do you think it's always true for ideals of $\mathbb{Z}$? It is not true in all rings.

Chapter 2

Basic results in group theory

2.1 Cosets and some applications

While there are algebraic objects with fewer defining properties than groups (monoids, semigroups, groupoids), the notion of a group is where we start our review of standard results.

One of the most fundamental structure theorems is the theorem of Lagrange.

Theorem 2.1.1 Lagrange's theorem. *Let G be a finite group, and H a subgroup of G. Then*

$$|G| = [G : H] \cdot |H|,$$

where $[G : H]$ is the number of cosets in G/H.

Proof. Recall that the proof is very straightforward. We know that the (left) cosets of H in G form a partition of G. Since everything is finite let's enumerate the cosets $G/H = \{g_kH \mid k = 1, \ldots, r\}$, so $r = [G : H]$. Then G is the disjoint union of the cosets meaning both

$$G = \bigcup_{k=1}^{r} g_kH \text{ but also } |G| = \sum_{k=1}^{r} |g_kH|.$$

However any two cosets have the same cardinality, $|H|$, so

$$|G| = \sum_{k=1}^{r} |H| = r|H| = [G : H]|H|.$$

■

Corollary 2.1.2 *Lagrange's theorem not only says that the order of any subgroup divides the order of a group, but also that the order any element in a group divides the order of the group, the later since the order of an element $x \in G$ equals the order of the cyclic subgroup $H = \langle x \rangle$ it generates.*

Proposition 2.1.3 *Let H, K be finite subgroups of a group G. Then the cardinality of the set $HK = \{hk \mid h \in H, k \in K\}$ is given by*

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

Proof. Whether or not HK is a subgroup of G, we can view HK as a union of cosets:

$$HK = \{hk \mid h \in H, k \in K\} = \bigcup_{h \in H} hK.$$

Now all the cosets hK have the same size as K, so

$$|HK| = |K| \cdot \text{number of distinct cosets.}$$

We see that for $h_1, h_2 \in H$,

$$h_1K = h_2K \iff h_2^{-1}h_1 \in K \iff h_2^{-1}h_1 \in H \cap K \iff h_1H \cap K = h_2H \cap K.$$

Thus the number of distinct cosets in $\{hK \mid h \in H\}$ equals the number of distinct cosets in $\{h(H \cap K) \mid h \in H\}$. But H is a group and $H \cap K$ a subgroup, so by Lagrange's theorem, we know that the number of cosets of $H \cap K$ in H is

$$|H/H \cap K| = |H|/|H \cap K|.$$

Thus

$$|HK| = |K| \cdot \text{number of distinct cosets} = |K|\frac{|H|}{|H \cap K|}.$$

■

Remark 2.1.4 In the special case that HK is a subgroup of G (see Theorem 1.5.3), we shall see this result follows from the second isomorphism theorem for groups.

Example 2.1.5 Let G be the symmetric group S_3, and let $H = \langle(1\ 2)\rangle$ and $K = \langle(2\ 3)\rangle$ be cyclic groups of order 2 generated by the given transpositions. It is clear by inspection that $H \cap K = \{1\}$, so $|HK| = \frac{2 \cdot 2}{1} = 4$. By Corollary 2.1.2, since $4 \nmid 6$, HK cannot be a subgroup of G. □

Example 2.1.6 In a different direction, and still with $G = S_3$, if $H = \langle(1\ 2)\rangle$ and $K = \langle(1\ 2\ 3)\rangle$, then H and K have orders 2 and 3 respectively. It follows that $|H \cap K| = 1$ since by Lagrange $|H \cap K|$ must divide $|H| = 2$ and $|K| = 3$, hence must divide their gcd which is 1. Thus HK has order 6, so $HK = G$. In particular HK is a group. □

2.2 Understanding quotients and further isomorphism theorems

Proposition 1.3.6 gives equivalent conditions for a subgroup H of a group G to be normal. We also know that the kernel of any group homomorphism is a normal subgroup; indeed being normal is equivalent to being the kernel of some homomorphism, though that fact is not in and of itself all that useful. On the other hand, both conditions are themselves useful as we shall see in examples.

Example 2.2.1 Let F be a field, and $G = GL_n(F)$, the **general linear** group (of invertible $n \times n$ matrices with entries in F). From the point of view of rings, if $R = M_n(F)$ is the ring of $n \times n$ matrices with entries in F, then $G = R^\times$, the unit group of the ring R. Let $K = SL_n(F)$, the **special linear group**, the subset of invertible matrices whose determinant is one.

Give two proofs that K is a normal subgroup of G, one using Proposition 1.3.6, and the other characterizing K as the kernel of a group homomorphism.

Solution 1. Directly, one can use determinants both to show that K is a subgroup, but also that it is normal. For the normality part, let $g \in G$ and $k \in K$. To check that $gkg^{-1} \in K$, one needs only observe that

$$\det(gkg^{-1}) = \det(g)\det(k)\det(g)^{-1} = \det(k) = 1$$

to show $gkg^{-1} \in K$. We have used the fact that $\deg(g^{-1}) = \det(g)^{-1}$, and of course that $\det g, \det k$ are nonzero scalars in F, which commute.

Solution 2. A second solution is to recognize K as the kernel of a homomorphism. One that comes to mind is

$$\varphi : GL_n(F) \to F^\times$$

given by $\varphi(g) = \det g$. Then K is obviously the kernel. □

Example 2.2.2 With the notation as above, show that $GL_n(F)/SL_n(F)$ is an abelian group.

Solution. Since $K = SL_n(F)$ is a normal subgroup of $G = GL_n(F)$, we know at least that the quotient is a group. One could show it is abelian directly by showing that $gKg'K = g'KgK$ since

$$gKg'K = g'KgK \iff gg'K = g'gK \iff g^{-1}g'^{-1}gg' \in K$$

via determinants, but that is a bit grungy, and does not really leave us with a sense of what G/K looks like.

We return to the homomorphism

$$\varphi : GL_n(F) \to F^\times$$

given by $\varphi(g) = \det g$. We already know the kernel, and a moment's thought convinces you that φ is surjective for given $\alpha \in F^\times$, φ takes the diagonal matrix $\operatorname{diag}(\alpha, 1, \ldots, 1)$ to α. Thus via the first isomorphism theorem,

$$GL_n(F)/SL_n(F) \cong F^\times$$

the multiplicative group of a field which is certainly an abelian group. □

Theorem 2.2.3 Second Isomorphism Theorem. *Let H, K be subgroups of a group G, and suppose that $H \subseteq N_G(K)$ (e.g., if $K \trianglelefteq G$). Then $HK \leq G$, $K \trianglelefteq HK$ and*

$$HK/K \cong H/(H \cap K).$$

Proof. By Theorem 1.5.3, we have seen the condition $H \subseteq N_G(K)$ proves that HK is a subgroup of G, and also easily shows that $K \trianglelefteq HK$. For the rest, we try to let our theorems do the work.

Consider the homomorphism

$$\varphi : H \to HK/K$$

which is the composition of natural homomorphisms given by

$$H \to HK \text{ via } h \mapsto h \cdot 1 \in HK$$

and the projection

$$HK \to HK/K \text{ via } hk \mapsto hkK,$$

so $\varphi(h) = hK$.

Since the coset $hkK = hK$, we see the map φ is surjective. For the kernel, we see that $\varphi(h) = hK = K$ iff $h \in H \cap K$, and the first isomorphism theorem gives the result. ■

Remark 2.2.4 Note that when H, K are finite subgroups, this theorem is stronger than Proposition 2.1.3 which computed the cardinality of HK.

Exercise 2.2.1 Let F be a field, $G = GL_n(F)$, $K = SL_n(F)$, and $H = D_n(F)$ the subgroup of diagonal matrices in G. You have already observed that K is a normal subgroup of G.

(a) Show that $H = D_n(F)$ is not a normal subgroup of G.

Solution. Recall that the process of diagonalizing a matrix A (if possible) is one of finding an invertible matrix P so that $P^{-1}AP = D$ is diagonal, which means (generally) $PDP^{-1} = A$ is not diagonal. For a specific example,

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}.$$

(b) Show that $G = HK$.

Solution. Let $A \in G = GL_n(F)$, and let $\alpha = \det A \in F^\times$. Then

$$\mathrm{diag}(\alpha^{-1}, 1, \ldots, 1)A \in SL_n(F),$$

so

$$G = GL_n(F) \subseteq HK,$$

and of course $HK \subseteq G$, which finishes the argument.

(c) Apply the second isomorphism theorem to H, K.

Solution. Since we know that $K \trianglelefteq G$, the theorem applies, so we have

$$GL_n(F)/SL_n(F) = HK/K \cong H/(H \cap K) \cong F^\times.$$

Of course we could have proven this directly with the determinant map using domain H.

Theorem 2.2.5 Third Isomorphism Theorem. *Let G be a group with normal subgroups H, K, and suppose that $H \leq K$. Then $K/H \trianglelefteq G/H$, and*

$$(G/H)/(K/H) \cong G/K.$$

Proof. It is simply refreshing to let our theorems do all the work in the proof of this theorem with a seemingly complicated statement. With H, K normal in G, it is immediate that $H \trianglelefteq K$, so all the quotients G/H, G/K, and K/H are groups. Now we let the theorems take over.

By the fundamental homomorphism theorem, the natural surjective map (projection) $\pi : G \to G/K$ **factors through** the quotient G/H, so essentially for free we are handed a natural (well-defined) surjective homomorphism

$$\pi_* : G/H \to G/K$$

defined by $\pi_*(gH) = gK$. We need only ask for its kernel and apply the first isomorphism theorem. But

$$\ker \pi_* = \{gH \in G/H \mid gK = K\} = \{gH \in G/H \mid g \in K\} = K/H.$$

■

Example 2.2.6 A simple is to take $G = \mathbb{Z}$, so all subgroups are normal, and for positive integers m, n, let $K = m\mathbb{Z}$ and $H = mn\mathbb{Z}$. Thus

$$(\mathbb{Z}/mn\mathbb{Z})/(m\mathbb{Z}/mn\mathbb{Z}) \cong \mathbb{Z}/m\mathbb{Z}.$$

□

Theorem 2.2.7 Fourth Isomorphism/Correspondence Theorem. *Let $\varphi : G \to G'$ be a surjective homomorphism between groups G, G' having kernel K. Then there is a one-to-one correspondence between the subgroups of G' and those of G which contain K. Moreover, under the correspondence, normal subgroups* *correspond to normal subgroups.*

Sketch. The correspondence is straightforward:

$$\begin{aligned} H' \leq G' &\mapsto \varphi^{-1}(H') = \{g \in G \mid \varphi(g) \in H'\} \leq G, \\ H \leq G &\mapsto \varphi(H). \end{aligned}$$

The one important fact to prove in order to establish the correspondence is one-to-one is to note that for $H \leq G$,

$$\varphi^{-1}(\varphi(H)) = HK,$$

but since $K \leq H$, we have $HK = H$. ■

Example 2.2.8 What are the subgroups of $\mathbb{Z}/24\mathbb{Z}$?

Answer. Consider the natural surjective homomorphism $\varphi : \mathbb{Z} \to \mathbb{Z}/24\mathbb{Z}$, and use the correspondence theorem. There is a one-to-one correspondence between the subgroups of $\mathbb{Z}/24\mathbb{Z}$ and the subgroups of $\mathbb{Z}$ which contain $24\mathbb{Z}$ (the kernel). But every subgroup of $\mathbb{Z}$ is of the form $n\mathbb{Z}$ for some n, and $n\mathbb{Z} \supseteq 24\mathbb{Z}$ if and only if $n \mid 24$. So the subgroups of $\mathbb{Z}$ containing $24\mathbb{Z}$ are

$$d\mathbb{Z} \text{ for } d = 1, 2, 3, 4, 6, 8, 12, 24,$$

so for these same d the subgroups of $\mathbb{Z}/24\mathbb{Z}$ are $d\mathbb{Z}/24\mathbb{Z}$. □

2.3 Group Actions and applications

A powerful tool both within algebra and other areas of mathematics is the notion of a **group action.** There are two equivalent ways in which the characterize a group action. We begin with the more “constructive” one.

Definition 2.3.1 Let G be a group and X a set. We say that G **acts on the set** X if there is a map $G \times X \to X$, denoted $(g, x) \mapsto g \cdot x$, satisfying two properties:

- $e \cdot x = x$ for all $x \in X$. Here e is the identity of the group G.
- $g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x$ for all $g_1, g_2 \in G$ and $x \in X$.

◊

Example 2.3.2 We say that the group G **acts on itself by conjugation** if $X = G$ and the map $G \times G \to G$ is given by $(g, x) \mapsto gxg^{-1}$. □

Example 2.3.3 Let G be a group, H a subgroup, and $X = G/H$ the set of left cosets. We say that the group G **acts on** G/H **by left translation** if the map $G \times G/H \to G/H$ is given by $(g, xH) \mapsto gxH$. □

Example 2.3.4 Let G be a finite group and for a prime p, suppose that $p^m \mid |G|$. Let

$$X = \{H \leq G \mid |H| = p^m\}.$$

The Sylow theorems tell us that X is a non-empty set. We let the group G **act on X by conjugation** via

$$(g, H) \mapsto gHg^{-1}.$$

We note that this makes sense since for a given $g \in G$, the map $\varphi_g : G \to G$ given by $\varphi_g(x) = gxg^{-1}$ is an isomorphism, and as such

$$gHg^{-1} = \varphi_g(H)$$

is a subgroup of G having the same cardinality at H, so is again in X. □

There is an equivalent way in which to think of a group action G acting on a set X, and that is as a group homomorphism $\varphi : G \to Per(X)$, where $Per(X)$ is the set of bijections $X \to X$ viewed as a group under function composition.

Proposition 2.3.5 Equivalent notions of group actions. *There is a one-to-one correspondence between group actions of a group G on a set X described as a map $G \times X \to X$, and homomorphisms $G \to Per(X)$ given by:*

- *Given an action denoted by $(g, x) \mapsto g \cdot x$, define the homomorphism $\varphi : G \to Per(X)$ by*

$$\varphi(g) = \varphi_g \in Per(X) \text{ where } \varphi_g(x) = g \cdot x.$$

- *Conversely, given a homomorphism $\varphi : G \to Per(X)$ given by $\varphi(g) = \varphi_g \in Per(X)$, define the map $G \times X \to X$ by*

$$(g, x) \mapsto \varphi_g(x).$$

These correspondences are inverse to one another.

Remark 2.3.6 This may seem a bit complicated at first glance. Simply realize that the homomorphism $\varphi : G \to Per(X)$ has as its codomain the group of bijective functions $X \to X$. These are simply bijections as in general X has no algebraic structure. On the other hand, this explains our notation: Since $\varphi(g)$ is a function, it makes intuitive sense to name it as such, so we set $\varphi(g) = \varphi_g$ where $\varphi_g : X \to X$ is a bijective map.

There are very good reasons for utilizing this correspondence. Group actions viewed as maps $G \times X \to X$ are easy to describe, but have less obvious group-theoretic implications. However, an action described as a homomorphism $\varphi : G \to Per(X)$ has obvious algebraic objects related to it, such as its kernel.

Example 2.3.7 Let a group G act on itself by left translation. That means there is a map $G \times G \to G$ given by $(g, x) \mapsto gx$ where the juxtaposition gx is the product in the group.

The associated permutation representation is

$$\varphi : G \to Per(G) \text{ given by } \varphi(g) = \varphi_g \text{ with } \varphi_g(x) = gx,$$

and is called the **left regular representation** of G.

What is the kernel of φ? By definition,

$$\ker \varphi = \{g \in G \mid \varphi(g) = \varphi_g = e_{Per(G)}\},$$

the identity bijection. That means that $\varphi_g(x) = gx = x$ for all $x \in X = G$. Simply taking $x = e_G$ tells us that $g = e$, so the kernel is trivial and this permutation representation is injective. □

As a corollary of this observation we obtain Cayley's theorem.

Theorem 2.3.8 Cayley's theorem. *Every group G is isomorphic to subgroup of a permutation group, in particular $Per(G)$. If G is finite with $|G| = n$, then G is isomorphic to a subgroup of S_n.*

To gain further insight into group actions, we need to define two more notions, an **orbit** and a **stabilizer**.

Definition 2.3.9 Let G act on a set X and let $x \in X$.

- The **stabilizer** or **isotropy** subgroup of x is

$$G_x := \{g \in G \mid g \cdot x = x\}.$$

- The **orbit** of x is

$$Gx = G \cdot x := \{g \cdot x \mid g \in G\} = \{y \in X \mid y = g \cdot x \text{ for some } g \in G\}.$$

Note that G_x is a subgroup of G, while Gx is a subset of X. ♢

Consider a couple of examples to anchor these definitions.

Example 2.3.10 Let G act on itself by conjugation, and let $x \in X$. Then

$$Gx = \{gxg^{-1} \mid g \in G\}$$

is the **conjugacy class** of x, while

$$G_x = \{g \in G \mid gxg^{-1} = x\}$$

is the **centralizer** of x. □

Example 2.3.11 Let G be a group, and $X = \{H \leq G\}$ be the set of subgroups of G. G acts on X by conjugation. Let $x = H \in X$. Then

$$Gx = \{gHg^{-1} \mid g \in G\}$$

is the **conjugacy class** of H, while

$$G_x = \{g \in G \mid gHg^{-1} = H\}$$

is the normalizer of H. □

Exercise 2.3.1 Let G act by conjugation on the set X of all subgroups of G as in the second example above.

(a) What does it mean when the size of the orbit is one, meaning the orbit consists only of H? How is an orbit of size one related to the stabilizer?

Answer. An orbit of size one says $H = gHg^{-1}$ for all $g \in G$, that is H is a normal subgroup. It follows that in this case, the isotropy subgroup (normalizer of H) is all of G, so

$$|Gx| = 1 \iff G_x = G.$$

We shall generalize this below.

(b) Let $G = S_3$ and X all the subgroups of S_3. For each $x \in X$, compute the orbit and stabilizer.

Answer. All proper subgroups of S_3 are cyclic which makes our notation easier. Let e be the identity, t be any transposition: (1 2), (2 3), or (1 3), and T be either 3-cycle: (1 2 3) or (1 3 2).

$$\begin{array}{lll} x = \{e\} & Gx = \{e\} & G_x = G. \\ x = \langle t \rangle & Gx = \{\langle(1\ 2)\rangle, \langle(2\ 3)\rangle, \langle(1\ 3)\rangle\} & G_x = \langle t \rangle. \\ x = \langle T \rangle & Gx = \{x\} & G_x = G. \\ x = S_3 & Gx = \{x\} & G_x = G. \end{array}$$

Theorem 2.3.12 Orbit-Stabilizer theorem. *Let G act on a set X, and let $x \in X$. Then there is a bijection*

$$G/G_x \leftrightarrow Gx$$

between the set of left cosets G/G_x and the orbit Gx given by

$$gG_x \leftrightarrow gx.$$

As a consequence, we have that

$$[G : G_x] = |Gx|,$$

the index of the stabilizer equals the size of the orbit. Of course this statement is only really useful when the quantities are finite.

Proof. Let's show that the map $gG_x \mapsto gx$ is well-defined. Suppose that $gG_x = hG_x$. By Proposition 1.3.4, the two cosets are equal iff $h^{-1}g \in G_x$, the stabilizer.

So $gG_x = hG_x$ iff $h^{-1}gx = x$ iff $gx = hx$ (using the basic properties of a group action). Thus the map is not only well-defined but also one-to-one. It is also clear that it is surjective since given $gx \in Gx$, the coset gG_x maps onto it.

The final statement is immediate since

$$|Gx| = |G/G_x| = [G : G_x],$$

the last equality by definition. ■

Since we know that that cosets of G_x in G partition G, it should come as no surprise that the orbits partition X.

Proposition 2.3.13 *Let G act on a set X. Then the set of orbits $\{Gx \mid x \in X\}$ form a partition of X.*

Proof. By Definition 1.2.1, we need only check that every element of X is in some orbit, and that two orbits are either disjoint or equal.

By the property of a group action that requires $e \cdot x = x$ for all $x \in X$, we see that $x \in Gx$ (no matter what G is).

Suppose that $z \in Gx \cap Gy$. We must show that $Gx = Gy$. By symmetry, it is enough to show $Gx \subseteq Gy$. Since $z \in Gx \cap Gy$, we may write

$$z = g_1x = g_2y.$$

But that means that $x = (g_1^{-1}g_2)y$, so for any $g \in G$,

$$gx = (gg_1^{-1}g_2)y \in Gy$$

which completes the proof. ■

Finally, let's prove a couple of nontrivial results with group actions. For the first, we learned early on that a subgroup H for which the index $[G : H] = 2$ is normal (since there are only two (left or right) cosets one of which is H). The following is a good deal stronger.

Proposition 2.3.14 *Let G be a finite group, and H a subgroup whose index, $[G : H] = p$, is the smallest prime which divides the order of G. Then H is a normal subgroup of G.*

We use a lemma to simply the proof of the proposition, but which is useful in its own right.

Lemma 2.3.15 *Let G be a group, H a subgroup, and let G act on the coset space G/H by left translation. Let $\varphi : G \to Per(G/H)$ be the associated permutation representation. Then $\ker\varphi \leq H$.*

Proof. The permutation representation φ acts by

$$\varphi(g) = \varphi_g \text{ where } \varphi_g(xH) = gxH.$$

We note that $k \in \ker\varphi$ if and only if $\varphi(k) = \varphi_k$ is the identity map. So if $k \in \ker\varphi$, then

$$\varphi_k(xH) = kxH = xH$$

for all $x \in G$. But simply choosing $x = e$ gives us that $kH = H$ which means $k \in H$. ■

Now we return to our proposition.

Proof of Proposition 2.3.14. Let G be a finite group, and H a subgroup whose index, $[G : H] = p$, is the smallest prime which divides the order of G. Let G act on G/H by left translation. By the lemma, the kernel $K = \ker\varphi \subseteq H$. To show that H is normal, we do it indirectly, by showing that $H = K$ which is known to be a normal subgroup!

We know that $K \subseteq H \subseteq G$ and that $G/K \cong \operatorname{Im}\varphi \leq Per(G/H)$ by the first isomorphism theorem. By Lagrange's theorem

$$|G/K| = [G : K] \mid [G : H]! = |Per(G/H)| = p!.$$

By a simple homework type exercise we know that

$$[G : K] = [G : H][H : K] = p[H : K].$$

Putting things together we infer that

$$[H : K] \mid (p - 1)!.$$

But again by Lagrange,

$$[H : K] \mid |H| \mid |G|,$$

so

$$[H : K] \mid \gcd(|G|, (p - 1)!) = 1$$

since p was the smallest prime dividing $|G|$. Since $[H : K] = 1$, $H = K$ is a normal subgroup of G. ■

Another standard application of group actions provides the **class equation** which we use to show that every group of order p^2 (p prime) is abelian. We of course know that every group of order p is cyclic, hence abelian.

As preface, given a finite group G we let G act on itself by conjugation. Thus the orbits are conjugacy classes of elements, and stabilizers are centralizers. An element $z \in G$ lies in the **center**, $Z(G)$, of G if and only if its conjugacy class consists solely of the element z, and hence its centralizer, $C_G(z)$, is the entire group G. Because we know that that orbits partition the set (in this case the group G), we can write G as the disjoint union of orbits (conjugacy classes)

$$G = Gg_1 \sqcup Gg_2 \sqcup \cdots \sqcup Gg_t,$$

recalling that the classes of size one correspond to the elements in the center of G.

Theorem 2.3.16 Class Equation. *Let G be a finite group, and $g_1, \ldots, g_r$ representatives of the distinct conjugacy classes of G which are not contained in*

$Z(G)$. Then

$$|G| = |Z(G)| + \sum_{i=1}^{r} [G : G_{g_i}] = |Z(G)| + \sum_{i=1}^{r} [G : C_G(g_i)].$$

Proof. As indicated above, we can write G as the disjoint union of orbits:

$$G = Gh_1 \sqcup \cdots \sqcup Gh_t \sqcup Gg_1 \sqcup Gg_2 \sqcup \cdots \sqcup Gg_r$$

where the h_i are representatives of the conjugacy classes of size one, and the g_i represent those classes of size greater than one. As this is a disjoint union, it follows that

$$|G| = \sum_{i=1}^{t} |Gh_i| + \sum_{i=1}^{r} |Gg_i|.$$

The first sum is the order of the center, while each summand in the second

$$|Gg_i| = [G : G_{g_i}] = [G : C_G(g_i)]$$

by the orbit-stabilizer theorem, hence the result. ∎

Exercise 2.3.2 For a prime p, a **p-group** is a finite group whose order is a power of p.

(a) Let G be any p-group. Show that G has a non-trivial center. Specifically show that $p \mid |Z(G)|$.

Answer. This follows from the class equation:

$$|G| = |Z(G)| + \sum_{i=1}^{r} [G : C_G(g_i)],$$

where $g_1, \dots, g_r$ are representatives of the conjugacy classes of size greater than one. But that means that each $[G : C_G(g_i)] > 1$ and since $[G : C_G(g_i)] \mid |G|$ by Lagrange, each $[G : C_G(g_i)]$ is a power of p, in particular divisible by p. Thus

$$|Z(G)| \equiv |G| - \sum_{i=1}^{r} [G : C_G(g_i)] \equiv 0 \pmod{p}.$$

(b) Let G be a group, and suppose that $G/Z(G)$ is cyclic. Show that G is abelian.

Answer. Let's write Z for $Z(G)$. G/Z cyclic means that $G/Z = \langle xZ \rangle$ for some $x \in G$. To show that G is abelian, choose $g, h \in G$ and we shall show they commute.

By our assumption, $gZ = x^m Z$ and $hZ = x^n Z$ for some integers m, n. This means that $g = x^m z_1$ and $h = x^n z_2$ for some elements $z_i \in Z$. It follows that

$$gh = x^m z_1 x^n z_2 = z_1 x^{m+n} z_2 = z_2 x^{n+m} z_1 = z_2 x^n x^m z_1 = x^n z_2 x^m z_1 = hg$$

using that the z_i commute with all elements in the group and $x^m x^n = x^n x^m$.

(c) Let G be a group of order p^2. Show that G is abelian.

Answer. From the first exercise, we know that $|Z(G)| = p$ or p^2. If $|Z(G)| = p^2$, then $G = Z(G)$, so is abelian. Assume to the contrary, $|Z(G)| = p$ (so G is not abelian). But $Z(G) \trianglelefteq G$, and since $G/Z(G)$ has order p, it is cyclic. But the previous exercise shows that G is abelian, a contradiction.

2.4 Some structure and classification theorems

Here we state with few proofs some structure theorems which advance the goal of classifying finite groups. We also include a few examples. We begin with a major result whose proof relies on group actions.

We have defined a p-group as any group whose order is a power of a prime p. Suppose that G is a finite group of order n and p is prime dividing n. Write $n = p^k n_0$ where $p \nmid n_0$, so k is the largest exponent so that $p^k \mid n$. A **p-subgroup** H of G is simply any subgroup which is a p-group. A **Sylow p-subgroup** of G is a subgroup whose order is the largest power of p dividing the order of the group, in this case p^k. We are about to state a result showing that such subgroups always exist.

Theorem 2.4.1 Sylow theorems. *Let G be a group of order $n = p^k n_0$ with p a prime and $p \nmid n_0$. Then*

- *There exist subgroups of G of all orders p^ℓ with $1 \le \ell \le k$. In particular, Sylow p-subgroups exist for all primes p dividing $|G|$. Every p-subgroup of G is contained in a Sylow p-subgroup of G.*

- *For a fixed prime p, if P and Q are two Sylow p-subgroups of G, they are conjugate, i.e., there exists $g \in G$ with $P = gQg^{-1}$.*

- *Let n_p equal the number of Sylow p-subgroups of G. Then $n_p \equiv 1 \pmod{p}$ and $n_p \mid n_0 = |G|/p^k$.}*

Underlying concepts. Let p be a prime dividing $|G|$, and let X be the set of all Sylow p-subgroups in G. The first Sylow theorem says that X is non-empty. Note that G acts on X by conjugation: if $P \in X$, then since $x \mapsto gxg^{-1}$ is an (inner) automorphism, gPg^{-1} is a subgroup of G having the same order as P, so is again an element of X.

Now that we have a group action, we can talk about orbits and stabilizers. So let $P \in X$ be a Sylow p-subgroup. The second Sylow theorem says that the orbit $G \cdot P = X$; G is said to act **transitively** on X.

But now n_p is the number of Sylow p-subgroups in G, but that is simply the size of X. Thus

$$n_p = |X| = |G \cdot P| = [G : N_G(P)],$$

where the last equality comes from the orbit-stabilizer theorem. Now we always have the inclusions

$$P \leq N_G(P) \leq G$$

and

$$n_0 = \frac{|G|}{|P|} = [G : P] = [G : N_G(P)][N_G(P) : P] = n_p[N_G(P) : P].$$

In particular $n_p \mid n_0$, part of the third Sylow theorem. ■

The normal subgroups of a group provide insight into its structure. A group G whose only normal subgroups are G and $\{e\}$ is called a **simple** group and are fundamental to the so-called Hölder program. For groups which are not simple, their normal subgroups often lead to their characterization as a product of smaller groups, which we investigate shortly. A corollary of the Sylow theorems provides a simple way to determine if a Sylow p-subgroup is normal.

Corollary 2.4.2 *In the notation of the Sylow theorem, $n_p = 1$ if and only if the Sylow p-subgroup is normal.*

Proof. Suppose that $n_p = 1$ and P is given Sylow p-subgroup. For any $g \in G$, gPg^{-1} is also a Sylow p-subgroup, and since there is only one, $P = gPg^{-1}$ for any $g \in G$, so $P \trianglelefteq G$.

Conversely, suppose that P is a normal Sylow p-subgroup, and let Q be any Sylow p-subgroup. By the second Sylow theorem, $Q = gPg^{-1}$ for some $g \in G$. But since P is normal, $Q = P$, hence $n_p = 1$. ■

Recall that given two groups G_1 and G_2, we can make their Cartesian product, $G_1 \times G_2$, of ordered pairs into a group under component-wise operations. What we would like is to characterize when a given group is isomorphic to a direct product of groups.

Proposition 2.4.3 *Let G be a group and H,K subgroups. Suppose that*

- *H and K are both normal subgroups.*
- *$H \cap K = \{e\}$*
- *$G = HK(= KH)$*

Then the map $H \times K \to HK = G$ given by $(h, k) \mapsto hk$ is an isomorphism, and G is called the ***(internal) direct product*** *of the subgroups H and K.*

Proof. Let $\varphi : H \times K \to G$ be defined by $\varphi((h, k)) = hk$. The map φ is surjective by the third assumption, and it is one-to-one by the second assumption:

$$hk = h'k' \iff (h')^{-1}h = k'k^{-1},$$

but since $H \cap K = \{e\}$, $(h')^{-1}h = k'k^{-1} = e$, thus $(h, k) = (h', k')$.

Showing that φ is a homomorphism is we need to exercise a bit of care.

$$\begin{aligned}\varphi((h_1,k_1)(h_2,k_2)) &= \varphi((h_1h_2,k_1k_2)) = h_1h_2k_1k_2 \\ \varphi((h_1,k_1))\varphi((h_2,k_2)) &= h_1k_1h_2k_2.\end{aligned}$$

So we need to show that

$$h_1h_2k_1k_2 = h_1k_1h_2k_2 \iff h_2k_1 = k_1h_2$$

for any $h_i \in H$ and $k_i \in K$. While in general $HK = KH$ being a subgroup of G does not imply that the elements commute, but when both subgroups are normal (and have trivial intersection) we gain some added power:

$$hk = kh \iff h^{-1}k^{-1}hk = e,$$

but using the normality of each subgroup, we see

$$\begin{aligned}h^{-1}k^{-1}hk &= (h^{-1}k^{-1}h)k \in K \\ h^{-1}k^{-1}hk &= h^{-1}(k^{-1}hk) \in H\end{aligned}$$

and since

$$H \cap K = \{e\},$$

the elements commute, and our map φ is a homomorphism. ■

Before giving an example, we state an important, but simple result which can be viewed as one version of the Chinese Remainder Theorem, though we give a direct proof.

Proposition 2.4.4 *Let Z_n denote a cyclic group of order n, so $Z_n \cong \mathbb{Z}/n\mathbb{Z}$.*

$$Z_m \times Z_n \cong Z_{mn} \text{ iff } \gcd(m,n) = 1.$$

Proof. Let $d = \gcd(m,n)$, and note that $\dfrac{mn}{d} = m\dfrac{n}{d} = n\dfrac{m}{d}$ is a product of integers. It follows that every element of $Z_m \times Z_n$ has exponent mn/d. So if $d > 1$, there is no element of order mn in $Z_m \times Z_n$, so the group is not cyclic.

Conversely, suppose that $d = 1$. Let $x \in Z_m$ have order m and $y \in Z_n$ have order n, and put $z = (x,y) \in Z_m \times Z_n$. Since

$$z = (x,y) = (x,e)(e,y) = (e,y)(x,e)$$

it is easy to see that z has exponent mn. We want to show that the order of z is mn.

So suppose that ℓ is any exponent for z. So

$$z^\ell = (x^\ell, y^\ell) = (e,e).$$

Since m is the order of x, we know that $m \mid \ell$, and since n is the order of y, we know that $n \mid \ell$. But $d = \gcd(m,n) = 1$ which implies that $mn \mid \ell$. Thus $\ell = mn$ is the smallest exponent, hence the order.

Thus $Z_m \times Z_n$ is cyclic of order mn, so

$$Z_m \times Z_n \cong Z_{mn}$$

as there is a unique cyclic group of any given order (up to isomorphism). ■

Example 2.4.5 Let $p < q$ be primes with $p \nmid (q-1)$. Then every group of order pq is cyclic.

Solution. We first apply the Sylow theorems to G. Let H_p and H_q be (respectively) Sylow p and q-subgroups of G. Because they have prime order, we know that $H_p \cong Z_p$ and $H_q \cong Z_q$. We want to know that G is the direct product of H_p and H_q. That they have trivial intersection is immediate from Lagrange since they have relatively prime orders.

Given their trivial intersection, Proposition 2.1.3 tells us that $|H_pH_q| = pq$, so necessarily $G = H_pH_q$. All the remains is for us to show that each of the Sylow subgroups is normal.

Proposition 2.3.14 tells us that since $[G : H_q] = p$ is the smallest prime dividing the order of G, that it must be normal, though we give an independent proof using the Sylow theorems.

In the notation of the Sylow theorems, the subgroups will both be normal iff $n_p = n_q = 1$. By the Sylow theorems, we know that

$$n_p \equiv 1 \pmod{p} \text{ and } n_p \mid q, \text{ while } n_q \equiv 1 \pmod{q} \text{ and } n_q \mid p.$$

Since $n_q \mid p$, $n_q = 1$ or p. But if $n_q = p$, then $n_q = p \equiv 1 \pmod{q}$ which says that $q \mid (p-1)$. But $p < q$ by assumption, so that is impossible. Thus $n_q = 1$ implying H_q is a normal subgroup.

Similarly, $n_p = 1$ or q. If $n_p = q$, then $n_p = q \equiv 1 \pmod{p}$, implying that $p \mid (q-1)$, contrary to assumption.

Thus both subgroups are normal, have trivial intersection, and their product is G, so by Proposition 2.4.3,

$$G \cong H_p \times H_q \cong Z_p \times Z_q \cong Z_{pq}$$

the last isomorphism by Proposition 2.4.4. □

Remark 2.4.6 The condition that $p \nmid (q-1)$ was absolutely critical in the previous example. Consider groups of order $6 = 2 \cdot 3$. Since $2 \mid (3-1)$ the argument we gave does not show the Sylow 2-subgroup is normal. Indeed it need not be.

If $G = S_3$, the symmetric group, we know that the Sylow group H_3 is normal (generated by either 3-cycle), however there are 3 Sylow 2-subgroups, each generated by a different transposition, so $n_2 = 3$, and Sylow 2-subgroups are not normal.

Moreover, it is clear that $S_3 \ncong H_2 \times H_3$ since the later is abelian, while S_3 is not. Of course if the group had been $G = \mathbb{Z}/6\mathbb{Z}$, both Sylow subgroups would be normal, and $G \cong H_2 \times H_3$.

Finally, this is not an isolated situation. After all for any odd prime q, $2 \mid (q-1)$, so any group of order $2q$ can have this issue. Indeed, we know the problem will arise, since there are dihedral groups (non-abelian of order $2n$) for every $n \geq 3$.

Remark 2.4.7 It is also worth noting that when a group G has subgroups H, K with $G = HK$, $H \cap K = \{e\}$ and *only one* of H or K is normal, there is still something that can be said, namely that G is a **semi-direct** product of H and K. The structure is more complicated since that map $\varphi : H \times K \to G$ given by $(h,k) \mapsto hk$ is *not* a homomorphism.

One of the remarkably pretty results from group theory is the classification of finite abelian groups. While relatively easy to state, the proof is rather long. You will find direct proofs in textbooks focused on just on groups, or more general proofs which apply to finitely generated modules over PIDs of which finite abelian groups are a special case.

We begin with an intermediate result which we shall use to give the full result.

Theorem 2.4.8 *Let G be a finite abelian group whose order n has prime factorization $n = p_1^{e_1} \cdots p_r^{e_r}$. Let H_i be a Sylow p_i-subgroup of G. Then*

$$G \cong H_1 \times \cdots \times H_r.$$

Proof. The proof is by induction on r. First note that all subgroups of G are normal since G is abelian. If $r = 1$ there is nothing to prove, and the case of $r = 2$ is a direct application of Proposition 2.1.3 and Proposition 2.4.3.

Now consider $r = 3$. By Proposition 2.4.3, $H := H_1H_2 \cong H_1 \times H_2$. Moreover H is normal since G is abelian, and $H \cap H_3 = \{e\}$ by Lagrange. Again by Proposition 2.1.3 and Proposition 2.4.3, we have

$$G = HH_3 \cong H \times H_3 \cong H_1 \times H_2 \times H_3.$$

Now assume that $r \geq 4$, and as above we have constructed

$$H := H_1 \cdots H_{r-1} \cong H_1 \times \cdots \times H_{r-1}.$$

The same arguments now show that $G = HH_r \cong H \times H_r$ which finishes the proof. ■

Having reduced the structure of a finite abelian group to a direct product of its Sylow p-subgroups, we now characterize all the isomorphism types of an abelian p-group.

Theorem 2.4.9 *Let G be a finite abelian group of order p^n for some prime p and $n \geq 1$. Then $G \cong Z_{p^{a_1}} \times Z_{p^{a_2}} \times \cdots \times Z_{p^{a_r}}$ with $a_1 \geq a_2 \geq \cdots \geq a_r \geq 1$ and $\sum_{i=1}^r a_i = n$.*

*Moreover, if $H \cong Z_{p^{b_1}} \times Z_{p^{b_2}} \times \cdots \times Z_{p^{b_s}}$ with $b_1 \geq b_2 \geq \cdots \geq b_s \geq 1$ and $\sum_{i=1}^s b_i = n$, then $G \cong H$ iff $r = s$ and $a_i = b_i$ for all $1 \leq i \leq r$. The powers, p^{a_i}, are called the **elementary divisors** of G.*

Remark 2.4.10 The integers $a_1 \geq a_2 \geq \cdots \geq a_r \geq 1$ with $\sum_{i=1}^{r} a_i = n$ is said to form a **partition** of n.

One defines the **partition function**, $p(n)$, which counts the number of partitions of the positive integer n. The first few values are easy to compute:

$$p(0) := 1; \quad p(1) = 1; \quad p(2) = 2; \quad p(5) = 7; \quad p(10) = 42, \ldots .$$

On the other hand, larger values can be more challenging:

$$\begin{aligned} p(20) &= 627 \\ p(40) &= 37338 \\ p(100) &= 190,569,292 \\ p(200) &= 3,972,999,029,388 \end{aligned}$$

There are many theorems about the partition function including asymptotics:

$$p(n) \sim \frac{e^{\kappa\sqrt{n}}}{4n\sqrt{3}}$$

as $n \to \infty$ where $\kappa = \pi\sqrt{2/3}$, or generating functions:

$$\prod_{m=1}^{\infty} \left(\frac{1}{1 - x^m} \right) = \sum_{n=0}^{\infty} p(n) x^n.$$

There are also recurrence formulas which produce the exact numbers listed above.

Example 2.4.11 Up to isomorphism find all abelian groups of order p^5, that is find a set of representatives of all the isomorphism classes of abelian groups of order p^5.

Solution. We begin by listing the partitions of 5:

$$5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1$$

so $p(5) = 7$, so there will be 7 isomorphism classes. The integers in the partition correspond to the elementary divisors p^a in the decomposition.

Thus G is isomorphic to precisely one of the following abelian groups:

$$\begin{aligned} &Z_{p^5} \\ &Z_{p^4} \times Z_p \\ &Z_{p^3} \times Z_{p^2} \\ &Z_{p^3} \times Z_p \times Z_p \\ &Z_{p^2} \times Z_{p^2} \times Z_p \\ &Z_{p^2} \times Z_p \times Z_p \times Z_p \\ &Z_p \times Z_p \times Z_p \times Z_p \times Z_p. \end{aligned}$$

□

Now we would like to combine Theorem 2.4.8 and Theorem 2.4.9 into one theorem, which characterizes finite abelian groups by their **invariant factors.**

Theorem 2.4.12 *Every finite abelian group is isomorphic to exactly one group of the form* $Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_r}$ *where and* $n_1 \mid n_2 \mid \cdots \mid n_r \geq 2$. *It follows that* $|G| = n_1 n_2 \cdots n_r$. *The* n_i *are called the* ***invariant factors*** *of* G.
(Main Idea). We know that every finite abelian group is a direct product of its Sylow p-subgroups, each one of which has a decomposition in terms of elementary divisors. To combine these products we use the Chinese Remainder theorem for groups joining the largest powers of elementary divisors for each of the primes into the first invariant factor, then the second largest into the second invariant factors, and so on. ■

An example translating between the two types should make things clear.

Example 2.4.13 To go from the invariant factor decomposition to the product of Sylow subgroups:

$$\begin{aligned} Z_{20} \times Z_{300} &\cong (Z_4 \times Z_5) \times (Z_4 \times Z_3 \times Z_{25}) \\ &\cong (Z_4 \times Z_4) \times Z_3 \times (Z_5 \times Z_{25}) \\ &\cong H_2 \times H_3 \times H_5 \text{ (Sylow subgroups)} \end{aligned}$$

Now we go from Sylow/elementary divisor to invariant factor decomposition collecting the largest powers of elementary divisors, then second largest and so on.

$$\begin{aligned} (Z_p \times Z_{p^2} \times Z_{p^3}) \times (Z_q \times Z_{q^5}) &\cong (Z_p) \times (Z_{p^2} \times Z_q) \times (Z_{p^3} \times Z_{q^5}) \\ &\cong Z_p \times Z_{p^2 q} \times Z_{p^3 q^5} = Z_{n_1} \times Z_{n_2} \times Z_{n_3}. \end{aligned}$$

□

Remark 2.4.14 While the Sylow/elementary divisor method may seem more natural, and is certainly the easier way to list all the isomorphism classes of an abelian groups of a given order, the invariant factor decomposition reveals deeper structure of the group. For example, while we know that if

$$G \cong Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_r}$$

with $n_1 \mid n_2 \mid \cdots \mid n_r \geq 2$, then the group has order $n = n_1 \cdots n_r$ and that every element in the group has order dividing n. The invariant factor decomposition tells us more, namely that the largest order of an element in the group is n_r.

We do one last example listing the isomorphism classes in both elementary divisor and invariant factor forms

Example 2.4.15 Up to isomorphism, classify all abelian groups of order $6125 = 5^3 \cdot 7^2$.

Solution. First we decompose G into a product of its Sylow subgroups: $G \cong H_5 \times H_7$ with $|H_5| = 5^3$ and $|H_7| = 7^2$. Next for each Sylow subgroup we need to compute its possible elementary divisors, and to do so, we need to compute the partitions of 2 and 3:

$$\begin{aligned}&\text{Partitions of 3: } 3 = 2+1 = 1+1+1\\&\text{Partitions of 2: } 2 = 1+1\end{aligned}$$

This translates to the following possibilities for the elementary divisor decomposition of each Sylow subgroup:

$$\begin{aligned}&H_5 \cong Z_{5^3} \text{ or } Z_{5^2} \times Z_5, \text{ or } Z_5 \times Z_5 \times Z_5\\&H_7 \cong Z_{7^2} \text{ or } Z_7 \times Z_7.\end{aligned}$$

Finally we assemble all the data we have computed to arrive at all the possible isomorphism classes both in terms of elementary divisors and invariant factors.

$\mathbb{Z}_{5^3} \times \mathbb{Z}_{7^2}$	$\mathbb{Z}_{6125}$
$\mathbb{Z}_{5^3} \times \mathbb{Z}_7 \times \mathbb{Z}_7$	$\mathbb{Z}_7 \times \mathbb{Z}_{875}$
$\mathbb{Z}_5 \times \mathbb{Z}_{5^2} \times \mathbb{Z}_{7^2}$	$\mathbb{Z}_5 \times \mathbb{Z}_{1225}$
$\mathbb{Z}_5 \times \mathbb{Z}_{5^2} \times \mathbb{Z}_7 \times \mathbb{Z}_7$	$\mathbb{Z}_{35} \times \mathbb{Z}_{175}$
$\mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_{7^2}$	$\mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_{245}$
$\mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_7 \times \mathbb{Z}_7$	$\mathbb{Z}_5 \times \mathbb{Z}_{35} \times \mathbb{Z}_{35}$
elementary divisors	invariant factors

□

Checkpoint 2.4.16 The group $G = Z_{25} \times Z_{245}$ is also an abelian group of order 6125, but does not appear in the lists given in the example above. To which groups in the above list is it isomorphic?

Hint. Decompose the given direct product into elementary divisors

2.5 The Symmetric Group

For a set X, denote by S_X the set of bijective maps $f : X \to X$. We make S_X into a group with composition of functions as the binary operation: $(f, g) \mapsto f \circ g$, so $f \circ g(x) = f(g(x))$.

The identity element is the identity map $id_X : X \to X$ defined by $id_X(x) = x$ for all $x \in X$. Every element $f \in S_X$ has an inverse since each f is one-to-one and onto. The operation is associative since composition of maps is.

When X is finite, say $|X| = n$, we assume $X = \{1, 2, \ldots, n\}$ and write S_n for S_X. It is easy to check that $|S_n| = n!$. The group S_n is called the **symmetric group on n letters**. Any subgroup of S_n is called a **permutation group**.

Note that $S_1 = \{id\}$ and $S_2 = \{id, \sigma\}$ (where σ interchanges 1 and 2) are cyclic groups. For $n \geq 3$, it is easy to check that S_n is nonabelian.

Consider $\sigma \in S_{11}$ described somewhat cumbersomely by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 5 & 1 & 3 & 2 & 7 & 6 & 4 & 9 & 8 & 11 & 10 \end{pmatrix}$$

where σ takes an element in the top row to the element directly below it. It is more convenient to write σ as a product of disjoint cycles.

Recall that a **cycle of length** k (or a k-**cycle**) is a permutation, $\tau \in S_X$, for which there exist $a_1, \ldots, a_k \in X$ satisfying

$$\tau(a_i) = a_{i+1} \text{ for } 1 \leq i \leq k,\ \tau(a_k) = a_1, \text{ and } \tau(x) = x \text{ for all other } x \in X.$$

The cycle τ is written as $\tau = (a_1\ a_2\ \cdots\ a_k)$.

Example 2.5.1 For example, we can write

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 5 & 1 & 3 & 2 & 7 & 6 & 4 & 9 & 8 & 11 & 10 \end{pmatrix}$$

as as a **product of disjoint cycles**:

$$\sigma = \underbrace{(1\ 5\ 7\ 4\ 2)}_{5-cycle}\ \underbrace{(3)(6)}_{1-cycles}\ \underbrace{(8\ 9)(10\ 11)}_{2-cycles}.$$

□

Recall the the following properties:

Proposition 2.5.2

1. *Every permutation can be written uniquely as the product of disjoint cycles.*

2. *The cycle can be represented as*

$$(a_1\ \ldots\ a_r) = (a_2\ a_3\ \ldots a_r\ a_1) = (a_3\ a_4\ \ldots,\ a_r\ a_1\ a_2) = \cdots$$

3. *Disjoint cycles commute.*

The process by which we write a permutation as a product of disjoint cycles can be presented as algorithmic, but it also has a natural interpretation in terms of group actions.

Suppose we are given a permutation σ in S_n and want to find its representation as a product of disjoint cycles. Let $G = \langle\sigma\rangle$ be the cyclic group generated by σ. Then G acts on $X = \{1, \ldots, n\}$ by $(\tau, k) \mapsto \tau(k)$. Given a group action, we know that the set X is partitioned into orbits. Each orbit corresponds to one of the disjoint cycles in the decomposition:

$$G \cdot k = \text{Orbit of } (k) = \{\tau(k) \mid \tau \in \langle\sigma\rangle\}.$$

The only difference between the orbit and the cycle, is that the elements in the cycle are ordered to convey a bit more information:

$$(k\ \sigma(k)\ \sigma^2(k)\ \cdots\ \sigma^{\ell-1}(k))$$

where ℓ is the smallest positive integer such that $\sigma^\ell(k) = k$. So now the proposition above is clearer: the cycles are uniquely determined because the orbits under the action of $\langle\sigma\rangle$ are, an orbit can be named by any element it contains since the group acts transitively on each orbit, and X being the disjoint union of orbits does not depend on the order in which you write the orbits in the union.

Exercise 2.5.1

(a) Show that the order (as a group element in S_n) of an r-cycle $\sigma = (a_1\ a_2\ \cdots\ a_r)$ is r.

Solution. σ takes a_i to a_{i+1} for $1 \le i \le r$ and a_r to a_1, it is straightforward to check by induction that σ^j takes a_i to a_{i+j} with $i + j$ read with least positive residues modulo r. It follows that the order of σ is r.

(b) Write a permutation $\sigma \in S_n$ as a product of disjoint cycles: $\sigma = \sigma_1 \cdots \sigma_s$. Show that the order of σ is the least common multiple of the lengths of the cycles σ_i.

Solution. Let m_j be the length (order) of the cycle σ_j, and let $M = \operatorname{lcm}\{m_1, \ldots, m_s\}$. Then, since disjoint cycles commute,

$$\sigma^M = (\sigma_1 \cdots \sigma_s)^M = \sigma_1^M \cdots \sigma_s^M = 1$$

since each m_j is a multiple of M. It follows from Lagrange that the order of σ divides M. Suppose that $N = |\sigma|$, then since disjoint cycles commute,

$$\sigma^N = (\sigma_1 \cdots \sigma_s)^N = \sigma_1^N \cdots \sigma_s^N = 1,$$

hence

$$\sigma_1^{-N} = \sigma_2^n \cdots \sigma_t^N.$$

Recall that the σ_j are disjoint cycles so as permutations only move sets which are disjoint from one another. That is, σ_1 cannot move any integer in the sets moved by $\sigma_2, \ldots, \sigma_s$, yet their Nth powers are equal. The only resolution is that $\sigma_1^{-N} = 1 = \sigma_1^N$. This implies $m_1 \mid N$. Since the cycles commute, we may do the same trick for all the σ_j to conclude $m_j \mid N$ for all j, and hence $M \mid N$. Thus $|\sigma| = M$.

Exercise 2.5.2 Write a permutation $\sigma \in S_n$ as the product of disjoint cycles —including the 1-cycles:

$$\sigma = \sigma_1 \cdots \sigma_r,$$

and let a_i denote the length of σ_i. Because disjoint cycles commute, we may assume that the cycles are arranged in so their lengths appear in descending order and since we include the 1-cycles their lengths form a partition of the integer n.

If $\sigma = \sigma_1 \cdots \sigma_r$ has lengths $a_1 \ge \cdots \ge a_r \ge 1$, then we call the r-tuple $(a_1, \ldots, a_r)$ the **cycle type** of σ.

(a) Let $\tau = (a_1\ a_2\ \cdots\ a_k)$ be any k-cycle. Show that for any permutation σ,

$$\sigma\tau\sigma^{-1} = (\sigma(a_1)\ \sigma(a_2)\ \cdots\ \sigma(a_k)),$$

so in particular, conjugation takes a k-cycle to another k-cycle.

Solution. Let's first see where we send an integer of the form $\ell = \sigma(a_i)$ for some i. We see that \[\sigma\tau\sigma^{-1}(\sigma(a_i)) = \sigma\tau(a_i) = \sigma(a_{i+1}),\] where we read the subscripts modulo k, so we might conjecture that $\sigma\tau\sigma^{-1} = (\sigma(a_1)\,\sigma(a_2)\,\cdots\,\sigma(a_k))$, however to be sure, we must show that $\sigma\tau\sigma^{-1}$ fixes all other integers. So now let ℓ be an integer, with $\ell \neq \sigma(a_i)$ for any i. Then $\sigma^{-1}(\ell) \neq a_i$ for any i, so τ leaves it unchanged, so that σ takes it back to ℓ. So no integer other than the $\sigma(a_i)$ is moved by $\sigma\tau\sigma^{-1}$, and our decomposition is complete.

(b) Let $\tau, \tau' \in S_n$ be any two cycles of length k. Show that there is an element $\sigma \in S_n$ so that

$$\tau' = \sigma\tau\sigma^{-1}.$$

This means that for each k-cycles in S_n form their own and complete conjugacy class.

Solution. Let $\tau = (a_1\,a_2\,\cdots\,a_k)$ and $\mu = (b_1\,b_2\,\cdots\,b_k)$. Then for any permutation σ which takes $a_i \mapsto b_i$ we have the desired equality.

Use the exercises above to deduce the following proposition.

Proposition 2.5.3 *There is a one-to-one correspondence between conjugacy classes of elements in S_n are partitions of the integer n.*

Proof. Let $\sigma \in S_n$ and consider its conjugacy class. Write σ as the product of disjoint cycles (including the 1-cycles):

$$\sigma = \sigma_1 \cdots \sigma_r \text{ with } m_i = |\sigma_i|.$$

We have observed that (arranged in descending order since disjoint cycles commute), $m_1, \ldots, m_r$ is a partition of n. Since conjugation is an (inner) automorphism,

$$\tau\sigma\tau^{-1} = \tau\sigma\tau^{-1} \cdots \tau\sigma_r\tau^{-1},$$

and by the exercises above $\tau\sigma_j\tau^{-1}$ is again a cycle of length m_j, so the resulting cycles in $\tau\sigma\tau^{-1}$ again produce the same partition of n.

Given a partition of n, there is certainly a permutation with that cycle type, just by listing the integers from 1 to n, and grouping them into cycles of the desired length.

So we have a surjective map from permutations to partitions which is well-defined on conjugacy classes. We have another exercise which shows that two cycles of a given length are always conjugate and this extends to a product of disjoint cycles. ■

Permutations are divided into **even** and **odd** permutations according to one of the following two equivalent schemes. The virtue of the first definition is that it is clearly well-defined.

Define a function sgn : $S_n \to \{\pm 1\}$ as follows: for a permutation $\sigma \in S_n$, write σ as the product of disjoint cycles (including the 1-cycles), say $\sigma = \sigma_1 \cdots \sigma_t$. Then we define

$$\operatorname{sgn}(\sigma) = (-1)^{n-t}.$$

It follows that $\operatorname{sgn}(1) = 1 = (-1)^{n-n}$, and $\operatorname{sgn}((a\ b)) = (-1)^{n-(n-1)} = -1$. For a permutation σ, the value $\operatorname{sgn}(\sigma)$ is called the **sign of the permutation.**

For the second definition, note that any cycle can be written as the product of transpositions:

$$(a_1\ a_2\ \cdots\ a_r) = (a_1\ a_r) \cdots (a_1\ a_3)(a_1\ a_2),$$

hence so can any permutation (though not in a unique way). If

$$\sigma = \tau_1 \cdots \tau_r,$$

where all the τ_i are transpositions, then it is also true that

$$\operatorname{sgn}(\sigma) = (-1)^r.$$

Proposition 2.5.4 *The definitions are the sign function given above are equivalent. Moreover,* sgn : $S_n \to \{\pm 1\}$ *is a group homomorphism, and is surjective for* $n \geq 2$. *It's kernel is called the alternating group, denoted* A_n.

Remark 2.5.5 For $n \geq 2$, it follows from the first isomorphism theorem that A_n is a normal subgroup of index 2 in S_n. For $n \geq 3$, it is generated by the 3-cycles, and for $n \geq 5$, it is simple group.

Chapter 3

Basic results in ring theory

3.1 Basic definitions and motivations

Like most algebraic objects, certain adjectives can be applied to rings to refine their properties. Some familiar examples:

- **Non-commutative rings**: $n \times n$ matrices ($n \geq 2$), Hamilton's quaternions: $\mathbb{H}$, the four-dimensional vector space of $\mathbb{R}$ with basis $\{1, i, j, k\}$ subject to the relations $ij = k = -ji$ and $i^2 = j^2 = k^2 = -1$.
- **Commutative rings**: fields, $\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$, polynomial rings with coefficients in a commutative ring.
- **Integral domains**: (also called **entire rings**) are commutative rings with identity and no zero divisors such as $\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$ (iff n is prime), polynomial rings whose coefficient ring is an integral domain.

Much of ring theory evolved to accommodate generalizations of properties of the integers to more general rings. Attempts to extend the notion of unique factorization in the integers led both to the notion and importance of ideals and to entire subjects like algebraic number theory and algebraic geometry.

3.2 Factoring in integral domains

Let's review some ideas surrounding the concept of factorization in rings to remind ourselves of how certain terminology became relevant. Let's begin with factorization in the integers $\mathbb{Z}$. The Fundamental theorem of arithmetic is often phrased as

Theorem 3.2.1 Fundamental Theorem of Arithmetic. *Every integer $n > 1$ is either prime or can be factored as a product of primes. And such a factorization is unique up to a rearrangement of the factors.*

Remark 3.2.2 We simply take the statement of the fundamental theorem in this context at face value, but as soon as we begin to poke at the edges, all sorts of questions come up.

- So why is there this restriction to integers greater than one? Well, ok, we are smart enough to avoid 0 and 1, and perhaps even -1, but what's the matter with the integers $n < -1$?

- Then there is the word *prime* which in the integers plays two roles. A prime in $\mathbb{Z}$ acts as both an irreducible element and a prime element, but how exactly are those roles evidenced?

- Moreover, a prime in $\mathbb{Z}$ seems to be slightly more restrictive in its meaning than in general (an integer $p \geq 2$ whose only *positive* divisors are 1 and itself). Why those restrictions?

There are many properties that make the integers special, but the facts that the unit group consists of only two elements, ± 1, and that irreducibles and primes are the same, strongly influence how the Fundamental Theorem is stated when compared to a statement describing unique factorization in a more general integral domain.

Let's start with an arbitrary integral domain R, and consider what we might mean by unique factorization. Certainly we must begin with some statement which says we have factored the given element and it can't be factored anymore. We should exclude trying to factor 0 or units in the ring, and we shouldn't fuss about the order of factors nor associates. For example, we don't want to distinguish factorizations like

$$6 = 2 \cdot 3 = 3 \cdot 2 \cdot 1 = (-2) \cdot (-3).$$

But as we said in $\mathbb{Z}$, this is easy to control since there are only two units.

In $\mathbb{Q}[x]$, we would like to say that $f(x) = 2x$ cannot be factored anymore (is irreducible) even though

$$2x = 2 \cdot x = (2/3) \cdot (3x) = \cdots .$$

On the other hand, in $\mathbb{Z}[x]$, the polynomial $2x$ is not irreducible since $f(x) = 2x = 2 \cdot x$ and neither 2 nor x are units in $\mathbb{Z}[x]$.

So factoring into irreducibles seems like a pretty natural notion, though we still have to deal with associates.

Example 3.2.3 So let's take a nonzero, non-unit element a in an integral domain R and try to factor it into irreducibles. What works and what might go wrong?

Solution. Factoring is a simple process. We ask if the given (nonzero, non-unit) element a is irreducible. If it is, we are done. However if it is not, that means there is a nontrivial factorization, that is as a product

$$a = a_1 b_1$$

where neither factor is a unit. Now we iterate.

If both a_1, b_1 are irreducibles, we are done, otherwise assume a_1 is not. Then a_1 has a nontrivial factorization

$$a_1 = a_2 b_2$$

where neither factor is a unit. And we iterate. So what's the problem? There does not seem to be any mechanism to force this procedure to terminate. □

In the example above, we see that it is natural to try to factor in any integral domain, but what is lacking is a mechanism to terminate the procedure. While not often discussed in a first algebra course, this leads quite naturally to the notion of a **Noetherian ring**, but before we define it, let's make the transition from elements to ideals seem natural.

We are talking about the factorization of elements in a ring R. We say that the element b **divides** the element a (written $b \mid a$) if there exists a $c \in R$ so that $a = bc$, in other words if b is a **factor** of a. Now one issue is that if $b \mid a$, then so does bu for any unit u since

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

Notice that $b \mid a$ if and only if the principal ideals $(a) \subseteq (b)$, and in an integral domain, $(b) = (b')$ if and only if b and b' are associates. So principal ideals capture the notion of factoring and of associates all in one concept.

Returning to our example of factoring, if a factored non-trivially as $a = a_1 b_1$, and a_1 was not irreducible, then $a_1 = a_2 b_2$ with neither a_2, b_2 units. In terms of ideals we would have

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

This means that until at some point in the factorization an irreducible appeared in the factorization, this **ascending chain of ideals** would go on indefinitely. Addressing this issue is one of the conditions which define a Noetherian ring.

Theorem 3.2.4 *Let R be a commutative ring with identity. The following three conditions are equivalent and define what it means for the ring to be **Noetherian.***

1. *R satisfies the **ascending chain condition** (ACC), meaning that given any ascending chain of ideals*

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

in R, there exists an index $r \geq 1$ so that

$$I_r = I_{r+1} = I_{r+2} = \cdots .$$

2. *Every ideal $I \subseteq R$ is finitely generated.*

3. *Every non-empty collection of ideals has a maximal element.*

You certainly know of Noetherian integral domains since every PID is automatically Noetherian since every ideal is generated by a single element. The main takeaway here is the following result.

Theorem 3.2.5 *Let R be a Noetherian integral domain. Then every nonzero, non-unit in R can be factored as a finite product of irreducibles.*

Proof. From the example and discussion above, the Noetherian condition tells us that as we begin to fact a nonzero, non-unit an irreducible must eventually appear as one of its factors.

So if $a = \pi_1 a_1$ where π_1 is irreducible, we ask if a_1 is a unit. If it is, we are done. Otherwise $a_1 = \pi_2 a_2$ with π_2 irreducible. Again if a_2 is a unit we are done. Otherwise $a_3 = \pi_3 a_3$. So again at this stage we are building another ascending chain of ideals:

$$(a) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots$$

which must terminate, but terminating means at the last stage $a_{r-1} = \pi_r a_r$ with π_r irreducible and a_r a unit, which terminates the factorization. ■

Example 3.2.6 The ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$ is a Noetherian integral domain, so every nonzero, non-unit factors as a finite product of irreducibles, just like in $\mathbb{Z}$. But where $\mathbb{Z}$ enjoys unique factorization, $\mathbb{Z}[\sqrt{-5}]$ does not. So primes in $\mathbb{Z}$ are somehow different than irreducibles in $\mathbb{Z}[\sqrt{-5}]$. We investigate how.

Solution. That the ring $R = \mathbb{Z}[\sqrt{-5}]$ is Noetherian follows from somewhat more advanced knowledge: Since $\mathbb{Z}$ is a PID, it is Noetherian, and so by the **Hilbert basis theorem**, so is the polynomial ring $\mathbb{Z}[x]$. Now an easier exercise is that that homomorphic image of a Noetherian ring is Noetherian, and our ring R is the homomorphic image of $\mathbb{Z}[x]$ under the evaluation map induced by $x \mapsto \sqrt{-5}$. The ring is an integral domain since it is a subring of the field $\mathbb{C}$.

Now to come to the more relevant part. We can write down a sentence like

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

and claim this shows the ring does not have unique factorization, but that is just an assertion without much proof.

What does it mean to say this is a counterexample to unique factorization? Well we would have to know that $2, 3, 1 \pm \sqrt{-5}$ are all irreducible and not associates.

Both of these questions can be answered using the **norm** map $N : \mathbb{Z}[\sqrt{-5}] \to \mathbb{Z}$ given by

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

It is easy to see that since $N(\alpha) = \alpha\overline{\alpha}$, the product of the element and its complex conjugate, that $N(\alpha\beta) = N(\alpha)N(\beta)$ for any $\alpha, \beta \in R = \mathbb{Z}[\sqrt{-5}]$.

The first lemma to prove is that α is a unit in $\mathbb{Z}[\sqrt{-5}]$ if and only if $N(\alpha) = \pm 1$ (actually $+1$ in our case since $a^2 + 5b^2 \geq 0$ for all a, b). It follows that the units of $\mathbb{Z}[\sqrt{-5}] = \{\pm 1\}$, so it is at least clear that none of the factors in the two factorizations of 6 are associates.

To show that $2, 3, 1 \pm \sqrt{-5}$ are all irreducible is done case by case. We show that 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$. Suppose not. Then

$$2 = \alpha\beta$$

where neither α, β are units, i.e., have norm 1. But since the norm is multiplicative,

$$4 = N(2) = N(\alpha)N(\beta),$$

where now this equation is an equation in $\mathbb{Z}$ where we have unique factorization. So the only possibilities are that $N(\alpha) = 1, 2,$ or 4. We cannot have $N(\alpha) = 1$ since that means that α is a unit, nor can be have $N(\alpha) = 4$ since that forces β to be a unit. So our only chance for a nontrivial factorization is if $N(\alpha) = N(\beta) = 2$. But since $a^2 + 5b^2 = 2$ has no solutions for $a, b \in \mathbb{Z}$, we are forced to conclude that one of α or β is a unit, meaning that 2 is irreducible in $\mathbb{Z}[\sqrt{-5}]$. The other cases are analogous. □

The failure of unique factorization in the ring $\mathbb{Z}[\sqrt{-5}]$ is a consequence of the fact that not all irreducibles are primes. Having shown that they are irreducible, it is easy to show the elements are not prime. Recall, we have

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

so that means that $2, 3, 1 \pm \sqrt{-5}$ all divide 6, so in particular

$$2 \mid (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

We need to show that 2 does not divide either $1+\sqrt{-5}$ nor its conjugate. But an easy lemma is that in an integral domain, if $\pi_1 \mid \pi_2$ where the π_i are irreducible, then they must be associates. Since we know the units of the ring are only ± 1, it is clear they are not associate. So these 4 elements are irreducible, but not prime in $\mathbb{Z}[\sqrt{-5}]$.

The theorem we can now prove is

Theorem 3.2.7 *Let R be a Noetherian integral domain in which every irreducible element is a prime element. Then R is a unique factorization domain, meaning that every nonzero, non-unit in R has a factorization into a finite product of irreducibles which is unique in the sense that if*

$$a = \pi_1 \cdots \pi_r = \pi_1' \cdots \pi_s'$$

are two factorizations of a into irreducibles, then $r = s$, and (after a possible reordering) π_j and π_j' are associates for all $j = 1, \ldots, r$.

(sketch). Only the uniqueness statement remains to be shown, and the proof here is exactly as it is in $\mathbb{Z}$, so we simply sketch the argument.

Proceed by induction on r. If $r = 1$, then by the definition of an irreducible, it can have no nontrivial factorizations which makes $s = 1$ and $\pi_1 = \pi_1'$. For $r \geq 2$, we know that $\pi_1 \mid a$, so

$$\pi_1 \mid \pi_1' \cdots \pi_s'.$$

Since the irreducible π_1 is prime, it must divide one of the factors on the right hand side, so without loss of generality (since products commute), we can say

$\pi_1 \mid \pi_1'$. Again, both being irreducibles, they must be associate, so we may write

$$\pi_1 \cdots \pi_r = \pi_1' \cdots \pi_s' = \pi_1 \pi_2'' \pi_3' \cdots \pi_s',$$

where π_2'' is an associate of π_2'.

Since we are in an integral domain, the cancellation law holds yielding

$$\pi_2 \cdots \pi_r = \pi_2'' \pi_3' \cdots \pi_s',$$

and the argument finishes by induction. ■

The problem of course is how do we know if we are in a ring which has the property that every irreducible is prime? You may recall you spent time learning about Euclidean domains, PIDs, and UFDs and the relationship between them. What you leveraged to produce examples is that Euclidean domains are all PIDs, and that all PIDs are UFDs, but identifying a PID which was not Euclidean or a UFD which is not a PID was not immediately obvious. Let's first collect a few more tools and then dive into those issues.

3.3 Ideals and quotients

Let R, S be rings. A **ring homomorphism** $\varphi : R \to S$ is (as usual) a structure-preserving map, in this case taking sums to sums and products to products. It is in particular a homomorphism of the additive groups of the rings, so its **kernel** is $\ker \varphi := \{r \in R \mid \varphi(r) = 0\}$, and is an ideal of the ring R.

The fundamental homomorphism theorem for rings says that given a ring homomorphism $\varphi : R \to S$ and any ideal $I \subseteq \ker \varphi$, there is a well-defined ring homomorphism $\varphi_* : R/I \to S$ with $\varphi_* \circ \pi = \varphi$. Here $\pi : R \to R/I$ is the usual projection. In particular the **first isomorphism theorem** says

$$R/\ker \varphi \cong \operatorname{Im} \varphi.$$

Exercise 3.3.1 Let R, S be rings with identities and $\varphi : R \to S$ a (nontrivial) ring homomorphism.

(a) Show that $\varphi(1_R)$ an idempotent in S, so $\varphi(1_R) = 1_S$ or is a zero divisor in S.

Answer. $s := \varphi(1_R) = \varphi(1_R \cdot 1_R) = \varphi(1_R)\varphi(1_R) = s^2$ so $\varphi(1_R)$ is an idempotent. Rearranging we see that

$$\varphi(1_R)(\varphi(1_R) - 1_S) = 0$$

from which the conclusion follows.

(b) Conclude that if S is an integral domain, then $\varphi(1_R) = 1_S$.

Let A be a commutative ring with identity, and let $A[x]$ denote the ring of polynomials in the variable x with coefficients in A.

Proposition 3.3.1 *Let $a \in A$.*

- *Then every element $p \in A[x]$ has an expression of the form $p(x) = c_0 + c_1(x-a) + \cdots + c_n(x-a)^n$ for uniquely determined $c_i \in A$.*
- *$A[x]/(x-a) \cong A$*

Proof. Let $p(x) = b_0 + b_1 x + \cdots + b_n x^n$ be the usual expression for the polynomial. Consider the polynomial $q(x) = p(x+a)$. After a bit of algebra, the polynomial $q(x)$ has the expression $q(x) = c_0 + c_1 x + \cdots + c_n x^n$, for some (uniquely determined) $c_i \in A$. But then

$$p(x) = q(x-a) = c_0 + c_1(x-a) + \cdots + c_n(x-a)^n.$$

For the second statement, consider the evaluation homomorphism $ev_a : A[x] \to A$ given by $p(x) \mapsto p(a)$. It is immediate that ev_a is a surjective homomorphism. To compute its kernel, let $p \in A[x]$ and write p as $p(x) = c_0 + c_1(x-a) + \cdots + c_n(x-a)^n$. Then

$$ev_a(p) = 0 \iff p(a) = 0 \iff c_0 = 0.$$

It follows that $\ker ev_a = (x-a)$, and the result follows from the first isomorphism theorem. ■

We would like to understand finitely generated ideals as well as their quotient rings. Recall a simple but important idea. Suppose that S, T are subsets of a commutative ring R, and we wish to compare the ideals (S) and (T). Then

$$(S) = (T) \iff S \subseteq (T) \text{ and } T \subseteq (S),$$

which follows from the simpler $(S) \subseteq (T)$ if and only if $S \subset (T)$. Analogous statements hold for groups generated by set or subspaces generated by a collection of elements. Consider a few examples.

Example 3.3.2 Let $f(x) = x - 3$ and $g(x) = (x-3)(x-5) + 7$ be polynomials in $\mathbb{Z}[x]$. Compare the ideal $I = (f, g)$ in $\mathbb{Z}[x]$ versus $\mathbb{Q}[x]$.

Solution. Often it is useful to replace one set of generators of an ideal, by a simpler set of generators using the observation above. We claim that in either ring, $\mathbb{Z}[x]$ or $\mathbb{Q}[x]$

$$I = (f, g) = (f, 7).$$

We need only check that $f, g \in (f, 7)$ and $f, 7 \in (f, g)$. But of course $g = f(x-5) + 7 \in (f, 7)$ and $7 = g - f(x-5) \in (f, g)$

So now we simply consider the ideal $I = (x-3, 7)$. Since $7 \in \mathbb{Q}^\times = \mathbb{Q}[x]^\times$, viewed as an ideal in $\mathbb{Q}[x]$, $I = \mathbb{Q}[x]$.

Viewed as an ideal in $\mathbb{Z}[x]$, I is a proper ideal, indeed a maximal ideal, as we shall show a bit later by proving $\mathbb{Z}[x]/I = \mathbb{Z}[x]/(x-3, 7) \cong \mathbb{Z}/7\mathbb{Z}$.

For now, if you wish to prove it is a proper ideal, it suffices to show that 1 cannot be written as $h \cdot 7 + h' \cdot (x-3)$ for $h, h' \in \mathbb{Z}[x]$. □

Example 3.3.3 Every ideal in $\mathbb{Z}$ is principal: $I = n\mathbb{Z} = (n)$ for some $n \in \mathbb{Z}$.

Solution. Every ideal I is a subgroup of the additive group of $\mathbb{Z}$, a cyclic group, so $I = n\mathbb{Z}$ as a group, but this is also an ideal of $\mathbb{Z}$. □

Example 3.3.4 Let $n \in \mathbb{Z}$. The ideal $I = (n, x)$ is a proper ideal of $\mathbb{Z}[x]$ iff $n \neq \pm 1$.

Solution. If $n = \pm 1$, then I contains a unit in $\mathbb{Z}[x]$, so $I = \mathbb{Z}[x]$. For the converse, we assume that $n \neq \pm 1$ and show that $1 \notin I$. We proceed by contradiction and suppose that

$$1 = f \cdot n + g \cdot x$$

for some $f, g \in \mathbb{Z}[x]$.

Notice that $g \cdot x$ contributes zero to the constant term of $f \cdot n + g \cdot x$ no matter the choice of g, so if a_0 is the constant term of f, in order for $1 = f \cdot n + g \cdot x$, it is necessary that

$$1 = a_0 \cdot n.$$

But that demands that both a_0, n be units in $\mathbb{Z}[x]^\times = \mathbb{Z}^\times = \{\pm 1\}$ which is not true by assumption. □

Now we would like to consider quotients of rings and in particular, polynomial rings, but it is useful to recall a few definitions. First if I, J are two ideals of a commutative ring R with identity, it is useful to recall the meaning of $I + J$, IJ, and $I \cap J$ (see Definition 1.5.5).

Definition 3.3.5 Let R be a commutative ring with identity, and P an ideal of R. Then P is a **prime** ideal iff

- P is a proper ideal
- For every $a, b \in R$, if $ab \in P$, then either $a \in P$ or $b \in P$.

We remark that in a non-commutative ring, a different definition is required: P is a **prime** ideal iff P is proper and for any ideals $I, J \subset R$, $IJ \subseteq P$ implies $I \subseteq P$ or $J \subseteq P$. If the ring is commutative, this definition is equivalent to the previous one. ♢

Definition 3.3.6 Let R be a commutative ring with identity, and M an ideal of R. Then M is a **maximal** ideal iff

- M is a proper ideal
- Whenever I is an ideal of R with $M \subseteq I \subseteq R$, then either $I = M$ or R.

♢

Definition 3.3.7 Let R be a commutative ring with identity. Then two ideals I, J of R are said to be **comaximal** iff $I + J = R$. ♢

Example 3.3.8 If $M_1 \neq M_2$ are maximal ideals in a commutative ring R with identity, then they are comaximal ideals.

Solution. $M_i \subsetneq M_1 + M_2 \subseteq R$ and since the M_i are maximal, $M_1 + M_2 = R$. □

Exercise 3.3.2 Consider the ideals $I = (x-2)$, $J = (x+2)$ in $R[x]$ with $R = \mathbb{Z}$ or $\mathbb{Q}$.

(a) Show that I, J are comaximal ideals in $\mathbb{Q}[x]$, but not in $\mathbb{Z}[x]$.

Solution. It is easy to see that $4 \in I + J$. Since 4 is a unit in $\mathbb{Q}[x]$, $I + J = \mathbb{Q}[x]$. For $\mathbb{Z}[x]$, notice that every element of $I + J$ has an even constant term.

Exercise 3.3.3 Let R be a commutative ring with identity, with I, J ideals in R.

(a) Show that

$$(I + J)(I \cap J) \subseteq IJ \subseteq I \cap J$$

Solution. It is clear that $IJ \subset I \cap J$ since I, J are ideals of R. For the other inclusion, let $i \in I, j \in J, k \in I \cap J$. It is enough to check that $(i+j)k \in IJ$. Certainly $ik \in IJ$ since $k \in J$, and $jk \in JI$ since $k \in I$, but $IJ = JI$ since R is commutative.

(b) Show that if I, J are comaximal ideals, then

$$IJ = I \cap J$$

Solution. Since $I + J + R$, the above inclusions now read

$$I \cap J \subseteq IJ \subseteq I \cap J.$$

Recall the **Chinese Remainder Theorem** for rings.

Theorem 3.3.9 *Let R be a commutative ring with identity, and I, J comaximal ideals. Then*

$$R/IJ = R/(I \cap J) \cong R/I \times R/J.$$

(sketch). From the exercise above we know that $IJ = I \cap J$. Consider the natural projections

$$R \to R/I \times R/J \text{ given by } r \mapsto (r + I, r + J).$$

It is immediate to check this is a homomorphism with kernel $I \cap J$. The issue is surjectivity. Since I, J are comaximal, $I + J = R$, so choose $i \in I$, and $j \in J$ with $i + j = 1$. Choose an arbitrary element $(a + I, b + J)$ in the codomain, and put $r = bi + aj$. We claim that

$$r + I = a + I \text{ and } r + J = b + J.$$

We observe that

$$r + I = a + I \iff r - a \in I \iff (bi + aj) - a \in I \iff a(j-1) \in I$$

But $i + j = 1$, so $(j-1) = i \in I$. A similar argument shows the element maps onto $b + J$. ■

Example 3.3.10 Familiar examples include

- If m, n are relatively prime integers in $\mathbb{Z}$, then $m\mathbb{Z} + n\mathbb{Z} = \mathbb{Z}$, so
$$\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$
a result we have used for groups in talking about elementary divisors and invariant factors.

- From an exercise above and Proposition 3.3.1, you can conclude
$$\mathbb{Q}[x]/(x^2 - 4) \cong \mathbb{Q}[x]/(x - 2) \times \mathbb{Q}[x]/(x + 2) \cong \mathbb{Q} \times \mathbb{Q}.$$

- We still need to work out a more robust analog of the first example for polynomial rings, but that requires material from the next section.

□

Also recall how to characterize prime and maximal ideals via their induced quotients.

Proposition 3.3.11 *Let R be a commutative ring with identity and let P, M be proper ideals of R. Then*

- *P is a prime ideal iff R/P is an integral domain.*
- *M is a maximal ideal iff R/M is a field. In particular, maximal ideals are prime.*

Example 3.3.12

- (0) is a prime ideal in any integral domain. (0) is a maximal ideal in any field.
- If $n \neq 0$, then $n\mathbb{Z}$ is a prime ideal in $\mathbb{Z}$ iff $n\mathbb{Z}$ is a maximal ideal in $\mathbb{Z}$ iff $n\mathbb{Z} = p\mathbb{Z}$ where p is a prime.
- The principal ideal (x) is a prime ideal in $R[x]$ iff R is an integral domain, while (x) is a maximal ideal in $R[x]$ iff R is a field.

□

We mention another very useful proposition in dealing with quotients of polynomial rings. Here is the background. Given a commutative ring R with identity, fix an element $a \in R$. The natural projection $R \to R/(a)$ extends to one $R[x] \to R/(a)[x]$ sending $f(x) \in R[x]$ to $\overline{f}(x) \in R/(a)[x]$ by viewing coefficients of f in $R/(a)$, often stated as reducing them modulo the ideal.

Proposition 3.3.13 *Let R be a commutative ring with identity, $a \in R$, and $f(x) \in R[x]$. Then*
$$R[x]/(f(x), a) \cong (R/(a))[x]/(\overline{f}(x)).$$

Before proving the proposition, we should do an example to motivate what appears quite technical.

Example 3.3.14 Let $f(x) = x - 3$ and $g(x) = (x-3)(x-5) + 7$ in $\mathbb{Z}[x]$. Show that the ideal $I = (f, g)$ is maximal in $\mathbb{Z}[x]$.

Solution. In Example 3.3.2, we showed that $I = (f, g) = (f, 7)$. Now we use the proposition:

$$\mathbb{Z}[x]/I = \mathbb{Z}[x]/(f(x), 7) \cong (\mathbb{Z}/7\mathbb{Z})[x]/(\overline{f}(x)) = (\mathbb{Z}/7\mathbb{Z})[x]/(x - \overline{3}) \cong \mathbb{Z}/7\mathbb{Z}$$

the last by Proposition 3.3.1. Since $\mathbb{Z}[x]/I$ is a field, Proposition 3.3.11 tells us that I is maximal in $\mathbb{Z}[x]$. □

Proof of Proposition 3.3.13. Consider the composition of natural surjective homomorphisms φ:

$$R[x] \to (R/(a))[x] \to (R/(a))[x]/(\overline{f}(x)).$$

We need only show its kernel is (a, f) to complete the proof by the first isomorphism theorem. By walking the elements a, f through the compositions, it is easy to see that $(a, f) \subseteq \ker\varphi$. For the reverse, take $g(x) = b_0 + \cdots + b_n x^n \in R[x]$. Then $g \in \ker\varphi$ if and only if $\overline{g} \in (\overline{f}(x))$.

Write $f(x) = c_0 + \cdots + c_m x^m \in R[x]$, and suppose that $\overline{g}(x) = \overline{f}(x) \cdot \overline{h}(x)$, where $\overline{h}(x) = \overline{a}_0 + \cdots + \overline{a}_d x^d$. Then the coefficient of x^j in $\overline{g}$ is

$$\overline{b}_j = \sum_{k=0}^{j} \overline{c}_k \overline{a}_{j-k},$$

or consider the arithmetic in $R/(a)$,

$$b_j + aR = \sum_{k=0}^{j} c_k a_{j-k} + aR$$

So if we fix $h(x) = a_0 + \cdots + a_d x^d \in R[x]$ then $b_j - \sum_{k=0}^{j} c_k a_{j-k}$ is the jth coefficient of $f - gh$ is in aR. This implies that

$$g - fh = \sum_{j=0}^{n} [b_j - \sum_{k=0}^{j} c_k a_{j-k}] x^j \in aR[x],$$

or $g \in (a, f) \subseteq R[x]$. ■

It is useful to say a few words about polynomial rings, beginning with some elementary properties.

Proposition 3.3.15 *Let R be an integral domain, and $p, q \in R[x] \setminus \{0\}$. Then*

- $\deg(pq) = \deg(p) + \deg(q)$.
- $R[x]^\times = R^\times$.

- *$R[x]$ is an integral domain.*

Polynomial rings in several variables play a fundamental role in algebraic geometry, so we should say a few words about the different ways to view elements of those rings. For example, we have the natural corollary to the above result:

Corollary 3.3.16 *If R is an integral domain, then so is $R[x_1, \ldots, x_n]$.*

The idea behind the proof of the above corollary involves how to view the polynomial ring and its elements. We examine this in more detail below, but briefly we want to think of the polynomial ring, $R[x_1, \ldots, x_n]$, in n variables with coefficients in R as the polynomial ring in one variable, $S[x_n]$, with coefficients in $S = R[x_1, \ldots, x_{n-1}]$. Given this view, the proof of the corollary is by induction on n with the base case being Proposition 3.3.15.

The comments above apply to polynomial rings in any number of variables, but for simplicity of exposition, we consider the case of two variables.

One way to view elements of $R[x, y]$ is as finite sums of monomials $x^i y^j$ with coefficients $a_{ij} \in R$:

$$p(x, y) = \sum_{i,j} a_{ij} x^i y^j.$$

Alternatively, we can think of $p \in R[x, y]$ as an element of $(R[x])[y]$, so that we could write

$$p(x, y) = \sum_{i,j} a_{ij} x^i y^j = \sum_k \phi_k(x) y^k = \phi_0(x) + \phi_1(x) y + \cdots + \phi_n(x) y^n,$$

where the “coefficients”, $\phi_k(x) \in R[x]$.

Example 3.3.17 Different views of a polynomial in $\mathbb{Z}[x, y]$:

$$\begin{aligned} \mathbb{Z}[x, y] : &\quad 2 + 7y + 4xy + 3x^5 y \\ \mathbb{Z}[x][y] : &\quad 2 + (7 + 4x + 3x^5) y \\ \mathbb{Z}[y][x] : &\quad (2 + 7y) + (4y) x + (3y) x^5 \end{aligned}$$

We see that in the second representation, the polynomial is simply linear in y. □

Changing the perspective on how to view a polynomial affects both your intuition as well as the tools you might bring to bear to understand the polynomial. For example, we have some inkling on how to factor polynomials in one variable, but have almost no intuition when there is more than one variable.

Example 3.3.18 Viewed as polynomials in $\mathbb{Q}[x]$, we immediately know that $x^2 - 4$ factors non-trivially, but that $x^2 + 4$ does not (is irreducible).

Similarly, we see that $x^2 + 4y - y^2 - 1$ factors not by viewing it in $\mathbb{Q}[x, y]$, but thinking of it as analogous to $x^2 - 4$ in $\mathbb{Q}[y][x]$ as

$$x^2 - (y - 1)^2 = (x - (y - 1))(x + (y - 1)).$$

□

3.4 Euclidean domains, PIDs, UFDs and all that jazz

You may recall from Theorem 3.2.7, that a Noetherian integral domain in which every irreducible is prime enjoys unique factorization, but that determining those properties seems a formidable task. So we consider rings with stronger properties which imply those conditions, and also look for ways in which to build new UFDs from old ones.

The definition of a **Euclidean domain** varies a bit from source to source, but the main takeaway is that it is an integral domain that admits a division algorithm. The division algorithm can be leveraged to produce a Euclidean algorithm, and hence the notion of a greatest common divisor. We shall take the definition:

Definition 3.4.1 An integral domain R is a **Euclidean domain** if it is equipped with a function (norm) $d : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ so that given two elements $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ with $a = bq + r$, and either $r = 0$ or $d(r) < d(b)$. ♢

Example 3.4.2 The integers $\mathbb{Z}$ with the absolute value function $d(a) = |a|$ is a Euclidean domain. □

Example 3.4.3 If k is a field, then the polynomial ring $k[x]$ with the degree function as the function d is a Euclidean domain. □

Definition 3.4.4 An integral domain R is a **PID** (principal ideal domain) if every ideal in R is principal. An integral domain R is a **UFD** (unique factorization domain) if it satisfies the conclusion of Theorem 3.2.7. ♢

The following theorem presents the familiar relationship among these notions.

Theorem 3.4.5 *Every Euclidean domain is a PID, and every PID is a UFD.*

The proof of this theorem is standard, but we want to pull out a couple of propositions which are of independent value.

Proposition 3.4.6 *In an integral domain, every prime element is irreducible. In an UFD, every irreducible is prime.*

Proof. We first recall that prime and irreducible elements are nonzero and non-units. Let R be an integral domain, and $\pi \in R$ a prime element. To show π is irreducible, suppose that $\pi = ab$. Then of course $\pi \mid ab$ and since π is prime, it divides a or b, say $a = \pi a_0$. Then

$$\pi = ab = \pi a_0 b,$$

so

$$\pi(a_0 b - 1) = 0.$$

Since R is an integral domain and $\pi \neq 0$, we have $a_0 b = 1$ which implies $b \in R^\times$.

Let R be a UFD, and and $\pi \in R$ a irreducible element. To show that π is prime, suppose that $\pi \mid ab$. If $ab = 0$, then at least one of a or b equals zero, and since $\pi \mid 0$ we are done in this case, so suppose that $ab \neq 0$. Since $\pi \mid ab$, we know that ab is not a unit so ab can be factored into irreducibles:

$$ab = \pi_1 \cdots \pi_r \cdot \pi'_1 \cdots \pi'_s$$

with $\pi_i \mid a$, $\pi'_j \mid b$ and with the possibility that r or s could be zero (though not both). Since $\pi \mid ab$, unique factorization says that π is an associate of some π_i or π'_j implying $\pi \mid a$ or $\pi \mid b$. ■

Proposition 3.4.7 *In a PID, irreducibles generate maximal ideals.*
Proof. Let π be an irreducible in the PID R, and suppose that

$$(\pi) \subseteq I \subseteq R.$$

Since R is a PID, $I = (r)$ for some $r \in R$, and the inclusion $(\pi) \subseteq (r)$ implies that $\pi = rs$ for some $s \in R$. Since π is irreducible, either r or s is a unit. If r is the unit, then $I = R$. If s is the unit, then r and π are associates, so $I = (r) = (\pi)$. Thus (π) is maximal ideal. ■

Proposition 3.4.8 *In any integral domain, the principal ideal (π) is a prime ideal if and only if π is a prime element.*
Proof. Suppose that π is a prime element, so in particular, π is nonzero and a non-unit. Thus $P = (\pi)$ is a proper ideal. Suppose that $ab \in P = (\pi)$. Then $ab = \pi c$ for some $c \in R$. By Proposition 3.4.6, π is irreducible, so occurs in the factorization of a or b. But that means that $a \in P$ or $b \in P$.

Conversely, suppose that $P = (\pi)$ is a prime ideal. To show that π is a prime element, suppose that $\pi \mid ab$. Then $ab \in P = (\pi)$. Since P is a prime ideal, either a or b is in $P = (\pi)$, meaning $\pi \mid a$ or $\pi \mid b$. ■
Proof of Theorem 3.4.5. Suppose that R is a Euclidean domain, and let I be an ideal of R. Without loss, be may assume that $I \neq (0)$. Using the norm, d, with which the Euclidean domain is equipped, choose a nonzero element $b \in I$ so that $d(b)$ is minimal. We claim that $I = (b)$. Of course, $(b) \subseteq I$, so choose a nonzero element $a \in I$. Then we may write $a = bq + r$ with either $r = 0$ or $d(r) < d(b)$. As b was chosen to have $d(b)$ minimal, we must have $r = 0$ which implies that $a \in (b)$. Thus $I = (b)$, and R is a PID.

Suppose that R is a PID. By Theorem 3.2.4, R is a Noetherian integral domain. Let π be an irreducible element in R. By Proposition 3.4.7, the ideal (π) is a maximal ideal, hence by Proposition 3.3.11 a prime ideal, and finally Proposition 3.4.8 says that π is a prime element. That R is a UFD now follows from Theorem 3.2.7. ■

It is certainly easiest to recognize a ring as a PID by noting that it is Euclidean, and that it has unique factorization since it is a PID. But there are counterexamples which show that these classes of rings are distinct.

The harder of the two is to produce PIDs which are not Euclidean. The following are examples from algebraic number theory:

$$R = \mathbb{Z}[\frac{1+\sqrt{-d}}{2}] \text{ where } d = 19, 43, 67, 163$$

and the proofs are all similar in nature (and in standard references), but philosophically, the challenge in showing that something is not Euclidean is showing that there is no choice of norm d which satisfies the required division algorithm property, and not simply that some natural function fails.

Adding to the challenge of finding counter examples is that no polynomial ring $A[x]$ will work for we have the following proposition.

Proposition 3.4.9 *Let A be a commutative ring with identity, and assume that the polynomial ring $A[x]$ is a PID. Then A is a field, so that $A[x]$ is a Euclidean domain.*

Proof. Since $A[x]$ is an integral domain, so is A. Since by Proposition 3.3.1, $A[x]/(x) \cong A$, we infer by Proposition 3.3.11, that the ideal (x) is a prime ideal so that by Proposition 3.4.8, the element x is a prime, hence irreducible (Proposition 3.4.6) element of the ring. However, by Proposition 3.4.7, irreducibles in a PID generate maximal ideals, which once again by Proposition 3.3.11 shows that $A[x]/(x) \cong A$ is a field giving us that $A[x]$ is a Euclidean domain. ■

The following theorem is of great utility in producing new UFDs from old ones.

Theorem 3.4.10 *Let R be a UFD. Then the polynomial ring $R[x]$ is a UFD*

(Main idea). The proof of this result is not intrinsically difficult, but it does take some time and care to develop. The key is that associated to any integral domain (e.g., $\mathbb{Z}$) is its field of fractions (e.g., $\mathbb{Q}$). So associated to the UFD A is a field K which contains an isomorphic copy of A. Since $K[x]$ is a Euclidean domain, it is also a UFD and it is natural to compare factorizations of a polynomial with coefficients in A in the rings $A[x]$ versus $K[x]$.

The crucial result is **Gauss's lemma** which discusses how irreducibility changes as one views factorizations in $A[x]$ versus $K[x]$. For example, the polynomial $p(x) = 2x$ is irreducible in $\mathbb{Q}[x]$, but reducible in $\mathbb{Z}[x]$. Fortunately this example is as bad as things get and it is straightforward to address. ■

Remark 3.4.11 We now verify that it follows from Theorem 3.4.10 that both $\mathbb{Z}[x]$ and $\mathbb{Q}[x, y]$ are UFDs.

Since $\mathbb{Z}$ is a Euclidean domain, it is automatically a PID and UFD by Theorem 3.4.5, so that $\mathbb{Z}[x]$ is a UFD is immediate from Theorem 3.4.10.

Similarly, we know that $\mathbb{Q}[x]$ is a Euclidean domain, hence a UFD, so that $\mathbb{Q}[x, y] = \mathbb{Q}[x][y]$ is a UFD.

It then follows from the exercise below that $\mathbb{Z}[x]$ and $\mathbb{Q}[x, y]$ are not PIDs, providing our examples of UFDs which are not PIDs.

Exercise 3.4.1

(a) Show that the ideal $(2, x)$ is not principal in $\mathbb{Z}[x]$.

(b) Show that the ideal (x, y) is not principal in $\mathbb{Q}[x, y]$.

Remark 3.4.12 Something to ponder. Theorem 3.4.10 tells us that since $\mathbb{Z}$ and $\mathbb{Q}$ are UFDs, so are $\mathbb{Z}[x], \mathbb{Q}[x], \mathbb{Z}[x, y], \mathbb{Q}[x, y]$, as well as many others.

Suppose that $p(x) \in \mathbb{Z}[x]$. The question to consider is how does irreducibility of p change as we view in in these larger domains. For example,

- If p is irreducible in $\mathbb{Z}[x]$, is it irreducible in $\mathbb{Q}[x]$?
- If p is irreducible in $\mathbb{Q}[x]$, is it irreducible in $\mathbb{Z}[x]$?
- If p is irreducible in $\mathbb{Z}[x]$, is it irreducible in $\mathbb{Z}[x, y]$?

Consider why the last question is important. We know that $\mathbb{Z}[x]$ and $\mathbb{Z}[x, y]$ are both UFDs and we can view $\mathbb{Z}[x] \subset \mathbb{Z}[x, y]$. Are irreducibles in $\mathbb{Z}[x]$ still irreducible in $\mathbb{Z}[x, y]$, or are we forced to start from scratch in finding irreducibles in this larger domain?

3.5 Identifying irreducibles

In previous sections we have discussed the notion of unique factorization of elements into a product of irreducibles, but in no context other that the integers, $\mathbb{Z}$, do we have much experience identifying irreducibles. So we now take some time to explore tests for (ir)reducibility in polynomial rings $R[x]$, where R is a UFD.

To get very far, we really need to leverage the notion of a **greatest common divisor.** In $\mathbb{Z}$, this is a familiar notion based upon factorization and resulting in a unique positive integer when finding the gcd of nonzero integers. For example, we know that $\gcd(12, -30) = +6$, where somehow that sign of the integers does not matter. Of course we recognize this as a matter of units which can be far more extensive in general rings, so we take an approach that defines a gcd in a more general setting.

Definition 3.5.1 Let R be an integral domain, and $a, b \in R$, not both zero. A **greatest common divisor** of a, b is an element $d \in R$ satisfying

- $d \mid a$ and $d \mid b$ (i.e., d is a common divisor)
- If $d' \mid a$ and $d' \mid b$, then $d' \mid d$, meaning any other common divisor divides d, making d the greatest in terms of divisibility.

$\Diamond$

Remark 3.5.2 It is an easy exercise that in an integral domain R, any two gcds d, d' of elements a, b differ by a unit, that is $d = d' \cdot u$ for some $u \in R^\times$. Since $\mathbb{Z}$ has only two units, ± 1, this explains why in $\mathbb{Z}$, it is safe simply to choose the positive gcd.

Remark 3.5.3 Also in the integers, unless someone asked you to compute the gcd of large numbers, your most likely method of finding a gcd was to factor both numbers, and determine the largest power of each prime dividing both numbers.

But as it turns out, factoring in $\mathbb{Z}$ (and in other realms) is very hard, so hard that the security of certain cryptographic systems depends on that difficulty, so while in theory, it is a great way to write down a gcd, for practical purposes, it is not very good at all. We review how gcds manifest themselves in Euclidean domains, PIDs, and UFDs.

Exercise 3.5.1 Greatest common divisors are guaranteed to exist in any UFD, but how they manifest themselves differs depending on whether the UFD has any more robust properties.

(a) Let R be a Euclidean domain with norm $d : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$. The division algorithm on R gives rise to a Euclidean algorithm. Show that just as in the case of $\mathbb{Z}$, the last nonzero remainder from the Euclidean algorithm is a gcd of the given elements. Also note that backtracking through the Euclidean algorithm allows you to write $\gcd(a, b) = ax + by$ for some $x, y \in R$.

Hint. Just as in $\mathbb{Z}$, show that if $a = bq + r$, then

$$\gcd(a, b) = \gcd(b, r)$$

where by $\gcd(a, b)$ we mean any greatest common divisor.

(b) Let R be a PID, and $a, b \in R$, not both zero. Since the ideal (a, b) is a principal ideal, say

$$(d) = (a, b),$$

show that d is a gcd of a, b.

Hint. Observe that if the ideal $(a, b) = (d)$, then d is a common divisor of a, b. Moreover, recall that

$$(a, b) = \{ax + by \mid x, y \in R\},$$

so if $(d) = (a, b)$, then $d = ax_0 + by_0$ for some $x_0, y_0 \in R$. Use that to show that d satisfies the other requirement of being a gcd.

(c) Let R be a UFD, and $a, b \in R$, not both zero. Show how to use unique factorization to extract a gcd. In particular, after dispensing with the case that one of a or b may be zero, both a and b can be factored with the same set of irreducibles (up to units) if we allow exponents to be zero when an irreducible divides only one of a or b.

We focus now on polynomial rings, characterizing irreducibles in them and applications. One difficulty that arises relates to the context in which we view a polynomial. We posed a number of interesting questions in Remark 3.4.12 surrounding the universe in which we view things. We now take some time to explore them.

In thinking about unique factorization and irreducibles, units invariably insert themselves into the picture. We stated in Proposition 3.3.15 that for an integral domain R, the unit group of the polynomial ring is the same as the unit group of the coefficient ring ($R[x]^\times = R^\times$). So when R is a field, every nonzero constant is a unit, so irreducible polynomials with coefficients in a field must have degree at least one. For rings such as $\mathbb{Z}$, the unit group is much smaller and polynomials of degree zero can also be irreducible.

Because things are a bit simpler, we begin with polynomials with coefficients in a field F. Again by Proposition 3.3.15, any polynomial of degree one in $F[x]$ must be irreducible. As we increase the degree, and factor a polynomial, irreducible factors of degree one correspond to roots of the given polynomial. Recall the theorem:

Theorem 3.5.4 Roots and linear factors. *Let F be a field and $p \in F[x]$ a non-constant polynomial. Then $(x-a) \mid p$ if and only if $p(a) = 0$.*

(sketch). Apply the division algorithm in $F[x]$, dividing $(x-a)$ into $p(x)$, and use the evaluation homomorphism at a on the resulting equation. ■

It follows by a degree argument that for polynomials over a field, a polynomial of degree 2 or 3 is irreducible if and only if it has no roots. Since finding roots is of some interest, we remind the reader of the rational root test.

Theorem 3.5.5 Rational Root Test. *Let R be a UFD and let K be its field of fractions (e.g., $R = \mathbb{Z}$ and $K = \mathbb{Q}$). Let $p(x) = a_nx^n + \cdots + a_1x + a_0 \in R[x]$. Let $r, s \in R$ with $\gcd(r, s) = 1$. The element $\alpha = r/s \in K$ is a (rational) root of p only if $r \mid a_0$ and $s \mid a_n$. In particular, if p is monic, then its only possible rational roots belong to R.*

Proof. Assume that α is a root of p, and evaluate p at α and clear the resulting denominators to produce the equation:

$$a_nr^n + a_{n-1}r^{n-1}s + \cdots + a_1rs^{n-1} + a_0s^n = 0.$$

If we move one of the end terms to the other side of the equation, we acquire two equations

$$\begin{aligned} a_nr^n &= -\left(a_{n-1}r^{n-1}s + \cdots + a_1rs^{n-1} + a_0s^n\right) \\ a_0s^n &= -\left(a_nr^n + a_{n-1}r^{n-1}s + \cdots + a_1rs^{n-1}\right) \end{aligned}$$

By inspection of those equations,

$$s \mid a_nr^n \text{ and } r \mid a_0s^n.$$

However as $\gcd(r, s) = 1$, we must have $r \mid a_0$ and $s \mid a_n$. ■

We are often in the situation of the previous theorem, having a polynomial with coefficients in a UFD R and wanting to determine whether it is irreducible. Viewing the same polynomial in $K[x]$, where K is the field of fractions, affords us more tools since $K[x]$ is a Euclidean domain. Thus it is important to understand the relationship of irreducibles in $R[x]$ versus $K[x]$.

This need is amplified as it provides an essential tool to prove Theorem 3.4.10. This relationship is completely described in **Gauss's lemma.** To state the theorem, we need a definition.

Definition 3.5.6 Let R be a UFD, and $p(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. We say that p is a **primitive** polynomial if $\gcd(a_0, \ldots, a_n) = 1$, that is, there is no common divisor of all the coefficients except for units. It is immediate that for any $p \in R[x]$ we can write $p = C(p)p_0$ where p_0 is primitive and $C(p) \in R$; the constant $C(p)$ is usually referred to as the **content** of p. $\Diamond$

Theorem 3.5.7 Gauss's lemma. *Let R be a UFD, and K its field of fractions. Let $p(x) = a_0 + \cdots + a_nx^n$ be a primitive polynomial in $R[x]$. Suppose that p factors in $K[x]$ as the product of non-constant polynomials $p(x) = F(X)G(X)$. Then there exists $\alpha, \beta \in K^\times$ so that*

$$f = \alpha F \in R[x], g = \beta G \in R[x], \text{ and } p = fg.$$

Remark 3.5.8 In some texts, Gauss's lemma is phrased as the product of two primitive polynomials is primitive which implies that for two polynomials $p, q \in R[x]$, $C(pq) = C(p)C(q)$, the content is a multiplicative function.

This version of Gauss's lemma is necessary to prove the one above.

In the above notation, the following corollary clarifies the relationship between irreducibility of a polynomial in $R[x]$ versus $K[x]$.

Corollary 3.5.9 *Let R be a UFD, K its field of fractions, and $p \in R[x]$. Then p is irreducible in $R[x]$ if and only if p is primitive (in $R[x]$), and irreducible in $K[x]$.*

Proof. If p is irreducible in $R[x]$, then it is necessarily primitive. If it were reducible in $K[x]$, then $p = FG$ for non-constant polynomials in $K[x]$. By Gauss's lemma, there exists $\alpha, \beta \in K^\times$ so that

$$f = \alpha F \in R[x], g = \beta G \in R[x], \text{ and } p = fg.$$

Since $\deg f, \deg g \geq 1$, this would imply p is reducible in $R[x]$, a contradiction.

Conversely, suppose that p is primitive and irreducible in $K[x]$. If it were reducible in $R[x]$, then $p = fg$ for $f, g \in R[x]$ with $\deg f, \deg g \geq 1$ since p is primitive. Then neither f nor g are units in $K[x]$, so the factorization shows p is reducible in $K[x]$, a contradiction. ■

So now we need some theorems which help us actually determine whether a polynomial is irreducible. Remember we have the rational root test which is useful for pulling off linear factors via Theorem 3.5.4, and can help resolve irreducibility of polynomials defined over a field of degree at most 3. However, even in the case of small degree, it may not be the tool of choice.

Example 3.5.10 Consider the polynomial $p(x) = x^3 + 82x^2 + 3456782 \in \mathbb{Z}[x]$. The rational root test would have us check $p(r)$ for all $r \mid 3456782$. As it turns out, there are not that many divisors, but you would have to find them, and that looks painful. □

The following is a very general criterion, but some luck is involved in its use.

Theorem 3.5.11 Reduction criterion. *Let R be an integral domain, $p \in R[x]$ a monic, non-constant polynomial. Let I be any proper ideal in R, and consider the reduction map $R[x] \to (R/I)[x]$, denoted $p \mapsto \overline{p}$. If $\overline{p}$ cannot be factored into the product of two non-constant polynomials, then p is irreducible in $R[x]$.*

Proof. The proof is by contradiction. Suppose that $\overline{p}$ cannot be factored into the product of two non-constant polynomials, but p is reducible in $R[x]$, say $p = fg$. Since p is monic, the leading coefficients of f, g must be units, which forces $\deg f, \deg g \geq 1$ to provide a witness to reducibility. Set

$$f(x) = a_m x^m + \cdots + a_0 \text{ and } g(x) = b_n x^n + \cdots + b_0.$$

As we noted, $a_m, b_n \in R^\times$.

Consider the reduction modulo I :

$$p = fg \mapsto \overline{p} = \overline{fg} = \overline{f} \cdot \overline{g} \in (R/I)[x].$$

We note that

$$\overline{f} = (a_m + I)x^m + \cdots + (a_0 + I) \text{ and } \overline{g} = (b_n + I)x^n + \cdots + (b_0 + I).$$

Since $a_m, b_n \in R^\times$ and I is a proper ideal, we see $a_m + I \neq 0 + I$ and $b_n + I \neq 0 + I$, so $\deg \overline{f}, \deg \overline{g} \geq 1$ which gives a factorization of $\overline{p}$ into two non-constant polynomials, a contradiction. ■

Example 3.5.12 Let's try the reduction criterion on the polynomial $p(x) = x^3 + 82x^2 + 3456782 \in \mathbb{Z}[x]$ on which we chose not to use the root test.

As a first guess, let $I = 2\mathbb{Z}$. But then $\overline{p} = x^3 = x \cdot x^2$ which does not fit the hypotheses of the criterion, so we try again.

Let $I = 3\mathbb{Z}$. Then $\overline{p} = x^3 + x + 2 \in \mathbb{Z}/3\mathbb{Z}[x]$. It is easy to check that $\overline{p}$ has no roots in $\mathbb{Z}/3\mathbb{Z}$ and since its degree is 3 and $\mathbb{Z}/3\mathbb{Z}$ is a field, we conclude $\overline{p}$ is irreducible over $\mathbb{Z}/3\mathbb{Z}$, and so of course cannot be written as the product of two non-constant polynomials. By the reduction criterion, $p(x) = x^3+82x^2+3456782$ is irreducible in $\mathbb{Z}[x]$. □

The reduction criterion is surprisingly powerful. Consider the following example of a polynomial in two variables.

Example 3.5.13 Let $p(x, y) = -1 + x - y + x^2 y + y^3 \in \mathbb{Q}[x, y]$. We can view p as

$$p = y^3 + (x^2 - 1)y + (x - 1) \in \mathbb{Q}[x][y].$$

Let $R = \mathbb{Q}[x]$. So we see that p is monic in $R[y]$. Let $I = (x) \subsetneq R$. Now $R/I = \mathbb{Q}[x]/(x) \cong \mathbb{Q}$ via the evaluation map $x \mapsto 0$ (see Proposition 3.3.1), so $\overline{p} \in (R/I)[y] \cong \mathbb{Q}[y]$ is given by setting $x = 0$, so

$$\overline{p} = y^3 - y - 1 \in \mathbb{Q}[y].$$

Since the degree is 3, $\overline{p}$ will be irreducible if it has no roots, and the root test tells us the only possible roots are ±1, neither of which are roots. The reduction criterion now says that p is irreducible in $\mathbb{Q}[x, y]$.

We soon give a different proof that this polynomial is irreducible using Eisenstein's criterion. □

Theorem 3.5.14 Eisenstein's criterion. *Let R be a UFD and $p = a_nx^n + \cdots + a_0 \in R[x]$ a primitive polynomial. Let $\pi \in R$ be an irreducible (hence prime (Proposition 3.4.6)), with*

$$\pi \mid a_i, i = 0 \ldots, n-1, \quad \pi \nmid a_n, \text{ and } \pi^2 \nmid a_0.$$

Then p is irreducible in $R[x]$.

Proof. We proceed by contradiction and assume that p is reducible in $R[x]$. Since p is primitive, this means

$$p = fg \text{ where } f, g \in R[x], \text{ with } \deg f, \deg g \geq 1,$$

say

$$f = b_0 + \cdots + b_rx^r \text{ and } g = c_0 + \cdots + c_sx^s.$$

Recalling that π is a prime element, then since $\pi \mid a_0 = b_0c_0$, it must divide one of b_0 or c_0, but $\pi^2 \nmid a_0$ says it can divide precisely one, say $\pi \mid b_0$ and $\pi \nmid c_0$. Also since $\pi \nmid a_n = b_rc_s$, we see $\pi \nmid b_r$ and $\pi \nmid c_s$.

Since $\pi \mid c_0$, but $\pi \nmid c_s$, let ℓ be the smallest index so that $\pi \nmid c_\ell$. So

$$0 < \ell \leq s < n \text{ and } a_\ell = b_\ell c_0 + b_{\ell-1}c_1 + \cdots + b_0c_\ell.$$

Since $\ell < n$, we know that $\pi \mid a_\ell$, and by the choice of ℓ, $\pi \mid c_0, \ldots, c_{\ell-1}$, which implies $\pi \mid b_0c_\ell$. But π prime then says $\pi \mid b_0$ or $\pi \mid c_\ell$, a contradiction. ■

Example 3.5.15 There exist irreducible polynomials of all degrees $n \geq 1$ in $\mathbb{Z}[x]$.

Solution. Let p be a prime in $\mathbb{Z}$. Then by Eisenstein, $x^n \pm p$ is irreducible in $\mathbb{Z}[x]$. □

Example 3.5.16 Show that $x^{907} + 27x^3 + 15x^2 - 81x + 6$ is irreducible in $\mathbb{Z}[x]$.

Solution. It is irreducible by Eisenstein with $p = 3$. □

Example 3.5.17 Show that $p = -1 + x - y + x^2y + y^3$ is irreducible in $\mathbb{Q}[x, y]$.

Solution. We looked at this polynomial when considering the reduction criterion. So here $R = \mathbb{Q}[x]$ is our UFD and we write p as an element of $R[y]$ as

$$p = y^3 + (x^2 - 1)y + (x - 1).$$

Moreover, $\pi = (x - 1) \in R = \mathbb{Q}[x]$ is an irreducible. We check that Eisenstein's criterion applies, and we are done. □

Given an integral domain R, it is often possible to more easily check the irreducibility of a polynomial in $R[x]$ by first applying an isomorphism to the ring. That is, if $\varphi : R[x] \to R[x]$ is a ring isomorphism, we know that $p \in R[x]$ is irreducible if and only if $\varphi(p)$ is irreducible in $R[x] = \varphi(R[x])$. A simple isomorphism is given by $f(x) \mapsto f(x \pm a)$ for any $a \in R$, so $f(x)$ is irreducible iff $f(x \pm a)$ is irreducible.

Example 3.5.18 Let p be a prime in $\mathbb{Z}$. Then

$$f(x) = 1 + x + x^2 + \cdots + x^{p-1} = \frac{x^p - 1}{x - 1}$$

is irreducible in $\mathbb{Z}[x]$.

Solution. Let $g(x) = f(x+1)$. Then f is irreducible iff g is. We see that

$$\begin{aligned} g(x) =& f(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} \\ =& \frac{x^p + \binom{p}{1}x^{p-1} + \cdots + \binom{p}{p-1}x + 1 - 1}{x} \\ =& x^{p-1} + \binom{p}{1}x^{p-2} + \cdots + \binom{p}{p-1} \end{aligned}$$

which is irreducible in $\mathbb{Z}[x]$ using Eisenstein with the prime p since $p \mid \binom{p}{k}$ for $1 \leq k \leq p-1$ and $\binom{p}{p-1} = p$. □

3.6 Applications

If we consider the polynomial $p(x) = x^7 + 12x^3 + 2x + 6 \in \mathbb{Z}[x]$, we know it is irreducible by Eisenstein's criterion ($p = 2$), so in particular, p has no roots in $\mathbb{Q}$. Where are its roots? Is there a smallest field containing all the roots of p?

Perhaps this seems an uninteresting question because somewhere in your distant past, you learned that all the roots of p must be contained in $\mathbb{C}$ since $\mathbb{C}$ is **algebraically closed.** Well, for starters, the complex numbers is an enormous field (certainly uncountable), and are you really all that comfortable with how to construct it?

In fact all the roots of all the polynomials in $\mathbb{Q}[x]$ are contained in a subfield which is countable, and the smallest field containing the roots of any given polynomial is only a finite-dimensional vector space over the rationals.

Now add to that the following question. Consider the polynomial $x^2 - 3 \in (\mathbb{Z}/5\mathbb{Z})[x]$. It is irreducible since it has degree two and no roots in the field $\mathbb{Z}/5\mathbb{Z}$. Is there a field containing its roots? Certainly the real and complex numbers are of no help here since this field has characteristic 5. So the ability to construct such a field is really significant.

If K is a field and $p \in K[x]$ is irreducible, we know that (p) is a maximal ideal, and hence $K[x]/(p)$ is a field. We shall show that this is a field containing an isomorphic copy of K and a root of the polynomial p.

Let's consider a ***motivating example***. We want to construct a field K containing $\mathbb{Q}$ which also contains a root of the polynomial $p(x) = x^7 - 10$.

Eisenstein's criterion (with $p = 2, 5$) assures us this polynomial is irreducible in $\mathbb{Q}[x]$, so in particular has no roots in $\mathbb{Q}$. We claim that $\mathbb{Q}[x]/(x^7 - 10)$ is such a field K.

Now of course since this is motivation, we shall acknowledge the existence of the real and complex numbers, and let $\sqrt[7]{10}$ denote any such root in $\mathbb{C}$. The polynomial has one real and six complex roots. Consider the evaluation map

$$\varphi : \mathbb{Q}[x] \to \mathbb{Q}[\sqrt[7]{10}]$$

given by $\varphi(f(x)) = f(\sqrt[7]{10})$, and the figure below.

$$\begin{tikzcd}
\mathbb{Q}[x] \arrow[r, "\varphi"] \arrow[d, "\pi"] & \mathbb{Q}[\sqrt[7]{10}] \subset \mathbb{C} \\
\mathbb{Q}[x]/(x^7 - 10) \arrow[ur, dashed, "\varphi_*"'] &
\end{tikzcd}$$

Figure 3.6.1 A field containing a root of $x^7 - 10$

The map φ is a surjective homomorphism with image $\mathbb{Q}[\sqrt[7]{10}]$ which is certainly an integral domain as a subring of the field $\mathbb{C}$. Moreover, it is clear that the ideal $(x^7 - 10) \subseteq \ker\varphi$, but also we know $(x^7 - 10)$ is a maximal ideal since $x^7 - 10$ is irreducible in the PID $\mathbb{Q}[x]$ (see Proposition 3.4.7), and since φ is not the zero homomorphism, it follows that $\ker\varphi = (x^7 - 10)$. Thus we conclude

$$\mathbb{Q}[x]/(x^7 - 10) \cong \mathbb{Q}[\sqrt[7]{10}].$$

This means that

- $\mathbb{Q}[\sqrt[7]{10}]$ is a field
- Under the isomorphism φ_*, the element $\alpha = x + (x^7 - 10) \in \mathbb{Q}[x]/(x^7 - 10)$ corresponds to $\sqrt[7]{10} \in \mathbb{Q}[\sqrt[7]{10}]$. But what does that really mean?

First, let's understand how $\mathbb{Q}$ is a subfield of $\mathbb{Q}[x]/(x^7 - 10)$. Consider the map

$$\mathbb{Q} \to \mathbb{Q}[x] \to \mathbb{Q}[x]/(x^7 - 10)$$

which takes $a \in \mathbb{Q}$ to the coset $a + (x^7 - 10)$. In particular

$$10 \mapsto \text{ the coset } 10 + (x^7 - 10).$$

So what is $\alpha^7 - 10 \in \mathbb{Q}[x]/(x^7 - 10)$, meaning what is $\alpha^7 - [10 + (x^7 - 10)]$? For compact notation, let $I = (x^7 - 10)$.

$$\alpha^7 - [10 + I] = [x + I]^7 - [10 + I] = [x^7 + I] - [10 + I] = [x^7 - 10] + I = 0 + I$$

Which says the element $\alpha = x + I \in K = \mathbb{Q}[x]/I$ is the seventh root of 10.

Remark 3.6.2 The example above is entirely representative of the general situation. If F is any field, and $p \in F[x]$ is irreducible, then $K := F[x]/(p)$ is a field containing an isomorphic copy of F and a root $\alpha = x + (p(x))$ of the polynomial p, and to ease the notation, we generally denote the field K not as the quotient, but as $K = F(\alpha)$ which denotes the smallest field containing (an isomorphic copy of) F and the root α.

When first seeing this, one is probably more in their comfort zone seeing things like $\mathbb{Q}(\sqrt{2})$ or $\mathbb{Q}(\sqrt[7]{10})$, but when you think harder, if the polynomial is something like $p(x) = x^7 + 12x^3 + 2x + 6$ which is irreducible in $\mathbb{Q}[x]$, what better notation would there be than simply to write $\mathbb{Q}(\alpha)$, where α is some root of p. The ***point of the construction*** is that we know that we can construct such a field and an appropriate root.

Remark 3.6.3 One might have noticed that in the example of the construction, the notation $\mathbb{Q}[\sqrt[7]{10}]$ (square brackets) was used instead of the later notation such as $\mathbb{Q}(\sqrt{2})$ (parentheses).

There is a distinction between the notations $F[\alpha]$ (the smallest ring containing F and α) in contrast to $F(\alpha)$ (the smallest field containing F and α) which involves a discussion of whether α is **algebraic** or **transcendental** over F, but in the cases above they are equal.

There is however a useful observation to be made. First we observed that $F[\alpha] \cong F[x]/(p)$ was a field since (p) is maximal, so that $F[\alpha]$ actually equals $F(\alpha)$, the smallest field containing F and α.

The beauty of knowing that $F[\alpha]$ is a field is that every element of the field can be written as $a_0 + a_1\alpha + \cdots + a_n\alpha^n$ for some n and $a_i \in F$. In particular, an element like $1/\alpha$ has such a representation!

We conclude this section on applications by giving a characterization of $F[\alpha]$ when α is a root of some irreducible polynomial p with coefficients in F.

Theorem 3.6.4 *Let F be a field and $p \in F[x]$ an irreducible polynomial of degree d. Then $F[x]/(p)$ is a field which as a vector space over F has dimension d with basis $\{\overline{1}, \overline{x}, \ldots, \overline{x}^{d-1}\}$. Here, we use the notation $\overline{x}^j = x^j + (p)$.*

That means that when we write $F[\alpha] \cong F[x]/(p)$, every element of the field $F[\alpha]$ has a representation as $a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1}$ for uniquely determined $a_i \in F$.

Proof. By the division algorithm, every element $f(x) \in F[x]$ can be written as

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

where either $r = 0$ or $\deg r < \deg p = d$. This means that $f(x) + (p) = r(x) + (p)$. Thus it is clear that $\{\overline{1}, \overline{x}, \ldots, \overline{x}^{d-1}\}$ spans $F[x]/(p)$ as a vector space over F.

To show independence, suppose that

$$a_0 + a_1\overline{x} + \cdots + a_{d-1}\overline{x}^{d-1} = \overline{0} \text{ in } F[x]/(p).$$

This would imply that

$$a_0 + a_1 x + \cdots + a_{d-1}x^{d-1} \in (p).$$

But as $\deg p = d$, the only way that could happen is if all the coefficients $a_i = 0$. This proves independence. ■

Chapter 4

Definitions

Here we accumulate basic definitions and examples from a standard first course in abstract algebra.

4.1 Basic Definitions

Listed in alphabetical order.

Definition 4.1.1 Two elements a, b in a ring (with identity) R are called **associates** if $a = ubv$ for some units u, v in R. In a commutative ring, we can simply write $a = bv$. $\Diamond$

Definition 4.1.2 Let R be a commutative ring with identity. Then two ideals I, J of R are said to be **comaximal** iff $I + J = R$. $\Diamond$

Definition 4.1.3 Let X be a non-empty set. A **relation** on X is a subset $R \subseteq X \times X$, that is a collection of ordered pairs. Often instead of saying $(x, y) \in R$, write $x \sim y$ and say x is related to y.

An **equivalence relation** on X is a relation which satisfies three properties:

- $x \sim x$ (i.e., $(x, x) \in R$) for all $x \in X$. This is called the **reflexive** property of the relation.

- If $x \sim y$, then $y \sim x$, that is, whenever the ordered pair $(x, y) \in R$, then also $(y, x) \in R$. This is called the **symmetric** property of the relation.

- If $x \sim y$ and $y \sim z$, then $x \sim z$, that is, if $(x, y), (y, z) \in R$, then so it (x, z). This is called the **transitive** property of the relation.

$\Diamond$

Definition 4.1.4 An integral domain R is a **Euclidean domain** if it is equipped with a function (norm) $d : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ so that given two elements $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ with $a = bq + r$, and either $r = 0$ or $d(r) < d(b)$. $\Diamond$

Definition 4.1.5 Let G be a group, and $x \in G$. Any positive integer n for which $x^n = e$ is called an **exponent** for the element; the smallest exponent is called its order. The order may be infinite.

If there is a positive integer n so that $x^n = e$ for every $x \in G$, then n is called an exponent for the group. $\diamondsuit$

Definition 4.1.6 A **group** is a nonempty set G with a binary operation $*$ satisfying the properties:

- G is **closed** under the operation, that is for $a, b \in G$ we have $a * b \in G$.
- $*$ is an **associative** operation.
- There exists an **identity** element, that is there exists an element $e \in G$, so that $e * g = g * e = g$ for all $g \in G$.
- Every element has an **inverse**, that is, for every $g \in G$, there exists an $h \in G$ with $g * h = h * g = e$. One shows that the inverse of an element is unique, so we denote the inverse of g by g^{-1}.

$\diamondsuit$

Definition 4.1.7 Let R be an integral domain, and $a, b \in R$, not both zero. A **greatest common divisor** of a, b is an element $d \in R$ satisfying

- $d \mid a$ and $d \mid b$ (i.e., d is a common divisor)
- If $d' \mid a$ and $d' \mid b$, then $d' \mid d$, meaning any other common divisor divides d, making d the greatest in terms of divisibility.

$\diamondsuit$

Definition 4.1.8 Let R be a ring. A subset $I \subseteq R$ is called an **ideal** if I is an additive subgroup of R, with the property that $R \cdot I \subseteq I$ and $I \cdot R \subset I$. This is also called a **two-sided ideal.** There are also left and right ideals in which only one condition holds. Of course in a commutative ring, all ideals are two-sided.

Recall that the property of being a (two-sided) ideal is precisely the condition required to make the quotient R/I a ring with well-defined operations on the cosets. $\diamondsuit$

Definition 4.1.9 Let R be a ring. An element $s \in R$ is an **idempotent** if $s^2 = s$. $\diamondsuit$

Definition 4.1.10 A function $f : X \to Y$ between sets X and Y is **injective** if for every $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$. $\diamondsuit$

Definition 4.1.11 Let F denote the field of real or complex numbers. For $z = a + bi \in \mathbb{C}$ $(a, b \in \mathbb{R}$ and $i^2 = -1)$, we have the notion of the **complex conjugate** of z, denoted $\overline{z} = a - bi$. Note that when $z \in \mathbb{R}$, that is $z = a = a + 0i \in \mathbb{C}$, we have $z = \overline{z}$. The **magnitude** (**norm**, *absolutevalue*) of $z = a + bi$ is $|z| = \sqrt{a^2 + b^2}$.

Let V be a vector space over the field F. An **inner product** is a function:

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

so that for all $u, v, w \in V$ and $\lambda \in F$:

1. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
2. $\langle \lambda v, w\rangle = \lambda\langle v, w\rangle$
3. $\overline{\langle v, w\rangle} = \langle w, v\rangle$, where the bar denotes complex conjugate.
4. $\langle v, v\rangle$ is a positive real number for all $v \neq 0$.

$\Diamond$

Definition 4.1.12 An **inner product space** is a vector space V defined over a field $F = \mathbb{R}$ or $\mathbb{C}$ to which is associated an inner product. If $F = \mathbb{R}$, V is called a **real inner product space**, and if $F = \mathbb{C}$, then V is called a **complex inner product space.** $\Diamond$

Definition 4.1.13 Let R be a commutative ring with identity. An element $\pi \in R$ is said to be **irreducible** if $\pi \neq 0$, $\pi \notin R^\times$, and whenever we write $\pi = ab$, either a or b is a unit in R. $\Diamond$

Definition 4.1.14 The **Kronecker delta** is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

$\Diamond$

Definition 4.1.15 Let R be a commutative ring with identity, and M an ideal of R. Then M is a **maximal** ideal iff

- M is a proper ideal
- Whenever I is an ideal of R with $M \subseteq I \subseteq R$, then either $I = M$ or R.

$\Diamond$

Definition 4.1.16 Let G be a group and H a subgroup of G. The following conditions are equivalent and define what it means for H **normal** subgroup of G.

- $gHg^{-1} = H$ for all $g \in G$
- $gH = Hg$ for all $g \in G$
- $gHg^{-1} \subseteq H$ for all $g \in G$

See Proposition 1.3.6 for a proof of their equivalence. $\Diamond$

Definition 4.1.17 Let G be a group and H a subgroup of G. The **normalizer** of H in G is

$$N_G(H) = \{g \in G \mid gHg^{-1} = H\}.$$

In particular, K is a normal subgroup of G iff $N_G(K) = G$. $\Diamond$

Definition 4.1.18 Let X be a non-empty set. A **partition** of X is a collection $P = \{X_i \mid i \in I\}$ of nonempty subsets so that

- $X = \bigcup\limits_{i \in I} X_i$, and
- $X_i \cap X_j = \emptyset$ for all $i \neq j$.

$\Diamond$

Definition 4.1.19 Let R be a commutative ring with identity. An element $\pi \in R$ is said to be **prime** if $\pi \neq 0$, $\pi \notin R^\times$, and given $a, b \in R$ with $\pi \mid ab$, then $\pi \mid a$ or $\pi \mid b$. $\Diamond$

Definition 4.1.20 Let R be a commutative ring with identity, and P an ideal of R. Then P is a **prime** ideal iff

- P is a proper ideal
- For every $a, b \in R$, if $ab \in P$, then either $a \in P$ or $b \in P$.

We remark that in a noncommutative ring, a different definition is required: P is a **prime** ideal iff P is proper and for any ideals $I, J \subset R$, $IJ \subseteq P$ implies $I \subseteq P$ or $J \subseteq P$. If the ring is commutative, this definition is equivalent to the previous one. $\Diamond$

Definition 4.1.21 Let R be a UFD, and $p(x) = a_0 + a_1x + \cdots + a_nx^n \in R[x]$. We say that p is a **primitive** polynomial if $\gcd(a_1, \ldots, a_n) = 1$, that is there is no common divisor of all the coefficients except for units. It is immediate that for any $p \in R[x]$ that $p = c(p)p_0$ where p_0 is primitive and $c(p) \in R$ which is usually referred to as the **content** of p. $\Diamond$

Definition 4.1.22 A **ring** is a nonempty set R with two binary operations, $+,\times$ so that

- $(R, +)$ is an abelian group
- $\times$ is an associative operation
- $(a + b) \times c = (a \times c) + (b \times c)$
- $a \times (b + c) = (a \times b) + (a \times c)$

It is a **commutative** ring if $\times$ is commutative. The ring R has an **identity** if there is an element $1 \in R$ so that $1 \times r = r \times 1 = r$ for all $r \in R$. Generally, we write rs instead of $r \times s$.

For an element $r \in R$, $-r$ is its additive inverse, and r^{-1} is its multiplicative inverse (if it exists). We denote by $R^{\times}$ the unit group of a ring R with identity. $\lozenge$

Definition 4.1.23 A nonempty subset H of a group G is a **subgroup** of G if it is closed under products and inverses. More succinctly, it is a subgroup if for every $x, y \in H$, $xy^{-1} \in H$. It is usually denoted $H \leq G$ or $H < G$ for a proper subgroup. $\lozenge$

Definition 4.1.24 A function $f : X \to Y$ between sets X and Y is **surjective** if for every $y \in Y$, there exists an $x \in X$ such that $f(x) = y$. $\lozenge$

Definition 4.1.25 Let R be a ring with 1. An element $u \in R$ is a **unit** if there exists an inverse in R, that is there exists an element $v \in R$, with

$$uv = vu = 1.$$

$\lozenge$

Definition 4.1.26 A **vector space** is a non-empty set V and an associated field of scalars F, having operations of vector addition, denoted $+$, and scalar multiplication, denoted by juxtaposition, satisfying the following properties: For all vectors $u, v, w \in V$, and scalars $\lambda, \mu \in F$

closure under vector addition

$u + v \in V$

addition is commutative

$u + v = v + u$

addition is associative

$(u + v) + w = u + (v + w)$

additive identity

There is a vector $\mathbf{0} \in V$ so that $\mathbf{0} + u = u$.

additive inverses

For each $u \in V$, there is a vector denoted $-u \in V$ so that $u + -u = \mathbf{0}$.

closure under scalar multiplication

$\lambda u \in V$.

scalar multiplication distributes across vector addition

$\lambda(u + v) = \lambda u + \lambda v$

distributes over scalar addition

$(\lambda + \mu)v = \lambda v + \mu v$

scalar associativity

$(\lambda\mu)v = \lambda(\mu v)$

V**is unital**

The field element $1 \in F$ satisfies $1v = v$.

$\lozenge$

Definition 4.1.27 A nonzero element a in a ring R is a **zero divisor** if there is a nonzero $b \in R$ with $ab = 0$ or $ba = 0$. In noncommative rings, one condition may hold, but not the other (e.g., in matrix rings). If desired, one can talk about left and right zero divisors, though that will not be our focus. $\Diamond$

References and Suggested Readings

[**1**] D. Dummit and R. Foote*Abstract Algebra*. 3rd ed. John Wiley 2004